DICTIONARY OF PRIVACY, DATA PROTECTION AND INFORMATION SECURITY

Dedication
The authors would like to dedicate this Dictionary to
Lynne, Pietro, Nik, Tom and Rebecca.

Dictionary of Privacy, Data Protection and Information Security

Mark Elliot

Professor of Data Science, School of Social Sciences, University of Manchester, UK

Anna Maria Mandalari

Assistant Professor, Department of Electronic and Electrical Engineering, University College London, UK

Miranda Mourby

Researcher, Faculty of Law, University of Oxford, UK

Kieron O’Hara

Emeritus Fellow in Electronics and Computer Science, University of Southampton, UK

EE Edward Elgar
PUBLISHING

Cheltenham, UK • Northampton, MA, USA

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

Library of Congress Control Number: 2024936917

This book is available electronically in the Elgaronline
Law subject collection
https://dx.doi.org/10.4337/9781035300921

ISBN 978 1 0353 0091 4 (cased)
ISBN 978 1 0353 0092 1 (eBook)

Typeset by Cheshire Typesetting Ltd, Cuddington, Cheshire

Contents

Acknowledgements xxxviii
Preface xxxix
How to use this dictionary xlii

A

A29WP 1
Abortion 1
Abstract 2
Accessibility 2
Access Control 2
Access Control List (ACL) 2
Access Point 3
Accountability 3
Account Management 4
Account Take Over (ATO) 4
Accuracy 4
ACL 5
Additivity 5
Adequacy 5
Ad Exchange 6
ADF 6
Ad Hoc Network 7
Ad Network 7
Adtech 8
Advanced Electronic Signature 8
Advanced Encryption Standard (AES) 8
Adversary 9
Adware 9
AES 10
Affinity Analysis 10
Aggregation 10
Agreement 10
AI 11
Algorithm 11
American Data Privacy and Protection Act 11
Analogue Hole 11
Analysis Server 12
Analytical Completeness 12
Analytical Validity 12

Anomaly Detection 13
Anonymisation 13
Anonymisation Decision-Making Framework (ADF) 15
Anonymising Proxy 15
Anonymity 15
Anonymous Search Engine 16
Anti-Discrimination Law 16
Anti-Malware Software 16
Anti-Virus Software 17
APEC Privacy Principles 17
API 18
App 18
Application (App) 18
Application Layer Attack 19
Application Programming Interface (API) 19
Appropriate Safeguards 20
Appropriate Technical and Organisational Measures 20
Appropriation of Name or Likeness 21
AR 21
Article 29 Working Party (A29WP) 21
Artificial Intelligence (AI) 22
AS 23
Asset 23
Associational Privacy 24
Assured Data Deletion 24
Asymmetric Cryptography 24
Asymmetric Information 25
ATO 25
Attack 26
Attacker 26
Attack Model 26
Attack Surface 26
Attack Tree 27
Attack Vector 27
Attentional Privacy 27
Attention as a Resource 28
Attention Tracking 28
Attitude–Behaviour Gap 29
Attribute 29
Attribute Disclosure 30
Attribution 30
Audit Trail 31

Augmented Reality (AR) 31
Authentication 32
Authorisation 32
Automated Decision-Making 33
Autonomous System (AS) 34
Autonomy 34
Auxiliary Data 35
Auxiliary Information 35
Auxiliary Knowledge 35
Availability 36
Awareness 36

B

Backdoor 38
Back-Up 38
Barnardisation 39
BCI 39
BCR 40
Behavioural Advertising 40
Benefits of Privacy 40
Bicycle Attack 41
Big Brother 42
Big Data 42
Binary Variable 43
Binding Corporate Rules (BCR) 44
Biobank 44
Biometric Data 45
Biometrics 46
Black Hat Attack 46
Blacklist 46
Blackmail 47
Blinding 48
Blockchain 48
Block Cipher 50
Blocking Variable 50
Blocklist 50
Bluejacking 50
Bluesnarfing 51
Blue Team 51
Bluetooth 51
Bodily Privacy 52
Bot 52

Botnet 53
Boundary 54
Bounded Rationality 54
Bounds 55
Brain–Computer Interface (BCI) 55
Brain Implant 56
Brainwashing 56
Breach 57
Breach Disclosure 58
Breach of Confidence 59
Bring Your Own Device Policy (BYOD) 59
Browser Fingerprinting 60
Browsing History 60
Brussels Effect 61
Brute Force Attack 61
Buffer Overflow Attack 62
Bug 63
Business Case 63
Business Impact Level 64
BYOD 64

C

CA 65
Categorical Data 65
CCTV 65
Celebrity Privacy 66
Cell Suppression 66
Censorship 66
Census 67
Centralised Governance 67
Certification 68
Certification Authority (CA) 69
Chain of Trust 69
Challenge-Response 70
Charter of Fundamental Rights 70
Charter Rights 71
Checksum 71
Chief Privacy Officer 72
Children's Privacy 72
Chilling Effect 73
Chinese Wall 74
Choice Architecture 74

CIA Triad 74
Cipher 75
Ciphertext 76
Classified Information 76
Cleartext 76
Clickstream Data 77
Client Confidentiality 77
Client-Side Scanning 78
Closed Circuit Television (CCTV) 78
Cloud Computing 79
Cloud Storage 80
Code Audit 81
Code of Conduct 82
Code of Ethics 82
Code Review 83
Commodification 83
Common Law 84
Communication 84
Communication Privacy 85
Communication Privacy Management (CPM) Theory 86
Community Privacy 87
Compliance 88
Concentration Rule 88
Conditions for Processing 88
Confidence 89
Confidentiality 89
Confidentiality Club 91
Confidentiality Pledge 91
Conflict of Rights 92
Connected Place 92
Connectomics 93
Consent 93
Consent Form 94
Consequential Data 95
Consistency Mechanism 95
Consumer Information Markets 96
Consumer Preference Information 97
Content Data 97
Contextual Advertising 98
Contextual Integrity 98
Continuous Data 99
Controlled Rounding 99

Controlled Tabular Adjustment (CTA)	99
Convention 108	100
Cookie	100
Cooperation Mechanism	101
Co-Privacy	102
Correct Attribution Probability	102
Count Data	103
CPM	103
Credentials	103
Creepiness	103
Creepy Line, The	104
Crime Prevention Exemptions	104
CRM	105
Cross-Border Data Processing	105
Cross-Device Tracking	105
Cross-Site Request Forgery (CSRF)	106
Cross-Site Scripting (XSS)	107
Cryptanalysis	107
Cryptocurrency	108
Cryptographic Hash Function	109
Cryptographic Key	109
Cryptographic Protocol	110
Cryptography	110
Crypto-Shredding	111
Crypto Wars	112
CTA	112
Cultural Variation of Privacy	113
Currency	113
Customer Relationship Management (CRM)	114
Customer Tracking	114
Cybercrime	115
Cyber Insurance	116
Cyber Resilience	117
Cybersecurity	117
Cyberstalking	117
Cyberterrorism	118
Cyberwarfare	119
Cypher	120
Cypherpunk	120

D

Dark Pattern 121
Dark Web 121
Data 122
Data Abuse 123
Data Ageing 123
Data Aging 124
Data at Rest 124
Database 124
Database of Ruin 124
Data Breach 125
Data Breach Notification 126
Data Broker 126
Data Capture 127
Data Centre 127
Data Classification 128
Data Controller 128
Data Curation 129
Data Custodian 129
Data Degaussing 129
Data Destruction 130
Data Divergence 130
Data Dumping 130
Data Enclave 131
Data Environment 131
Data Environment Analysis 132
Data Ethics 132
Data Exhaust 133
Datafication 133
Data Flow 134
Data Flow Diagram 134
Data Governance 135
Data Harmonisation 136
Data Harvesting 136
Data in Motion 137
Data Intermediary 137
Data in Transit 137
Data Intruder 138
Data Intrusion Simulation 138
Data in Use 138
Data Lake 138
Data Lifecycle 139

Data Lifecycle Management 140
Data Linkage 140
Data Map 141
Data Minimisation 141
Data Minimisation Principle 142
Data Mining 142
Data Ownership 143
Data Portability 144
Data Privacy 145
Data Processing 145
Data Processor 145
Data Protection 145
Data Protection Authority 146
Data-Protection-by-Default 146
Data-Protection-by-Design 147
Data Protection Directive 147
Data Protection Impact Assessment (DPIA) 148
Data Protection Officer (DPO) 149
Data Protection Policy 150
Data Protection Principles 151
Data Provenance 151
Data Quality 152
Data Recipient 153
Data Release 153
Data Retention 154
Data Safe Haven 155
Data Sanitisation 156
Data Schema 156
Dataset 156
Data Sharing 157
Data Sharing Agreement (DSA) 158
Data Situation 158
Data Situation Audit 159
Data Sovereignty 159
Data Steward 160
Data Stewardship Organisation 161
Data Storage 161
Data Subject 162
Data Subject Access Request 163
Data Synthesis 163
Data Transfer 164
Data Trust 165

Data Unit 166
Data User 166
Data Utility 166
Dataveillance 167
Data Warehouse 168
DDOS 168
Deanonymisation 169
Decentralisation of the Web 169
Decisional Privacy 170
Declared Data 170
Decryption 171
Deepfake 171
Deep Learning 172
Deep Packet Inspection (DPI) 173
Deep Web 173
Defamation 173
Default Settings 174
De-Identification 175
Deletion 175
Delta 176
Demographic Advertising 176
Demonstration Attack 176
Denial of Service (DoS) 177
Deterministic Record Linkage 178
Device Fingerprinting 178
DICOM Standard (Digital Imaging and Communications in Medicine) 179
Differencing 179
Differential Identifiability 180
Differential Privacy 180
Digital Assistant 181
Digital Breadcrumbs 182
Digital Certificate 182
Digital Divide 183
Digital Economy 183
Digital Fingerprinting 183
Digital Footprint 184
Digital Footprint Eraser 185
Digital Hygiene 186
Digital Imaging and Communications in Medicine 186
Digital Identity 186

Digital Inequality	187
Digital Inheritance	188
Digital Literacy	188
Digital Persona	189
Digital Rights Management (DRM)	189
Digital Self-Determination	190
Digital Signature	190
Digital Twin	191
Digital Wallet	191
Dignity	192
Direct Access Attack	192
Direct Identifier	193
Direct Marketing	193
Directory Indexing	194
Disassociability	194
Disclosive Data	194
Disclosure	195
Disclosure and Barring (Check)	195
Disclosure Control Methods	195
Disclosure Risk	196
Discrete Data	196
Discretionary Access Control	197
Discretisation	197
Disguise	197
Distributed Denial of Service	197
Distributed Ledger	198
DNS Server	198
Domain Generalisation	198
Dominance Rule	199
Do Not Track (Protocol)	199
DOS	200
Doxxing	200
DPI	200
DPIA	200
DPO	200
D-Privacy	200
DRM	201
DSA	201
Duty of Confidence	201
Duty to Protect	201
Duty to Warn	201
Dyad	202

Dynamic Consent 202
Dynamic Data Situation 203

E

E2EE 204
E3 204
Eavesdropping 204
Eavesdropping Attack 204
ECHR 205
E-Commerce 205
Economics of Privacy 206
Edge Computing 207
EDPB 207
EDPS 207
EHR 208
Electronic Health Record (EHR) 208
EM Algorithm 208
Emotion Recognition 209
Employee Information 210
Encrypt-Everything-Everywhere (E3) 210
Encryption 211
Encryption Algorithm 211
Encryption Key 211
Endpoint Security 212
End-to-End Encryption (E2EE) 212
End-User Licence Agreement (EULA) 213
Engineering Ethics 213
Enhanced Privacy ID (EPID) 214
EPID 214
Eprivacy Directive 214
Eprivacy Regulation 215
Epsilon 215
Equivalence Class 216
Equivalence Class Structure 216
Erasure 216
Escrow 217
Ethical Hacking 218
Ethics 218
Ethics Committee 218
EULA 219
European Convention on Human Rights (ECHR) 219
European Data Protection Board (EDPB) 220

European Data Protection Supervisor (EDPS) 221
Exfiltration 221
Expectation-Maximization Algorithm 221
Explainable AI (XAI) 221
Explanatory Variable 222
Explicit Consent 222
Exposure 223
Express Consent 223
Extranet 224
Extrinsic Privacy 224

F

Face Recognition 225
Facial Recognition Technology 225
FAIR 226
Fair Information Practice Principles (FIPPS) 226
Fairness 227
Fair Processing Notice 228
Fake Profile 228
False Light 229
False Negative 230
False Positive 230
Family Resemblance Theory of Meaning 231
FARAS 231
Feature 231
Federal Trade Commission (FTC) 232
Federated Identity 232
Federated Learning 233
Feminist Critique of Privacy 233
FHE 234
Fiduciary Duty 234
Filing System 235
Financial Privacy 235
FIPPS 236
Firewall 236
Firmware 237
Fishing Attack 238
Five Safes 238
Flexible Output 238
Formal Anonymisation 239
Formal Privacy 239
Format Preserving Encryption (FPE) 240

Foundation Model 240
FPE 240
Freedom of Expression 241
Freedom of Information 241
Freely Given Consent 242
Frequency Data 242
FTC 243
Fully Automated Remote Analysis System (FARAS) 243
Fully Homomorphic Encryption (FHE) 243
Functional Anonymisation 244
Functional Unique Identifier 245
Function Creep 246
Fuzzing 246

G

Gait Recognition 247
Game Theory 247
Gatekeeper 248
GDPR (General Data Protection Regulation) 248
Gendered Spaces 250
General Data Protection Regulation 250
Generative AI 250
Genetic Data 251
Genetic Fingerprint 251
Genetic Privacy 252
Genomic Data 253
Geographical Resolution 253
Geoprivacy 254
Geo-Social Data 254
Geotagging 255
Globally Unique Identifier (GUID) 255
Global Privacy Control (GPC) 256
Global Recoding 256
Global Suppression 257
Gossip 257
GPC 257
Graduated Security 258
Grey Hat Attack 258
Group Harms 258
Group Privacy 259
GUID 260

H
Hacking 261
Harassment 261
Harm 262
Hashing 263
Header Information 263
Health Information Exchange (HIE) 264
Health Insurance Portability and Accountability Act (HIPAA) 265
Hellinger Distance 265
Hidden Service 266
HIE 266
Hierarchical Data 266
HIPAA 266
History of Privacy 266
Homomorphic Encryption 268
Honeypot 268
HRIA 269
HTTP 269
HTTPS 269
Human-Centred Cybersecurity 269
Human Rights Impact Assessment (HRIA) 270
Hypertext Transfer Protocol (HTTP) 270
Hypertext Transfer Protocol Secure (HTTPS) 271

I
I2P 272
ID Card 272
Idem-Identity 272
Identifiability 273
Identifiable Data 273
Identifiable Individual 273
Identifiable Natural Person 273
Identification Card 274
Identification File 275
Identified Data 276
Identified Natural Person 276
Identifier 277
Identity 277
Identity Assurance 279
Identity Cloning 280
Identity Disclosure 280
Identity Documents 281

Identity Management 281
Identity Provider 282
Identity Theft 283
Ideological Privacy 283
IDS 283
Impact Management 284
Impersonation 284
Implicit Consent 284
Imputation 285
Inadvertent Disclosure 285
Incognito Mode 286
Incremental Authorisation 286
Indirect Identifier 286
Inference 287
Inference Attack 288
Inferential Disclosure 288
Inferred Data 289
Inforgs 289
Information 289
Informational Privacy 289
Informational Self-Determination 291
Information Broker 291
Information Classification Table 291
Information Ethics 292
Information Governance 292
Information Lifecycle Management 293
Information Loss 293
Information Ownership 293
Information Security 294
Informed Consent 294
Infosphere 295
Inherence 296
Input Privacy 296
Input Statistical Discloure Control 297
Integrity 297
Intellectual Privacy 298
Intellectual Property (IP) 298
Intentional Data 300
Intention–Behaviour Gap 300
Interference 300
Internal Security Testing 301
International Transfer 301

Internet 301
Internet of Humans 302
Internet of People 302
Internet of Things 302
Internet Protocol (IP) 303
Interoperability 304
Interval Publication 305
Intimacy 305
Intranet 306
Intruder 306
Intruder Testing 307
Intrusion 307
Intrusion Detection System (IDS) 308
Intrusion Prevention System (IPS) 308
Intrusion upon Seclusion 309
Invasive BCI 309
Inversion Attack 310
Inviolate Personality 310
Invisible Computing 311
Invisible Internet Project 311
IP 311
IP Address 312
IPS 312
IPSE-Identity 312
Iris Scanning 312
Irreversibility 313
ISO27001 313
ISO27002 314
Isolation 314

J

Jensen–Shannon Divergence 316
Jigsaw Identification 316
Joint Data Controller 316
Jurisdiction 317
Just-In-Time Consent 317
Just-In-Time Notice 318

K

K-Anonymity 319
Key Disclosure 319
Key Logger 319

Key Logging 319
Key Variable 320
Knowledge Economy 320
Kompromat 320

L

Laplace Noise 322
Large Language Model 322
Lawful Basis 322
Lawfulness 323
Layered Notice 324
Layered Security Model 324
LBS 325
L-Diversity 325
Lead Supervisory Authority 325
Least Privilege 326
Legal Basis for Processing 326
Legitimate Interest 326
Libel 327
Licence Agreement 327
Lifecasting 328
Lifelogging 328
Lifestreaming 329
Linkability 329
Linkable Information 330
Linkage 330
Linkage Attack 330
Link Encryption 331
Local Shared Object (LSO) 331
Local Suppression 331
Locational Privacy 332
Location-Based Service (LBS) 333
Location Data 333
Location Tracking 334
Logic Bomb 334
Longitudinal Data 335
Loyalty Card 335
LSO 336

M

MAC 337
Machine Learning (ML) 337

Magnitude Data	339
Main Establishment	339
Male Gaze, The	339
Malicious Proxy Server	340
Malware	340
Management Information System	340
Mandatory Access Control (MAC)	341
Mandatory Decryption	341
Mandatory Key Disclosure	342
Man-in-the-Middle Attack	342
Manual Key Transport	343
Mash Attack	343
Mask	343
Masking	344
Masquerade	344
Matching	344
Material Scope	345
Material Transfer Agreement	345
Maximum Knowledge Intruder	346
Media Access Control (MAC) Address	346
Medical Record	347
Membership Inference Attack	347
Mental Capacity	347
Mental Privacy	348
Mesh Network	349
Message Digest	349
Metadata	350
Metadata-Level Controls	350
Metaverse	351
MFA	351
Microaggregation	351
Microdata	352
Minimal Unique	352
Missing Data	353
Mission Creep	353
Misuse of Private Information	353
ML	354
Mobility Traces	354
Model Inversion Attack	355
Molka	355
Monetary Equivalent Burden (of Privacy)	355
Mosaic Identification	356

Motivated Intruder 356
Motivated Intruder Test 356
M-Probability 356
Multi-Factor Authentication (MFA) 357
Multimodal De-Identification 357
Multiple Imputation 357
Multi-Vector Attacks 358
Mutual Assistance 358
Mutual Authentication 358

N

National Security 360
Natural Person 360
Necessity 361
Need to Know 361
Negative Externalities of Disclosed Data 362
Negative Externalities of Privacy 362
Negligence 363
Network 364
Network Encryption 364
Network Layer Attack 364
Network Security 365
Network Segmentation 365
Neural Prosthesis 366
Neurocapitalism 366
Neurodata 367
Neuroethics 367
Neuroprivacy 368
Neuroprosthetics 368
Neurotechnology 368
(N,K) Rule 369
Noise Addition 369
Nom de Guerre 369
Nom de Plume 369
Nominal Data 370
Non-Disclosure Agreements 370
Non-Discrimination Law 370
Non-Invasive BCI 371
Notice and Consent 371
Nudge Theory 372
Nudging 372

O

Obfuscation 373
Objective Harm 373
Oblivious Transfer (OT) 373
Obscurity 374
Obtrusion 375
OECD Guidelines (On Privacy) 375
Offline Dictionary Attack 376
One-Stop-Shop 377
One-Way Hash Function 377
Onion Routing 378
Online Vetting 378
Ontological Security 378
Onward Transfer 379
OPE 380
Open Access 380
Open Data 381
Open Source 382
Opt-In 383
Opt-Out 383
Order-Preserving Encryption 384
Ordinal Data 385
Orwell Attack 385
OT 385
Other 386
Outing 386
Outlier 386
Output Checking 387
Output Privacy 387
Output Statistical Disclosure Control 388
Overimputation 389

P

P3P 390
Packet Filter 390
Packet Sniffing 390
Panel Data 391
Panopticon 391
Paparazzi 392
Paparazzi Attack 392
Parental Controls 393
Partially Homomorphic Encryption 393

Participant Information Sheets 393
Participatory Surveillance 394
Passive Collection 394
Password 395
Password Manager 395
Patch 395
Patch Management 396
PDS 396
Peeping Tom 396
Penetralia 396
Penetration Test 397
Persistent Cookie 397
Persistent Pseudonym 397
Person 398
Personal Data 398
Personal Data Cloud 399
Personal Data Store (PDS) 399
Personal Identification Number (PIN) 400
Personal Information 400
Personal Information Management System (PIMS) 401
Personalisation 402
Personalisation Reactance 403
Personalised Medicine 403
Personalised Services 404
Personally Identifiable Information (PII) 404
Personal Space 405
Personhood 405
Perturbation 406
Pervasive Computing 406
PET 406
Pharming 407
PHE 407
Philosophy of Information 407
Phishing 409
Phone Hacking 409
Physical Privacy 409
PIA 410
PII 410
PIMS 410
PIN 410
Pixelisation 410
Plaintext 411

Platform for Privacy Preferences (P3P) 411
Poisoning Attack 412
Population 412
Population Unique 412
Population Unit 413
Port Scan 413
Positive Externalities of Disclosed Data 414
Post Quantum Cryptography (PQC) 414
Post Randomisation (PRAM) 415
PPDA 415
PPDM 415
PPML 415
P/Q Rule 415
PRAM 416
Predictive Analytics 416
Predictive Modelling 417
Presence Detection 417
Price Discrimination 417
Primary Data 418
Prior Posterior Ambiguity Rule 418
Privacy 418
Privacy as Control 419
Privacy as Redistribution of Costs 420
Privacy Avatar 421
Privacy, Benefits of 421
Privacy Budget 422
Privacy-by-Design 422
Privacy Calculus 423
Privacy Concern 423
Privacy, Cultural Variation of 424
Privacy Elasticity 424
Privacy Engineering 424
Privacy-Enhancing Technology (PET) 425
Privacy First 425
Privacy Fundamentalists 426
Privacy Guarantee 426
Privacy Impact Assessment 427
Privacy Insurance 427
Privacy-Invasive Technology 427
Privacy Metric 428
Privacy Model 428
Privacy Notice 429

Privacy Officer 429
Privacy Paradox 430
Privacy Policy 431
Privacy Pragmatists 431
Privacy Preference 432
Privacy Premium 432
Privacy-Preserving Data Analytics (PPDA) 432
Privacy-Preserving Data Mining (PPDM) 433
Privacy-Preserving Data Publishing 433
Privacy-Preserving Machine Learning (PPML) 433
Privacy-Preserving Record Linkage (PPRL) 434
Privacy Risk 434
Privacy Screen 435
Privacy Seal 435
Privacy Settings 436
Privacy Threat 436
Privacy Tort 437
Privacy Trade-Off 438
Privacy Unconcerned 438
Privacy, Value of 439
Private Army 439
Private Biometrics 439
Private Browsing Mode 440
Private Enterprise 440
Private Key 440
Private Life 440
Private Parts 441
Private Property 442
Private School 442
Private Sector 443
Private Sphere 443
Probabilistic Record Linkage 444
Processing 445
Profile 445
Profiling 445
Prolepticon 446
Proportionate Security 446
Proportionality 447
Proprietary Privacy 447
Protected Characteristics 448
Protocol 448
Provable Security 449

Proxemics 450
Proxy 450
P% Rule 450
Pseudonym 450
Pseudonymisation 451
Pseudonym Reversal 453
Psychographic Advertising 453
Psychological Privacy 453
Public 454
Publication 455
Public Disclosure of Private Facts 455
Public Domain 456
Public Figure 456
Public Interest 457
Publicity 458
Public Key 459
Public-Key Cryptography 459
Public-Key Infrastructure (PKI) 460
Public Records 460
Public Sphere 461
Publishing 462
Purchase History 462
Purple Team 462
Purpose Limitation 463
Purpose Specification 464

Q

Quantitative Privacy 465
Quantum Computing 465
Quasi-Identifier 466
Query Logging 466
Query Overlap 467
Quishing 467

R

Radical Transparency 468
Radio Frequency Identification (RFID) 468
Randomised Response 469
Random Rounding 469
Random Unique 469
Ransomware 470
Rational Consumer 470

Reality Mining 471
Reasonable Expectation of Privacy 471
Reasonable Search 473
Recognition 473
Recommendation System 473
Recommender System 474
Reconstruction Attack 474
Record 474
Record Linkage 475
Records Management 476
Record Suppression 476
Rectification 476
Redaction 476
Red Team 477
Reference Monitor 477
Regulation 477
Regulators 478
Reidentification 478
Reidentification Attack 480
Relational Autonomy 480
Release and Forget 481
Reliability 481
Reliance 481
Reliance Authentication 481
Remailing 482
Remediation 482
Remedies 482
Remote Access 483
Remote Access Server 483
Remote Analysis Server 483
Remote Query 484
Replay Attack 484
Repurposing 485
Reputation 485
Reputation Management 485
Reserve 486
Resilience 486
Respondent 486
Response Knowledge 487
Response Variable 487
Responsible Disclosure 487
Restricted Access 488

Retention 488
Revenge Porn 488
Reverse Fishing Attack 489
Reversibility 489
Revocation 489
Right of Access 490
Right to an Explanation 490
Right to Be Forgotten 491
Right to Be Informed 492
Right to Be Let Alone 493
Right to Correct 494
Right to Data Portability 494
Right to Data Protection 494
Right to Deletion 495
Right to Object 495
Right to Privacy 496
Right to Rectification 496
Right to Restriction 496
Risk 497
Risk Assessment 498
Risk Management 498
Risk Tolerance 498
Risk–Utility Trade-Off 499
Roe v Wade 500
Role-Based Access Control 500
Rounding 500
RSA Encryption 501
R-U Map 501
Rumour 501

S

Safe Data 502
Safe Harbor 502
Safe Output 503
Safe People 503
Safe Projects 503
Safe Settings 503
Safety 504
Salt 504
SAML 505
Sample Unique 505
Sample Unit 505

Sampling 505
Sampling Fraction 505
Sandbox 506
Sandboxing 506
Scenario Analysis 506
Schrems 507
Scraping 507
Scrutiny 507
SDC 508
SDL 508
SDLC 508
Search 508
Search Engine 510
Seclusion 510
Secondary Data 510
Secondary Differentiation 511
Secondary Use 511
Secrecy 511
Secret 512
Secret Ballot 512
Secret Sharing 513
Secure Communication 514
Secure Messaging 514
Secure Multi-Party Computation 514
Secure Sockets Layer (SSL) 515
Secure Web Gateway (SWG) 515
Secure Web Platform 516
Security 516
Security Assertion Markup Language (SAML) 517
Security Audit 518
Security-By-Design 518
Security-By-Obscurity 518
Security Information Management (SIM) 519
Security Parameter 519
Security Posture 520
Security Requirement 520
Security Token 520
Self-Archiving 521
Self-Control Security 521
Self-Disclosure 521
Self-Reflection 522
Self-Sovereign Identity (SSI) 522

Self, The 523
Semantic Security 524
Semi-Invasive BCI 524
Sensitive Data 524
Sensitive Variable 524
Sensitivity 525
Sensor 525
Serial Number 526
Service User Agreement 526
Sessional Cookie 526
Shoulder Surfing 526
SIM 527
Single Out 527
Single Sign-On 528
Singularity, The 528
Slander 529
Smart City 529
Smart Device 530
Smart Grid 530
Smart Implants 531
Smart Meter 531
Smishing 531
SNARK 532
Social Credit System 532
Social Engineering 532
Social Genome 533
Social Linked Data (SOLID) 533
Social Network 534
Social Network Analysis 536
Social Profiling 536
Social Steganography 536
Sock Puppet 536
Software 536
Software Development Lifecycle (SDLC) 537
SOLID 537
Solitude 537
Sousveillance 538
Spam 538
Spam Filter 539
Spatial Cloaking 539
Spatial Privacy 539
Spear Phishing 540

Special Category Data 540
Special Unique 540
Speech Recognition 541
Split Tunnelling 541
Spontaneous Recognition 542
Spoofing Attack 542
Spyware 543
SQL 543
SQL Injection 543
SSI 544
SSL 544
Stakeholder 544
Stalking 545
Standard 545
Standard Contractual Clauses 546
Standard Model Clauses 547
Static Key 547
Statistical Disclosure 547
Statistical Disclosure Control (SDC) 548
Statistical Disclosure Limitation (SDL) 549
Steganography 549
Storage Limitation 549
Streisand Effect 550
Structural Zero 550
Structured Query Language (SQL) 551
Subject Access Request 551
Subjective Harm 551
Subliminal Advertising 552
Subtraction Attack 552
Succinct Non-Interactive Zero-Knowledge Proof (SNARK) 553
Super Cookie 553
Supervisory Authority 554
Suppression 554
Surname Attack 555
Surveillance 555
Surveillance Capitalism 556
Swapping Key 557
SWG 558
Symmetric Cryptography 558
Symmetric Key Encryption 558
Synthetic Data 558

T

Table Redesign	559
Tabular Data	559
Tagging	559
Target Dataset	560
Targeted Advertising	560
Target Variable	560
TCB	560
T-Closeness	560
TCP	561
Technical And Organisational Measures	561
TEE	561
Telephone Tapping	561
Terms of Service	561
Territorial Privacy	562
Territorial Scope	562
Text Anonymisation	563
TFA	563
The Onion Router	564
Therapeutic Alliance	564
Thermal Imaging	564
Third Party	565
Third Party Doctrine	566
Thought Police	566
Threat Modelling	566
Threshold Rule	567
Time Bomb	567
Time Series	568
TIPS (Trust, Identity, Privacy, Security)	568
TLS	569
Tokenisation	569
Topcoding	569
TOR (The Onion Router)	569
Tracing	570
Tracker	570
Tracker Blocker	571
Tracking	571
Traffic Data	571
Transmission Control Protocol (TCP)	572
Transparency	572
Transparency Notice	573
Transport Layer Security (TLS)	573

Trapdoor 574
Trespass 574
Trojan Horse 574
Trust 575
Trusted Computing Base (TCB) 576
Trusted Execution Environment (TEE) 576
Trusted Research Environment 577
Trusted Third Party 577
Trustworthy Digital Identity 578
Tunnel Encryption 578
Two Factor Authentication (TFA) 578

U

Ubicomp 579
Ubiquitous Computing (UBICOMP) 579
UK GDPR 579
Unambiguous Consent 580
Unicity 580
Uniform Resource Identifier (URI) 580
Unique Identifier 581
Uniqueness 581
Unreasonable Search 581
Untraceability 582
U-Probability 583
URI 583
User 583
User-Centred Design 584
User Modelling 584
Username 585
US Privacy Laws 585
Utility First 586

V

Value–Action Gap 587
Value of Data 587
Value of Privacy 588
Value Sensitive Design (VSD) 589
Veil 589
Verifiable Secret Sharing (VSS) 590
Vicarious Liability 590
Virality 591
Virtual Machine (VM) 592

Virtual Private Network (VPN)	592
Virus	593
Vishing	593
Vital Interests	593
VM	594
Voice over Internet Protocol (VoIP)	594
VoIP	595
Voyeurism	595
VPN	595
VSD	595
VSS	595
Vulnerability	595
Vulnerability Management	596
Vulnerable Population	597

W

Wearable Computing	598
Wearable Tech	599
Web 2.0	599
Web 3.0	599
Web Beacon	600
Web Bug	601
Web of Trust	601
Web Profiling	602
Web Skimming Attack	602
Whistleblowing	602
White Box Testing	603
White Hat Attack	604
Wiretapping	604
World Wide Web (WWW)	604
Worm	606
WWW	606

X

XAI	607
XSS	607
X Variable	607

Y

Y Chromosome	608
Y Variable	608

Z

Zero Day Attack 609
Zero Knowledge 609
Zero Knowledge Proof 610
Zero Trust Security 610

Acknowledgements

Many people have been helpful in the creation of the *Dictionary* you hold in your hand or see on your screen. We would like to acknowledge the contributions of: Emma Barrett, Danny Dresner, Andre Freitas, Hamed Haddadi, Naomi Hawkins, and Mark Taylor, for specific advice on and input to some entries; Harry Fabian, our commissioning editor from Edward Elgar Publishing; the participants in a Dictionary workshop at the SPRITE+ community meeting in Belfast in 2023, who gave valuable feedback on the list of terms and style; and the various members of the UKAN Anonymisation Network, whose input to earlier works by Elliot and O’Hara provided essential background to the task of creating this *Dictionary*.

We would like to acknowledge the SPRITE+ management team for agreeing to fund the open access fee. Miranda Mourby would also like to acknowledge funding from the Bundesministerium für Bildung und Forschung project awarded to Professor Fruzsina Molnar-Gabor, *Datenschutzrechtliches Reallabor für eine Datentreuhand in der Netzwerkmedizin – TrustDNA*; the EU-STANDS4PM consortium (www.eustands4pm.eu) that was funded by the European Union Horizon2020 framework programme of the European Commission under Grant Agreement #825843; and the School of Law of the University of Sheffield.

Preface

Digital technology has revolutionised many areas, ranging across science, defence, education, sport and leisure, entertainment, policymaking, civil society, finance, defence and policing. It is widely appreciated that these revolutions depend on access to abundant data about people and their behaviour. Many of the great benefits of the modern world are therefore bought at some cost to privacy.

With increasing concern about the potential threats to privacy, many experts are addressing these often thorny issues, across disciplines including anthropology, artificial intelligence and machine learning, computer science, cybersecurity, economics, ethics, ethnography, history, law, management, medicine and medical science, philosophy, psychology, sociology, statistics, technology design and many others.

The cross-fertilisation of these disciplines has been immensely fruitful. However, as they each contain different methodologies and tools to engineer different types of outcome, there has been an inevitable increase in complexity when we talk about privacy. The common problem has not led to a common language. Different argots compete, serving different imperatives. Some projects are conceptual, while others aim to create privacy-supportive architectures. Some are aimed at enabling people to make choices and achieve their preferences, while others try to support the contextual integrity of social norms. Legal and moral rights compete against less idealistic regulatory and political constraints. For some, privacy is achieved by the implementation of an algorithm; for others, it requires enforcement of rights and freedoms; still others see it as a constant process of negotiation of interests.

The danger is that such a heterogeneous study space creates an inchoate cacophony of misunderstanding, in which various vices can flourish, ranging from technological solutionism, to legalism and casuistry, to abject surrender to the disclosure of our very identities. To prevent this, we need resources to enable productive discussion across disciplines and sectors of society.

There have been many successful ad hoc attempts to do this in the context of particular research projects. However, until now there has not been a single resource that documents the breadth of vocabulary of privacy studies, such that individual researchers, entrepreneurs, regulators, lawyers, policymakers and students can find the terminology and assumptions of the varying disciplines set out for inspection and comparison. This Dictionary of Privacy, developed by four researchers who represent disciplines including statistics, cybersecurity, law, computer science and philosophy, is an attempt to create such a resource to fill this gap.

Our method was straightforward. We seeded the Dictionary with the terms in a number of well-known glossaries and lists of key terms and brainstormed more. Major publications were scoured for key words. We did gap analyses where we could, and plenty of them appeared (and were filled) as we wrote. Doubtless there are more gaps to be discovered, for which we apologise in advance. Doubtless our own areas of expertise are more expertly covered, to the detriment of others, and we apologise again. Ultimately, we had a word limit and a deadline, and so what ideally would have been a never-ending process of infinitely large output, pleasing everyone, had to be reduced to a finite process with finite output which with any luck pleases some people some of the time.

Because the disciplines themselves often differ in their basic vocabulary, we have had to make some choices. For instance, in statistics, the term 'variable' is used to stand for the operationalisation of a construct in data; in computer science, this is often called an 'attribute'. But the n 'attribute' may also stand for something attributed (for instance, by the analysis of data). We have, therefore, followed the statistics usage of the term 'variable'.

We have always been consistent within an article, but we have not felt the same pressure to be consistent across the Dictionary. One bugbear is the term 'data', which some use in the plural (in statistics and law, for instance), and others as a singular mass noun (as in computer science, and arguably more common in layperson's English). Unfortunately, one cannot publish a major survey of privacy without mentioning data. We ourselves had differing views on this; one of us, more pedantic than the other three, has even published a blog about it. We decided in the end not to enforce any particular usage, to keep an uneasy peace.

Pronouns are another area of friction, and we have tried to be as neutral and inclusive as possible; we have used 'they/them' throughout. It will also be noticed that this is a very European effort: we are jointly citizens of three European countries, and resident in two. Our expertise is inevitably shaped by that. Sometimes it may only be a matter of preferring one spelling to another, but, especially in law, geography counts. The Dictionary is therefore admittedly Eurocentric, has a few discussions of the privacy situation in the United States (which has been disproportionately influential on the literature), and provides very little indeed that specifically references issues raised in the Global South. This is a matter of scope and pragmatics, rather than an attempt to exclude.

With more than 1000 terms meticulously set out, described and cross-referenced, the Dictionary of Privacy explains, in simple and straightforward language, complex technical terms, legal concepts, privacy management techniques and conceptual matters, alongside the 'common sense' vocabulary that informs public debate. We believe that no other guide to privacy covers a comparable disciplinary range or addresses such a broad audience. While the field is fast-moving, the Dictionary takes a longer view, abstracting away from the details of today's problems, technology and law, to the wider principles that underlie privacy discourse. In that way, it is hoped that the Dictionary will remain relevant to privacy research for many years to come.

How to use this dictionary

As a dictionary, of course the articles are in alphabetical order. However, that still leaves some decisions to be made, and some choices about the conventions. Numbers and punctuation are counted as prior to letters, so ‘A29WP’ is the first entry under ‘A’. We did not count spaces or hyphens, looking only at the letters in an entry’s title, so that ‘ADEQUACY’ comes before ‘AD EXCHANGE’. For the same reason, acronyms are counted as ‘words’, so that ‘ACL’ appears between ‘ACCURACY’ and ‘ADDITIVITY’. Spelling is British English, so ‘GREY HAT ATTACK’ and not ‘GRAY HAT ATTACK’. Emphasis in the articles is shown by *italics*.

Except in rare cases where the acronym of a term is very widely known and/or the full name rarely used (such as ‘GDPR’ and ‘TOR’, instead of ‘GENERAL DATA PROTECTION REGULATION’ and ‘THE ONION ROUTER’), articles about the entity are placed under the full name, to which the acronym cross-refers. Thus, the full entry for ‘ACL’ is ‘*See*: ACCESS CONTROL LIST’, and the informative article is under the title ‘ACCESS CONTROL LIST (ACL)’. Where an entity has an acronym, it will appear, bracketed, in the title of the full article. Someone who needs to look up the acronym will be cross-referred to it, and so is able to find the article even if they did not initially know what the acronym stood for.

Cross-references are signalled in four ways. First, where two or more names refer to the same entity, one name will carry the full entry, and the other names will simply refer, as with ‘ACL’ above.

Second, where a term is mentioned in an article, its first mention will appear in **bold type**, and further information on that entity may be looked up.

Third, where the name of the entity does not appear in the text of an article, a further list of cross-references will be given at the end of the article, under the heading ‘*See also*:’

Fourth, in the online version, hyperlinks will be available.

Many articles contain suggestions for further reading; these may be more general texts, surveys, or specific standard articles where a research result or concept was first described. Many of these are labelled with Web links; these links were checked as working and correct in October 2023.

A

A29WP

See: ARTICLE 29 WORKING PARTY

Abortion

Healthcare decisions engage rights to **privacy** in most countries. The termination of a pregnancy will also engage rights to privacy as an aspect of reproductive healthcare. However, under the **European Convention of Human Rights** restrictions on abortion will not necessarily constitute a breach of privacy rights if passed by a democratically elected body (highlighting the distinction between **interference** with and **breach** of a right). In the United States, the right to elect a medical termination of pregnancy as an aspect of constitutional privacy rights was established in the 1973 US Supreme Court decision in *Roe v Wade*. This interpretation of privacy rights was upheld in 1992 by *Planned Parenthood v Casey*, but ultimately overturned by *Dobbs v Jackson Women's Health Organization* in 2022, which transferred the right to determine the legality of abortion back to individual States.

Roe v Wade followed the example of *Griswold v Connecticut* of 1965 (which established the legality of the purchase of contraceptives by married couples) in addressing family, sexual and other intimate matters as privacy issues.

Further reading:

Cosentino, C., 2015. Safe and legal abortion: an emerging human right? The long-lasting dispute with state sovereignty in ECHR jurisprudence. *Human Rights Law Review*, 15(3), 569–89, https://doi.org/10.1093/hrlr/ngv013.

Perry, M.J., 1976. Abortion, the public morals, and the police power: the ethical function of substantive due process. *UCLA Law Review*, 23(4), 689–736. https://heinonline.org/HOL/LandingPage?handle=hein.journals/uclalr23&div=46&id=&page=.

See also: BODILY PRIVACY, INTIMACY, DECISIONAL PRIVACY

Abstract

The notion of summarising **information**al content (often in the form of text). Abstraction can be a mechanism for reducing the amount of **personal data** shared or the **identifiability** of individuals within some text.

See also: NEED TO KNOW, TEXT ANONYMISATION

Accessibility

The ease with which a given system or data resource can be accessed. Accessibility is often traded off for **security** and **privacy** assurance.

Access control

Access control is implemented to regulate access within a **network**, or to a resource. Various techniques are used for implementing access control, including **authentication**, **authorisation**, **password**s, identifying and verifying **user**s, assigning privileges and permissions, and **tracking** and monitoring access to resources. Access control can be implemented using **software**, hardware or a combination of both, and is an essential component of any comprehensive **cybersecurity** strategy.

Further reading:
Sandhu, R.S. and Samarati, P., 1994. Access control: principle and practice. *IEEE Communications Magazine*, 32(9), 40–8, https://doi.org/10.1109/35.312842.

See also: APPLICATION, INTERNET

Access Control List (ACL)

An Access Control List (ACL) is used for **access control** to a specific resource (e.g., files, devices, systems or physical spaces), to specify which **user**s have been granted access to the resource and under what conditions. An ACL administrator, who will manage the permissions to ensure that only authorised users gain access to the resource, is usually empowered to add or remove permissions and to monitor activity to detect **security** and/or privacy **breach**es.

Further reading:

Gollmann, D., 2010. Computer security. *Wiley Interdisciplinary Reviews*: *Computational Statistics, 2*(5), 544–54, https://doi.org/10.1002/wics.106.

Access Point

An access point is a piece of hardware that allows devices to connect to the **Internet** and/or local **network**. Access points usually connect to a wired network using an Ethernet cable and provide Wi-Fi coverage to a specific area, such as a home, office or public space. An access point can also be used to extend the range of an existing Wi-Fi network. Access points may contain additional **security** features, such as **encryption**, network management and quality of service measurement.

Further reading:

Gupta, A. and Jha, R.K., 2015. Security threats of wireless networks: a survey. *In*: *International Conference on Computing, Communication & Automation*, IEEE, 389–95, https://doi.org/10.1109/CCAA.2015.7148407.

Accountability

The idea of liability to provide an account of oneself – particularly to a given authority or public – has found expression in this term since the eighteenth century. The more specific principle of accountability entered the **data protection** lexicon in the 1980 **OECD Guidelines** and has featured in numerous legislative regimes ever since.

Accountability can be seen as the overall spirit of **compliance** with privacy and data protection laws, as well as finding expression in some of their more concrete requirements. Examples of discrete accountability mechanisms include **certification** with an accountability agent (under the **APEC Privacy Principles**), requirements to keep adequate documentation, mandatory reporting of **breach**es and **Data Protection Impact Assessment**s.

Further reading:

Demetzou, K., 2019. Data Protection Impact Assessment: a tool for accountability and the unclarified concept of ‘high risk’ in the General Data Protection Regulation. *Computer Law and Security Review*, 35(6), 105342, https://doi.org/10.1016/j.clsr.2019.105342.

European Data Protection Supervisor, 2015. *Opinion 3/2015*: *Europe’s big opportunity.* Available from: https://edps.europa.eu/sites/edp/files/publication/15-10-09_gdpr_with_addendum_en.pdf.

Guagnin, D., 2012. *Managing privacy through accountability.* Basingstoke: Palgrave Macmillan.

Account Management

Account management is a process used for creating, monitoring and maintaining the accounts of **user**s of a system or online service. Through account management, an administrator can ensure that each user has the right permissions to perform the tasks that they need to perform within the system or service.

Account management also enables system administrators to conduct **authorisation** and **access control** and ensures that only users with the right permissions have access to **sensitive data** (including, where relevant, **personal data**).

Account Take Over (ATO)

A form of **identity theft** focused on gaining access to an individual's account on an online system. This could be because the account has value (either financial or for the **personal data** that it contains) or because the account allows access to the larger network. Attack vectors include **credential** surfing, **replay attack**s and **phishing**.

Further reading:
Gao, M., 2022. Account takeover detection on e-commerce platforms. *In*: *2022 IEEE International Conference on Smart Computing (SMARTCOMP),* 196–7. https://doi.org/10.1109/SMARTCOMP55677.2022.00052.

See also: AUTHENTICATION, PASSWORD

Accuracy

A term used specifically in **machine learning** to refer to how closely a predicted value for some **data** matches the actual value. Typically, the ratio of correctly predicted observations to all the observations in the **dataset** is used as the accuracy metric. For instance, the accuracy of a binary classification model would be 90 per cent if it correctly predicted 90 out of 100 samples.

When assessing the privacy implications of a machine learning model, accuracy can be a factor to consider. 100 per cent accuracy might be undesirable in some circumstances as it could imply that the model can reliably infer private or **sensitive** information about people. This also means that accuracy information might itself be disclosive and might need to be subject to **disclosure control**.

Further reading:

Article 29 Data Protection Working Party, 2014. *Opinion 05/2014 on anonymisation techniques*. Available from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

Yin, M., Wortman, V.J. and Wallach, H., 2019. Understanding the effect of accuracy on trust in machine learning models. *In*: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300509.

See also: INFERENCE

ACL

See: ACCESS CONTROL LIST

Additivity

Additivity is the property of a set of summary statistics which respect the arithmetic relationships implicit in their construction. The standard example is in tables of counts where row and column totals should equal the sum of the cells in those rows and columns.

Some perturbative **disclosure control methods** such as **random rounding** may give results that are non-additive. Non-additivity may be seen as problematic by **user**s; it may also reveal **information** about the pre-disclosure-controlled data, as the plausible real values may be constrained by the intersection of non-additive values.

Further reading:

Shlomo, N., 2007. Statistical disclosure control methods for census frequency tables. *International Statistical Review*, 75(2), 199–217, https://doi.org/10.1111/j.1751-5823.2007.00010.x.

See also: PERTURBATION, TABULAR DATA

Adequacy

The European Union refers to the adequacy of countries outside its jurisdiction, meaning the sufficiency of that country's safeguards to protect **personal data**. This encompasses the robustness of that jurisdiction's privacy laws, and the extent to which these provisions can be overruled for other purposes. For example, opaque and wide-ranging exemptions for law enforcement purposes can be seen as posing a risk to EU **data subject**s which is incompatible with the rights and freedoms they should expect when their **personal information** is used by others.

An adequacy decision is thus a formal, legally binding statement from the European Commission that personal data can be shared with a third country without the need for further safeguards, such as **standard contractual clauses** or **binding corporate rules**. Although the criteria the Commission should consider are set out in the **GDPR**, the process leading up to an adequacy decision is not prescribed. This procedural freedom has led some commentators to criticise the opacity and inconsistency of the commission's decision-making. Its adequacy decisions with respect to the United States have been successfully challenged (twice) in the **Schrems** litigation.

Further reading:

Stoddart, J., Chan, B. and Joly, Y. 2016. The European Union's adequacy approach to privacy and international data sharing in health research. *The Journal of Law, Medicine & Ethics*, 44(1), 143–55. https://doi.org/10.1177/1073110516644205.

See also: CROSS-BORDER DATA PROCESSING

Ad Exchange

An ad exchange is a platform where publishers can auction slots for online advertisements, for which advertisers can bid, often facilitated by an **ad network**. For instance, after a **user** clicks on a link, the publisher may set up an auction for a banner on the webpage, using **data** about the user to enable effective **targeted advertising**. The auction will be concluded in the short period while the page is being rendered for the user, and the advert will appear instantly.

See also: BEHAVIOURAL ADVERTISING

ADF

See: ANONYMISATION DECISION-MAKING FRAMEWORK

Ad Hoc Network

In an ad hoc network, devices communicate directly with each other with no need for a centralised infrastructure (e.g., an **access point**). Ad hoc networks are wireless – with devices establishing connections to one another dynamically – and after establishment are self-organised and self-configuring. Ad hoc networking is typically used in situations where there is no **network** infrastructure available, for example in rural communities or during military operations.

Ad hoc networks can enhance privacy because they can be used for peer-to-peer (P2P) communication (without the need for a central server), enhancing decentralisation, **end-to-end encryption** and **anonymity**. However, they can still have **security** vulnerabilities: since **data** is shared directly between peers, there is a risk that content can be altered by malicious nodes.

Further reading:
Ramanathan, R. and Redi, J., 2002. A brief overview of ad hoc networks: challenges and directions. *IEEE Communications Magazine*, 40, 20–2. https://doi.org/10.1109/MCOM.2002.1006968.

Ad Network

A service, typically run by a **third-party** provider, that connects advertisers with websites that have advertising space. The ad network allows advertisers to reach larger audiences by placing their ads on multiple sites at the same time. Ad networks can offer many options to advertisers, such as display, video and native ads. Publishers usually provide advertisers with the means for **targeted advertising**.

Ad networks provide a revenue stream for publishers. Advertisers usually pay the third party per click, that is, every time **user**s click on their ad.

To deliver the ad to the right audience, ad networks collect a large amount of user data, sometimes without the user's **consent**, enabling **tracking** and **profiling**. Some ad networks have also been vehicles for the distribution of **malware** by malicious advertisers.

Further reading:
Hannak, A., Sapiezynski, P., Molavi, K.A., Krishnamurthy, B., Lazer, D., Mislove, A., Wilson, C., 2013. Measuring personalization of web search. *In*: *Proceedings of the 22nd International Conference on World Wide Web*, 527–38, https://doi.org/10.1145/2488388.2488435.

Li, Z., Zhang, K., Xie, Y., Yu, F. and Wang, X., 2012. Knowing your enemy: understanding and detecting malicious web advertising. *In*: *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, 674–86, https://doi.org/10.1145/2382196.2382267.

See also: BEHAVIOURAL ADVERTISING

Adtech

A collective term for a diverse set of **software** systems which assist advertisers and publishers in management of advertising streams. Such systems can and do process **personal data**, particularly to create profiles to be used in real time bidding systems.

Further reading:
ICO, 2019. *Update report into adtech and real time bidding.* Available from: https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906-dl191220.pdf.
Veale, M. and Borgesius, F.Z., 2022. Adtech and real-time bidding under European data protection law. *German Law Journal*, 23(2), 226–56, https://doi.org/10.1017/glj.2022.18.

See also: BEHAVIOURAL ADVERTISING, TARGETED ADVERTISING, PROFILING

Advanced Electronic Signature

See: DIGITAL SIGNATURE

Advanced Encryption Standard (AES)

The Advanced Encryption Standard (AES), invented by Joan Daemen and Vincent Rijmen, is the most widely used **encryption algorithm** in the world. AES’s predecessor was the Data Encryption Standard (DES). Unlike DES, AES uses a fixed block size of 128 bits and supports key lengths of 128, 192 and 256 bits. With AES, it is very difficult for an adversary to recover the **plaintext** from the **ciphertext** without knowing the **encryption key**. However, it is still vulnerable to side-channel attacks that can exploit **information** leakage from a cryptographic system through variations in power consumption or timing to obtain the **cryptographic key**.

Further reading:
Heron, S., 2009. Advanced Encryption Standard (AES). *Network Security,* 12, 8–12, https://doi.org/10.1016/S1353-4858(10)70006-4.

See also: SYMMETRIC KEY ENCRYPTION, TRANSPORT LAYER SECURITY

Adversary

In general, the term is used to refer to an entity that deliberately attempts to breach security, confidentiality and/or privacy.

In scenarios describing a **breach** of **informational privacy**, the adversary is conceived as a person who desires to gain access to **information** to which they are not entitled, whose motives are malicious. Scenarios focus on the skills and computational resources the adversary would need to compromise a system. Also called a **motivated intruder**, opponent, enemy, snooper or attacker.

Further reading:
Information Commissioner's Office, 2012. *Anonymisation*: *managing data protection risk code of practice*. https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

See also: INTRUDER

Adware

Adware is **software** that creates advertisements automatically to be presented to **user**s as part of their online experience. Adverts may appear at a point on a webpage, in a box or banner, or may open a new window (a 'pop up'). Particularly intrusive ones may take over the full screen. Some definitions also add that the adware should install itself without the user's knowledge or permission, and so is a kind of **malware**. Although ad-blockers or pop-up blockers exist, some adware will attempt to evade or dismantle such programs.

Some adware also gathers **information** about users' **browsing history**, enabling **profiling** for **targeted advertising** and the use or sale of **personal data**.

Further reading:
Aycock, J., 2011. *Spyware and adware*. New York: Springer, https://doi.org/10.1007/978-0-387-77741-2.

See also: BEHAVIOURAL ADVERTISING, CONTEXTUAL ADVERTISING, SPYWARE

AES

See: ADVANCED ENCRYPTION STANDARD

Affinity Analysis

A set of **data mining** techniques for associating items according to their co-presence within sets of items. It is most heavily used in retail to analyse transactions to determine when things are likely to be bought together. This in turn underpins online **recommendation systems** and physical layout in shops and supermarkets. This application arguably impacts **decisional privacy**.

Aggregation

Any grouping of **data** which reduces its granularity.

See: ABSTRACT, GLOBAL RECODING

Agreement

According to Elliot et al., agreement is one of the two core concepts, the other being **awareness,** that make it easier to pragmatically understand higher order concepts such **informed consent** and **transparency**. **Data subject**s can be aware of **processing** of their data without having agreed to it or vice versa.

The term is also used more formally to refer to the contractual underpinnings for data access arrangements such **user agreement**s and **data sharing agreement**s.

Further reading:
Elliot, M., Mackey, E. and O'Hara, K., 2020. The Anonymisation Decision-Making Framework: European practitioners' guide, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

AI

See: ARTIFICIAL INTELLIGENCE

Algorithm

An algorithm is a procedure, a finite sequence of fully specified instructions that takes an input and produces an output (to perform a particular task).

Further reading:
Hill, R.K., 2016. What an algorithm is. *Philosophy & Technology,* 29, 35–59, https://doi.org/10.1145/2093548.2093549.

See also: DATA MINING, MACHINE LEARNING

American Data Privacy and Protection Act

See: US PRIVACY LAWS

Analogue Hole

The analogue hole is a concept rooted in **digital rights management**, as a weakness in **end-to-end encryption**. Even an encrypted message must be read, watched or listened to, and so must be decrypted for consumption. An **adversary** may simply target the endpoint; for example, an encrypted movie might simply be recorded off the screen when it is watched.

But this might also be a **privacy** matter, where an encrypted message is copied at the point at which it is read. For instance, the Pegasus **spyware** system works around the end-to-end encryption of messaging systems such as WhatsApp or Signal by reading the message directly off the smartphone screen.

Further reading:

Chawla, A., 2021. Pegasus spyware – 'a privacy killer'. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3890657.

Sicker, D.C., Ohm, P. and Gunaji, S., 2006. The analog hole and the price of music: an empirical study. *Journal on Telecommunications & High Tech Law*, 5, 573–87. Available from: www.jthtl.org/content/articles/V5I3/JTHTLv5i3_SickerOhmGunaji.PDF.

See also: COMMUNICATION PRIVACY

Analysis Server

See: REMOTE ANALYSIS SERVER

Analytical Completeness

The capacity of a **dataset** that has been subjected to **suppression** to support the same analysis as an untreated version of the data. Some **disclosure control methods**, particularly ones that **aggregate** categories, mean that analyses that might have been conducted with the untreated data can no longer be carried out. An example is the use of geographical aggregation with smaller areas being merged, preventing analysis of small administrative units using the dataset without analytical compromises.

Further reading:

Purdam, K. and Elliot, M. (2007). A case study of the impact of statistical disclosure control on data quality in the individual UK samples of anonymised records. *Environment and Planning A*, 39(5), 1101–18, https://doi.org/10.1068/a38335.

See also: ANALYTICAL VALIDITY, GLOBAL RECODING, STATISTICAL DISCLOSURE CONTROL

Analytical Validity

The capacity of a **dataset** to lead to the same **inference**s being drawn before and after treatment using **disclosure control methods**, when the same analysis is conducted.

A companion to **analytical completeness**, loss of analytical validity is more critical because of its insidious nature. Technically, loss of validity can be said to occur when a disclosure control method has changed a dataset to the point where a **user** reaches a different conclusion from the same analysis. This typically happens with **perturbative** disclosure control techniques such as **microaggregation**, **local suppression**, **post randomisation** or **noise addition**. Concerns about the utility impacts of **differential privacy** are essentially concerns about analytical validity.

Further reading:
Purdam, K. and Elliot, M. (2007). A case study of the impact of statistical disclosure control on data quality in the individual UK samples of anonymised records. *Environment and Planning A*, 39(5), 1101–18, https://doi.org/10.1068/a38335.

See also: STATISTICAL DISCLOSURE CONTROL

Anomaly Detection

The practice of identifying an unusual or unexpected signal in data. Anomaly detection has numerous uses such as fault diagnosis, fraud and **intrusion** detection.

In a privacy context, anomaly detection is an approach to targeted surveillance which enables machine learning models to analyse video feeds in real time. Anomaly detection techniques could also be used by adversaries looking for unusual patterns as the basis of a **fishing attack**.

Further reading:
Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: a survey. *ACM Computing Surveys (CSUR)*, *41*(3), 1–58, https://doi.org/10.1145/1541880.1541882.

See also: SURVEILLANCE, INTRUSION DETECTION SYSTEM, SPECIAL UNIQUE

Anonymisation

Anonymisation is the practice of transforming **identifiable data** into non-identifiable or non-informative data, most commonly by reducing the amount of **information** that either is present in the **dataset** or inferable from it. In practice, **identifier**s should be removed, altered, **aggregated** or otherwise obscured. Note that identification does not just mean that a person can be named; it also covers the case where **information** about a person can be attached with certainty to them. Anonymisation is also known in some jurisdictions as **de-identification**.

Under the **GDPR**, fully anonymised data (called anonymous **information**) is not treated as **personal data**, because individuals cannot be identified in it. However, the expansive definition of 'identifiable' in GDPR means that anonymisation, under its definition, must be irreversible. This requires, for instance, that the original, identifiable dataset should be **deleted**. This gives the term 'anonymisation' an absolute meaning.

A more intuitive definition of anonymisation focuses on the risk of **reidentification** of persons represented within the anonymised dataset. This means that the anonymisation process is intended, not to make anonymisation irreversible, but rather to lower the **risk** of reidentification (or de-anonymisation) to acceptable (e.g., negligible) levels. Under this conception, anonymisation is a risk management process.

It can be shown mathematically that anonymisation in this latter sense is reversible because an **adversary** could use **auxiliary knowledge** to identify **data subject**s in the anonymised dataset, and it can never be known in advance what data is available to the adversary. Several real-world scandals have occurred in which inadequately anonymised datasets have been released, and reidentification has taken place (for example, with AOL in 2006, Netflix in 2007 and data about New York yellow cabs released as part of a **Freedom of Information** request in 2014). These have been argued (for instance by Paul Ohm) to show that anonymisation is not an adequate defence of privacy, although in each case reidentification was possible principally because the anonymisation was inadequately planned and executed.

A more sophisticated approach called **functional anonymisation** considers that the risk lies in the relationship between data and the **data environment**. This opens up the possibility that, rather than altering the data, the data environment may be controlled, so that the ability of outsiders to interrogate the dataset is limited, for example by **access control**s or restrictions on linking to auxiliary datasets.

Further reading:

Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework*: *European practitioners' guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

Elliot, M., O'Hara, K., Raab, C., O'Keeffe, C.M., Mackey, E., Dibben, C., Gowans, H., Purdam, K. and McCullagh, K., 2018. Functional anonymisation: personal data and the data environment, *Computer Law and Security Review*, 34(2), 204–21, https://doi.org/10.1016/j.clsr.2018.02.001.

Hintze, M. and El Emam, K., 2018. Comparing the benefits of pseudonymisation and anonymisation under the GDPR. *Journal of Data Protection and Privacy*, 2(2), 145–58.

Ohm, P., 2010. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701–77, https://heinonline.org/HOL/LandingPage?handle=hein.journals/uclalr57&div=48&id=&page=.

See also: INFERENCE, PSEUDONYMISATION

Anonymisation Decision-Making Framework (ADF)

The Anonymisation Decision-Making Framework (ADF) is a practical guide to **anonymisation** intended to provide operational support to **data controller**s anonymising **personal data** (or other **sensitive data**), which is consistent with **code**s **of conduct** such as the UK Information Commissioner's *Anonymisation Code of Practice*. The ADF was originally written by researchers from the UK Anonymisation Network (UKAN), a network of **data custodian**s, and exists in three forms: a version consistent with the EU **Data Protection Directive** in 2016, a version consistent with **GDPR** in 2020, and a version (the De-Identification Decision-Making Framework, DDF), adapted by researchers at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for Australian law.

The ADF is intended to implement **functional anonymisation**, which manages the **risk** of **reidentification** of anonymised data with measures specific to the context in which it is held. It consists of ten components, divided into a **data situation audit** (to frame the relevant data context), **risk assessment** and control, and **impact management** (to consider measures to ensure risk remains negligible going forward, as well as to plan for **security breach**es).

Further reading:
Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework*: *European practitioners' guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

Anonymising Proxy

See: PROXY

Anonymity

The word 'anonymity' is sometimes used to mean simply 'not named', such as when an individual's name is not published in media reports. However, authors such as Nissenbaum have argued that the term denotes a more precisely calibrated state of non-**identifiability**. While the term is not explicitly defined in law, a more rigorous definition would be a state in which an individual is no longer identifiable from formerly **personal data**, at least within a particular defined context.

An example of the term within EU law is the concept of 'donor anonymity', which is assured by tissue and cell donation regulations. The donor must not be identifiable from any means reasonably likely to be used, but this cannot mean 'non-identifiable by anyone'. Clearly, the clinicians facilitating the donation will know the identity of the donor, and thus 'anonymity' in its legal sense should not be understood in absolute terms.

Further reading:

Mourby, M., 2020. Anonymity in EU healthcare law: not an alternative to information governance. *Medical Law Review*, 28(3), 478–501. https://doi.org/10.1093/medlaw/fwaa010.

Nissenbaum, H., 1999. The meaning of anonymity in an information age. *The Information Society*, 15(2), 141–4. https://doi.org/10.1080/019722499128592.

See also: ANONYMISATION, FUNCTIONAL ANONYMISATION, GDPR

Anonymous Search Engine

See: SEARCH

Anti-Discrimination Law

See: NON-DISCRIMINATION LAW

Anti-Malware Software

Anti-malware **software** is designed to protect **user**s from a broad range of malicious software or **malware**. Anti-malware can be installed on mobile devices, standalone computers or on whole networks. Similar to **anti-virus software**, anti-malware protects against computer **virus**es, **worm**s and **trojan horse**s, but may also protect against other threats such as **spyware**, **adware**, and **ransomware** attacks. Depending on the design, it may use a combination of deep packet inspection, signature detection, behaviour fingerprinting and/or **sandbox**ing to identify and prevent **security** attacks.

Further reading:

Rieck, K., Holz, T., Willems, C., Dussel, P. and Laskov, P., 2008. Learning and classification of malware behavior. *In*: Zamboni, D. ed. *Detection of intrusions and malware, and vulnerability assessment*, 108–25. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-70542-0_6.

Talal, M., Zaidan, A.A., Zaidan, B.B., Albahri, O.S., Alsalem, M.A., Albahri, A.S., Alamoodi, A.H., Kiah, M.L.M., Jumaah, F.M. and Alaa, M., 2019. Comprehensive review and analysis of anti-malware apps for smartphones. *Telecommunication Systems*, 72, 285–337. https://doi.org/10.1007/s11235-019-00575-7.

Anti-Virus Software

Anti-virus **software** helps to detect, prevent, and remove **virus**es from a computer. Typically, anti-virus software is also able to detect **worms**, **trojan horse**s and some **spyware**. The primary purpose of anti-virus software is to prevent unauthorised access to private resources, providing **security** against these types of threats.

Since anti-virus software necessarily monitors the **user**'s system and online activity and may open **backdoor**s that allow for **remote access**, it is important to read the user **agreement** of the software carefully before installing it. It is also essential to use a provider with a good reputation and proven track record.

Further reading:

Rieck, K., Holz, T., Willems, C., Dussel, P. and Laskov, P., 2008. Learning and classification of malware behavior. *In*: Zamboni, D. ed. *Detection of intrusions and malware, and vulnerability assessment*, 108–25. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-70542-0_6.

See also: MALWARE

APEC Privacy Principles

In 2004, the Asia-Pacific Economic Cooperation, a forum for nations on the Pacific Rim, published a voluntary privacy framework intended to improve **informational privacy** standards in Asia. Its nine principles were: preventing **harm**; **notice**; collection limitation; use of **personal information**; choice; **integrity** of personal **information**; **security** safeguards; access and correction; and **accountability**. The APEC Principles have been criticised on the grounds that they are voluntary, not geared towards EU compatibility and largely based on **OECD Guidelines** from the 1970s, with little modernisation.

Further reading:
Burri, M., 2021. Interfacing privacy and trade. *Case Western Reserve Journal of International Law*, 53, 35–85.
Greenleaf, G., 2009. Five years of the APEC Privacy Framework: failure or promise? *Computer Law and Security Review*, 25(1), 28–43, https://doi.org/10.1016/j.clsr.2008.12.002.

See also: SAFE HARBOR

API

See: APPLICATION PROGRAMMING INTERFACE

APP

See: APPLICATION

Application (App)

An application or app is a program, usually provided by a **user** interface, designed to perform a specific task on a device (smartphone, laptop, tablet). Apps are widely used for various purposes, including communication, entertainment and productivity. They are becoming increasingly central to the operation of the **digital economy**.

However, the app market is not regulated, and so precaution is advised before downloading and installing any new app. Privacy concerns include data collection; apps may need to collect and store **personal information** on the user, such as **IP address**, **location** and usage habits, to deliver their function. Others may collect such data as a hidden payment for delivering a free service (which the user has signed up to by agreeing to the **terms of service**) and still others may do so without any form of **consent** at all. Such apps may also contain third-party **tracker**s to collect **information** about users' behaviour. Additionally, apps are not guaranteed to have strong **security** protection in place (and some do use weakly encrypted connections or are vulnerable to interception). Some apps can be malicious, containing malware or other malicious **software** by design, or they may do so without the intent of the app designer.

Further reading:

Feal, A., Calciati, P., Vallina-Rodriguez, N., Troncoso, C. and Gorla, A., 2020. Angel or devil? A privacy study of mobile parental control apps. *In*: *Proceedings on Privacy Enhancing Technologies,* 2, 314–35. https://doi.org/10.2478/popets-2020-0029.

Ren, J., Rao, A., Lindorfer, M., Legout, A. and Choffnes, D., 2016. ReCon: revealing and controlling PII leaks in mobile network traffic. *In*: *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, New York: ACM, 361–74. https://doi.org/10.1145/2906388.2906392.

Application Layer Attack

An application layer attack is a kind of attack that involves the network application layer. Examples of application layer attacks include **Distributed Denial of Service** (DDoS) attacks, HTTP floods, **SQL injection**s, **cross-site scripting** and parameter tampering.

To prevent these attacks, most organisations have a number of application-level **security** protections in place, such as web application **firewall**s (WAFs), **secure web gateway** services and other protective mechanisms.

Further reading:

AbdAllah, E.G., Hassanein, H.S. and Zulkernine, M., 2015. A survey of security attacks in information-centric networking. *IEEE Communications Surveys & Tutorials*, 17, 1441–54. https://doi.org/10.1109/COMST.2015.2392629.

Cao, L., Jiang, X., Zhao, Y., Wang, S., You, D. and Xu, X., 2020. A survey of network attacks on cyber-physical systems. *IEEE Access*, 8, 44219–27. https://doi.org/10.1109/ACCESS.2020.2977423.

See also: DENIAL OF SERVICE, HYPERTEXT TRANSFER PROTOCOL SECURE

Application Programming Interface (API)

A set of protocols and tools for building **software** and **application**s. An API defines the way that different software components should interact, allowing for **communication** and data exchange between them. APIs provide a standard way for applications to request services from a system or software library, making it easier for developers to build new products and services using existing functionality.

Some potential privacy risks related to APIs are due to the fact that they allow data exchange between applications. Through APIs it is possible to collect **browsing history** and other sensitive information.

Further reading:

Bloch, J., 2006. How to design a good API and why it matters. *In*: *Companion to the 21st ACM SIGPLAN Symposium on Object-Oriented Programming Systems, Languages, and Applications*, New York: ACM, 506–7. https://doi.org/10.1145/1176617.1176622.

Sharon, T., 2021. Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, 23, 45–57. https://doi.org/10.1007/s10676-020-09547-x.

Appropriate Safeguards

The GDPR uses the term '**appropriate technical and organisational measures**' in several contexts. It is commonly shortened to 'appropriate safeguards'.

Appropriate Technical and Organisational Measures

Appropriate technical and organisational measures are generally required of **data controller**s to protect **personal data**. The phrase highlights the two key elements of **data protection** practice: appropriate use of technology (given the state of the art, resources and **risk**s involved), as well as consideration of human behaviour. For example, this covers not only the **information security software** used, but also who is **trusted** with access to the data. If a controller uses the services of a **data processor**, the latter should also provide written assurance of appropriate technical and organisational measures to protect **data**.

ISO standards **ISO27001** and **ISO27002** have become an important touchstone for many organisations wishing to demonstrate implementation of appropriate safeguards.

Further reading:

Calder, A. 2020. *EU GDPR – an international guide to compliance*. Ely: IT Governance. Available from: www.itgovernancepublishing.co.uk/product/eu-gdpr-an-international-guide-to-compliance.

See also: ACCOUNTABILITY, DATA-PROTECTION-BY-DESIGN, DATA PROTECTION POLICY

Appropriation of Name or Likeness

Appropriation is the fourth of William Prosser's four **privacy tort**s. In an influential paper of 1960, Prosser argued against the salience of the **right to be let alone**, traced in American law by Warren and Brandeis. He claimed instead that the privacy torts that actually existed in law did not furnish a broader principle of integrated coverage of a right to be let alone but were instead a set of four discrete and discontinuous protections.

Appropriation involves the defendant using the plaintiff's name or likeness, or other aspect of their identity, for his or her (commercial or other) advantage without consent. Examples would be using someone's image in an advertisement, or their name as a personal endorsement during a job interview. Defences include that the defendant was illustrating a news item or commentary (without placing the plaintiff in a **false light**), or that the likeness was altered for artistic reasons, resulting in a creative work.

Further reading:
Prosser, W.L., 1960. Privacy. *California Law Review*, 48, 383–423.
Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220.

See also: CELEBRITY PRIVACY, IDENTITY THEFT

AR

See: AUGMENTED REALITY

Article 29 Working Party (A29WP)

The Article 29 Working Party was the pan-EU body responsible for coordinating **data protection** guidance and policy at the European level. It consisted of representatives from the national supervisory authorities (i.e., the data protection regulators in each member state), as well as representatives of the EU bodies and the European Commission. Although it had secretarial support from the Commission, its guidance was devised independently and did not officially represent the position of the European Commission.

The A29WP, launched in 1996, ceased to exist as of 25 May 2018, when the **GDPR** came into force. It has now been replaced by the (very similar) **European Data Protection Board** (EDPB). The guidance, or 'Opinions', of the A29WP remains influential, and an important means of coordinating interpretation of data protection law across the EU. Many of these Opinions have been publicly endorsed by the EDPB, although the Opinions of the A29WP are not legally binding.

Further reading:
European Commission, 2023. *Article 29 Working Party*. Available from: https://ec.europa.eu/newsroom/article29/items/itemType/1358.

Artificial Intelligence (AI)

The term artificial intelligence refers to computer systems that can perform tasks that ordinarily require natural (usually human) intelligence, such as speech recognition, decision-making, visual perception and language translation. AI systems can be programmed to learn from data inputs and adapt, which allows them to perform better over time.

One privacy **risk** with AI is the possibility of **sensitive** or **personal data** being misused or exposed by AI systems. Large **dataset**s, including personal data such as names, addresses and other identifying **information**, are frequently collected, and processed by AI systems. If these data are not adequately protected, they may be susceptible to theft, **hacking** or unauthorised access. Additionally, AI systems may be used to make decisions that may affect people's privacy, such as when hiring people or using **automated decision-making** processes in other delicate circumstances.

Recent developments in AI, such as **deep learning** and foundational **large language model**s, are regularly producing output that challenges human ideas of what machines are capable of. Their language comprehension and production capabilities far exceed that of earlier generations of AI methods, while their ability to craft images and **deepfake**s concerns many. Most important for privacy, their data assimilation and pattern recognition are sufficiently powerful for very weak signals to be detected in noisy data (sometimes beyond the capabilities of human, social or organisational processes), so that AI techniques now enable far more information to be extracted from a given dataset, opening more possibilities of unanticipated privacy **breach**es.

As AI increases in capacity a new privacy angle emerges. If AI becomes sentient, then it will enter our moral universe as an agent (not just as a tool). We therefore may and probably will become concerned about what we reveal to an AI system, and not just in terms of how the outputs of that system might be used by another human.

Further reading:

Hagendorff, T., 2020. The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines,* 30, 99–120, https://doi.org/10.1007/s11023-020-09517-8.

Spindler, G., 2021. Algorithms, credit scoring, and the new proposals of the EU for an AI Act and on a Consumer Credit Directive. *Law and Financial Markets Review*, 15(3), 239–61, https://doi.org/10.1080/17521440.2023.2168940.

See also: DEEPFAKE, DEEP LEARNING, EXPLAINABLE AI, MACHINE LEARNING, RECOMMENDATION SYSTEM

AS

See: AUTONOMOUS SYSTEM

Asset

An asset is property that has value, and may be used to meet commitments, pay debts or generate income streams. *Asset privacy* is the practice of concealing ownership of assets, for example by transferring ownership to a company whose ownership is itself opaque, while retaining control. Asset privacy has been seen as a major facilitator of corruption and money laundering, and so many countries are seeking to promote **transparency** with *registers of beneficial ownership*, which make clear which human individuals are the ultimate beneficiaries of a business' activity.

Data, as a source of future value, can also be seen and managed as an asset. In that case, asset management techniques will seek to manage **risk**, facilitate data **security** and ensure **data quality**.

Further reading:

Berkhout, R. and Fernando, F., 2022. *Unmasking control*: *a guide to beneficial ownership transparency*. Washington DC: International Monetary Fund. Available from: www.elibrary.imf.org/display/book/9798400208041/9798400208041.xml?cid=web-com-TBOIGPEA.

Leonelli, S., 2019. Data – from objects to assets. *Nature*, 574, 317–20, https://doi.org/10.1038/d41586-019-03062-w.

See also: DATA STEWARD, FINANCIAL PRIVACY, PRIVATE PROPERTY

Associational Privacy

Associational privacy is the ability of individuals to form and join the groups they wish to, combined with the right of the group to withhold **information** about the individuals within it from outsiders, including the state, thereby restricting the ability of outsiders to act against it. Associational privacy is a keystone of law in many nations and was a key factor in the US civil rights movement, protecting organisations such as the National Association for the Advancement of Colored People from aggressive policing in some Southern US states.

However, associational privacy may clash with **non-discrimination law** or equality principles, as for example with a club that wished to bar ethnic minority members, or a women's sports event that wished to exclude trans women.

Further reading:

Allen, A.L., 2011. Associational privacy and the First Amendment: NAACP v. Alabama, privacy and data protection. *Alabama Civil Rights and Civil Liberties Law Review*, 1(1), 1–13.

See also: GROUP PRIVACY

Assured Data Deletion

See: ERASURE

Asymmetric Cryptography

Asymmetric cryptography, first proposed by Diffie and Hellman in the 1970s, is used to guarantee **confidential** and **secure communication** online. It uses two different **cryptographic key**s, one for **encryption** and another for **decryption**. Each party has a pair of keys (a **private key** and a **public key**); the public key can be shared with anyone, while the private key is kept secret. If a sender sends a message encrypted with the recipient's public key the recipient can then decrypt it with their corresponding private key. On the other hand, if a sender encrypts a message with their own private key, it can be decrypted using the sender's public key. This later use case increases verifiability and therefore the trustworthiness of the sender's communications.

This differs from **symmetric key encryption**, which uses the same key for both purposes. Asymmetric cryptography does not require the exchange of secret keys in advance, which considerably reduces interception risks. However, to achieve the same level of **security** an asymmetric key needs to be an order of magnitude larger (in terms of bit sizes) and is considerably slower to process.

Asymmetric cryptography is commonly used in secure Internet protocols, such as **Transport Layer Security (TLS)**, **digital signature**s and encrypted email.

Further reading:

Diffie, W. and Hellman, M., 1976. New directions in cryptography. *IEEE Transactions on information theory*, *22*(6), 644-654, https://doi.org/10.1145/3549993.3550007.

Kessler, G.C., 2003. *An overview of cryptography*. www.garykessler.net/library/crypto.html.

See also: CRYPTOGRAPHY, ENCRYPTION KEY

Asymmetric Information

An **information** imbalance between two parties in a transaction. The concept was developed within economics (primarily considering imbalances between buyers and sellers) but has been more recently applied to privacy.

For instance, a website may gather **user**s' **personal data** without disclosing to them how the information will be used or who will have access to it. As a result, there may be an imbalance of power between the user and the data collector, with the latter having greater control over how the user's personal data is used and shared.

Another example is the power obtained by viewing somebody's social media profile before meeting them in person and then using the information obtained to manipulate the interaction.

Further reading:

Hancock, J.T., Toma, C.L. and Fenner, K., 2008. I know something you don't: the use of asymmetric personal information for interpersonal advantage. *In: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 413–16, https://doi.org/10.1145/1460563.1460629.

ATO

See: ACCOUNT TAKE OVER

Attack

A deliberate attempt to breach privacy or security.

See also: APPLICATION LAYER ATTACK, ATTACK SURFACE, ATTACK TREE, BICYCLE ATTACK, BLACK HAT ATTACK, BRUTE FORCE ATTACK, BUFFER OVERFLOW ATTACK, DEMONSTRATION ATTACK, DENIAL OF SERVICE, DIRECT ACCESS ATTACK, EAVESDROPPING ATTACK, FISHING ATTACK, GREY HAT ATTACK, INFERENCE ATTACK, INVERSION ATTACK, LINKAGE ATTACK, MAN-IN-THE-MIDDLE ATTACK, MASH ATTACK, MEMBERSHIP INFERENCE ATTACK, MODEL INVERSION ATTACK, MULTI-VECTOR ATTACKS, NETWORK LAYER ATTACK, OFFLINE DICTIONARY ATTACK, ORWELL ATTACK, PAPARAZZI ATTACK, POISONING ATTACK, RECONSTRUCTION ATTACK, REIDENTIFICATION ATTACK, REPLAY ATTACK, REVERSE FISHING ATTACK, SPOOFING ATTACK, SUBTRACTION ATTACK, SURNAME ATTACK, WEB SKIMMING ATTACK, WHITE HAT ATTACK, ZERO DAY ATTACK

Attacker

See: ADVERSARY

Attack Model

A framework for understanding how privacy or security of a system might be breached.

See also: ATTACK VECTOR, SCENARIO ANALYSIS, THREAT MODEL

Attack Surface

The attack surface is the digital surface area of a system or organisation that is exposed to potential cyber-attacks. In other words, it refers to the set of entry ports, touchpoints and **vulnerabilities** that can be exploited by an **adversary** to enter a system and cause damage to it. The larger the attack surface, the higher the **risk** of a successful attack. For this reason, it is important to minimise the attack surface of a system by eliminating vulnerabilities and limiting access only to those who need it.

Further reading:
Manadhata, P.K. and Wing, J.M., 2010. An attack surface metric. *IEEE Transactions on Software Engineering*, 37(3), 371–86, https://doi.org/10.1109/TSE.2010.60.

See also: VULNERABILITY MANAGEMENT

Attack Tree

By identifying the various steps an **adversary** could take to compromise a system, an attack tree is used to analyse and assess the **security** of a system. The initial attack goal is at the root with potential attack paths branching out from there. The branches correspond to the different ways an adversary could complete each node, representing a particular attack step. The attack scenarios or techniques that an adversary might employ are represented by the tree’s leaves. An attack tree is often used to evaluate a system’s security.

Further reading:
Manadhata, P.K. and Wing, J.M., 2010. An attack surface metric. *IEEE Transactions on Software Engineering*, 37(3), 371–86, https://doi.org/10.1109/TSE.2010.60.

See also: ATTACK SURFACE

Attack Vector

The means by which an **attack** takes place; this will refer to both the resources employed by the **adversary** (e.g., **auxiliary data**) and the route or pathway for the attack (e.g., a **phishing email**).

Attentional Privacy

Attentional privacy is the state of being shielded from being the subject of attention, which covers much of what Warren and Brandeis considered under the **right to be let alone**. O’Hara suggests that one has attentional privacy when: (i) one’s *behaviour* is not under **surveillance**, and particularly not being noted or recorded ready for recall and examination in the future; (ii) one's *appearance* is not subject to **scrutiny**; (iii) one is not *questioned*, held to **account** or interrogated; (iv) one's *speech* is free from eavesdropping, and by extension, *communications* are free from interception (**communication privacy**); (v) one is free from **publicity**; (vi) one is not the subject of *discussion*, speculation or **gossip** (whether true or false) by others. A breach of attentional privacy could be effected by direct perception, or via some technological means allowing asynchronous scrutiny of a photograph, recording or video.

Further reading:
O'Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Attention as a Resource

The *attention economy* is a set of economic relations where the attention of individuals is a scarce resource, and content providers compete to gain access to it. Attention is defined by Davenport and Beck as focused mental engagement with a particular item of **information**. **User** interfaces, thumbnails, headlines and menus are therefore designed to draw the viewer's eye to particular items, sometimes as a device to privilege sensational content. When these are misleading, they are known as *clickbait*. The attention economy can undermine privacy, either by creating incentives for content providers to find out about people – to customise the content they are offered – or by invading their privacy from outside with **spam** or other types of information pollution.

Further reading:
Davenport, T. and Beck, J.C., 2001. *Attention economy*: *understanding the new currency of business*. Cambridge MA: Harvard Business School Press.

See also: SURVEILLANCE CAPITALISM

Attention Tracking

Attention tracking is the practice of observing and noting where someone's attention is focused as they do a task, such as surfing a webpage. One of the most common forms of this is *eye tracking*, where movements of the eye are noted to determine what is of interest to the reader. Eye tracking can be used, for example, to test advertisements and optimise their attractiveness to the reader, or to record which shelves a customer's gaze falls upon, and which they ignore, as they walk through a retail outlet.

Further reading:

Wedel, M. and Pieters, R., 2008. A review of eye-tracking research in marketing. *Review of Marketing Research*, 4, 123–47, https://doi.org/10.1108/S1548-6435(2008)0000004009.

See also: CUSTOMER TRACKING

Attitude–Behaviour Gap

In human psychology, the attitude–behaviour gap is a perceived disconnection between someone's attitudes and their actions, or alternatively between their first-order (immediate) **privacy preferences** and their second-order preferences (preferences about which preferences they should hold). A specific instance of this is the **privacy paradox**, the claim that many people have a strong positive attitude about their privacy which is belied by their cavalier treatment of their **personal data**.

Further reading:

Acquisti, A., Brandimarte, L. and Loewenstein, G. 2015. Privacy and human behavior in the age of information. *Science*, 347(6221), 509–14, https://doi.org.10.1126/science.aaa1465.

Godin, G., Connor, M. and Sheeran, P., 2005. Bridging the intention–behaviour gap: the role of moral norm. *British Journal for Social Psychology*, 44(4), 497–512, https://doi.org/10.1348/014466604X17452.

Attribute

An attribute is a property of a type of object that is a component of its representation in data, or the value of that property for a specific instance of the type. For example, 'area' is an attribute of 'nation', and 'area of 30,000km^2' is an attribute of Belgium.

See also: FEATURE

Attribute Disclosure

A type of inference where an agent learns a piece of information about an entity from some data with or without **reidentification** of that individual in the data. For example, if a **dataset** tells me that all people living a certain neighbourhood earn less than a certain amount each year and I know that you live in that neighbourhood then I have learnt something about your income.

In principle, the **attribution** could be deterministic (if X is true then Y is always true as well) or it could be probabilistic (if X is true then Y is more likely to be true than if X is not). The latter is essentially the same as **inference** and in some parts of the literature a distinction is made between attribute disclosure (deterministic) and inferential disclosure (probabilistic).

Further reading:

Hittmeir, M., Mayer, R. and Ekelhart, A., 2020. A baseline for attribute disclosure risk in synthetic data. In *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, 133–43, https://doi.org/10.1145/3374664.3375722.

Rubinstein, I.S. and Hartzog, W., 2016. Anonymization and risk. *Washington Law Review*, 91, 703–60, https://digitalcommons.law.uw.edu/wlr/vol91/iss2/18/.

Smith, D. and Elliot, M., 2008. A measure of disclosure risk for tables of counts. *Transactions on Data Privacy*,1(1), 34–52, www.tdp.cat/issues/tdp.a003a08.pdf.

See also: ATTRIBUTE, INFERENCE ATTACK

Attribution

The process of associating a particular piece of data with a particular entity (person, household, business etc). Note that attribution can happen without **reidentification** (if for example all members of a group share a common **attribute**).

The **GDPR** only discusses attribution in the context of **pseudonymisation**. While **personal data** relate directly to an individual, pseudonymised data can only be ‘attributed’ to an individual with the use of auxiliary information (which must be held separately).

See also: ATTRIBUTE DISCLOSURE, IDENTIFIABLE DATA

Audit Trail

A **record** or set of records providing evidence of activities which impact (or may impact) a system, entity or process.

In an **information security** context, the term is employed to mean a record of system activities which enable the examination of **security** events.

The term is not often commonly used in legislation, but a carefully preserved system of documentation is often a requirement to demonstrate **compliance** in practice. The trail of filed **information** should allow an external auditor (for example, a **supervisory authority**) to re-trace the steps taken to protect **personal data**. For example, in the event of a **breach** of personal data, a trail of documents showing prompt **remedial** action and timely data **breach notification** of affected **data subject**s can be a mitigating factor when a regulator assesses a **data controller**'s responsibility for the lapse.

Further reading:

Buchanan, S. and Gibb, F., 2008. The information audit: theory versus practice. *International Journal of Information Management*, 28(3), 150–60, https://doi.org/10.1016/j.ijinfomgt.2007.09.003.

Calder, A. 2020. *EU GDPR – an international guide to compliance*. Ely: IT Governance. Available from: www.itgovernancepublishing.co.uk/product/eu-gdpr-an-international-guide-to-compliance.

See also: ACCOUNTABILITY

Augmented Reality (AR)

By overlaying computer-generated images and data on the physical environment, augmented reality (AR) technology enhances the **user**'s perception of the outside world. This can be accomplished with a variety of devices, including smartphones, tablets and specialised AR headsets.

AR poses some privacy risks, such as the collection of data. AR **application**s and devices frequently gather a variety of **data**, including **location**, images and audio, which could be used to identify people or follow their movements. Without the user's knowledge or permission, this data may be kept and used for marketing or other purposes. Physical **surveillance** is a concern because AR devices with cameras or **sensor**s could be used to **track** people's whereabouts and potentially violate their privacy rights.

Further reading:

Billingshurst, M., Clark, A. and Lee, G., 2015. A survey of augmented reality. *Foundations and Trends in Human–Computer Interaction*, 8(2–3), 73–272, https://doi.org/10.1561/1100000049.

See also: ARTIFICIAL INTELLIGENCE, MACHINE LEARNING

Authentication

Authentication is the process by which the **identity** of a **user** or entity attempting to access a system is verified. In other words, it is a process of confirming the identity of a user or entity to ensure that only **authorised** users can access protected resources.

There are several authentication methods, such as using a **username** and **password**, using **digital certificate**s, using **security token**s or **biometrics** such as fingerprints or voice recognition.

Further reading:

Burrows, M., Abadi, M. and Needham, R., 1990. A logic of authentication. *ACM Transactions on Computer Systems*, 8(1), 18–36, https://doi.org/10.1145/77648.77649.

Authorisation

Authorisation is a process of verifying that an agent can legitimately take some action, such as gaining access to a resource, editing a document, entering a building or making a payment. An administrative authority must determine whether there are sufficient grounds for authorising the action.

Authorisation has two related meanings. The first is that the administrative authority confers on an agent a set of privileges to access resources or take actions; the agent becomes authorised to use a system. The second meaning is that, when an authorised agent wishes to use the system directly, they present their **credentials**, and an authorisation process grants immediate access. As an example, in the first meaning, a customer will be authorised to use an online banking system. In the second meaning, the customer is authorised, perhaps via a **password**, **biometric data** or banking card, to perform some concrete action, such as withdrawing some money.

Authorisation is a key concept in **security**, and in preventing **hacking**.

Further reading:

De Capitani di Vimercati, S., Foresti, S., Samarati, P. and Jajodia, S., 2007. Access control policies and languages. *International Journal of Computational Science and Engineering*, 3(2), 94–102, https://doi.org/10.1504/IJCSE.2007.015739.

See also: ACCESS CONTROL, AUTHENTICATION, CERTIFICATION AUTHORITY

Automated Decision-Making

The use of computational systems to make decisions with or without human involvement. These systems will take **data** as an input and produce a decision or recommendation. Where automated decision making is embodied in an actuated system or device (such as a driverless car) then the decisions are translated directly into automatic actions.

The EU’s **GDPR** has specific provisions for solely automated decision-making. This has been defined by the **Article 29 Working Party** as ‘the ability to make decisions by technological means without human involvement’. Human involvement in the decision-making is sufficient to prevent the application of Article 22 GDPR rights to **information** about the logic of the automated processing involved. However, the guidance quoted emphasises that such involvement must be meaningful, from a person with sufficient expertise and authority to overrule the **algorithm** if necessary.

Automated decision-making is also notable for providing a context for the much-contested **right to an explanation** under the GDPR, which has opened wider debates about **explainable AI** and whether such explanations help individuals safeguard their rights, or whether a more systemic approach towards **accountability** is required.

Further reading:

Marabelli, M., Newell, S. and Handunge, V., 2021. The lifecycle of algorithmic decision-making systems: organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, 30(3), 101683, https://doi.org/10.1016/j.jsis.2021.101683.

Selbst, A.D. and Powles, J. 2017. Meaningful information and the right to explanation. *International Data Privacy Law*, 7(4), 233–42, https://par.nsf.gov/servlets/purl/10074338.

Kaminski, M.E., Malgieri, G.,2021. Algorithmic impact assessments under the GDPR: producing multi-layered explanations. *International Data Privacy Law*, 11(2), 125–44, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224.

See also: RIGHT TO PRIVACY, TRANSPARENCY, US PRIVACY LAWS

Autonomous System (AS)

An autonomous system is a collection of interconnected networks and devices that operate under a common administrative domain and share routing policies. The main goal of an AS is to ensure that the routing of traffic within the system is done in a consistent and efficient manner. The routers in an AS use a routing protocol called BGP to exchange routing **information** with other routers and determine the best path.

Devices under the same AS operate under the same administrative domain and have a common routing policy. This is often used in enterprise networks, service providers and in some cases, government networks.

Privacy implications of an AS are similar to those of the **Internet Protocol**. They can be used to monitor a **user**'s **Internet** activity. If an **adversary** gains access to the routing information within an AS, it can provide a detailed view of the network **traffic data** and provide insights into the users.

Additionally, routing information exchanged between ASs can reveal the relationships between different networks and devices, which could be used to infer information about the organisations or individuals that control them. This can be a serious concern in scenarios where an AS is being used to support a critical infrastructure or sensitive operations.

Further reading:
Tozal, M.E., 2016. The Internet: a system of interconnected autonomous systems. *In*: *2016 Annual IEEE Systems Conference (SysCon)*, 1–8, https://doi.org/10.1109/SYSCON.2016.7490628.

Autonomy

Autonomy is the capacity of an actor to make an informed and uncoerced decision and to act on that decision. Autonomous agents are pieces of intelligent **software** that serve the interests of a **user** without direct input from the user at the time of action.

The concept of autonomy arises in three places relating to privacy. First, **consent** (to the use of private **information** or **personal data**) is a key basis for **data processing** under many information laws. Its aim is to support the autonomy of **data subject**s by allowing them to decide whether and when information about them can be used. A common objection, made for example by Solomon Barocas and Helen Nissenbaum, is that consent does not confer autonomy because (a) it is rarely **informed consent**, as the uses of information are opaque, and (b) it is commonly coerced, for example by requiring consent to an intrusive **privacy policy** to receive a service. A reply, made by Daniel Solove among others, is that regulating how information may be used directly, overriding consent in the data subjects' interests, allows no input from data subjects at all, and so arguably renders them less rather than more autonomous.

A second discussion of autonomy criticises the idea of **decisional privacy** (the ability to make decisions without interference), with some, such as Judith Jarvis Thomson, claiming that the latter, rather than being any kind of privacy at all, is only another name for autonomy. Third, commentators such as Beate Rössler argue that privacy is a prerequisite for autonomy, so that individuals can defend a space in which they can make their uncoerced decisions.

Further reading:

Barocas, S. and Nissenbaum, H., 2009. On notice: the trouble with notice and consent. *In*: *Proceedings of the Engaging Data Forum*: *The First International Forum on the Application and Management of Personal Electronic Information*. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2567409.

Rössler, B., 2005. *The value of privacy*. Cambridge: Polity Press.

Solove, D.J., 2013. Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7), 1880–1903. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224.

Thomson, J.J., 1975. The right to privacy. *Philosophy and Public Affairs*, 4(4), 295–314.

See also: RELATIONAL AUTONOMY

Auxiliary Data

See: AUXILIARY KNOWLEDGE

Auxiliary Information

See: AUXILIARY KNOWLEDGE

Auxiliary Knowledge

When **direct identifier**s are removed or pseudonymised, **data subject**s are no longer **identifiable** using only the resources available in the dataset itself. However, if an **adversary** brings auxiliary knowledge to the dataset, then such knowledge can provide more context to allow identifications. For instance, suppose a spreadsheet of hospital admissions is formally anonymised. Auxiliary knowledge about a particular individual of interest (for example, their age, when they entered and left hospital, etc.) may allow an adversary to **single out** the medical record of that individual in the dataset. Because **data controller**s can never know in advance what auxiliary knowledge an intruder might possess, it follows that no perfect **anonymisation** process is possible.

A particular type of auxiliary **information** is the **key** to a **cipher**. Possession of the key enables the **decryption** of an encrypted message.

Further reading:

Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework*: *European practitioners' guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

Narayanan, A. and Shmatikov, V., 2010. Myths and fallacies of 'Personally Identifiable Information'. *Communications of the ACM*, 53(6), 24–6, https://doi.org/10.1145/1743546.1743558.

See also: RESPONSE KNOWLEDGE

AVAILABILITY

A common **information security** principle is that a person or organisation should be able to access the **personal data** for which they are responsible or **accountable**. The EU's **GDPR** refers to data **security** as the ability to ensure the 'ongoing **confidentiality**, **integrity**, availability and **resilience**' of **data processing** systems, as well as the need to secure the 'availability, authenticity, integrity and confidentiality' of personal data. This builds on established industry standards – often summarised in the **CIA Triad** model – in which availability is a key element of secure system design.

See also: DATA CONTROLLER

Awareness

According to Elliot et al., awareness is one of the two core concepts, the other being **agreement**, that make it easier to pragmatically understand higher order concepts such **informed consent** and **transparency**. **Data**

subjects can be aware of **processing** of their data without having agreed to it or vice versa.

Further reading:

Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners' guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

B

Backdoor

A backdoor is a low visibility **access point** to a system or **network**, which allows a **user** to bypass the standard **authentication** procedure. Backdoors might be benign in intent; for example, being created by system designers as safeguards against system failures that have caused the standard authentication system to malfunction. But they might also be malign in intent; for example, a classic **Trojan horse** programme could infiltrate a network to set up a backdoor allowing hackers to enter undetected. Certain types of **worm** software might install a backdoor to enable an **adversary** to use an infected machine to send **spam** email. Or hackers who have obtained entry through another form of attack could install a backdoor to maintain access even after the initial attack has been discovered and resolved.

As it is difficult to detect backdoors, the best solution may be prevention. It is important to update **software** on a regular basis, regularly monitor systems for unusual activity and install **anti-virus software** and other **intrusion prevention system**s. Backdoor attacks should also be included in the suite of **attack**s to be simulated by **red team**s.

Further reading:

AbdAllah, E.G., Hassanein, H.S. and Zulkernine, M., 2015. A survey of security attacks in information-centric networking. *IEEE Communications Surveys & Tutorials*, 17(3), 1441–54, https://doi.org/10.1109/COMST.2015.2392629.

Cao, L., Jiang, X., Zhao, Y., Wang, S., You, D. and Xu, X., 2020. A survey of network attacks on cyber-physical systems. *IEEE Access*, 8, 44219–27, https://doi.org/10.1109/ACCESS.2020.2977423.

See also: SECURITY

Back-Up

A copy of **data** usually stored in a different location to the original (virtually and sometimes physically) that can be used to restore data or systems after a data loss or **cybersecurity** incident. Back-ups can be created by both individual **users** and whole systems or **network**s. In the latter case this might, for example, be used to return the system to the state it was in before a **breach**.

In a privacy context, there needs to be consideration of protection of sensitive and **personal information** during the back-up process. The user needs to ensure that the back-ups are securely stored, encrypted, and accessible to only authorised people. For example, the not uncommon practice of users backing up to portable devices needs to be monitored and managed.

In terms of **security**, back-up refers to the measures taken to prevent theft or loss of the backed-up data. In particular, the practice of back-up and forget is to be avoided. The security of back-up systems needs to be managed to the same degree as live systems (for example through the installation of security **patch**es to **software**) and the back-up system itself needs to be regularly tested.

Back-ups can themselves be targets of a sophisticated **adversary**. For example, to be effective a **ransomware** attack may need to infect, destroy or corrupt system back-ups as well as infecting the live system.

Further reading:

Chervenak, A., Vellanki, V. and Kurmas, Z., 1998. Protecting file systems: a survey of backup techniques. *In*: *Joint NASA and IEEE Mass Storage*, 99, www.storageconference.us/1998/papers/a1-2-CHERVE.pdf.

See also: INTEGRITY

Barnardisation

A **statistical disclosure control** technique – attributed to the statistician George Barnard – applied to a **table of counts**, in which cell counts are adjusted by adding or subtracting 1 at a given (usually small) probability. In some uses, cells are paired to ensure that additions and subtractions balance out and **additivity** is ensured. The technique was only ever popular in the UK and has fallen into disuse since a review in 2011 by the Office for National Statistics whose analysis showed there were reasons to doubt its efficacy.

Further reading:

Hakim, C., 1979. Census confidentiality in Britain. *In*: Bulmer, M.ed., *Censuses, surveys and privacy*. London: Palgrave, 132–57.

SDC UKCDMAC Subgroup. *Statistical Disclosure Control (SDC) methods short-listed for 2011 UK Census tabular outputs*. Paper 1. Office for National Statistics, www.ons.gov.uk/file?uri=/census/2011census/howourcensusworks/howwetookthe2011census/howweplannedfordatadelivery/protectingconfidentialitywithstatisticaldisclosurecontrol/sdcsubpaper1_tcm77-189745.pdf.

BCI

See: BRAIN-COMPUTER INTERFACE

BCR

See: BINDING CORPORATE RULES

Behavioural Advertising

A method for delivering advertisements to consumers which targets **user**s using **data** collected from their online activity. It works by creating a user’s **profile** by collecting **browsing history**, **information** about websites visited, **search** terms, purchases, and so on, using mechanisms such as **persistent cookie**s. Under the EU’s **ePrivacy Directive** the use of such **cookie**s for profiling is only permissible with the user’s explicit **consent**.

From the advertiser’s point of view behavioural advertising is believed to be more effective than traditional online ads because it is based on the user’s actual behaviour, rather than their **demographic** or other general characteristics. From the user’s point of view, it can deliver more relevant ads and therefore more **personalisation**, albeit at the cost of surrendering **personal data** to the advertiser. However, privacy advocates have expressed concern that these ‘relevant’ ads can cumulatively entrench or exacerbate adverse user interests, such as self-harm, radical views and disordered eating, or be based on a harmful false perception (e.g., that a user is still pregnant).

Further reading:

Prince, A.E.R., 2022. Can you hide your pregnancy in the era of Big Data? *The Atlantic,* www.theatlantic.com/ideas/archive/2022/10/can-you-hide-your-pregnancy-era-big-data/671692/.

Si, C., Yajun, W., Fengyi, D. and Kuiyun, Z., 2023. How does ad relevance affect consumers’ attitudes toward personalized advertisements and social media platforms? The role of information co-ownership, vulnerability, and privacy cynicism. *Journal of Retailing and Consumer Services*, 73, 103336, https://doi.org/https://doi.org/10.1016/j.jretconser.2023.103336.

See also: BIG DATA, DATA USER, TARGETED ADVERTISING, TRACKER, WEB PROFILING

Benefits of Privacy

Privacy is sometimes understood as an intrinsic good, that is, good in its own right. However, it also has instrumental value – it is desired because it supports other goods.

Some of these goods contribute to the **value of privacy**, which may accrue to society (such as aiding democracy) or to individuals (such as enabling **intimacy**). These legitimate goods are often used as justifications for defending privacy against **privacy threat**s.

However, other benefits of privacy for the individual are less benign. Privacy may allow people to plan antisocial actions, or to behave selfishly. For instance, private communications may allow the planning of terrorist or criminal actions. Privacy and **confidentiality** in healthcare matters may allow individuals to pass on communicable diseases. Privacy and confidentiality about criminal records may allow offenders to continue offending unchallenged. **Financial privacy** may support tax evasion or money laundering. Privacy of the household may be used to conceal abuse within the family. In all these ways, privacy benefits the individual at a potential cost to other individuals and/or society. These types of individual benefits are the basis of the perceived tension between privacy and other important values such as **security** and **safety**.

Further reading:

Etzioni, A., 1999. *The limits of privacy*. New York: Basic Books.

O'Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Utloff, C., 2016. Technology and the tension between security and privacy. *US Cybersecurity Magazine*, Spring 2016, www.uscybersecurity.net/csmag/technology-and-the-tension-between-security-and-privacy/.

See also: FEMINIST CRITIQUE OF PRIVACY, ONTOLOGICAL SECURITY

Bicycle Attack

An **attack** to determine the length of passwords. The **adversary** intercepts encrypted packets (usually transmitted over **HTTPS**), deducing the length of the **password** by subtracting the known header lengths from the total length of the request.

Further reading:

Harsha, B., Morton, R., Blocki, J., Springer, J. and Dark, M., 2021. Bicycle attacks considered harmful: quantifying the damage of widespread password length leakage. *Computers & Security*, 100, 102068, https://doi.org/10.1016/j.cose.2020.102068.

Big Brother

In George Orwell's novel *Nineteen Eighty-Four*, Big Brother is the otherwise unnamed leader of the state of Oceania, where the action takes place. The government places all citizens above the lowest social strata under **surveillance**, through ubiquitous telescreens as well as by networks of spies.

The term has entered the general lexicon to denote perceived perpetual surveillance, that is, the principle of the **panopticon**, in which one is aware that there is always a possibility that one is being observed, without knowing whether one is under surveillance at any particular time. One therefore *self-censors* one's actions and speech. In the novel, reminders of the surveillance system are prominently displayed, in the form of posters which read 'Big Brother is Watching You'.

Further reading:

Orwell, G., 1949. *Nineteen eighty-four*. London: Martin Secker & Warburg.

Power, D.J., 2016. 'Big Brother' can watch us. *Journal of Decision Systems*, 25, 578–88, https://doi.org/10.1080/12460125.2016.1187420.

Big Data

Big data is a loosely defined term, which in general refers to large amounts of structured and unstructured **data** collected from or generated by various sources and processes, including **social media**, **sensor**s, system logs, commercial transactions and more. The amount of data that is generated/collected daily is massive and continues to grow rapidly. The speed at which data is generated and processed is increasing, with new sources of data and faster processing technologies. Many definitions focusing on these features refer to the three Vs: 'volume', 'velocity' and 'variety', but other features include the increasing use of **data linkage** to bring together data from different sources and real-time automated data generation. The size of big data sources enables them to be used to train **machine learning** and **artificial intelligence** systems.

Big data are used to provide valuable insights for businesses and organisations, such as detecting trends and patterns in consumer behaviour, predicting customer needs and discovering new business opportunities. To manage and analyse big data, it is necessary to use technologies and tools that support large scale processing, such as computer clusters, distributed **databases**, real-time data analysis systems and machine learning **algorithms**. Furthermore, it often requires data to have extensive cleaning and validation before it can be used.

One of the main privacy concerns with big data is the vast amount of **personal information** that can potentially be collected, stored and analysed. This might include sensitive information such as medical records, financial transactions and **browsing history**. Consequently, big data can be used for malicious purposes, such as **mash attack**s, **membership inference attack**s or **identity theft**.

Another concern is that big data can be used for detailed **profiling** of individuals, which may be used for **surveillance** or to discriminate against certain groups of people. For example, an insurance company could use big data to identify certain individuals as high-risk customers and charge them higher insurance rates (or refuse the insurance). Moreover, analysis of big data can (unintentionally) reinforce existing biases or even create new ones, as the data is not designed to be representative and has been generated out of (biased) human processes.

Although in general it is standard practice to implement **security** measures to protect **personal data** from unauthorised access, with big data this may be very difficult to achieve in practice because of its scale and the complexity of its data generating processes.

It is often observed that it is important to implement machine learning in a manner that ensures the decisions are fair and **transparent**. But with complex data drawn from multiple sources, possibly being updated in real time, even understanding how to evaluate **fairness** may become impossible.

Further reading:

Mayer-Schönberger, V. and Cukier, K., 2013. *Big data: a revolution that will transform how we live, work, and think.* New York: Houghton Mifflin Harcourt.

Jain, P., Gyanchandani, M. and Khare, N., 2016. Big data privacy: a technological perspective and review. *Journal of Big Data*, 3, 1–25, https://doi.org/10.1186/s40537-016-0059-y.

See also: DATAFICATION, DATA IN USE, DATA PROCESSING

Binary Variable

A type of **categorical data**, also called dichotomous, which has two categories which are usually coded as '0' (attribute is absent) and '1' (attribute is present). For example, *Female* would be coded as '1' for female persons and '0' otherwise.

From a disclosure risk viewpoint, binary variables are often viewed as low **risk** because of their coarseness. But consider the variable *has HIV.* A value of 1 on this variable would be regarded by most as sensitive, and that value is also potentially disclosive because it is rare.

Binding Corporate Rules (BCR)

The EU’s **GDPR** establishes a hierarchy of options for the transfer of **personal data** to third countries, that is, countries outside the European Economic Area. BCRs are one of these options.

BCRs are a set of internal rules that companies within a supranational structure must follow to lawfully exchange personal data which includes that of EU residents. The BCRs must be legally binding on all entities within the undertaking and approved by the competent national **supervisory authority** under the **consistency mechanism** (most likely the EU country where the relevant company is based). It can be best practice for an organisation to adopt separate BCRs for their employee and customer **data**, as the rights and interests in these data may differ. The BCRs need to ensure that **data subject**s have legally enforceable rights equivalent to those bestowed by the GDPR, in case data are shared with a jurisdiction which does not guarantee such rights under its own law.

Further reading:
Moerel, L., 2012. *Binding corporate rules*. Oxford: Oxford University Press.
Phillips, M., 2018. International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR). *Human Genetics,* 137(8), 575–82, https://doi.org/10.1007/s00439-018-1919-7.

See also: ADEQUACY, CROSS-BORDER DATA PROCESSING, DATA PROTECTION, DATA TRANSFER, JURISDICTION, PROCESSING, TERRITORIAL SCOPE

Biobank

A biobank stores samples of tissue, usually human, for medical research, often linked to **information** about their donors’ medical histories. Biobanks are regarded as promising means for drug discovery, to pursue **personalised medicine** and in particular the study of the long-term effects of genomes. As both their collections of samples and the associated medical histories grow, biobanks are becoming rich sources of **data**, invaluable for finding associations between genomes and health outcomes.

Although some biobanks store non-human material, clearly privacy issues arise only with those focusing on humans. One issue is the far-reaching implications of storing genetic material, research from which would implicate not only the donor but also their relatives. Samples are **anonymised** as far as possible but – given the context of a rich selection of genetic material (and the likelihood of data-sharing between biobanks for research purposes) – this may be hard to do effectively. Another complication is the principle of biobanking that donors should have a right to see results of research relevant to them, which conflicts with the requirement for anonymisation. Donors’ **consent** needs to be managed, with donors also given rights to withdraw.

Further reading:
Pascuzzi, G., Izzo, U. and Macilotti, M., eds, 2013. *Comparative issues in the governance of research biobanks: property, privacy, intellectual property, and the role of technology*. Berlin: Springer.

See also: CONSENT, GENETIC PRIVACY, GENOMIC DATA

Biometric Data

Biometrics are measurements of biological characteristics of an individual. Biometric data can be used for **authentication** and such use cases are becoming increasingly common in both **location** and personal device **security**.

The EU **GDPR** defines biometric data as ‘**personal data** resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data’. This was an update from the previous **Data Protection Directive**, which referred to **data** relating to health, but did not anticipate advances in identification technology which make individual physical variants increasingly valuable as personal data. Although **genomics data** also stem from ‘specific technical processing’ relating to a person’s physiological characteristics, they are classed as **special category data** under the GDPR.

There are similar definitions in Australia’s federally approved Biometrics Institute Privacy Code, and the proposed American Data Privacy and Protection Act. The draft ADPPA, however, excludes photographs (presumably even photographs of faces), illustrating that facial images can be a grey area in the scope of biometric data.

Protecting biometric data from a determined **adversary** is notoriously difficult; its functionality needs it to be uniquely associated with particular individuals and therefore necessarily it is personal data. There is however some interesting work being done on cancellable biometrics which may provide some protection.

Further reading:

Liu, N.Y., 2012. *Bio-privacy: privacy regulations and the challenge of biometrics*. Oxford: Routledge.

Manisha and Kumar, N., 2020. Cancelable biometrics: a comprehensive survey. *Artificial Intelligence Review*, 53, 3403–46, https://doi.org/10.1007/s10462-019-09767-8.

See also: DATA PROTECTION, FACIAL RECOGNITION TECHNOLOGY, GAIT RECOGNITION, IRIS SCANNING, NATURAL PERSON, SENSITIVITY, UNIQUE IDENTIFIER, US PRIVACY LAWS

Biometrics

See: BIOMETRIC DATA

Black Hat Attack

An **attack** on a computer, data store or network which is unethical in nature, perhaps with the intention of causing **harm** or stealing **sensitive** and **personal information**. **Hacking**, **malware**, DDoS attacks, **phishing**, and **social engineering** can all be considered black hat attacks.

Further reading:

AbdAllah, E.G., Hassanein, H.S. and Zulkernine, M., 2015. A survey of security attacks in information-centric networking. *IEEE Communications Surveys & Tutorials*, 17(3), 1441–54, https://doi.org/10.1109/COMST.2015.2392629.

Cao, L., Jiang, X., Zhao, Y., Wang, S., You, D. and Xu, X., 2020. A survey of network attacks on cyber-physical systems. *IEEE Access,* 8, 44219–27, https://doi.org/10.1109/ACCESS.2020.2977423.

See also: DENIAL OF SERVICE, GREY HAT ATTACK, WHITE HAT ATTACK

Blacklist

A blacklist is a list of **Internet** domains, **IP address**es or account emails that have been rated as malicious. Blacklist mechanisms provide a methodology for identifying and mitigating **security** threats and unwanted activities on the Internet. They also help maintain the security and privacy of systems and data.

Modern browsers support a list of IP addresses or web domains that are known to be malicious or contain malicious content such as **malware**, **virus**es or **spyware**. Usually, the browser warns the **user** and blocks access to the site. This also helps to speed up the browser stability and activity. The list can also be used by **tracker blocker**s for blocking Internet connections to malicious destinations.

The term has become less used as technology providers move towards inclusive language. Google, for example, now uses ‘blocklist’.

Further reading:

Felegyhazi, M., Kreibich, C. and Paxson, V., 2010. On the potential of proactive domain blacklisting. *In*: *Proceedings of the 3rd USENIX Conference on Large-Scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More*, USA: USENIX Association, www.usenix.org/legacy/event/leet10/tech/full_papers/Felegyhazi.pdf.

Blackmail

Richard Posner argued that the law should not protect against disclosures of embarrassing or discreditable **information**, because they are in the **public interest** by removing **asymmetric information**. On this reading, blackmail to extort money or favours in return for concealing true but compromising information is an antisocial type of privacy protection. The blackmailer’s threat is to do something that is legal, and arguably their duty, but they prefer to take the payoff. Some have defended blackmailers as straightforward economic actors, bargaining for the value of the information they hold; for example, Miceli put blackmail in a category with patents and **non-disclosure agreement**s.

A specific type of extortion in the digital context uses **ransomware** to encrypt digital assets, so that the blackmailer can demand a payment from the data owner for the decryption key.

Further reading:

Miceli, T.J., 2020. Trading in information: on the unlikely correspondence between patents and blackmail law. *Review of Industrial Organization*, 56(4), 637–50, https://doi.org/10.1007/s11151-020-09749-z.

Posner, R.A., 1993. Blackmail, privacy, and freedom of contract. *University of Pennsylvania Law Review*, 141(5), 1817–47, https://doi.org/10.2307/3312575.

See also: DATA OWNERSHIP, DUTY TO PROTECT, PRIVACY THREAT, PRIVACY TORT, PUBLIC DISCLOSURE OF PRIVATE FACTS, VALUE OF PRIVACY

Blinding

Blinding is a **cryptographic** technique in which a function is computed by an agent without the agent knowing either the input or the output. Broadly, the principal takes the input x and applies a bijective function C, unknown to the agent, to it. Because C is bijective, it maps each x to a unique y, and has an inverse function which maps each y to a unique x. The agent then computes the desired function F over C(x), and returns F(C(x)) to the principal. The principal then applies a decoding function D, such that D(F(C(x))) = F(x).

A *blind signature* uses such techniques, enabling a blinded message to be signed, such that the signature can be publicly **authenticated** against the original unblinded message. The person signing the message, however, is unaware of the content at the point of signing. An example use of this would be electronic voting, where an election official may need to sign a digital ballot to show that it was legitimate but should not be able to see the vote in a **secret ballot**.

Further reading:
Bleumer, G., 2011. Blinding techniques. *In*: van Tilborg, H.C.A. and Jajodia, S., eds, *Encyclopedia of cryptography and security*, vol.1. 2nd edition. New York: Springer, 150–2, https://doi.org/10.1007/978-1-4419-5906-5_182.

See also: DIGITAL SIGNATURE, SECURE MULTI-PARTY COMPUTATION

Blockchain

A blockchain, or *distributed ledger*, is a cryptographically secured **database** stored on a peer-to-peer **network** of computers. When a transaction or other change of the ledger occurs, it is timestamped and validated by all or most nodes on the network; when there is consensus that the transaction is legitimate, it is placed in a new block of transactions and added to the chain of previous blocks that constitute the blockchain. The resulting database stores all transactions. Because it is decentralised, it is expensive for any single actor or group to control. It is open to inspection by any node of the network, and so it cannot be altered secretly, facilitating data **integrity**. Furthermore, as a blockchain stores **data**, many of them can store and execute **software**, called *smart contracts*. Smart contract code is run whenever its preconditions are met and is difficult to interfere with or influence.

Many cryptocurrencies, hosted on blockchains (such as Bitcoin), allow anonymous transactions, which has made them attractive for criminal activities. As well as financial transactions, blockchains have been proposed to store legal records and registries (especially where such records are scarce or not trusted) and health records. The importance of timestamps and **transparency** makes blockchains potentially useful for keeping complex logistics and supply chains up to date and safer from fraud. It is also thought that some of the **security** issues underlying the **Internet of Things** could be mitigated with blockchain storage of the data.

Blockchains can be *permissionless* or *permissioned*. A permissionless blockchain is **public**, so that any computer can join the network. This opens up the data to view by anyone, as well as having no mechanism to prevent the blockchain being used for criminal purposes. Understanding the network's behaviour may aid privacy **breach**es. Following the activities of a particular node (or *wallet*) could enable an identification, if **auxiliary knowledge** was available.

A permissioned blockchain has some kind of centralised organisation (though it is still a peer-to-peer network with distributed **data storage**), but the organisation can act as a **gatekeeper** to prevent outsiders joining. The flip side of this is that it is easier for the organisation to change the ledger than with the permissionless structure. The permissionless blockchain is the purer blockchain vision, but perhaps more open to abuse; the permissioned blockchain is safer, provided that the gatekeeper is **trustworthy**.

The privacy advantages of blockchains have been widely discussed. Since there is no central authority with privileged or opaque access, the **information** on it cannot be interfered with, and a blockchain should be safe from data **breach**es, as all the data is, or can be, **encrypted**. Furthermore, it is possible for **data subjects** to enforce **access control**, via **public-key cryptography**. However, there are questions as to how **GDPR**-compliant they can be; encrypted data may still be regarded as **personal data**, and the GDPR's right to **erasure** would be hard to enforce, given their immutability. Perhaps even more importantly, it is also unclear who would be held responsible for any breaches. For example, anyone storing a blockchain has to process all the personal data in all the smart contracts across the entire chain every time the chain is updated (which may be frequently). They may also be in any **jurisdiction** across the globe, and the chain itself is held across jurisdictions, which renders enforcement near intractable. Blockchain can be said therefore to stress the limits of **data protection** orthodoxy.

Further reading:
Henry, R., Herzberg, A. and Kate, A., 2018. Blockchain access privacy: challenges and directions. *IEEE Security and Privacy*, 16(4), 38–45, https://doi.org/10.1109/MSP.2018.3111245.
Zhang, R., Xue, R. and Liu, L., 2020. Security and privacy on blockchain. *ACM Computing Surveys*, 52(3), article no.51, https://doi.org/10.1145/3316481.

Block Cipher

A **cryptographic** procedure that is used to **encrypt** and decode blocks of data that have a predetermined size. Using **symmetric key encryption**, it converts input **plaintext** blocks into **ciphertext** blocks of the same size.

Further reading:
Lai, X., 1992. *On the design and security of block ciphers.* Thesis (PhD). ETH Zurich. https://doi.org/10.3929/ethz-a-000646711.

See also: ADVANCED ENCRYPTION STANDARD

Blocking Variable

A term used in **probabilistic record linkage**. A blocking variable is one which must match for two **record**s to be linked. This is a heuristic method for reducing the search space of possible linkages and therefore the computational resources required. It will also create false negatives (because true links may be missed). Therefore, blocking variables should be used sparingly and only those with low **data divergence** should be used.

Further reading:
Blakely, T. and Salmond, C., 2002. Probabilistic record linkage and a method to calculate the positive predictive value. *International Journal of Epidemiology*, 31(6), 1246–52, https://doi.org/10.1093/ije/31.6.1246.

Blocklist

See: BLACKLIST

Bluejacking

See: BLUETOOTH

Bluesnarfing

See: BLUETOOTH

Blue Team

In a **cybersecurity** context, a blue team is an organisation's **security** team that defends the organisation against both real cyberattacks and simulated **adversaries**, called **red teams**, that test the effectiveness of their defensive systems. Red team tests differ from simple **penetration tests** as they tend to operate over longer time periods using an array of **attack** vectors, to be countered by a corresponding blue team. Red–blue team exercises therefore form part of the organisation's ongoing security infrastructure rather than being simply one-off tests.

Further reading:
Miessler, D., 2021. The difference between red, blue, and purple teams, *Unsupervised Learning.* https://danielmiessler.com/study/red-blue-purple-teams/.

See also: MOTIVATED INTRUDER TEST, PURPLE TEAM

Bluetooth

A wireless technology that connects and exchanges **data** between devices over short distances, usually up to about ten metres. Smartphones, tablets, laptops and other devices can be connected to peripherals such as headphones, speakers and smartwatches using Bluetooth.

Bluetooth privacy **risks** include those associated with unsecured connections. An **adversary** could intercept the Bluetooth signal and access the data being transmitted if a device is not properly secured or configured. In a 'bluejacking' attack, an adversary sends unwanted text messages or files to nearby Bluetooth devices. When a hacker gains unauthorised access to a Bluetooth device to steal **information** such as contacts, messages and other data, this is known as 'bluesnarfing'.

Bluetooth also has a **tracking vulnerability**. Users should ensure that their Bluetooth devices use the most recent security **protocols**, such as pairing and **encryption**, to reduce these risks. **Users** should turn off their Bluetooth when not in use and avoid connecting to unknown or untrusted devices. Additionally, developers should follow **security-by-design** guidelines and offer detailed instructions on configuring and securing Bluetooth connections.

Further reading:

Albazrqaoe, W., Huang, J. and Xing, G., 2016. Practical Bluetooth traffic sniffing: systems and privacy implications. *In*: *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, New York: ACM, 333–45. https://doi.org/10.1145/2906388.2906403.

Bodily Privacy

Bodily privacy is the protection against non-consensual **intrusion** into an individual's physical person. Such intrusions may include invasions of **personal space**, touching of the body, reaching into clothing, injections through the skin, inappropriate or forced healthcare interventions or entering bodily cavities. A common term for these forms of physical intrusion under European human rights caselaw is interference with a person's 'physical integrity'. Healthcare and other decisions relating to someone's physical body therefore engage rights to privacy.

As well as these **physical privacy** examples, other types of interference may be more **information**ally oriented, including testing blood or waste products, drug-testing, genetic/DNA testing and genome mapping.

Further reading:

Rao, R., 2000. Property, privacy, and the human body. *Boston University Law Review*, 80, 359–460.

See also: ABORTION, BIOMETRIC DATA, BOUNDARY, GENETIC PRIVACY, GENOMIC DATA, INTERFERENCE, SPATIAL PRIVACY, DIGNITY

Bot

A bot is **software** that automates activities on behalf of a **user**. Bots run without human intervention, and they are used for different human tasks such as responding to customer service inquiries, performing data entry or publishing content online. For example, chatbots are used to provide customer support.

Bots can be also used maliciously, such as the example of **botnet**s (spreading malicious activities), crawling content from websites for malicious purposes or performing **DDoS** attacks. To protect users against malicious bots, preventative measures include implementing bot detection techniques, keeping software up to date and implementing techniques of

traffic data monitoring, such as **Deep Packet Inspection (DPI)**, log analysis and cloud-based traffic monitoring analysis.

Further reading:

Beatson, O., Gibson, R., Cunill, M.C. and Elliot, M., 2023. Automation on Twitter: measuring the effectiveness of approaches to bot detection. *Social Science Computer Review*, 41(1), 181–200, https://doi.org/10.1177/08944393211034991.

Kudugunta, S. and Ferrara, E., 2018. Deep neural networks for bot detection. *Information Sciences*, 467, 312–22, https://doi.org/10.1016/j.ins.2018.08.019.

Orabi, M., Mouheb, D., Al Aghbari, Z. and Kamel, I., 2020. Detection of bots in social media: a systematic review. *Information Processing & Management*, 57(4), 102250, https://doi.org/10.1016/j.ipm.2020.102250.

See also: ACCESS CONTROL, SPAM

Botnet

A botnet is a **network** of compromised devices that are under the remote control of an **adversary**. Once compromised, the adversary uses the devices in the network to carry out a series of malicious activities, such as **distributed denial-of-service (DDoS)** attacks, **eavesdropping** of **personal information** and other **attack**s. Users of the individual computers in a botnet are often unaware that they are participating in the botnet. One of the most famous botnets developed during recent years is **MIRAI**.

Further reading:

Feily, M., Shahrestani, A. and Ramadass, S., 2009. A survey of botnet and botnet detection. *In*: *2009 Third International Conference on Emerging Security Information, Systems and Technologies*, 268–73, https://doi.org/10.1109/SECURWARE.2009.48.

Manos, A., Tim, A., Michael, B., Matt, B., Elie, B., Jaime, C., Zakir, D., J, A.H., Luca, I., Michalis, K., Deepak, K., Chaz, L., Zane, M., Joshua, M., Damian, M., Chad, S., Nick, S., Kurt, T. and Yi, Z., 2017. Understanding the Mirai botnet. *In*: *26th USENIX Security Symposium (USENIX Security 17)*, Vancouver, BC: USENIX Association, 1093–1110, www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis.

See also: MALWARE

Boundary

A line in physical, virtual or conceptual space demarcating a change in regions.

Boundaries are often used as demarcations of private space. For example, **personal space** will be demarked by a boundary – albeit a culturally, situationally and psychologically mediated one – the crossing of which will indicate that **bodily privacy** has been breached. **Confidentiality** can be viewed as a boundary state – a line which **data** are not to cross. Duties of **confidence**, confidentiality pledges and associated data governance and infrastructure can all be regarded as mechanisms for enforcing a boundary. The principle behind both **anonymisation** and **formal privacy** models can be viewed as allowing useful data to cross the boundary but leaving identifying **information** behind.

Further reading:
Lamont, M. and Molnár, V., 2002. The study of boundaries in the social sciences. *Annual Review of Sociology*, 28, 167–95, https://doi.org/10.1146/annurev.soc.28.110601.141107.

See also: INFORMATION GOVERNANCE, PSYCHOLOGICAL PRIVACY

Bounded Rationality

Whereas traditional economics posited the abstraction of ‘rational economic man’, real human beings labour under obvious constraints of time, reasoning power, memory and other resources, all of which limit their ability to be fully rational. The decisions they make tend to be acceptable heuristics sufficient to support action, rather than optimal. Rationality is therefore bounded, and their heuristic decision-making is sometimes called *satisficing*. Bounded rationality is one of the constraints on reasoning about privacy, for example whether to give **consent** to use of **data**. It has been argued that **informed consent** is not possible in a complex digital **information** environment, thanks to cognitive limitations and bounded rationality.

Further reading:
Barocas, S. and Nissenbaum, H., 2014. Big data’s end run around procedural privacy protections. *Communications of the ACM*, 57(11), 31–3, http://dx.doi.org/10.1145/2668897.

Simon, H.A., 1955. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99–118, https://doi.org/10.2307/1884852.
Simon, H.A., 1956. Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–38, https://psycnet.apa.org/doi/10.1037/h0042769.

See also: ECONOMICS OF PRIVACY, RATIONAL CONSUMER

Bounds

The constraint or limiting value on some metric.

In a **statistical disclosure control** context, bounds are the maximum and minimum possible values of some quantity given that has been employed to obscure or disguise the value. For example, if published cell counts in a table of **frequency data** have been rounded to base ten, an observer will know that the real values will lie between ±5 of the published values. Direct observations such as this are called the ‘trivial bounds’ and in practice it may also be possible to reduce the spread of the bounds using other pieces of **information**. For example, if an organisation publishes two univariate **tables of counts** drawn from the same **dataset**s, then it has implicitly published a set of bounds for the two-way cross-classification of the two variables. In other words, each of the interior cells in the two-way table has a maximum and minimum value for the observer of the univariate tables. As **data** become more complex the possibility of tightening bounds becomes greater. Tightening of bounds to reverse engineer disclosure controls is one of the possible aims of a **subtraction attack**.

In a **provable security** context it is bounds is used to indicate a provable limit on some privacy metric such as the maximum amount of information that might be leaked to an adversary.

Brain–Computer Interface (BCI)

A technology that enables brain-to-computer device communication. BCIs enable **user**s to operate computer programs or other devices by sending and receiving brain signals without making any physical movements. BCIs operate by measuring electrical signals, which are then decoded into commands by **algorithm**s and used to control a machine or **application**.

The accuracy and dependability of the signals as well as the privacy and **security** of the user’s brain data remain major technical and **neuroethics** challenges that BCI technology must still overcome. Despite these difficulties, BCI research is moving forward quickly and shows a lot of promise.

Further reading:

Graimann, B., Allison, B. and Pfurtscheller, G., 2010. Brain-computer interfaces: a gentle introduction. In: Graimann, B., Pfurtscheller, G. and Allison, B., eds, *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*, Berlin, Heidelberg: Springer, 1–27, https://doi.org/10.1007/978-3-642-02091-9_1.

Wolpaw, J.R., 2013. Brain–computer interfaces. *Handbook of Clinical Neurology*, 110, 67–74, https://doi.org/10.1016/B978-0-444-52901-5.00006-X.

See also: BRAIN IMPLANT, NEURODATA, NEUROTECHNOLOGY

Brain Implant

A specific form of **brain–computer interface** where some electronic device is attached directly to the subject's brain. The current use case for these is medical, primarily to augment or replace some organic function where the subject's brain is no longer (fully) functioning, perhaps after a stroke or brain injury. This includes sensory prosthesis (e.g., to overcome blindness). However, research is also being conducted on military applications as well as human augmentation.

These type of invasive BCIs raise the strongest **neuroethics** concerns with the possibility of chips being **hacked** or subject to **surveillance**.

Further reading:

Gilbert, F., 2015. A threat to autonomy? The intrusion of predictive brain implants. *AJOB Neuroscience*, 6(4), 4–11, https://doi.org/10.1080/21507740.2015.1076087.

Reinares-Lara, E., Olarte-Pascual, C. and Pelegrín-Borondo, J., 2018. Do you want to be a cyborg? The moderating effect of ethics on neural implant acceptance. *Computers in Human Behavior*, 85, 43–53, https://doi.org/10.1016/j.chb.2018.03.032.

See also: AUTONOMY

Brainwashing

The word 'brainwashing' entered the Western lexicon in the 1950s, an apparent Anglicisation of a term invented by the Chinese people to denote their attempted ideological correction by the Soviet authorities of the time. In its adoption within the United States, it conveyed some of the fear among the domestic population emanating from the ‘Red Scare’ of that era.

The infringement of individual **autonomy** and/or **decisional privacy** was not, at the point of its original adoption, the greatest concern inherent in the idea of brainwashing. The political power of a regime – particularly a foreign, hostile regime – to convert a citizen to its cause, to the point of blind faith and absolute loyalty, was the predominant threat of the perceived phenomenon (as in the 1962 film *The Manchurian Candidate*). The word has, however, since diffused across contexts, sectors and degrees of seriousness. Common usage would include a more informal, or even lighthearted, suggestion that an individual has been cumulatively influenced by advertisements and other consumer messaging to the point of developing an uncritical obedience. As such, the term has almost come full circle to denote the subversion of individual autonomy by 21st-century digital capitalism. As such, the term is now ideologically neutral, and can be used seriously (in the case of terrorist or cultish mind-control) or with a more flippant irony in the case of excessive consumerism.

Further reading:

Hunter, E., 1956. *Brainwashing: the story of men who defied it*. New York: Farrar, Strauss & Cudahy.

Seed, D., 2013. *Brainwashing: the fictions of mind control, a study of novels and films since World War II.* Kent: The Kent State University Press.

See also: NUDGE THEORY

Breach

Within professional **information governance** practice, breach is used in two main senses: to indicate that some **confidentiality** or privacy **boundary** has been crossed, or that some obligation, expectation or requirement has not been met or complied with. These two senses often co-occur but not always.

In the first sense we might say that a breach has occurred if someone has invaded our **personal space** or if a **hacker** has broken through a **cybersecurity** system. In the second sense an obligation, expectation or requirement does not need to be legal in nature for non-**compliance** to be described as a breach. Breaches may be procedural in that a rule or policy has been not complied with, but no **harm** has happened. However, the term ‘breach’ usually implies a level of seriousness in the relevant social, technical or ethical obligation, to the extent that non-conformity with its requirements is a more than trivial matter.

There are myriad laws that set the parameters for what could be termed a privacy breach. In different countries across the world, a ‘breach’ could be of a **common law duty of confidence**, of a constitutional right to **privacy**, of duties under the law of **privacy tort**, or of a statutory **data protection** obligation. Under the EU **GDPR**, Article 4(12) defines a ‘personal data breach’ as ‘a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data’. A breach of security is interpreted in line with **CIA Triad** principles, following guidance from the **Article 29 Working Party**.

However, a privacy breach should be understood as distinct from a **data breach**, which is more specifically used to refer to a failure of data **security**. A breach in a cybersecurity sense is used with more global consistency than a breach of privacy. If a privacy breach has occurred, this will invariably imply some specific harm to a specific individual. Some data breaches will also be privacy breaches; however, privacy breaches can occur without any data breach and vice versa.

Further reading:

Article 29 Working Party, 2017. *Guidelines on Personal Data Breach Notification under Regulation 2016/679*. Available from: ARTICLE29 – Guidelines on Personal data breach notification under Regulation 2016/679 (wp250rev.01).

See also: BREACH DISCLOSURE, DATA BREACH NOTIFICATION, DATA DESTRUCTION, DATA PROTECTION PRINCIPLES, PERSONAL DATA, RIGHT OF ACCESS

Breach Disclosure

A breach disclosure is the communication or **publication** of **information** about a **breach**. The **disclosure** might be mandatory or voluntary depending on the **jurisdiction** and the nature of the breach.

Under the EU **GDPR**, there are two categories of people to whom the fact of a **data breach** should be disclosed. The first is the **supervisory authority** (i.e., the national data protection regulator for the relevant country). The authority should be notified of the breach within 72 hours *unless* the **data controller** determines it is not likely to result in a **risk** to the rights and freedoms of **natural persons** (Article 33).

The second category is the **data subject**s to whom the affected **personal data** relates. Where it is likely that the breach of personal data will result in a high risk of **harm** (a higher threshold), data subjects should be informed without ‘undue delay’ (Article 34).

The EU's Network and Information Systems ('NIS') Directive, passed the same year as the GDPR, also contains **data breach notification** obligations, but these are less concerned with risks to rights and freedoms of natural people (such as **privacy risk**s) and more about compromise of the **integrity** of key networks. The NIS Directive does not apply generally to all actors, but rather to Internet service providers and other key digital infrastructures listed in the Directive.

Further reading:
Schmitz-Berndt, S., and Schiffner, S., 2021. Don't tell them now (or at all) – responsible disclosure of security incidents under NIS Directive and GDPR. *International Review of Law, Computers & Technology*, 35(2), 101–15, https://doi.org/10.1080/13600869.2021.1885103.

Breach of Confidence

Breach of confidence is a term used in some **common law jurisdiction**s to indicate that a legal **duty of confidence** has been violated. It has its origins in the English law of equity but has slowly evolved into a cause of action itself. **Data protection** law has also introduced **confidentiality** as a statutory requirement within and beyond common law systems.

Bring Your Own Device Policy (BYOD)

An organisation may establish a Bring Your Own Device (BYOD) policy to control how employees (and guests) use their own devices for activities related with their work. The policy should outline the rights and obligations of both the **user** and the organisation, as well as the **security** requirements that the user must meet to guarantee the protection of corporate systems and **data**.

A BYOD policy's main objective is to strike a balance between the productivity advantages of staff using personal devices for work and the security **risk**s of allowing less regulated hardware access to systems.

Further reading:
Barlette, Y., Jaouen, A. and Baillette, P., 2021. Bring Your Own Device (BYOD) as reversed IT adoption: insights into managers' coping strategies. *International Journal of Information Management*, 56, 102212, https://doi.org/10.1016/j.ijinfomgt.2020.102212.

See also: APPROPRIATE TECHNICAL AND ORGANISATIONAL MEASURES

Browser Fingerprinting

Browser fingerprinting is a technique used to collect **information** about a device and its **user** by identifying the characteristics of the web browser they are using. The information collected can include things like the browser version, installed fonts, browser extensions and settings, as well as information about the device itself, such as its screen resolution, time zone and installed languages.

This information can be used (using **cookie**s and **cross device tracking**) by online advertisers and analytic services to track an individual's activity across the **Internet**, even if they are using different devices or attempting to remain **anonymous** by using a **VPN** (to hide their **IP address**) or other **privacy-enhancing tools** (e.g., an **anonymous search engine**).

Possible mitigations are to use browser add-ons or plugins to change the browser's identity and thereby make it more difficult to track. Also, the user should be aware of the **privacy policies** of the websites they visit, outlining how websites collect, use and share **personal data**, to avoid unwanted personal data collection.

Further reading:
Laperdrix, P., Bielova, N., Baudry, B. and Avoine, G., 2020. Browser fingerprinting: a survey. *ACM Transactions on the Web*, 14(2), article no.8, https://doi.org/10.1145/3386040.

See also: TARGETED ADVERTISING, TRACKING

Browsing History

A record of the websites, web pages, and online content that a **user** has visited or accessed on their device, usually through a web browser.

A web browser can save a user's browsing history, giving them quick and simple access to frequently visited websites. Based on the user's prior browsing behaviour and **privacy settings**, browsers may also use browsing history to recommend relevant content or **targeted advertising**. However, a user's online activities, interests and preferences can be revealed by looking at their browsing history, so could be a valuable source of **profiling information** for third parties.

Users can limit data collection, **data sharing** and **privacy risk**s by changing their browser settings, clearing their browsing history on a regular basis and using private browsing options or a **virtual private network**.

Further reading:
Laperdrix, P., Bielova, N., Baudry, B. and Avoine, G., 2020. Browser fingerprinting: a survey. *ACM Transactions on the Web*, 14(2), article no.8, https://doi.org/10.1145/3386040.

See also: BROWSER FINGERPRINTING, PROFILING, TRACKER, VIRTUAL PRIVATE NETWORK

Brussels Effect

The European Union has in recent years regulated privacy and **data protection** increasingly stringently, with strong top-down measures culminating in the **GDPR** and the inclusion of a **right to data protection** (alongside a separate **right to privacy**) in the **Charter of Fundamental Rights** of the European Union, which came into effect with the Lisbon Treaty of 2009. Meanwhile, its Court of Justice has been described as taking a more activist line on privacy, in cases such as the **Right to be Forgotten** case against Google Spain in 2014 and the **Schrems** case arguing that the EU–US **Safe Harbor** principles could not provide adequate protection of Europeans' **personal data**.

Given the size of the EU as a market, its regulatory capacity and expertise and the global reach of its laws, some non-EU companies trading in Europe have found it easier to conform globally to EU privacy and data protection law (as well as law in other areas, such as antitrust and environmental law) than to maintain different standards across jurisdictions. Although measuring this effect has proved difficult, legal scholar Anu Bradford called this *de facto* global regulation the Brussels effect, analogous to the California effect, whereby traders across America find it efficient to follow the regulations of the most stringent state, which is usually California.

Further reading:
Bradford, A., 2020. *The Brussels effect: how the European Union rules the world.* New York: Oxford University Press.

See also: CHARTER RIGHTS, DATA PROTECTION PRINCIPLES

Brute Force Attack

A brute force attack is a type of cyberattack in which an **adversary** uses automated tools to try many combinations of characters or words repeatedly to guess the correct **password** or **key** for a given system or **application**.

The adversary will systematically try every combination of characters to find the right one.

To protect against brute force attacks, it is important to use strong passwords and to enable **multi-factor authentication** where possible. Also, users can use **software** and tools to detect and block potential attacks, limit the number of login attempts or lock an account after a set number of failed login attempts. Organisations can also implement **firewalls**, **intrusion detection systems** and other types of security software to detect and block brute force attacks.

Further reading:
Raza, M., Iqbal, M., Sharif, M. and Haider, W., 2012. A survey of password attacks and comparative analysis on methods for secure authentication. *World Applied Sciences Journal*, 19(4), 439–44, https://idosi.org/wasj/wasj19(4)12/1.pdf.

See also: AUTHENTICATION, CYBERSECURITY, RIGHT OF ACCESS

Buffer Overflow Attack

Buffer overflow occurs when a program or **user** attempts to write more **data** to a temporary storage space than the buffer can accommodate. As a result, the extra data may overwrite other regions of memory, altering the data there and perhaps resulting in a program crash or other anomalous behaviour.

A buffer overflow attack uses some input to overrun the buffer and overwrite memory regions with malicious code. Typically, the **adversary** provides the data to a vulnerable program. They can then run this code, perhaps stealing **confidential** data or compromising the system.

Mitigating buffer overflow attacks involves using techniques such as Address Space Randomisation (ASLR) to randomise data region addresses, making it difficult for adversaries to locate executable code. Data Execution Prevention (DEP) marks memory areas as non-executable. Structured Exception Handler Overwrite Protection (SEHOP) could safeguard against attacks, by making it harder to compromise **software** exceptions.

Further reading:
Kuperman, B.A., Brodley, C.E., Ozdoganoglu, H., Vijaykumar, T.N. and Jalote, A., 2005. Detection and prevention of stack buffer overflow attacks. *Communications of the ACM*, 48(11), 50–6, https://doi.org/10.1145/1096000.1096004.

See also: DATA STORAGE

Bug

A bug is an error in the code of a computer program. Bugs can have different kinds of impact, from zero to minor business efficiency consequences to catastrophic **security** and system **integrity** issues.

Bugs are a fact of life for programmers. Numerous studies have reported numbers between 10 and 70 bugs per 1000 lines of code. One way of reducing the prevalence of bugs is to avoid the usage of deprecated functions and through validation of **user** inputs. Developers should regularly test the **software**, checking for (new) vulnerabilities. Any discovered bugs should be quickly fixed and **patch**es released to users. Users should consequently reduce **security** and privacy risks by keeping software up to date by applying periodical patches released by software developers.

Further reading:
Castro, M., Costa, M. and Martin, J.P., 2008. Better bug reporting with better privacy. *SIGOPS Operating Systems Review*, 42(2), 319–28. https://doi.org/10.1145/1353535.1346322.

See also: SOFTWARE DEVELOPMENT LIFECYCLE, VULNERABILITY

Business Case

A company's business case is the justification (usually but not exclusively financial) for a particular action or project, arguing that whatever resources expended will be adequately compensated in profit or other outcomes of interest, and that any **risk**s taken on will be manageable. Aspects such as the costs of **compliance** with privacy regulations or of adequate and secure **data storage** will often need to be included in the case.

The **Anonymisation Decision-Making Framework (ADF)** includes as a first step the setting out of the use case for a **data share**. This is in effect the core of a business case for a data sharing project and is characterised by (i) the rationale for wanting to share the data, (ii) partners and organisations who are also involved in the project, and (iii) the use case for the data, that is, what the shared data might be used for in the intended **data environment**.

Further reading:
Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners' guide*, 2nd edition. United Kingdom Anonymisation Network. Available from: https://ukanon.net/framework/.

Business Impact Level

A UK government system of **data classification** consisting of seven levels (BIL0: No impact; BIL1: Unclassified BIL2: Protect BIL3; Restricted; BIL4: Confidential; BIL5: Secret; BIL6: Top Secret), which is framed around the likely impact of a **data breach**. Each level has an associated set of negative outcomes which are deemed likely, mixing **security** and privacy implications with material damage caused. The secondary labels give some indication of how the data in question should be handled. This was part of the **Information** Assurance Standard – IA Standard no.6 and has now been largely superseded by international **standard**s such as the ISO27000 series.

See also: ISO27001, ISO27002

BYOD

See: BRING YOUR OWN DEVICE POLICY

C

CA

See: CERTIFICATION AUTHORITY

Categorical Data

Categorical (or *nominal*) data divides the **data unit**s into a finite (usually small) number of categories.

On the surface the **disclosure risk** associated with categorical data is lower than that associated with **continuous data**. However, in practice the situation is more nuanced. Elliot and Dale describe the properties of differentiation (how many categories) and skew (how evenly spread is the **population** across the categories) as being the key properties which determine the intrinsic **riskiness** of a categorical variable. Even the simplest of categorical variables could be problematic. For example, a dichotomous indicator *has AIDS* would be skewed (in most populations) and therefore produce a small, more readily identifiable group. Also, membership of that group would usually be regarded as **sensitive**. So, this simple variable could be both a **key variable** and a **target variable**. The final point is that although singly a small categorical variable might seem innocuous, combinations of them can throw up unusual people – this is known as the **special unique**s problem.

Further reading:
Elliot, M. and Dale, A., 1999. Scenarios of attack: the data intruder’s perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring), 6–10, www.researchgate.net/profile/Mark-Elliot/publication/343963431_Scenarios_of_attack_the_data_intruder’s_perspective_on_statistical_disclosure_risk/links/5f4a4568299bf13c505020fd/Scenarios-of-attack-the-data-intruders-perspective-on-statistical-disclosure-risk.pdf#page=6.

See also: DISCRETE DATA

CCTV

See: CLOSED CIRCUIT TELEVISION

Celebrity Privacy

Whilst celebrities do not (in theory) enjoy a specific level of protection of privacy in law, they do face challenges to their privacy and, crucially, have the resources to bring legal action when these impositions go too far. For all that showbusiness news and **gossip** is often seen as a frivolous commodity, the celebrity industry has in fact played a key role in reshaping privacy doctrine.

Further reading:

Palmer, C., 2019. Celebrity privacy: how France solves privacy problems celebrities face in the United States. *California Western International Law Journal*, 50(1), 245–70, https://scholarlycommons.law.cwsl.edu/cgi/viewcontent.cgi?article=1563&context=cwilj.

Rowbottom, J., 2015. A landmark at a turning point: Campbell and the use of privacy law to constrain media power. *The Journal of Media Law*, 7(2), 170–95, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2711670.

See also: COMMON LAW, PAPARAZZI, PUBLIC FIGURE, PUBLIC INTEREST, STALKING, PUBLIC

Cell Suppression

A **statistical disclosure control** process where some cells (e.g., those containing low counts) in a table are redacted.

Cell suppression tends to be disliked by analysts who like complete data. It also needs to be done with care, as **bounds** can be placed on the **suppressed** cells in multidimensional tables which may reveal an approximate or even an exact value for a cell. This means that additional or secondary suppression must invariably be employed, whereby cells which are not themselves disclosive are also suppressed to increase the uncertainty about the disclosive cells.

Further reading:

Cox, L.H., 1980. Suppression methodology and statistical disclosure control. *Journal of the American Statistical Association*, 75(370), 377–85, https://doi.org/10.1080/01621459.1980.10477481.

Censorship

Censorship is the deliberate, and usually systematic, suppression of public expression, whether through speech, writing or broadcasting. Censorship is typically performed by governments, but anyone with power over a communication channel can censor. Where censorship is legalised and legitimate, there is often an official overseeing the process, called a *censor*. Censorship may help support privacy by suppressing disclosive expression, but is more likely in areas of political, religious and sexual discourse, suppressing discriminatory or hate speech and protecting **national security**.

Self-censorship occurs when someone censors their own output. This can happen because they do not wish to cause offence, or it may be that threats of punishment have a **chilling effect**. In the latter case, the coercion that leads to self-censorship breaches their **decisional privacy**.

Further reading:
Berkowitz, E., 2021. *Dangerous ideas: a brief history of censorship in the West from the ancients to fake news*. London: Westbourne Press.

See also: PUBLICATION, FREEDOM OF EXPRESSION, PUBLIC SPHERE, SECRET

Census

A form of data collection in which all members of a **population** are surveyed. Censuses have the property of being non-consensual and therefore present difficulties from a **data protection** point of view. Consequently, a lot of time is typically spent thinking about and determining appropriate levels of **statistical disclosure control** to place on the census outputs.

See also: CONSENT, STATISTICAL DISCLOSURE

Centralised Governance

While decentralised governance can take many different forms, centralised governance is more obviously identifiable as a model of regulation or organisation which revolves around a central decision-making authority. Examples would include companies within conventional market economies, as well as **data steward**s of large repositories of **personal data** (such as healthcare providers). Advocates of **privacy as control** tend to regard centralised forms of **information governance** as less compatible (or even incompatible) with individual control of **information** about them. This has led some commentators to call for new models of **data governance** to give individuals more decision-making power over their information, away from a central **data controller** who would otherwise decide the purpose and means of **data processing**.

Further reading:

Delacroix, S. and Lawrence, N.D., 2019. Bottom-up data trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, 9(4), 236–52, https://doi.org/10.1093/idpl/ipz014.

Kish, L. and Topol, E., 2015. Unpatients: why patients should own their medical data. *Nature Biotechnology*, 33, 921–4, https://doi.org/10.1038/nbt.3340.

Certification

A company can demonstrate its compliance with privacy regulation or best practice via certification schemes, whose criteria may be approved by **regulators**. Certification is usually voluntary, but will provide evidence of **compliance** for regulators, the public and other businesses.

The International Organization for Standardisation (ISO) and the International Electrotechnical Commission (IEC) are two of the most prominent international certification authorities with the authority to certify private bodies as compliant with best practice in industry **standards**. This can include implementation of adequate **privacy** and **data protection** safeguards by default within an organisation's day-to-day practice. Standards such as **ISO27001** and **ISO27002** on **information security** controls can provide an internationally recognised, independent seal of approval by an expert authority which should thus enhance the **trust** of third parties sharing **personal information** with the certified organisation. The EU **GDPR** gives national regulators (Supervisory Authorities) in member states the ability to accredit certification bodies, so that new data protection certification mechanisms can be established in those states.

Further reading:

Kamara, I., Leenes, R., Lachaud, E., et al., 2019. *Data Protection Certification Mechanisms*. Luxembourg: Publications Office. https://op.europa.eu/en/publication-detail/-/publication/4a30d394-8030-11e9-9f05-01aa75ed71a1/language-en.

Hornung, G. and Bauer, S., 2019. Privacy through certification: the new certification scheme of the General Data Protection Regulation. *In:* Rott, P., ed., *Certification: trust, accountability, liability*. Cham: Springer, 109–31.

See also: ACCOUNTABILITY

Certification Authority (CA)

In **public-key infrastructure** (PKI), it is essential that parties can rely on a **digital signature**, that is, that the signature could only be produced by a particular agent. **Cryptographically**, this means that the agent is the owner of the **public key** used to make the signature. In a PKI, this is established by a **certification authority**, a **trusted third party** that issues public key certificates asserting which named agent is the owner of the relevant key.

The International Telecommunication Union has issued a standard, X.509, defining the format of public key certificates.

Further reading:
International Telecommunications Union, 2021. *X.509: information technology – open systems interconnection – the directory: public-key and attribute certificate frameworks*. www.itu.int/rec/T-REC-X.509.

Chain of Trust

One problem with **public-key infrastructure** (PKI) is to ensure that the asserted link between an agent and their **public key** is authentic. A centralised solution to this is to **have certification authorities** hold **database**s of links, but flexibility may be facilitated with a hierarchical approach.

Where a **certificate** is produced and self-signed by a trusted authority within a PKI, this is called the *trust anchor* or the *root* **certification authority** (CA) certificate. There may be multiple intermediate CA certificates, each certifying another entity along the chain. The final intermediate CA certificate will in turn be used to sign the *end-entity CA certificate* which is issued to a website domain or other **public**-facing entity. Those interacting with the end entity will verify the chain of CA certificates, in which each CA certificate is certified by another certification authority, until the chain bottoms out at an accepted trust anchor. Flexibility is facilitated by the fact that any **user** can become a CA, issuing certificates, if they are underpinned by a chain of trust back to a trust anchor. Each authority is guaranteed by the previous one on the chain. The chain is necessarily finite, and so its inclusion of a **trust** anchor is determinable.

Further reading:
Martin, A., 2008. *The ten-page introduction to trusted computing*. Oxford: Software Engineering Group, Oxford University Computing Laboratory, https://ora.ox.ac.uk/objects/uuid:a4a7ae67-7b2a-4516-801d-9379d613bab4.

See also: PUBLIC-KEY CRYPTOGRAPHY, WEB OF TRUST

Challenge-Response

An approach to security called challenge-response is used to **authenticate** or confirm the **identity** of a **user**. It performs authentication by creating a ‘challenge’, a random string of characters or **data**, and then sending it to the user or system that needs to be authenticated. The user then creates a ‘response’ that is sent back to the system that issued the challenge using a secret key or **password**. The challenge-issuing system then compares the response to an expected value; if the responses match, the authentication is successful.

Several **security** applications including user authentication and **access control** use challenge-response mechanisms. The simplest form is simply the challenge ‘what is your password?’, but more sophisticated challenges involve **ciphers**; the system sends a **security token** – such as a string of alphanumeric characters – and the user must send the correct token back which will only be possible if they know the cipher. Another type of challenge-response is where the challenge is a puzzle that only a bona fide user will be able to solve (the principle behind Captcha systems whereby the user proves they are a human rather than a **bot**).

Challenge-response systems are not impenetrable, though, and can be subject to **replay attacks**, where a **hacker** intercepts a response and uses it to gain unauthorised access. To offer a stronger defence against attacks, challenge-response mechanisms are frequently used in conjunction with other security measures, such as **encryption**.

Further reading:
Kushwaha, P., Sonkar, H., Altaf, F. and Maity, S., 2021. A brief survey of challenge-response authentication mechanisms. *In*: *ICT analysis and applications: proceedings of ICT4SD*, volume 2, 573–81, https://doi.org/10.1007/978-981-15-8354-4_57.

Charter of Fundamental Rights

The Charter of Fundamental Rights of the European Union was signed on 7 December 2000 and brought many of the fundamental rights available under the **European Convention on Human Rights** (ECHR) into the scope of EU law. Furthermore, a specific **right to data protection** was introduced by the Charter, distinct from the **right to privacy** established by the ECHR. Some commentators, such as Lynskey, have argued that the introduction as a distinct right is appropriate, as the right to data protection gives individuals more rights in more types of **data**.

Kokott and Sobotta point to the **right to be forgotten** as a particular example of the wider scope of the right to data protection established by the Charter, as the **information** sought to be removed is not private (i.e., it has already been made **public**), but it is nonetheless **personal data**, and should not continue to adversely affect an individual without justification in **data protection** law.

Further reading:

Kokott, J. and Sobotta, C., 2013. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. *International Data Privacy Law*, 3(4), 222–8, https://doi.org/10.1093/idpl/ipt017.

Lynskey, O., 2014. Deconstructing data protection: the ‘added-value’ of a right to data protection in the EU legal order. *International and Comparative Law Quarterly*, 63(3) 569–97, https://doi.org/10.1017/S0020589314000244.

See also: CHARTER RIGHTS

Charter Rights

Rights established under the **Charter of Fundamental Rights**, including the **right to data protection**.

See also: DATA PROTECTION

Checksum

A checksum is a number generated by applying a function to a collection of **data** and is used to detect errors that might have occurred during data entry, transmission or **processing**.

In one typical example a checksum for data is created at the source and travels with the data; the receiver (who also knows the checksum function) then computes the checksum for the data that has been received. The likelihood that the data is accurate and was not corrupted during transmission or storage is high if the two checksums agree. The data may have been altered or corrupted during transmission if the checksums do not match, in which case appropriate steps can be taken to identify the problem and fix it.

Further reading:

Stone, J., Greenwald, M., Partridge, C. and Hughes, J., 1998. Performance of checksums and CRCs over real data. *IEEE/ACM Transactions on Networking*, 6(5), 529–43, https://doi.org/10.1109/90.731187.

See also: DATA STORAGE, INTEGRITY, TRANSPORT CONTROL PROTOCOL

Chief Privacy Officer

A chief privacy officer (CPO) is the senior executive in a corporation, organisation or government agency responsible for managing **privacy risk**s, setting **information** policy and strategy and coordinating responses with other senior executives.

‘Privacy officer’ can, in many jurisdictions, simply be a job title, indicating responsibility for an organisation’s internal measures to safeguard **personal data**, but lacking any specific statutory definition. In the United States, however, federal agencies are required by law to appoint a Chief Privacy Officer to assume primary responsibility for privacy and **data protection policy**. This is a continuation of the approach taken in the Privacy Act 1974, which governs only federal public bodies and not the **private sector**.

In the EU, the **GDPR** requires all organisations to appoint a **Data Protection Officer** if they process large volumes of **special category** personal data.

Further reading:

Fusaro, R., 2000. Chief privacy officer. *Harvard Business Review*, Nov–Dec 2000, https://hbr.org/2000/11/chief-privacy-officer.

See also: DATA CONTROLLER, DATA PROTECTION OFFICER, IMPACT MANAGEMENT, REPUTATION MANAGEMENT

Children’s Privacy

Children (i.e., young people aged under 18) can usually expect a higher degree of protection under privacy law and ethics. For example, the UK’s legally binding Children’s code provides age-appropriate design guidelines for online services likely to be accessed by children. This has inspired the California Age-Appropriate Design Code Act 2022, which (*inter alia*) introduced a duty of care on developers to prioritise children’s interests over their own commercial profit (where the two might conflict).

Under the (EU) **GDPR**, children should be able to expect greater safeguards around processing of their personal data as a default. This is particularly pertinent in the context of commercial online services such as **social networks**. Children under the age of 13 cannot provide **consent** for their data to be processed under the GDPR, which has led to popular applications such as TikTok and Instagram having a minimum user age of 13 in all countries, with higher ages of digital consent applicable in others.

It is widespread practice in the US and Europe for media organisations to blur unconsented images of children's faces prior to **publication**, as children are more likely to have a **reasonable expectation of privacy** than adults. In the UK, this dates in part from the decision in *Murray v Express Newspapers*, in which the author J.K. Rowling successfully obtained damages on behalf of her children. The children of famous and non-famous parents alike are judged entitled to walk around in **public** without having their actions publicised.

Further reading:

Duball, J., 2022. California Age-Appropriate Design Code final passage brings mixed reviews. https://iapp.org/news/a/california-age-appropriate-design-code-final-passage-brings-mixed-reviews/.

Information Commissioner's Office, 2022. *Age appropriate design: a code of practice for online services*. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/.

See also: CELEBRITY PRIVACY, MENTAL CAPACITY

Chilling Effect

The chilling effect is an ancillary phenomenon to the concept of privacy, but one which highlights its value. The chill in question is the unease in the civil population generated by an (actual or perceived) architecture of **surveillance** or **censorship**, which in turn stifles **freedom of expression**, because of the threat of the use of **information** against them. First coined by the US Supreme Court in 1952, the term has been used more recently by the Court of Justice of the European Union in their judgment in the *Digital Rights Ireland* case. As digital technologies generate ever greater volumes of **data** from interpersonal communication, privacy as a value which preserves the integrity of these information flows becomes even more socially and politically precious. However, authors such as Bedi have queried whether the chilling effect is a real social phenomenon, in which people are genuinely inhibited by fear of surveillance, or if this impact is in fact an imaginary **risk** codified by the US Courts as a legal principle.

Further reading:

Bedi, S., 2021. The myth of the chilling effect. *Harvard Journal of Law & Technology.* 35(1), 267–307, https://jolt.law.harvard.edu/assets/articlePDFs/v35/Bedi-The-Myth-of-the-Chilling-Effect.pdf.

Murray, D. and Fussey, P., 2019. Bulk surveillance in the digital age: rethinking the human rights law approach to bulk monitoring of communications data. *Israel Law Review*, 52(1), 31–60, https://doi.org/10.1017/S0021223718000304.

See also: DATAVEILLANCE

Chinese Wall

A Chinese wall is a barrier to the flow of **information** within an organisation, system or computer **network**, in order to prevent conflicts of interest or the **inadvertent disclosure** of **confidential** information or identifying information. Other terms such as '**firewall**' are now more usually used, as some have complained about the cultural insensitivity of the term.

Further reading:
Brewer, D.F.C. and Nash, M.J., 1989. The Chinese Wall security policy. *In: Proceedings of the 1989 IEEE Symposium on Security and Privacy*, 206–14, https://doi.org/10.1109/SECPRI.1989.36295.

Choice Architecture

An approach to the design of how choices are presented to individuals making decisions (often consumers or citizens), allowing subtle manipulation.

Further reading:
Thaler, R.H., Sunstein, C.R. and Balz, J.P., 2013. Choice architecture. *In:* Shafir, E., ed., *The behavioral foundations of public policy*. Princeton: Princeton University Press, 428–39.

See also: DECISIONAL PRIVACY, NUDGE THEORY

CIA Triad

CIA stands for Confidentiality, Integrity and Availability, defining the CIA triad, a model used to describe the three key aspects of **information security**.

Confidentiality refers to the capacity of making accessible **sensitive data** to **authorised users**. Measures such as **encryption**, **access control** and **data classification** enhance confidentiality.

Measures such as data validation, error checking and **checksum** usage enhance **integrity**. In this case, it is ensured that **information** is not corrupted when stored or transmitted.

Availability is defined as the ability of authorised users to access information when they need it. To improve availability, it is important to implement redundancy, **back-up** and **risk** recovery assessment to ensure that **data** is accessible even in case of unexpected events.

The origin of the construct is unclear but an early mention of it can be found in a paper by Neumann et al.

Further reading:

Neumann, A.J., Statland, N. and Webb, R.D., 1977. Post-processing audit tools and techniques. *In*: *Proceedings of the NBS Invitational Workshop*, 11–3, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nbsspecialpublication500-19.pdf.

Cipher

A cipher or cypher is a procedure for either the **encryption** or **decryption** of a message. The decrypted message is referred to as **plaintext** or *in the clear*. Having been encrypted, it should not be easily decrypted again without possession of the cipher procedure. The procedure typically will generate alternatives to replace the symbols in the plaintext message. A **brute force attack** on a cipher, which involves trying all possible decryption procedures, will eventually succeed, but if the cipher is complex enough, it cannot be guaranteed to work in an acceptable amount of time.

Many ciphers exploit auxiliary **data** in the form of an **encryption key** or *cryptovariable*. The key is used as part of the encryption/decryption procedure, so that an intruder must know both the procedure and the key. Different keys will result in different alternatives being generated, and so the key must be fixed before initial encryption. The key must be known to the receiver and the sender of the encrypted message and protected from the intruder. It is possible to use two keys, one for encryption and one for decryption. If it is hard or impossible to deduce one key from the other, then only one key need be kept private, the principle behind **public-key cryptography**.

Further reading:

Katz, J. and Lindell, Y., 2008. *Introduction to modern cryptography*. Boca Raton, FL: Chapman & Hall/CRC.

Ciphertext

Ciphertext is the result of applying an **encryption algorithm** to **plaintext**. It is unreadable by any **third party** until decrypted back into the original plaintext. Once encrypted the text can be sent over the Internet or stored at minimal **risk**. Only if the recipient has the right **encryption key**, using the same encryption algorithm, will they be able to decrypt the ciphertext back to the original form. The strength of the encryption depends on the complexity of the encryption algorithm and the length of the key used.

Further reading:
Bethencourt, J., Sahai, A. and Waters, B., 2007. Ciphertext-policy attribute-based encryption. *In: IEEE Symposium on Security and Privacy*, 321–34, https://doi.org/10.1109/SP.2007.11.

See also: CIPHER, ENCRYPTION

Classified Information

Information that a government or state identifies as sensitive is typically referred to as classified information. When information is classified, it is usually the case that access to it is limited to those with a sufficient level of **security** clearance, with criminal penalties for unauthorised access. Typical classification levels include **restricted**, **confidential**, **secret** and top secret.

Further reading:
Goldman, J. and Maret, S.L., 2016. *Intelligence and information policy for national security: key terms and concepts*. Lanham, MD: Rowman & Littlefield.
Maret, S.L. and Goldman, J., eds, 2009. *Government secrecy: classic and contemporary readings*. Westport, CT: Libraries Unlimited.

See also: INFORMATION SECURITY, NATIONAL SECURITY

Cleartext

See: PLAINTEXT

Clickstream Data

The electronic log of **user** activity on a website is referred to as clickstream data. It can, for instance, include: the order in which users interact with digital content and services; the time and date of the interactions; the user's **location data**; the pages viewed; the length of the visit; the **search** terms entered; and any actions the user took, such as filling out a form, making a purchase or clicking a link.

Predictive analytics tools frequently gather clickstream data, which can then be examined using a variety of methods, including **data mining**, **machine learning** and statistical analysis, to find patterns, trends and correlations that can help businesses improve their marketing strategies. Businesses frequently use clickstream data to better understand the behaviour, preferences and interests of their customers.

Further reading:

Baumann, A., Haupt, J., Gebert, F. and Lessmann, S., 2019. The price of privacy: an evaluation of the economic value of collecting clickstream data. *Business & Information Systems Engineering*, 61, 413–31, https://doi.org/10.1007/s12599-018-0528-2.

Olbrich, R. and Holsing, C., 2011. Modeling consumer purchasing behavior in social shopping communities with clickstream data. *International Journal of Electronic Commerce*, 16(2), 15–40, https://doi.org/10.2753/JEC1086-4415160202.

See also: BEHAVIOURAL ADVERTISING, ECONOMICS OF PRIVACY

Client Confidentiality

The **common law** duty of **confidentiality**, a descendent from equitable duties of confidence, has long recognised certain professionals as owing a **duty of confidence** to their clients. While doctors are referred to as bound by patient confidentiality, the general duty of non-disclosure owed by non-medical professionals (such as lawyers or counsellors) is similar.

However, the concept of legal professional **confidence** (and the linked doctrine of legal professional privilege) has a much older lineage and predates the expansion of the law of confidentiality as a more general obligation (see Toulson and Phipps). The terms 'legal privilege' and 'client confidentiality' are distinct: the former refers more specifically to an exemption from disclosure in court proceedings, whereas confidentiality is a more general prohibition on sharing **information** with the wider world (subject to exceptions in the **public interest**).

Further reading:

Imwinkelried, E.J., 2011. The dangerous trend blurring the distinction between a reasonable expectation of confidentiality in privilege law and a reasonable expectation of privacy in Fourth Amendment jurisprudence. *Loyola Law Review*, 57(1), 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1721820.

Toulson, R.G., and Phipps, C.M., 2012. *Confidentiality*, 3rd ed. London: Sweet & Maxwell/Thomson Reuters, Chapters 16 and 18.

See also: BREACH OF CONFIDENCE, REASONABLE EXPECTATION OF PRIVACY

Client-Side Scanning

The process of scanning and analysing data or applications on the client's device. It entails utilising **security** tools installed on the device to identify **malware** or what is referred to as 'objectionable content' (e.g., material relating to terrorist activity or child sexual abuse). The mechanism works by comparing a hashed form of the content against known hashes of objectionable content.

However, it is also a source of concern as the mechanism might also be used for **surveillance** or **censorship**. Client-side scanners have also been shown to represent a security **risk** as they are incompatible with **end-to-end encryption**.

Further reading:

Abelson, H., Anderson, R., Bellovin, S.M., Benaloh, J., Blaze, M., Callas, J., Diffie, W., Landau, S., Neumann, P.G., Rivest, R.L. and Schiller, J.I., 2021. Bugs in our pockets: the risks of client-side scanning, *arXiv*, https://doi.org/10.48550/arXiv.2110.07450.

Jain, S., Crețu, A. and de Montjoye, Y.-A., 2022. Adversarial detection avoidance attacks: evaluating the robustness of perceptual hashing-based client-side scanning. *In*: *31st USENIX Security Symposium (USENIX Security 22)*, 2317–34, www.usenix.org/conference/usenixsecurity22/presentation/jain.

Closed Circuit Television (CCTV)

A CCTV system consists of video cameras placed in an environment, delivering images of that environment to a central monitoring point. This enables the **surveillance** of the environment, which could for example be a specified geographical location (ranging from a single building to a neighbourhood), a public transport network, a road network, a school or a workplace. Its purposes may include monitoring automated industrial processes, crime prevention or solution, **security**, optimising parameters such as traffic flow or workflow, prevention of bullying and vandalism in schools, or, at the extreme, suppressing political freedoms and preventing organised political action.

CCTV has become more ubiquitous, almost routine, across the world in recent years. Its capabilities improve alongside technology. In its early days, CCTV pictures were often displayed on monitors for real-time observation by individuals such as security guards. Video recording allowed asynchronous playback by the 1970s, and present-day systems benefit from the miniaturisation of cameras, Wi-Fi connections across the Internet, object tracking, **face recognition** systems, and other types of video analytics. CCTV used to be the preserve of governments and large corporations, but lower prices mean it is increasingly used by private citizens, for example to monitor their homes.

CCTV has been a more prominent issue for activists than for the wider **public** in most nations. The utilitarian benefits of CCTV are ranged against claims of **intrusion**s into the **dignity** of those under surveillance. Privacy may be considered **breach**ed either by the presence of cameras in a **public** space, or by access being provided to recordings later. Legal regulation of CCTV therefore should specify who is allowed access, under what circumstances, and whether there is a requirement to destroy the images after a certain period. Images of identifiable individuals are defined as **personal data** under EU **data protection** law.

Further reading:

Slobogin, C., 2002. Public privacy: camera surveillance of public places and the right to anonymity. *Mississippi Law Journal*, 72, 213–315, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=364600.

Yao, Y., Basdeo, J.R., Mcdonough, O.R. and Wang, Y., 2019. Privacy perceptions and designs of bystanders in smart homes. *In: Proceedings of the ACM on Human–Computer Interaction*, 3, article no. 59, https://doi.org/10.1145/3359161.

See also: PUBLIC SPHERE

Cloud Computing

A model for providing on-demand **Internet** access to a shared pool of computing resources, including servers, storage and applications. Although cloud computing has many advantages including access to large **cloud storage** capacity and high-performance computing, there are also privacy and **security** risks. There is an increased **risk** that malicious actors could intercept, steal or compromise cloud data because it is kept on remote servers and accessed online. Strong **encryption** and **access control** measures such as **multi-factor authentication** and **role-based access control**, can help to reduce this risk.

Data from multiple customers may be stored on the same servers because cloud service providers frequently use shared infrastructure and resources. Due to this, there is a risk that a **user** may unintentionally or maliciously gain access to another's data. Strong data segregation and access control procedures, as well as checking that the cloud service provider has the necessary security and privacy policies in place could help mitigating these risks. Finally, there are concerns associated with **data sovereignty** and regulatory **compliance** when using cloud computing; the location of the servers where data are stored and processed may be out of the user's **jurisdiction**.

Further reading:

Chen, D. and Zhao, H., 2012. Data security and privacy protection issues in cloud computing. *In: 2012 International Conference on Computer Science and Electronics Engineering*, 1, 647–51, https://doi.org/10.1109/ICCSEE.2012.193.

Xiao, Z. and Xiao, Y., 2012. Security and privacy in cloud computing. *IEEE Communications Surveys & Tutorials*, 15(2), 843–59, https://doi.org/10.1109/SURV.2012.060912.00182.

See also: DATA STORAGE, REMOTE ACCESS

Cloud Storage

Cloud computing is an architecture for data storage and software supply in which computing resources are distributed across one or more servers owned by a cloud host. A typical cloud provider will store data in more than one location, to introduce redundancy for extra **security**. The data may be encrypted, so that the provider is unable to access it in the clear.

However, despite encryption, it does come with some privacy and **security** concerns. First, while 'the cloud' is a deliberately intangible metaphor, its servers are located in **data centres** in specific **jurisdiction**s which will affect the cloud provider's legal responsibilities to its clients. The clients would be wise to ensure that, for instance, the provider is compliant with relevant regulations. The jurisdiction will also affect how accessible the data is to government and law enforcement agencies.

Second, the responsibility for data security is transferred from organisations and individuals to the cloud provider. This has the advantage that the provider has incentives to invest in state-of-the-art security expertise to reap economies of scale, but the disadvantage that centralised **data storage** is a target for **hacking**. Furthermore, redundancy in **data storage** means that the data will be replicated, increasing the chances that the servers upon which it sits might be attacked. Cloud providers also typically employ many more people (who then have access to the data and have no reason to be loyal to the client), who may be bribed or become targets of **social engineering** attacks.

Third, the data is usually accessed by users over standard **Internet** connections, rather than secure in-house local area **network**s.

Some of these issues can be mitigated by large organisations by developing their own private cloud storage, rather than using a public cloud provider.

Further reading:

Pal, S., Le, D.-N. and Pattnaik, P.K., eds, 2022. *Cloud computing solutions: architecture, data storage, implementation, and security*. Hoboken, NJ: John Wiley & Sons.

Code Audit

A code audit is an examination of **software** code to identify bugs, or **security** problems that can jeopardise the security, dependability or functionality of a program. Programming mistakes such as buffer overflow vulnerabilities or code injection weaknesses can be detected with the aid of a code audit. Moreover, it can spot broader issues like ineffective coding, bad design decisions and significant performance bottlenecks.

A code audit can be static (assessment of code before the software is run) or dynamic (assessment of the code during or after it is run) and used standalone or in combination with other tools such as **penetration testing** and dependency analysis.

The term ‘code review’ is sometimes used as a synonym for ‘code audit’, although the consensus in the literature is that reviews are more informal processes that may be smaller in scale (perhaps covering just part of a piece of software), whereas audits are more formal and included in due diligence processes that should encompass the entirety of a software project.

Further reading:

Bacchelli, A. and Bird, C., 2013. Expectations, outcomes, and challenges of modern code review. *In: Proceedings of 35th International Conference on Software Engineering*, 712–21, https://doi.org/10.1109/ICSE.2013.6606617.

Edmundson, A., Holtkamp, B., Rivera, E., Finifter, M., Mettler, A. and Wagner, D., 2013. An empirical study on the effectiveness of security code review. *In: Proceedings of Engineering Secure Software and Systems: 5th International Symposium*, 197–212, https://doi.org/10.1007/978-3-642-36563-8_14.

See also: BUFFER OVERFLOW ATTACK, INTERNAL SECURITY TESTING, SECURITY AUDIT, SQL INJECTION

Code of Conduct

While **information security** standards such as **ISO27001** and **ISO27002** provide general guidance on internationally recognised best practice, codes of conduct tend to be more sector-specific. Key actors and stakeholders within (for example) the **cloud computing** industry, or the **genomics** research landscape, can come together to agree principles and guidelines for **data protection** in these limited contexts.

Under Article 40 of the EU **GDPR** it is now possible for a national **data protection authority** to approve codes of conduct, which can then form the basis for sharing **personal data** with **third parti**es outside the EU who also comply with the code.

Adherence to an approved code of conduct can also be a way for a **data controller** to ascertain whether a **data processor** they instruct has sufficient **security** measures in place, and to demonstrate the sufficiency of their own security measures more generally (see Article 32, GDPR). As a means of demonstrating **compliance**, codes of conduct differ from **certification** mechanisms (such as ISO standards) partly due to their sector-specificity, but also because they do not require an authorised body to certify compliance. Both, however, can be ways of demonstrating compliance with the GDPR.

Further reading:

Calder, A., 2021. *EU code of conduct for cloud service providers: a guide to compliance*. IT Governance Publishing, https://doi.org/10.2307/j.ctv22d4zj7.

Knoppers, B.M., Harris, J.R., Tassé, A.M., Budin-Ljøsne, I., Kaye, J., Deschênes, M. and Zawati, M.N.H., 2011. Towards a data sharing Code of Conduct for international genomic research. *Genome Medicine*, 3, 1–4, https://doi.org/10.1186/gm262.

See also: ACCOUNTABILITY, DATA-PROTECTION-BY-DESIGN, SUPERVISORY AUTHORITY

Code of Ethics

A code of ethics or *ethical code* is a statement of the values of an organisation intended to help its members, employees and agents in the ethical aspects of their decision-making. In that respect, it is similar to a **code of conduct** or code of practice, with a more prominent focus on moral values. A code of ethics would normally be publicly available, to help collaborators and customers as well as being, more pragmatically, a public relations tool. As the organisation is already bound to be **compliant** with existing regulations, the code should commit it to going beyond its legal responsibilities.

Organisations may use codes of ethics to announce policies that affect privacy positively, such as voluntarily committing to refraining from using **personal data** in certain ways, even if legal. However, the phrase ‘**data ethics**’ is widely used, often without precision, in ways that can overlap with legal compliance. As the legal principles within **data protection** are so broad, some have argued that ethical codes can help organisations interpret and apply concepts such as ‘fairness’ and the ‘**public interest**.’

Further reading:

Adams, J.S., Tashchian, A. and Shore, T.H., 2001. Codes of ethics as signals for ethical behavior. *Journal of Business Ethics*, 29(3), 199–211, https://doi.org/10.1023/A:1026576421399.

Rochel, J., 2021. Ethics in the GDPR: a blueprint for applied legal theory. *International Data Privacy Law*, 11(2), 209–23. https://doi.org/10.1093/idpl/ipab007.

See also: INFORMATION ETHICS, PRIVACY POLICY

Code Review

See: CODE AUDIT

Commodification

Commodification is the process of adapting an object or service for exchange, sale or exploitation under a capitalist system. Many aspects of the **digital economy** rest on the commodification of information, particularly **personal data**. This has created interest in increasing and making more efficient the information flow about **data subject**s, counterbalancing and sometimes outweighing the privacy interests of individuals.

Furthermore, the ability of **cookie**s and devices to **track** activity means that many kinds of behaviour can be rendered as **data**, thus making them similarly subject to commodification. So-called **surveillance capitalism** is an adaptation of capitalism where data about individuals’ behaviour has become fungible and tradable at scale.

Further reading:

Zuboff, S., 2019. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. London: Profile.

See also: ECONOMICS OF PRIVACY

Common Law

The common law is made up of by precedents set by the courts, rather than laws written in statute. While almost all legal systems have courts and judges, in common law **jurisdiction**s a judge’s decision can bind future judgments under the doctrine of precedent, particularly if that judge sits in a higher court. In many common law systems, for example, there is no **right to privacy** which has been drafted by the national legislature, and instead most privacy law has come from judge-made law.

Common law was introduced by Henry II in 12th-century England, and subsequently spread throughout the British Empire and Commonwealth. Historically, the common law courts administered the ‘King’s law’, whereas the Chancery Courts (or ‘Courts of Equity’) exercised the Lord Chancellor’s discretion to remedy any unfairness stemming from a gap in the common law. The **duty of confidence** (an early ancestor to the modern-day right to privacy) originally came from the courts of equity. Even in countries which still have equity law, however, the distinction between equitable duties of confidence and common law privacy rights has become less clear since the 20th century. The introduction of the Human Rights Act 1998, which incorporated within UK law the rights contained in the **European Convention on Human Rights**, has made the **reasonable expectation of privacy** a more important touchstone for privacy rights in England. Other common law systems have developed their own privacy rights, either through an explicit constitution (as in the United States) or through precedents set within the Courts.

Further reading:

Moreham, N.A., 2021. Conversations with the common law: exposure, privacy and societal change. *Wellington Law Review*, 52(3), 563–77, https://doi.org/10.26686/vuwlr.v52i3.7332.

Potter, H., 2015. *Law, liberty and the constitution: a brief history of the common law.* Rochester, NY: Boydell & Brewer.

See also: PRIVACY TORT

Communication

Communication between parties is the act of sending **information** from one to the other, and possibly back again. This covers an enormous range of behaviour, from termites leaving chemical signals for their fellow colonists, to human **publication** of sophisticated novels; from the adoption of facial expressions, to painting signs on the roadside. As a result, several disciplines study communication, including information theory, logic, semiotics, (sociological) communications studies and biosemiotics.

From the perspective of privacy, the key relevance of communication is that sometimes, communicating parties want to send messages that they do not want to be intercepted and understood by a **third party**. Communication, which takes place within a medium through which symbols are passed, must therefore consider the opportunities for interception. Either the medium must be closed to eavesdroppers, or the message itself must be scrambled, encrypted or **disguised** (for instance through **steganography**). Messages may also be decomposed so that any eavesdropper must devote more resources to capturing them, and their costs exceed the benefits. The medium itself will also add background noise to the message, which will further complicate the task of the eavesdropper.

The discipline most suited to this study is that of information theory, which studies the storage and transfer of (usually digitally) represented information, as developed by Claude Shannon and others in the 1940s, which understands communication as a series of processes: *encoding* the intended message; *transmitting* it through the medium; *receiving* it at the other end; *decoding* and *interpreting* it. Communication is successful when the interpreted message is the same as the encoded one. **Encryption** methods can be used to render the **communication** secure. The eavesdropper is successful if they can produce the intended interpretation. Interception of the message is called a **man-in-the-middle attack**.

Further reading:

Blahut, R.E., 2014. *Cryptography and secure communication*. Cambridge: Cambridge University Press.

Holden, J., 2017. *The mathematics of secrets: cryptography from Caesar ciphers to digital encryption*. Princeton: Princeton University Press.

See also: COMMUNICATION PRIVACY, CRYPTOGRAPHY, SECURE COMMUNICATION

Communication Privacy

If two agents communicate with each other, whether by speech, sign, letter, telephone, email or online messaging, they have communication **privacy** when no **third party** can intercept or amend the **communication**. This type of privacy is relatively easy to formalise and is the focus of many approaches to **cryptography**. A common scenario for **secure communication** is that Alice has a message that she wishes to send securely to Bob, without **eavesdropping** Eve being able to understand it.

Cryptography and related approaches to communication privacy aim to render the communication impenetrable to an interceptor, or alternatively to make it impossible for the interceptor to amend undetectably. Most obviously this would involve **encrypting** the message, and codes and **cipher**s were invented relatively early in history, but there are alternatives, such as **steganography** and **obfuscation**. A particular issue with encryption is that the sender not only needs to send the message to the recipient, but they also need to communicate the **encryption key** securely to allow the recipient to decrypt the message; **public-key cryptography** was invented to address that problem. Where the communication is oral and face-to-face, the communicators might whisper. At sporting events where cameras are ubiquitous, coaches often discuss tactics with players with their hands in front of their mouths, to prevent lip-reading.

A different approach is to disguise the fact that communication has taken place at all, either by using anonymous methods, such as unregistered mobile phones or public call boxes, drop addresses for written communications or complex onion routing methods such as The Onion Router (**TOR**) for online communication.

Further reading:

Blahut, R.E., 2014. *Cryptography and secure communication*. Cambridge: Cambridge University Press.

Singh, S., 1999. *The code book: the secret history of codes and code-breaking*. London: Fourth Estate.

See also: HISTORY OF PRIVACY, SECURE MESSAGING, THIRD PARTY

Communication Privacy Management (CPM) Theory

The CPM theory is a framework that clarifies how people handle the privacy of their **personal data** when engaging with others.

The principles governing CPM are **data ownership**, control and **self-disclosure**. The framework also establishes that the **boundari**es on **information** flow are subject to negotiation through interpersonal interaction. Agents in CPM must find a balance between their need for privacy and their need for social connection and are envisaged as weighing up the costs and benefits of any **disclosure** to another. As O’Hara observes, cultural norms and personal **privacy preference**s mediate the perception of this trade-off and technological developments from the printing press to **social media** also impact on the operation of CPM at the individual level.

CPM conceptually ties **informational privacy** to **confidentiality**, with both being grounded in the notion of boundaries in information flows, but with the critical distinction being that in the CPM model of privacy both the boundaries and the flows are controlled by the **data subject** rather than another party. This makes it a **privacy** framework, rather than an aid to confidentiality.

Further reading:

Hollenbaugh, Erin E., 2019. Privacy management among social media natives: an exploratory study of Facebook and Snapchat. *Social Media + Society*, 5(3), https://doi.org/10.1177/2056305119855144.

Petronio, S., 2010. Communication privacy management theory: what do we know about family privacy regulation? *Journal of Family Theory & Review*, 2(3), 175–96, https://doi.org/10.1111/j.1756-2589.2010.00052.x.

See also: BENEFITS OF PRIVACY, COMMUNICATION, COMMUNICATION PRIVACY, PRIVACY AS CONTROL

Community Privacy

Privacy is often understood to be a property of individuals or groups. A specific type of privacy that is also discussed is community privacy, where an organisation or group (such as a company or a political group) has requirements for **confidentiality** that it needs to defend. **Information** may be sensitive to the community but not sensitive for individuals within it. For instance, calendars of meetings or the work-related emails of employees may affect outside perceptions of the organisation. Some dispute that requirements for confidentiality are genuine cases of privacy; the case that organisations *do* have privacy requirements was first made by Alan Westin in his survey of privacy in the information age.

Further reading:

Codio, S., Kafura, D., Pérez-Quiñones, M., Gracanin, D. and Kavanaugh, A., 2012. A case study of community privacy. *In: 2012 International Conference on Social Informatics*. IEEE, https://doi.org/10.1109/SocialInformatics.2012.30.

Westin, A., 1967. *Privacy and freedom*. New York: Ig Publishing.

See also: GROUP PRIVACY, TRANSPARENCY

Compliance

Compliance refers not only to the success state of adherence to the law, but also collectively to the strategies, policies and other steps taken to achieve this state of adherence. An organisation may, through choice or requirement, employ a Compliance Officer, with the responsibility for ensuring compliance with relevant laws, regulations and policies. When the focus of compliance is primarily on **information governance** laws this role may also be called a **Privacy Officer** or **Chief Privacy Officer**.

See also: ACCOUNTABILITY

Concentration Rule

See: DOMINANCE RULE

Conditions for Processing

Under the EU **GDPR**, all uses of **personal data** must satisfy a **legal basis for processing** under Article 6 (e.g., **consent**, **legitimate interest** or **public interest**). Some personal data is considered particularly sensitive and included within **special category data**. The special categories include personal data concerning health, **genomics data** and **information** relating to trade union membership, or to a person’s sex life or sexual orientation.

The use of this special category personal data must also satisfy a **condition for processing** under Article 9 GDPR. This is because this data is deemed higher risk to data subjects’ rights, and therefore requires additional justification.

The conditions which can justify the use of such data include explicit consent (a higher evidential bar than the unambiguous consent required as a legal basis for processing under Article 6), as well as reasons of substantial public interest as set out in EU or Member State law. Most EU Member States have **data protection** legislation setting out these substantial public interest conditions.

Further reading:

Information Commissioner’s Office, n.d. *Special category data*. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/#scd3.

Confidence

In relation to privacy, confidence has two different meanings.

First, in a technical sense, it is the level of certainty with which an **adversary** conducting a **disclosure** attack such as **reidentification** is confident that the resulting disclosure is correct. As such, it is likely to include a subjective element. A subsidiary aim of **privacy risk** management is to ensure that, even if identifications cannot be prevented, that the adversary will have low confidence in them. A match is rarely perfect, and the adversary's confidence will depend on such matters as how typical the putatively identified subject is of the **population**, whether it is known that the subject is in the **dataset** (**response knowledge**), how close the match is between the **data** and what is already known about the subject, and so on.

Second, in a legal sense, **information** may be given to someone *in confidence*. This means that the giver of information places **trust** in the receiver not to relay it further. Where there is a **duty of confidence**, the giver may demand compensation and remedy from the receiver for **breach of confidence**. Information given under such conditions is referred to as **confidential**.

Further reading:

Elliot, M., Mackey, E., O'Shea, S., Tudor, C. and Spicer, K., 2016. End user licence to open government data? A simulated penetration attack on two social survey datasets. *Journal of official statistics*, 32(2), 329–48. https://doi.org/10.1515/jos-2016-0019.

Phipps, C.M., Harman, W.R. and Teasdale, S.T., 2020. *Toulson and Phipps on confidentiality*, 4th edition. London: Sweet & Maxwell.

See also: DISCLOSURE RISK, STATISTICAL DISCLOSURE

Confidentiality

Confidentiality is a state in which a **boundary** delimits the flow of **information** or data. This quality of confidentiality (and corresponding **duty of confidence** to treat information as confidential) arises from multiple legal and ethical obligations and is delivered through personal, organisational or technical measures which constrain the **data flow**s so that they are consistent with those legal and ethical obligations. The obligations could arise from a term of an employment contract, a requirement under a professional **code of conduct**, a private-law obligation, an informal **agreement** between two citizens or an obligation the state may owe a citizen in respect of some services.

The idea of an obligation of confidence in England dates back to the 16th century, with the obligations of professional lawyers being one of the earliest species of the duty. The term was popularised in the 18th century, in which some correspondence could be deemed ‘confidential’ because of the **secrecy** and **intimacy** between the parties. The term speaks of the **trust** and **confidence** which form a cornerstone of the relationship in question. This connotation survives today but has been assimilated into the lexicon of professional ethics, hence the sub-types of confidentiality: doctor–patient, lawyer–client, and general **client confidentiality**.

A specific kind of relationship is not necessary, however, for **communications** between parties to be deemed confidential. Particularly following the judgment of the European Court of Human Rights in *von Hanover v Germany*, confidentiality has become associated with the concept of a **reasonable expectation of privacy**. This means that an external observer, such as a court, would look to all the circumstances of the case (the connection between the parties, the nature of the communication, the **risk** of **harm** from **disclosure** and any notice given of disclosure requirements) to decide whether disclosure constitutes a **breach of confidence**.

The element of exclusivity of communication nonetheless survives in the 21st-century understanding of confidentiality. Most notable in Europe is the requirement of the **ePrivacy Regulation** that both the content and ‘external elements’ of communications (i.e., **metadata**) are not revealed to anyone other than the parties involved, except in strictly prescribed circumstances (such as for law enforcement purposes under national law). The **GDPR** associates confidentiality with technical **security** measures to maintain this **integrity** of information, illustrating the journey the concept has taken from its origins in notions of human trust and intimacy. Faith in an individual confidante is no longer the sole consideration: confided information now forms part of a complex digital ecosystem which must be kept secure from unauthorised or unexpected actors for an individual’s confidence to be well founded.

Further reading:

Duncan, G.T., Elliot, M. and Salazar-González, J.J., 2011. *Statistical confidentiality: principles and practice*, New York: Springer.

Toulson, R.G., and Phipps, C.M., 2012. *Confidentiality*, 3rd edition. London: Sweet & Maxwell/Thomson Reuters.

See also: COMMUNICATION PRIVACY, PRIVACY NOTICE, PRIVACY TORT

Confidentiality Club

In the ordinary course of civil litigation, both parties must publicly **disclose** relevant documents even if they contain **confidential information** (unless they are legally privileged, e.g., created in anticipation of litigation). In exceptional circumstances, however, the Courts in England and Wales have the discretion to impose a confidentiality club which specifies who can view certain documents and in what circumstances. Following case law from the **European Court of Human Rights**, this should only be imposed where it is a proportionate response to an evidenced **risk** to life, limb or property. In practice, it is risk to **intellectual property** which forms the most common ground for imposing confidentiality clubs.

An *external eyes-only club* (EEO club) is an exceptionally strict version of this, where only those who are external to the litigants (e.g., counsel, external solicitors, independent consultants) can access the documents.

Further reading:
Garbett, K. and Preston-Jones, R., 2016. In the club: confidentiality clubs. *New Law Journal*, 166(7702), 11, www.newlawjournal.co.uk/content/club.
Nnachi, R., 2021. Non-party disclosure and confidentiality clubs: Bugsby Property LLC v LGIM Commercial Leasing Ltd and another. *Practical Law Dispute Resolution Blog*, Thomson Reuters, http://disputeresolutionblog.practicallaw.com/non-party-disclosure-and-confidentiality-clubs-bugsby-property-llc-v-lgim-commercial-leasing-ltd-and-another/.

See also: COMMON LAW

Confidentiality Pledge

A formal assurance given to **data subject**s that their **data** will be **processed** in a manner that maintains **confidentiality**. This is particularly important where the data are collected without **consent**, such as in national **censuses**. A functional pledge will go beyond vague assurances and will list specific actions to be taken. Here is an example from the 2011 UK census: ‘The **information** you provided to us in the 2011 Census is confidential and protected by law. The confidentiality of personal information is a top priority for the census. Your personal census information is not shared with any other government department, local councils or marketing companies.’

Confidentiality pledges are a legal requirement for federal statistics agencies in the United States. The term has also been used by the UK Office for National Statistics, although it does not carry the same legal connotations outside the US. Even if a precise pledge is not required in other **jurisdictions**, national statistics agencies will still need to secure public **trust** and protect confidential information.

Further reading:
Redline, C. and Tuttle, A.D., 2022. In an era of enhanced cybersecurity: the effect of disclosing a third party's role in confidentiality pledges on response propensity. *Journal of Survey Statistics and Methodology*, 10(3), 500–17, https://doi.org/10.1093/jssam/smac009.

Conflict of Rights

The idea of a conflict of rights should be understood as distinct from a conflict of laws, which is where a set of facts could be governed by the laws of more than one **jurisdiction**. A conflict of rights, on the other hand, is a term commonly used in the context of EU or human rights law when a case falls within the scope of more than one legal right, each of which could lead to a different outcome if it formed the basis of adjudication by a court.

In the context of **privacy**, a common conflict is between an individual's privacy and another individual's **freedom of expression**. This conflict has been seen in **defamation** cases but has more recently gained new prominence following the inception of the **right to be forgotten** in the EU. Taylor, however, cautions against false dichotomies in our understanding of fundamental rights, and has emphasised the mutual interest of both the individual and the **public** of protecting **confidentiality** *and* carefully using data for health research.

Further reading:
Taylor, M., 2017. Protecting confidentiality and improving care: not a zero sum game. www.gov.uk/government/speeches/protecting-confidentiality-and-improving-care-not-a-zero-sum-game.

See also: PUBLIC INTEREST, RIGHT TO PRIVACY

Connected Place

A generalisation of the **smart city** concept which acknowledges that a location does not have to be a city to be smart.

Further reading:
NCSC, 2021. *Connected places: cyber security principles*, Cheltenham: National Cyber Security Centre, www.ncsc.gov.uk/files/NCSC-Connected-Places-security-principles-May-2021.pdf.

See also: SECURITY

Connectomics

The study of neural connections and specifically the construction and interpretation of maps of neural connections at multiple scales from individual synapses to functional units such as the retina. These processes result in large **dataset**s, some of which have been made **publicly** available. The potential value of these in health research is a big driver for the field; by comparing diseased and healthy connectomes, researchers hope to gain insight into structural and functional substrates of many conditions.

Although arguably not the most pressing **neuroethics** issue, this type of mapping has many of the properties of **genomics data**, a health-led **big data** initiative with difficult to anticipate but significant downstream impacts on individual privacy.

Further reading:
Laird, A.R., 2021. Large, open datasets for human connectomics research: considerations for reproducible and responsible data use. *NeuroImage*, 244, https://doi.org/10.1016/j.neuroimage.2021.118579.

See also: NEUROPRIVACY

Consent

Consent is commonly characterised as a paradigmatic aspect of **autonomy** and individual **privacy**, a fundamental way in which an individual can regulate incursions into any aspects of themselves or their property which could (legally, socially or ethically) be considered private. Inherent within the concept, therefore, is a vexed question as to the scope of activity individuals (with different ages, health conditions and levels of social power) can legitimately be expected to deliberate and regulate. Even if the scope of phenomena genuinely legitimated by individual consent can be agreed, the level of evidence required for consent can still be nuanced and subject to change over time.

Under the EU's **GDPR**, consent means any freely given, specific, informed and unambiguous indication of the **data subject**'s wishes by which they, by a statement or a clear affirmative action, signify **agreement** to the processing of **personal data** relating to them. 'Free' in this context means without fear of suffering detriment. The GDPR requires an appropriate balance of power between the **data controller** and data subject for the consent to be freely given, suggesting that many public authorities will not be able to rely on consent when processing citizens' **information**.

The consent requirements of the GDPR are in some senses a gold standard, preventing exploitation in reliance on an individual's apparent acceptance. It is, however, a difficult standard to meet in many contexts: even where consent is freely given, specificity can be difficult to maintain if **data processing** changes over time. The 'without fear of detriment' criterion is difficult to satisfy in many consent relationships – in particular, for consent to medical treatment, where fear of detriment can be a key motivator in a patient's risk–benefit analysis.

Consent is therefore best understood as a legitimating act of individual acceptance, the scope and nature of which will vary according to the nature of the proposed intervention.

Further reading:

Sheldon, S. and Thomson, M., 1998. *Feminist perspectives on healthcare law*. London: Cavendish.

Stroud, F., 1890. *Stroud's juridicial dictionary of words and phrases*, 10th edition. London: Sweet & Maxwell.

See also: DATA PROTECTION, DECISIONAL PRIVACY, DYNAMIC CONSENT, EXPLICIT CONSENT, EXPRESS CONSENT, IMPLICIT CONSENT, INFORMED CONSENT, JUST-IN-TIME CONSENT, NOTICE AND CONSENT, REVOCATION

Consent Form

Documentary evidence of **consent** can be important for interventions carrying a higher legal or ethical **risk** and is often collected within more bureaucratic contexts. Consent forms are therefore part of standard practice in (for example) human-subject research, more invasive forms of medical treatment and the sharing of children's **personal data**.

The subject of the intervention is typically asked to sign a formal, written document, which confirms the **information** they have received and the terms of their consent. This may be particularly appropriate when

data protection or **confidentiality** law requires **informed consent** to data use. Most consent forms will not be legal contracts, and any consent used as a **lawful basis** for **data processing** in EU data protection law must be revocable. Therefore, where the consent in question is a **GDPR** basis for processing, the party collecting the information will be bound by the terms of the consent form, but the **data subject** should be free to withdraw or amend their **agreement** at any time.

See also: CHILDREN'S PRIVACY

Consequential Data

Consequential data are **data** which are generated as a by-product of some other process. For example, to access a service, a **user** may have to provide their name and address, phone number, payment details and perhaps some other **information**. The data are not collected for their own sake but in service of another function.

This is a **privacy** concern because so many of our transactions now create consequential data that individuals can easily and unwittingly build a large **digital footprint**.

Further reading:
Purdam, K. and Elliot, M., 2015. The changing social data landscape. *In:* Halfpenny, P. and Proctor, R., eds, *Innovations in digital social research methods*. London: Sage, 25–58.

See also: DECLARED DATA, INTENTIONAL DATA

Consistency Mechanism

Under the EU **GDPR**, each Member State has its own national regulator, known as a **Supervisory Authority**. To ensure that these authorities enforce the GDPR in a consistent way, it provides for a consistency mechanism. This is particularly important when dealing with large organisations that process **personal data** from people across national boundaries, as the decision of one regulator will have implications for the **data subject**s in other countries (see Recital 135, GDPR).

The two key elements of the consistency mechanism are the power of the EU-wide **European Data Protection Board (EDPB)** to issue opinions, and the potential for dispute resolution in the event of a conflict.

The consistency mechanism is the stick to the carrot of the GDPR’s **cooperation mechanism**. Ideally, the national Supervisory Authorities will cooperate on enforcement in transnational processing cases, but if this cooperation breaks down the EDPB can arbitrate any disagreements.

Further reading:
Gentile, G., and Lynskey, O., 2022. Deficient by design? The transnational enforcement of the GDPR. *International and Comparative Law Quarterly*, 71(4), 799–830, https://doi.org/10.1017/S0020589322000355.

See also: DATA PROTECTION AUTHORITY

Consumer Information Markets

With the growth of **information** being collected about consumers through **e-commerce**, or digital payments systems, markets have grown up around this information in terms of both the information itself and the inferences that can be drawn from it (often called ‘insights’). The **data** may come from websites’ **traffic data**; from **social media**, including information provided by consumers themselves; or from mining a publicly available **dataset**. Market structures are becoming increasingly rich, with many **data intermediari**es adding value to the raw data gleaned at the time of transaction. Example markets include credit ratings, financial data, social media, **search engine**s, private medical and **genetic data**, e-commerce and analyses of emails.

There are several distinct types of consumer information market. The seller may collect information on already identified individuals on behalf of the buyer (**data linkage** across **database**s), or alternatively may present the buyer with new prospects, or anonymised ones. The seller may collect and sell information about a specific consumer segment (for instance, an event company might sell lists of those who attended particular events, or a news provider might sell lists of those who have downloaded articles on a particular topic or searched on a particular key word).

Relatedly, some data intermediaries sell value-added services, based on the information they have accumulated, such as **targeted advertising**. A search engine leverages its record of the search terms that **user**s have entered, together with anything extra it has via its understanding of the **identity** of the user, to recommend the user to advertisers based on their revealed preferences. Such value-added services do not sell information about consumers directly to buyers, but rather insights gleaned from the information.

Further reading:

Bergemann, D. and Bonatti, A., 2019. Markets for information: an introduction. *Annual Review of Economics*, 11, 85–107, https://doi.org/10.1146/annurev-economics-080315-015439.

See also: AD EXCHANGE, CUSTOMER TRACKING, DATA BROKER, DATA MINING, TRACKING, VALUE OF DATA

Consumer Preference Information

Consumer preference **information** is **data** about what an individual consumer wants that goes beyond their needs. Sellers get this data through surveys, forms for feedback, market studies, harvesting of **clickstream data**, AB split testing, and product placement analyses. They utilise their insights to develop marketing tactics that include **targeted advertising**.

Further reading:

Guo, M., Liao, X., Liu, J. and Zhang, Q., 2020. Consumer preference analysis: a data-driven multiple criteria approach integrating online information. *Omega*, 96, 102074, https://doi.org/10.1016/j.omega.2019.05.010.

Okazaki, S., Li, H. and Hirose, M., 2009. Consumer privacy concerns and preference for degree of regulatory control. *Journal of Advertising*, 38(4), 63–77, https://doi.org/10.2753/JOA0091-3367380405.

See also: BEHAVIOURAL ADVERTISING, DATA HARVESTING

Content Data

Data that make up the foundation of a specific piece of digital content, such as a webpage, document, video, or audio file, are called content data. In contrast to technical or descriptive data that describes or supports the content, content data refers to the substantive **information** or media contained within a digital **asset**. Text, images, audio, video, animations and other digital media that make up a digital asset's main content can all be considered content data. A wide range of digital tools and **software** can be used to create and share this data, and a wide range of digital hardware and software can be used to store and access it.

Digital tools and techniques are frequently used to analyse and process content data to extract meaning, spot patterns or carry out additional tasks such as sentiment analysis, image recognition or **natural language processing**. Businesses frequently use this analysis to enhance their content marketing and digital media strategies because it can offer insightful **customer tracking** information about **user** behaviour, preferences and interests.

See also: BEHAVIOURAL ADVERTISING, CONTEXTUAL ADVERTISING, TARGETED ADVERTISING

Contextual Advertising

A form of **targeted advertising** whereby advertisements appear on websites or **social media** and are selected and served by automated systems based on the context of what a **user** is looking at. Most commonly the system searches for keywords within a website's text and returns advertisements to the webpage based on those keywords. This is arguably a **breach** of **attentional privacy**.

Further reading:
Zhang, K. and Katona, Z., 2012. Contextual advertising. *Marketing Science*, 31(6), 980–94, https://doi.org/10.1287/mksc.1120.0740.

See also: RECOMMENDATION SYSTEM

Contextual Integrity

A theory of **informational privacy** developed by Nissenbaum, who argues that expectations of privacy no longer fall either side of the classic binary of public and private spheres. Instead, modern **surveillance** technologies mean we can be tracked, observed and evaluated whether at home or in public. Whether we expect **information** to be shared about us depends not on whether we are at home or in the street, but on the social norms associated with a more diverse range of contexts. The factors at play in how widely we will expect our information to be divulged may be relational, temporal, geographical, institutional, ethical or political. The situational nuance of the theory of contextual integrity can be helpful when the **common law** invites consideration of 'all the circumstances' in determining whether someone has a **reasonable expectation of privacy**.

Further reading:
Nissenbaum, H., 2010. *Privacy in context: technology, policy, and the integrity of social life*. Stanford: Stanford University Press.

Continuous Data

Continuous data is data represented on a scale to which standard arithmetic operators can be validly applied, and which is theoretically infinitely subdividable. The latter property means that continuous data presents a type of **disclosure risk** that is not present with categorical data. As an example, take a person’s height; if this is measured to sufficient precision this will be unique to that person, and therefore theoretically height is a **unique identifier**. In practice, issues of **accuracy** come into play and continuous data is subject to precision-based measurement error within both a **target dataset** and an **adversary**’s **information**, leading to **data divergence**.

Controlled Rounding

A **disclosure control method** for **tabular data** whereby the values of all cells are replaced by a value from a finite set. Typically, this set contains two values, which are the numbers rounding up and down the original cell value to a multiple of a given base number (for example, 10). Controlled rounding is more complex to implement than **random rounding**, but it has the advantage that **additivity** is maintained.

Further reading:
Salazar-González, J.J., 2006. Controlled rounding and cell perturbation: statistical disclosure limitation methods for tabular data. *Mathematical Programming*, 105(2), 583–603, https://doi.org/10.1007/s10107-005-0666-4.

See also: PERTURBATION, STATISTICAL DISCLOSURE

Controlled Tabular Adjustment (CTA)

A **statistical disclosure control** method for **tabular data** developed by Cox et al. CTA perturbs cell values within defined protection ranges while respecting tabular constraints, such as **additivity**, and minimising **information loss** as measured by a linear measure of overall data distortion, such as the sum of the absolute values of the individual cell value adjustments. CTA replaces each risky cell by either of the two endpoints of its protection range. Certain non-risky cell values are adjusted by small amounts to restore additivity.

Further reading:

Cox, L.H., Kelly, J.P. and Patil, R., 2004. Balancing quality and confidentiality for multivariate tabular data. *In:* Domingo-Ferrer, J. and Torra, V., eds, *Privacy in statistical databases*. Lecture Notes in Computer Science, 3050, New York: Springer, 87–98. https://doi.org/10.1007/978-3-540-25955-8_7.

See also: CONTROLLED ROUNDING, DATA QUALITY, PERTURBATION

Convention 108

The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108) is commonly known as ‘Convention 108’. Opened for signature in 1981, it was the first legally binding international instrument for **data protection**. It also defined core **data protection principles** and concepts, such as **personal data**, in a way which is still used in contemporary European data protection law, albeit with some modifications.

Convention 108 comes from the **Council of Europe**, a European body distinct from the European Union. As such, it is open for signature by non-European countries. The Convention was updated in 2018 to reflect the challenges of more sophisticated **data processing** and will become ‘Convention 108+’. Greenleaf has argued that Convention 108+ is the only feasible basis for the globalisation of data protection, as it is the only binding data protection **agreement** open to global accession.

Further reading:

Council of Europe, 2023. *Convention 108 and protocols*. www.coe.int/en/web/data-protection/convention108-and-protocol.

Greenleaf, G., 2021. How far can Convention 108 ‘globalise’? Prospects for Asian accessions. *The Computer Law and Security Report*, 40, 105414, https://doi.org/10.1016/j.clsr.2020.105414.

See also: EUROPEAN CONVENTION ON HUMAN RIGHTS, GDPR

Cookie

A cookie is a small text file that a website stores on a **user**’s device when they visit a website. Cookies are created to give websites a way to identify a user’s browser and remember things about them, such as their preferences, login **information** and **browsing history**. For instance, a website may use cookies to remember a user's login information so they do not have to enter it each time they visit the site, or to **track** which pages the user visits so the website can **personalise** their experience.

Cookies come in different varieties. First-party cookies are placed by the website the user is currently visiting, while third-party cookies are placed by outside advertisers or tracking firms. Some cookies are **persistent cookie**s and can stay on a user's device for a longer time, while **sessional cookie**s expire after a specific amount of time. The use of cookies to track user activities across various websites and services has raised **privacy concern**s. However, many websites depend on cookies to offer customised experiences and to gather analytics information to enhance their offerings. Websites should provide information about their cookie policies, and most web browsers allow users to control or disable cookies.

In the EU, cookies are regulated by the **ePrivacy Directive** or 'Cookie Directive', which at the time of writing is due to be replaced by the **ePrivacy Regulation**, which aims to simplify cookie banners. Cookie **consent** is notorious for being an ineffective form of **notice and consent**, whereby cookie banners frequently pop up with information that is seldom read, creating more of an illusion of **informed consent** than an actual choice for the average **Internet** user.

Further reading:

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F. and Holz, T., 2018. We value your privacy … now take some cookies: measuring the GDPR's impact on web privacy. *arXiv*, https://doi.org/10.48550/arXiv.1808.05096.

Smit, E.G., Van, N.G. and Voorveld, H.A., 2014. Understanding online behavioural advertising: user knowledge, privacy concerns and online coping behaviour in Europe. *Computers in Human Behavior*, 32, 15–22, https://doi.org/10.1016/j.chb.2013.11.008.

See also: BROWSER FINGERPRINTING, BROWSING HISTORY, DO NOT TRACK, SUPER COOKIE

Cooperation Mechanism

The EU **GDPR** introduced a new framework for national data **regulators** to work together on international enforcement. As large, commercial enterprises often engage in **cross-border data processing**, cooperation between privacy regulators is essential to ensure global **accountability**, as well as consistent protection for EU citizens of different nationalities.

When the **data processing** affects people in multiple EU countries, the relevant national regulators should support a single **lead supervisory**

authority. The lead authority is the regulator based in the country where the **data controller** responsible for the processing has its **main establishment**. The lead must, under Article 60 GDPR, share information and draft decisions with the other concerned supervisory authorities.

Further reading:
Gentile, G. and Lynskey, O., 2022. Deficient by design? The transnational enforcement of the GDPR. *International and Comparative Law Quarterly*, 71(4), 799–830, https://doi.org/10.1017/S0020589322000355.

See also: CONSISTENCY MECHANISM, DATA PROTECTION, DATA PROTECTION AUTHORITY, EUROPEAN DATA PROTECTION BOARD, ONE-STOP-SHOP, SUPERVISORY AUTHORITY

Co-Privacy

A shortening of co-operative privacy; the notion developed by Josep Domingo-Ferrer rooted in **game theory** to denote a situation wherein, for a rational player, the best strategy for protecting one's own privacy is to help others to protect theirs. The concept captures the paradox that although privacy is often conceived as grounded in individual **identiti**es, it is inherently a construct that relies on a degree of co-operation with others.

Further reading:
Domingo-Ferrer, J., 2011. Coprivacy: an introduction to the theory and applications of co-operative privacy. *SORT*, 25–40. www.idescat.cat/serveis/biblioteca/docs/bib/publicacions/r00262011specialissueprivacy.pdf.

Correct Attribution Probability

A measure of **disclosure risk** developed by Taub et al., an estimate of the probability that an **adversary** who attempts an **attribute disclosure** attack against a **dataset** correctly infers the value of an **attribute** of a **population unit**.

Further reading:
Taub, J., Elliot, M., Pampaka, M. and Smith, D., 2018. Differential correct attribution probability for synthetic data: an exploration. *In*: *Privacy in statistical databases*. Cham: Springer, 122–37, www.springerprofessional.de/en/differential-correct-attribution-probability-for-synthetic-data-/16106956.

Count Data

A form of **data** which are produced by counting items. Values of count variables will therefore always be positive integers, although summary statistics such as means and standard deviations can be produced with non-integer values. Count variables will often have names starting with ‘number of’. It is common for count variables to have skewed distributions with low counts including zeroes containing the bulk of the distribution (consider the number of children in a household, or the number of cigarettes smoked per day). These skewed distributions may create **outlier**s which can be **disclosive** and may need to be dealt with by, for example, **topcoding**.

CPM

See: COMMUNICATION PRIVACY MANAGEMENT

Credentials

Credentials are the details used to confirm a **user**’s **identity** within a computer system, **network** or **application**. A **username** or user ID is typically combined with a **password** or another form of **authentication**, like a **biometric** factor or a **security token**. Credentials are an essential component of the authentication and **authorisation** process in computer systems and are used for **access control** to resources. When a user tries to access a resource, they must supply the correct credentials to authenticate their identity and demonstrate that they have the required access rights. To prevent interception or tampering, credentials are typically kept in a secure **database** or directory and sent over **encrypted** channels.

See also: NETWORK SECURITY

Creepiness

The desire to provide salient and timely **information** and services to **users** under **surveillance capitalism** has led to **application**s, practices and functionalities of technology which, while perfectly legal, are regarded as ‘creepy’ by users. Such practices typically make manifest in unexpected ways the fact that the **data** infrastructure has a greater knowledge of the user than a human equivalent might have (for instance, the app 'knows' where the user is located, or what they have just **search**ed for).

Creepiness is a negative affect that detracts from the experience of using the app, even if it is performing its stated function. Practices cited as creepy by Tene and Polonetsky include ambient social apps, social listening (analysing **social media** content for sentiment), **personalised** analytics, data-driven marketing (including a famous case where the retailer Target 'knew' that a young girl was pregnant before her family did) and the launching of new products that challenge social norms.

Many information-based practices are regarded by users as creepy when they inadvertently reveal the depth of information held about them. The tension between the incentives to produce effective apps and to avoid creepiness was expressed by Google's then-CEO Eric Schmidt as: 'There is what I call the creepy line. The Google policy on a lot of things is to get right up to the creepy line and not cross it.' Unfortunately, this sentiment itself might be considered a creepy one.

Further reading:

Tene, O. and Polonetsky, J., 2013. A theory of creepy: technology, privacy and shifting social norms. *Yale Journal of Law and Technology*, 16, 59–102, www.yjolt.org/sites/default/files/theory_of_creepy_1_0.pdf.

Saint, N., 2010. Eric Schmidt: Google's policy is to 'get right up to the creepy line and not cross it'. *Business Insider*, 1 Oct 2010, www.businessinsider.com/eric-schmidt-googles-policy-is-to-get-right-up-to-the-creepy-line-and-not-cross-it-2010-10?international=true&r=US&IR=T.

See also: BEHAVIOURAL ADVERTISING, CONTEXTUAL ADVERTISING, CONTEXTUAL INTEGRITY, SURVEILLANCE

Creepy Line, The

See: CREEPINESS

Crime Prevention Exemptions

Most **jurisdiction**s which have **data protection** laws will also have exemptions from some of their requirements for the purposes of preventing or investigating crime. Under the UK Data Protection Act 2018, for example, there is an exemption from the **right to be informed** of **data processing**, for the obvious reason that otherwise law enforcement agencies would have to make suspects aware of their investigations. It is an example of a qualification of individual rights under data protection law to safeguard the wider **public interest**.

Further reading:
Information Commissioner's Office, 2023. *A guide to the data protection exemptions*. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/exemptions/.

CRM

See: CUSTOMER RELATIONSHIP MANAGEMENT

Cross-Border Data Processing

Under the EU **GDPR**, cross-border data processing refers to **data processing** activities which either take place in more than one member state of the EU or take place in one member state but affect **data subject**s in multiple member states. This is distinct from a transfer of **personal data** to a country outside the EU, which is commonly referred to as a **data transfer**.

Cross-border data processing within the EU is one of the reasons why national regulators need to cooperate to enforce the GDPR across national borders. The map of data processing on the **World Wide Web** is not necessarily a mirror image of the geographical borders between countries, and nation-state regulators have needed to adjust accordingly. This adds to the argument for the globalisation of **data protection** law, through agreements such as **Convention 108+**.

Further reading:
European Commission, n.d.. What happens if my company processes data in different EU Member States? https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/enforcement-and-sanctions/enforcement/what-happens-if-my-company-processes-data-different-eu-member-states_en.

See also: COOPERATION MECHANISM

Cross-Device Tracking

Cross-device tracking is a technique used by advertisers to monitor and gather **information** on **user**s across multiple devices (laptops, tablets, phones, etc.). This is accomplished by building a user **profile** from **data** gathered from their various devices, including **IP addresses**, **cookie**s, device **identifier**s and other information. To deliver more **targeted advertising** or **personalisation** across all of a user’s devices, cross-device tracking aims to gain a more thorough understanding of their behaviour and preferences. Cross-device tracking can be used, for instance, by an advertiser to show an ad for a product the user recently looked up on their phone while they are using a desktop computer to browse the **Internet**.

Cross-device tracking creates **privacy concerns** because it enables businesses to create comprehensive profiles of users without permission or **awareness**. Users may find it challenging to opt out of cross-device tracking because doing so requires them to manage their **privacy settings** across various platforms and devices. Users can take precautions like deleting their browser cookies, using ad-blocking software or using **virtual private network**s **(VPN**s**)** to hide their IP address and location to prevent cross-device tracking. Inbuilt privacy features are also available in some web browsers, which can help prevent cross-device tracking.

Further reading:
Brookman, J., Rouge, P., Alva, A. and Yeung, C., 2017. Cross-device tracking: measurement and disclosures. *In: Proceedings of Privacy Enhancing Technology*, 2017(2), 133–48, https://doi.org/10.1515/popets-2017-0020.

See also: AD NETWORK, BEHAVIOURAL ADVERTISING, LOCATION TRACKING, PRIVACY-ENHANCING TECHNOLOGY, TRACKER, TRACKING

Cross-Site Request Forgery (CSRF)

An attack that tricks a web application user to execute actions they did not intend. Usually employing social engineering, a successful CSRF attack can cause a user to change their email address, transfer funds, and other state changing actions. If the victim is a system administrator, a successful CSRF might compromise the entire web application.

Further reading:
Sudhodanan, A., Carbone, R., Compagna, L., Dolgin, N., Armando, A. and Morelli, U., 2017. Large-scale analysis & detection of authentication cross-site request forgeries. In: *2017 IEEE European symposium on security and privacy*, IEEE, 350–65.

See also: PHISHING

Cross-Site Scripting (XSS)

A cross-site scripting (XSS) attack is where an **adversary** injects malicious code into a webpage, which is subsequently executed by the **user**. This might lead to, for example, **malware** being installed on the user's device or the loss of **personal data** including **username**s and **password**s. Successful adversaries may be able to **masquerade** as the user, carry out any action that the user is able to perform and read any **data** that the user is able to access.

There are different cross-site scripting **attack**s: reflected XSS, where the script comes from the user's HTTP request; cached XSS, where the script is stored in the website's **database**, and DOM-based XSS, where the **vulnerability** exists in client-side code. Web developers should implement mitigations, such as input validation, and appropriate handling of user input.

Further reading:

Gupta, S. and Gupta, B.B., 2017. Cross-site scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art. *International Journal of System Assurance Engineering and Management*, 8, 512–30, https://doi.org/10.1007/s13198-015-0376-0.

Vogt, P., Nentwich, F., Jovanovic, N., Kirda, E., Kruegel, C. and Vigna, G., 2007. Cross site scripting prevention with dynamic data tainting and static analysis. *In: NDSS*, 12, https://people.scs.carleton.ca/~soma/id-2007w/readings/ndss07_xssprevent.pdf.

See also: SQL INJECTION

Cryptanalysis

Cryptanalysis is the process of deciphering and analysing codes to gain access to or reveal the encrypted data's original **plaintext**. Cryptanalysis studies the structure and characteristics of **encryption algorithm**s with the goal of identifying vulnerabilities that may be exploited to decrypt or access the encrypted data. It is frequently used to evaluate the robustness and efficiency of encryption systems, to spot **security** flaws and vulnerabilities, and to create new encryption methods and techniques.

The study of traditional encryption techniques like substitution **cipher**s and transposition ciphers falls under the category of classical cryptanalysis, while modern cryptanalysis studies contemporary encryption techniques like **block cipher**s, stream ciphers and **public key cryptography**. Techniques for cryptanalysis can be applied both defensively and offensively. Offensive cryptanalysis involves attempting to crack encrypted **communication**s to gain unauthorised access, whereas defensive cryptanalysis focuses on testing and assessing the strength of encryption systems to ensure they are effective.

Further reading:

Heys, H.M., 2002. A tutorial on linear and differential cryptanalysis. *Cryptologia*, 26(3), 189–221, https://doi.org/10.1080/0161-110291890885.

Knudsen, L. and Wagner, D., 2002. Integral cryptanalysis. *In:* Daemen, J. and Rijmen, V., eds, *Fast Software Encryption. FSE 2002.* Berlin: Springer, 112–27, https://doi.org/10.1007/3-540-45661-9_9.

See also: ASYMMETRIC CRYPTOGRAPHY, AUTHENTICATION, CRYPTOGRAPHY

Cryptocurrency

A cryptocurrency is a digital **currency** that is created and stored on a decentralised computer **network**, using distributed ledger technology sitting on a **blockchain**, where the ledger records transactions **transparently**. **Cryptographic** protocols ensure that any new transaction can be verified on the ledger, so that it is certain that the payer has the necessary funds in the cryptocurrency; no **trusted third party** is needed to verify this, and so transparent cryptocurrencies have given rise to the concept of trustless **trust**. The cryptocurrency is not managed by a bank, but the rules of the blockchain can be used to ensure certain properties hold (such as restrictions on the amount of the currency that can be coined, or parity with other currencies or resources). The first decentralised currency was Bitcoin; many have since followed.

Bitcoin, and many other cryptocurrencies, facilitate **financial privacy** by providing means to remain anonymous (or at least pseudonymous). Transactions of bitcoin are from wallet (a container of a **private key**) to wallet, not person to person. An individual can have as many wallets as they like, and the extent to which they need **identify** themselves to own a wallet depends on the currency (bitcoin owners need not identify themselves at all). The anonymity afforded has resulted in cryptocurrency getting a **reputation** for being especially valuable in **cybercriminal** contexts. It also presents holders of such cryptocurrencies with a dilemma: either they create no record of their ownership of their wallets, in which case they lose their holdings if the key is lost (including through the death of the wallet-holder), or they do create a **record**, in which case it is at least possible that they may be identified.

Further reading:
McDonald, O., 2021. *Cryptocurrencies: money, trust and regulation*. Newcastle: Agenda.

Cryptographic Hash Function

A cryptographic hash function is a hash function with important **cryptographic** properties that can be used in **cybersecurity** applications (and others, such as **blockchain**), where the hash can function as a kind of **digital fingerprint** of the input data. As the function is deterministic, **data** will always produce the same hash with the same hash function. The function should be complex enough to be infeasible to reverse (i.e., given a hash, to compute the data that produced it), so it is a *one-way* function. Although a cryptographic hash function has an infinite domain and finite codomain, and so cannot produce unique hashes, the chances of two pieces of data having the same hash should be negligible. Minor changes to the input should produce randomly drastic changes to the hash (i.e., similar inputs are unlikely to produce similar hashes).

Some cryptographic hash functions have been created as **standard**s, such as the Secure Hash Algorithms (SHA) published and certified by the US National Institute of Standards and Technology. For example, the hash function SHA3-256 is from the third release of SHA standards in 2015 and maps the input onto a 256-bit hash (in effect, a non-negative integer $<2^{256}$ expressed in binary notation); it is used by the Ethereum blockchain.

Further reading:
Mittelbach, A. and Fischlin, M., 2021. *The theory of hash functions and random oracles: an approach to modern cryptography*. Cham: Springer.

See also: DIGITAL SIGNATURE

Cryptographic Key

A **cryptographic** system **encrypt**s and decrypts **data** using a string of letters or numbers known as a cryptographic key. A key is a code that is used to scramble and unscramble **information**. This makes unauthorised access or tampering difficult as the encrypted data can only be accessed or decrypted by someone who knows the correct key.

Symmetric and asymmetric are the two main categories of cryptographic keys. **Symmetric encryption** systems, in which the same key is used for both encryption and decryption, use symmetric keys. **Asymmetric encryption** systems use a pair of keys, one for encryption and the other for decryption.

Further reading:
Lenstra, A.K. and Verheul, E.R., 2001. Selecting cryptographic key sizes. *Journal of Cryptology*, 14, 255–93, https://doi.org/10.1007/978-3-540-46588-1_30.

See also: AUTHENTICATION, ENCRYPTION KEY, PUBLIC-KEY CRYPTOGRAPHY

Cryptographic Protocol

A set of guidelines and practices called a cryptographic protocol is used to exchange **data** securely over between two or more parties. Many **security** systems and applications depend on cryptographic protocols because they offer a means of ensuring the **confidentiality**, **integrity** and **authenticity** of data transmitted over a **network**. To safeguard security and **communication privacy**, cryptographic protocols typically use **encryption algorithm**s, **digital signature**s and **encryption key** exchange protocols. These **protocol**s offer secure **authentication** and verification of the parties involved in the **communication** while also guarding against unauthorised access, interception and tampering.

One major cryptographic protocol is **Transport Layer Security**, which is used to secure web traffic and other network communications. Others are used to protect network traffic and **remote access** to systems and devices, including IPsec (Internet Protocol Security) and SSH (Secure Shell).

Further reading:
Goldreich, O., Micali, S. and Wigderson, A., 2019. Proofs that yield nothing but their validity and a methodology of cryptographic protocol design. *In:* Goldreich, O., ed., *Providing sound foundations for cryptography: on the work of Shafi Goldwasser and Silvio Micali*, 285–306, https://doi.org/10.1145/3335741.3335754.

See also: ASYMMETRIC CRYPTOGRAPHY, CRYPTOGRAPHIC KEY, CRYPTOGRAPHY, PUBLIC-KEY

Cryptography

The practice of protecting **information** and communications by putting **data** and **communication**s into a format that can only be accessed or read by those with the proper **authorisation**. The process converts **plaintext** (unencrypted data) into **ciphertext** (encrypted data) (and usually back again) using a **cryptographic key**. Numerous uses for cryptography exist, including secure **data storage**, **authentication** and **secure communication**.

Cryptographic **algorithm**s come in a variety of forms, such as **hashing** algorithms, **symmetric-key** algorithms and **asymmetric-key** algorithms. Asymmetric-key algorithms use two **encryption key**s: a **public key** for **encryption** and a **private key** for decryption; symmetric-key algorithms use the same key for both encryption and decryption. Fixed-length message digests are produced by hashing algorithms and are used for message authentication and data **integrity**.

Information security, computer science and mathematics expertise are all needed in the complex and rapidly developing field of cryptography. To ensure their efficacy and **security**, cryptographic techniques and algorithms must be carefully created and rigorously tested. They also need to be updated frequently and improved to fight against new privacy **threats** and security **attacks**.

Further reading:

Diffie, W. and Hellman, M.E., 2022. New directions in cryptography. *In: Democratizing Ccryptography: the work of Whitfield Diffie and Martin Hellman*, 365–90, https://doi.org/10.1145/3549993.3550007.

Katz, J. and Lindell, Y., 2020. *Introduction to modern cryptography*. New York: CRC Press, https://doi.org/10.1201/b17668

See also: PUBLIC-KEY CRYPTOGRAPHY

Crypto-Shredding

Crypto-shredding is the process of using **cryptography** to permanently scramble **data** before securely erasing sensitive information from a system. This method is also sometimes referred to as *secure erasure*. Using specialised **software** or techniques, it is still possible to recover data that has been deleted using conventional methods, such as deleting a file or re-formatting a hard drive. By using cryptographic **algorithm**s to scramble the data in a way that cannot be recovered, crypto-shredding solves this issue. Usually, to do this, the data is overwritten with random bits or a **cryptographic key**.

See also: ERASURE

Crypto Wars

From the early days of digital technology, **encryption** of **data** has been seen as a central protection of individuals' **privacy** from government **surveillance**. Meanwhile, governments (particularly the US government) had interests in (a) being able to read the **communication**s of citizens and non-citizens alike, for **national security** and espionage reasons, and (b) restricting access to cutting-edge **cryptography** to rival powers. The US's introduction of the Data Encryption Standard (DES) (the forerunner of the **Advanced Encryption Standard**) in 1975 was a catalyst for concerns that advanced **information** technology might be exported with DES installed, in effect handing its secrets to foreign powers.

This dilemma has led to a series of political, technological and legal struggles – dubbed the crypto wars – between governments and technology companies. The companies are concerned with retaining the **trust** of their customers and the **security** of their systems, while governments have sought the ability to access decrypted versions of encrypted data, through **backdoor**s or other methods. However, governments are necessarily conflicted, as their access must also be consistent with their ability to protect the **confidentiality** of their own communications. A controversial solution often touted is key **escrow**, where **private key**s required for decryption are held by a **third party** to allow access to governments, espionage agencies or law enforcement under specified circumstances. Crypto wars have tended to erupt when new encryption technologies have emerged.

A prominent crypto war concerned encryption on smartphones. Edward Snowden revealed in 2013 that the US government had legal routes to demand that encryption on specific Android and iOS smartphones be bypassed, and as a result Google and Apple redesigned their devices so that they were technically unable to comply with such demands, drawing government criticism. In a high-profile case of 2016, Apple refused to help the FBI unlock the work phone of a terrorist; the case was not resolved in court as the FBI gained access to the phone via another route.

Further reading:

Jarvis, C., 2021. *Crypto wars: the fight for privacy in the digital age – a political history of digital encryption*. Boca Raton, FL: CRC Press.

CTA

See: CONTROLLED TABULAR ADJUSTMENT

Cultural Variation of Privacy

The level of cultural variation about **privacy** is a matter of debate. Simple observation reveals major differences in privacy-relevant behaviour and norms across cultures, for example with respect to the amount of **personal space** an individual is comfortable with, or to the acceptability of interaction across the sexes. This cultural variation is detectable across space (different places at the same time), as well as time (different historical stages of the same culture).

However, those observed differences of *degree* do not entail that different cultures have different *ideas* about privacy; it may simply mean that different norms of privacy are operating, and within each culture the members of those cultures have different preferences about how and when their privacy is protected. Altman concluded a survey of cultural variation by maintaining that even in a culture with apparently little privacy, privacy-preserving mechanisms for regulating interpersonal relations could be discovered by observers.

O'Hara has argued (a) that the application of privacy conceptions to any society is possible, even one which exhibits very different attitudes and behaviour towards privacy, and (b) that, while cultural divergence does not necessarily entail conceptual difference, it is likely that any two societies remote from each other will struggle to understand each other's privacy practices. Observers need to ask not only what privacy conceptions of their own apply to a remote culture, but also what (different) conceptions the remote culture itself might consider, or have considered, salient.

Further reading:

Altman, I., 1975. *The environment and social behavior*: *privacy, personal space, territory, crowding*. Monterey: Brooks/Cole.

O'Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

See also: FAMILY RESEMBLANCE THEORY OF MEANING, HISTORY OF PRIVACY

Currency

A currency is a standard type of money in circulation, used for exchange (of goods, services and other currencies), as a holder of value, and as a unit of account. The term also covers the forms the money takes, such as banknotes, coins and digital tokens. **Financial privacy** concerns the desire to keep transactions denominated in a currency from outsiders, including tax authorities and law enforcement. Cash is anonymous (although high-denomination paper notes can be laboriously traced via their serial numbers), while bank and credit accounts are traceable, and digital payments leave data trails. Many central banks are experimenting with digital currencies, and these will, in the absence of privacy-preserving measures, centralise **information** about all transactions. At the other end of the scale, **cryptocurrenci**es based on **encryption** can allow anonymous or pseudonymous transactions to take place.

Further reading:
Davies, G. and Connors, D., 2016. *A history of money*, 4th edition. Cardiff: University of Wales Press.

Customer Relationship Management (CRM)

CRM is the practice of using **information** gathered about customers to improve services, retain custom and increase sales by raising customer satisfaction levels. Much of the information gathered will be **personal data** or **PII**, and so a CRM programme is likely to raise a company's **data protection compliance** costs.

'CRM' can also refer to the specific **software** system that is used for managing relationships.

Further reading:
Buttle, F. and Maklan, S., 2019. *Customer relationship management: concepts and technologies*, 4th edition. Abingdon: Routledge.

See also: CONSUMER INFORMATION MARKETS, CUSTOMER TRACKING, E-COMMERCE, LOYALTY CARD

Customer Tracking

Customer tracking is the practice of businesses to follow and understand their customers, including their demographics, where they live and how they shop, enabling their **customer relationship management**. Businesses can then **personalise** their services to the requirements of individual customers, ranging from payment types, marketing and **targeted advertising**, **price discrimination** and special offers, service delivery and product design.

There are several potential sources of **information** for customer tracking. **Purchase history**, **browsing history** and **clickstream data** are records of direct contact, as are email traffic with customers (rates of opening emails, resulting clickthroughs, immediate deletes and unsubscribe commands) and responses to **communications** (e.g., newsletters). Understanding the points at which customers abandon purchases (for instance, they put goods in their basket, but do not check out) can also be important, highlighting places where the process needs to improve. Downloads from **e-commerce** websites show what customers are searching for. **Social media** accounts usually contain analytics tools for further insights about engagement and **Web beacon**s will provide a view on the customer's browsing, including how they made it to the company website.

In a physical store, customers can be tracked using **CCTV**, **RFID** tags on shopping trolleys or hand-held self-scanners, and compared with a motion estimation **algorithm**. Such **tracking** can be used to aid retail space design.

Further reading:
Meyer, C. and Schwager, A., 2007. Understanding customer experience. *Harvard Business Review*, 85(2), 116–26, https://hbr.org/2007/02/understanding-customer-experience.

See also: CONSUMER INFORMATION MARKETS, LOYALTY CARD

Cybercrime

Cybercrime is a term of wide reference, used to refer to illegal activity that involves the use of the **Internet**, a computer **network** or other digital technology. Typically, cybercrime will involve a criminal **attack** on a computer system, also using a computer to gain access. **Cyberterrorism** is cybercrime for terrorist purposes. Cybercrime is countered by **cybersecurity** measures. Much cybercrime is facilitated by the concealed resources of the **Dark Web**, either in its perpetration or as a venue to sell the ill-gotten gains.

Many types of cybercrime are also **breach**es of privacy. *Fraud* may include **hacking** into private or **classified information**, to alter it, sell it on or use it to gain access to other systems (e.g., banking), or for the purposes of extortion. **Identity theft** involves stealing **data** that allows the criminal to **masquerade** as the victim. Certain types of real-world crime, particularly sex crimes but also assaults and even mass shootings, can be livestreamed for an audience. **Social media** can allow *cyberbullying*, **cyberstalking** and other types of **harassment** to take place. **Spam** can be invasive of the individual's time and computer resources.

Further reading:

Yar, M. and Steinmetz, K.F., 2019. *Cybercrime and society*, 3rd edition. London: Sage.

See also: RANSOMWARE

Cyber Insurance

Data breaches and hacks can be very expensive for companies, both financially and in terms of **reputation**. The **risk**s of these and other types of cyberattack are increasing, especially as legislation often forces victims of **data** breaches to inform their customers. In response to this, an industry in cyber insurance is emerging. Such insurance generally covers risks to privacy, the firm's own **security** and its operations. Policies might cover costs of forensic investigation, public relations, data repair, **ransomware** payments and lost revenue during a cyberattack. They are unlikely to cover lost future profits, or losses from theft of **intellectual property**.

The pricing of these risks is not simple. One difficulty is the possibility that a cyberattack is linked to a hostile state. Insurance companies typically avoid claims that are the result of an act of war, but much state-backed **hacking** may be seen as acts of **cyberwarfare** or cyberespionage, of uncertain status. Second, the quantification of potential losses is difficult, partly because of lack of quality historical data but partly because losses may vary dramatically from the trivial to the catastrophic (compare for example with the costs of a fire or flood, which will lie within a predictable range). Hence pricing premiums for a practical business model is difficult, and in such circumstances, insurance companies will tend to price high. Third, whereas most insurance risks are relatively independent of each other (fires and floods have little effect on other fires and floods, at least outside the local area), this is not true in **cybersecurity**. A flaw in a well-used operating system will increase the vulnerability of large numbers of firms and individuals simultaneously.

Further reading:

Talesh, S.A., 2018. Data breach, privacy, and cyber insurance: how insurance companies act as 'compliance managers' for businesses. *Law and Social Inquiry*, 43(2), 417–40, https://doi.org/10.1111/lsi.12303.

Cyber Resilience

A form of **resilience** which denotes the capacity of an organisation to anticipate, respond to and recover from cyberattacks.

See also: CYBERSECURITY, SECURITY POSTURE

Cybersecurity

Cybersecurity is the **risk**-based protection and recovery of computers, computer systems, automated systems, **network**s and connected devices from threats, including **virus**es and other **malware**, criminal **hacking** and other forms of **security breach**, to maintain acceptable levels of **confidentiality**, **integrity**, **availability** and non-repudiation. Cybersecurity also includes measures to protect **sensitive** data and **user privacy** during the **communication**, transmission and exchange of **information** over the **Internet** and other networks.

Cybersecurity has become increasingly important in recent years due to the increase in the use of the Internet and the spread of connected devices, such as smartphones, tablets and the **Internet of Things**. Organisations and individuals need to take appropriate **security** measures to protect hardware, **software** and **data** from cyber threats.

Cybersecurity measures may include data **encryption**, **network segmentation** and other secure architecture designs, user **authentication** and **authorisation**, protecting systems from viruses and other malware, and **vulnerability management** – especially through **patch management** and improving users' online behaviours.

Further reading:

Anderson, R., 2020. *Security engineering: a guide to building dependable distributed systems*, 3rd edition. Indianapolis: John Wiley.
Sarker, I.H., Kayes, A.S.M., Badsha, S., Alqahtani, H., Watters, P. and Ng, A., 2020. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, 7, 1–29, https://doi.org/10.1186/s40537-020-00318-5.
Skillicorn, D.B., 2021. *Cybersecurity for everyone*. Boca Raton, FL: CRC Press.

See also: INFORMATION SECURITY

Cyberstalking

Cyberstalking is the practice of **stalking** or **harassing** an individual or group using the **Internet** as the chief mode of access (it can take place alongside offline stalking). Cyberstalking is a continuous and targeted process, as opposed to a one-off or short-term interaction, which would be classified as *cyberbullying* or *trolling*. Cyberstalking, like offline stalking, is intended to discomfit the victim, for example by humiliating, embarrassing or scaring them. Typical behaviours include leaving insulting comments, sending abusive emails, posting libels or embarrassing truths, threatening physical violence, **doxxing** (publishing real-world addresses), **hacking** web resources, contacting friends or colleagues, ordering goods to be sent to the victim, posting **deepfake**s of the victim (for example, on pornography sites) or pretending to be the victim (for example, registering an **identity** on a dating site). The aim is to make the Internet and **social media** threatening and stressful spaces for the victim. Cyberstalking can be random, but is often gender-based, and may also focus on ex-partners, **celebriti**es or people who are the targets of online mobs and Twitterstorms. Many **jurisdiction**s have attempted to legislate against it, with varying degrees of success.

Further reading:
Parsons-Pollard, N. and Moriarty, L.J., 2009. Cyberstalking: utilizing what we do know. *Victims and Offenders*, 4(4), 435–41, https://doi.org/10.1080/15564880903227644.

See also: ATTENTIONAL PRIVACY, BLACKMAIL, CHILLING EFFECT, INTERNET OF PEOPLE, INTRUSION UPON SECLUSION, INVIOLATE PERSONALITY, PUBLISHING

Cyberterrorism

Cyberterrorism is the use of the **Internet** for terrorist acts, such as the destruction of infrastructure or the disruption of emergency services. Definitions vary as to how deeply the Internet has to be implicated in the action, and as to what makes a case of terrorism distinct from cases of **cyberwarfare** and **cybercrime**, although the aims of the act should be political or ideological, rather than for direct material gain. The privacy issue most strongly associated with cyberterrorism is the collection of data by law enforcement agencies, either to prosecute perpetrators, or alternatively to prevent action through intelligence. Courts must assess the **privacy risk**s of processing **personal data** against the **risk** of terror attack and make a judgment of **proportionality** about governments' actions.

Because of the danger of terrorism, such judgments may be unfavourable to privacy. But on the other hand, cyberterrorism has not yet been shown to be as dangerous to life as conventional terror attacks.

Furthermore, some terrorism scholars, such as Maura Conway, argue that cyberterrorism is a low risk, because (i) it requires technical expertise which few terrorists have in practice (and recruiting such expertise would be a **security** risk for them); (ii) it fails to provide the spectacle to guarantee widespread media coverage upon which terrorism feeds, and (iii) its outcomes might easily be seen as **software** failures, whereas terrorism needs to be seen as deliberate, premeditated and hostile.

Further reading:

Conway, M., 2011. Against cyberterrorism. *Communications of the ACM*, 54(2), 26–8, https://doi.org/10.1145/1897816.1897829.

Foggetti, N., 2009. Cyber-terrorism and the right to privacy in the third pillar perspective. *Masaryk University Journal of Law and Technology*, 3(3), 365–76. www.ceeol.com/search/article-detail?id=895174.

See also: CYBERCRIME, CYBERWARFARE, CYBERSECURITY, RISK, DATA DESTRUCTION, DATA IN USE, DATA PROCESSING, IDEOLOGICAL PRIVACY, INTERNET, PERSONAL DATA, PRIVACY RISK, PROPORTIONALITY, SECURITY, SOFTWARE

Cyberwarfare

Cyberwarfare is the use by states of computing resources, including **hacking**, to attack foreign states, for example for espionage or sabotage. The term has no agreed meaning internationally or in treaties, and can be stretched – for instance, Russian cyberattacks are sometimes performed by non-governmental nationalist hacktivist groups with tacit state support, while some Chinese state-backed cyberattacks are in the interests of Chinese business (industrial espionage) rather than the state itself directly.

While cyberwarfare usually focuses on infrastructure, it may have **privacy** implications. Many states are concerned with preventing enemy states, or companies based in enemy states, from getting access to **personal data** about their citizens (for instance, popular video sharing platform TikTok has been banned in some countries because of the possibility that personal data might be stored in a Chinese **cloud**). Cyberwarfare techniques also intrude into the **private sphere**, co-opting household appliances or **Internet of Things** devices into their **botnet**s.

Further reading:

Arquilla, J., 2021. *Bitskrieg: the new challenge of cyberwarfare*. Cambridge: Polity Press.

Lin, P., 2021. *TikTok vs Douyin: a security and privacy analysis*. Toronto: Citizen Lab, https://citizenlab.ca/2021/03/tiktok-vs-douyin-security-privacy-analysis/.

See also: CYBERCRIME, CYBERSECURITY, CYBERTERRORISM, SECURITY

Cypher

See: CIPHER

Cypherpunk

A cypherpunk is someone who promotes the usage of robust **cryptography** and **privacy-enhancing technology** as means of bringing about social change. The phrase first appeared in the 1990s in relation to the Internet and the developing field of cryptography and has since expanded to include a wide range of issues regarding **privacy**, **anonymity** and digital rights. A fundamental human right, according to cypherpunks, is the right to privacy, and in the digital age, strong cryptography is necessary to protect it. They typically see corporations and governments as **threat**s to people’s **autonomy** and privacy, and they work to give people the power to take control of their own **data** and **communications**.

Cypherpunks have been involved in a wide range of privacy and cryptography-related activities, such as creating and promoting encryption **software**, promoting the use of **Tor** and **I2P** as anonymous **communication** networks and promoting the use of **cryptocurrencies** such as Bitcoin. In addition to political activism, cypherpunks are active in a variety of social movements that address concerns about **surveillance**, **censorship** and digital rights.

Further reading:

Assange, J., Appelbaum, J., Muller-Maguhn, A. and Zimmermann, J., 2016. *Cypherpunks: freedom and the future of the Internet.* New York: OR Books.

Jarvis, C., 2022. Cypherpunk ideology: objectives, profiles, and influences (1992–1998). *Internet Histories*, 6(3), 315–42, https://doi.org/10.1080/24701475.2021.1935547.

See also: COMMUNICATION PRIVACY, PRIVACY AS CONTROL, RIGHT TO PRIVACY

D

Dark Pattern

Dark patterns are design elements or strategies for **user** interfaces with the goal of **nudging**, tricking or misleading users into acting in ways they otherwise would not. These design strategies take advantage of psychological tendencies and cognitive biases to influence user behaviour in ways that are advantageous to the designer or **third party** rather than the user.

Dark patterns can take many different forms, including misdirection (using visual or verbal cues to steer users to make a particular decision), forced actions (providing users with no choice or a false choice between options), hidden costs (hiding or obscuring the real cost of a product) and social proof (using social cues or testimonials to imply a false sense of consensus).

Further reading:
Kowalczyk, M., Gunawan, J.T., Choffnes, D., Dubois, D.J., Hartzog, W. and Wilson, C., 2023. Understanding dark patterns in home IoT devices. *In*: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ACM, article no.179, https://doi.org/10.1145/3544548.3581432.

See also: DECISIONAL PRIVACY, TRUST

Dark Web

The Dark Web refers to a portion of the **World Wide Web** that is deliberately segregated from the **public** Web and requires custom **software** to reach. The purpose of the Dark Web is to allow **anonymou**s and peer-to-peer **communication**. A *darknet*, a **network** which is part of the Dark Web, will typically be **encrypted**, with a specific communication **protocol**, and some kind of complex pathway (such as the layered servers used by the **TOR** onion router) which makes it extremely hard, if not impossible, to track users' **IP addresses** or geographical locations.

The anonymity of the Dark Web is of course valuable for criminals and extortionists, for collaborating on illegal activities such as match fixing and people trafficking, for hosting criminal marketplaces to buy and sell illegal products and services and disseminating hardcore and child pornography. However, it may also host the activities of political activists in autocracies,

whistleblowers and anti-**censorship** campaigners, and so is not an exclusively criminal space.

The Dark Web is a part of the *Deep Web*, which is that portion of the Web which is not indexed by standard Web crawlers, and therefore does not appear on **search engine** results.

Further reading:
Akhgar, B., Gercke, M., Vrochidis, S. and Gibson, H., eds, 2021. *Dark Web investigation*. Cham: Springer, https://doi.org/10.1007/978-3-030-55343-2.

See also: CYBERCRIME, CYBERTERRORISM

Data

Data refers to sets of symbols used to express **information**. It is usually seen as a relatively primitive type of expression generated by or collected from an instrument or **sensor**, computed by a mechanical or technical process, or otherwise recorded or read directly from an environment. It is therefore pre-analytic, and we might say the 'raw material' for scientific, statistical or computational reasoning (sometimes referred to as *raw data*). Data is therefore often assumed (controversially) to have a relatively objective status, somewhat less theory-laden, compared to the outputs of inference.

In computer science, data is uninterpreted, referring to the bits that are manipulated by the computing hardware, and which make up interpreted computational resources such as files, images and video. In other disciplines, 'data' usually refers to the primitive inputs to analytic processes, such as the output from sensors, experiments or surveys, in which case it will have an interpretation (a transfer function) based on the data acquisition methodology.

Data can be assembled in, for example, a **dataset**, **database** or *databank*. This imposes a structure on the data and improves **data utility** for analysis. It also facilitates **data linkage** and can therefore make the data more **vulnerable** to **statistical disclosure**.

The massive growth of computing power and storage has led to a corresponding increase in the quantity of data (leading to the use of the term **big data**), which has fuelled the so-called **digital economy**. Because much of this data expresses information about people, this has severe **privacy** implications, and was one of the drivers for the development of **data protection** law.

Further reading:
Kitchin, R., 2014. *The data revolution: big data, open data, data infrastructures and their consequences*. London: Sage.

See also: DATA ENVIRONMENT, DATA IN USE, DATA STORAGE, PERSONAL DATA, STATISTICAL DISCLOSURE CONTROL

Data Abuse

Misuse of data occurs when **personal data** is used inappropriately, against best practice or non-compliantly with relevant regulations – for example when it is used for purposes other than those for which it was acquired. Misuse is usually called abusive if there is malicious intent, and in particular when **harm** is caused to the **data subject**.

Further reading:
Privacy International, 2023. *Examples of abuse*, https://privacyinternational.org/examples.

See also: MISUSE OF PERSONAL INFORMATION, PURPOSE LIMITATION, PURPOSE SPECIFICATION

Data Ageing

As soon as it is created, **data** is immediately historical. As time passes the temporal distance between the data and the characteristic, event or phenomenon that it represents increases and so the data is said to age. Of course, data can be updated and also timestamped (and if so the mapping of the data onto the world should remain true).

Data ageing is one element of **data quality**. It also tends to increase **data divergence**, because as the data ages, its **linkability** to related data (including more current data) will tend to decrease.

Further reading:
Elliot, M. and Dale, A., 1999. Scenarios of attack: the data intruder's perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring), 6–10, www.researchgate.net/publication/343963431_Scenarios_of_attack_the_data_intruder's_perspective_on_statistical_disclosure_risk.

Data Aging

See: DATA AGEING

Data at Rest

One component of a tripartite scheme encapsulating the states of (digital) **data**. When data is at rest it is stored (in a computer system or other hardware) in a location that is not temporary, and it is not currently being processed. Note that this definition is neither formal nor precise and the boundary between **data in use** and data at rest is a matter of judgment about the frequency of use.

See also: DATA IN TRANSIT

Database

An often large collection of managed **data** that is stored and accessed electronically. Databases are typically used to support the multiple operational needs; they also therefore tend to be updated as new **information** arrives and, in these respects, they are distinct from **dataset**s, which tend to be collections of data that are fixed and have a single function of supporting statistical analysis and **machine learning**.

See also: DATASET

Database of Ruin

The **Database** of Ruin is a rhetorical description by legal scholar Paul Ohm of the consequences of the increase of **information** into the **public domain**. As more information about individuals is made accessible, or discovered by adversaries, linkage between **anonymised** databases becomes increasingly possible – in other words, the same individual can be picked out across databases. This creates a problem of accretion, because this linkage adds to the adversary’s information, and ultimately may enable them to **reidentify** entries in other anonymised databases. As this is irreversible, eventually, on Ohm’s dramatic metaphor, distinctions between databases will collapse, and it will be as if all anonymised **data** was available in the clear on a single overarching database. Ohm called an individual’s ‘Database of Ruin’ the set of compromising and **sensitive** data that would be associated with that individual in this overarching database.

If the construction of the Database of Ruin was possible, Ohm argued, this would discredit the **Personally Identifiable Information** (PII) approach to privacy protection. If only sensitive information were regulated, this would not prevent pieces of information that were individually non-sensitive being used to make the database linkages. Hence, on this picture, the **publication** of any information is potentially dangerous, in the sense that it could gradually and cumulatively lead an adversary to sensitive information about individuals.

While Ohm wrote as if the Database of Ruin was a genuine threat, he gave a largely theoretical justification, and did not specify how easily it could practically emerge.

Further reading:

Elliot, M., O'Hara, K., Raab, C., O'Keeffe, C.M., Mackey, E., Dibben, C., Gowans, H., Purdam, K. and McCullagh, K., 2018. Functional anonymisation: personal data and the data environment. *Computer Law and Security Review*, 34(2), 204–21, https://doi.org/10.1016/j.clsr.2018.02.001.

Ohm, P., 2010. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701–77. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006.

See also: DATA ENVIRONMENT, DATA LINKAGE, DIGITAL FOOTPRINT, FUNCTIONAL ANONYMISATION, LINKAGE ATTACK, PERSONAL DATA, SINGLE OUT

Data Breach

The term 'data breach' is often used to refer to a failure of **information security** that compromises **personal data**. Breaches of various laws – **data protection**, **privacy** and **confidentiality** – could result from unauthorised access to **personal information**, and so the term applies more to the fact of the failure than to its legal consequences.

The full term under the EU's **GDPR** is 'personal data breach', and the definition provided in Article 4(12) mirrors that given above. **Common law** confidentiality jurisprudence refers to a **breach of confidence**, which could equally result from a failure of information security. Other laws use equivalent terms: a breach of US health privacy legislation is instead known as a HIPAA violation.

Further reading:

Cheng, L., Liu, F. and Yao, D., 2017. Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5), e1211, https://doi.org/10.1002/widm.1211.

See also: HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT, US PRIVACY LAWS

Data Breach Notification

See: BREACH DISCLOSURE

Data Broker

A data broker or **information** broker is a person or company that collects **personal data** about people, or valuable **data** about companies, legally (from **public** sources, or via purchasing), to sell it to an interested **third party**.

They collect the information from a range of sources (e.g., bankruptcy **record**s, court records, warranty registrations, electoral registers, as well as from other data brokers), but not from individuals or companies themselves. The subjects of the information are therefore usually unaware of the dossier held about them, for sale on an open market. The broker may add value to the dossier by making **inference**s about subjects, categorising or clustering them. These categories may be used, for instance, for **targeted advertising** (either by the broker or by a third party).

Data brokering is a clear threat to the **privacy** of data subjects. Lack of **transparency** makes error correction hard. **Data storage** creates a **security** risk. However, data brokering does also have positive elements, such as allowing **identity** checks, making **identity theft** harder, supporting **personalisation** of services, and other kinds of **risk** mitigation.

Recently, **digital footprint** management services have entered the market, offering to contact lists of data brokers on behalf of their clients to demand removal of their personal data.

Further reading:

Federal Trade Commission, 2014. *Data brokers*: *a call for transparency and accountability*. Washington DC: Federal Trade Commission, www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

See also: AD EXCHANGE, AD NETWORK, DEMOGRAPHIC ADVERTISING, VALUE OF DATA

Data Capture

The first stage in the **data lifecycle** is its acquisition. Data capture is a term applied in various ways to the process of acquisition. It is sometimes used quite generally to refer to any process that produces **data** for analysis but is also used with more precision to cover the extraction of usable data from some kind of process, document or (usually electronic) device. For example, many processes begin with the manual filling in of paper forms. This is sometimes referred to as data capture, but more often the term is applied to the processes of digitising and storing the **information** (e.g., via optical character recognition). Other types of data capture are connected directly with the use of digital devices, such as the creation and storage of **sensor** data, or the gathering of data as a by-product of interaction with a website, or from the use of smart cards in a building. Automated data capture is supposedly less error-prone and cheaper than manual capture, and also produces a more standardised product (possibly including the automated generation of **metadata**). It is also likely to be harder for individuals to detect that data relevant to them is being captured.

Further reading:

Van den Eynden, V., 2020. The research data lifecycle. *In*: Corti, L., Van den Eynden, V., Bishop, L. and Woollard, M., eds, *Managing and sharing research data*: *a guide to good practice*, 2nd edition. London: Sage, 33–43, https://doi.org/10.25607/OBP-1540.

Wickramasuriya, J., Datt, M., Mehrotra, S. and Venkatasubramanian, N., 2004. Privacy protecting data collection in media spaces. *In*: *MULTIMEDIA '04*: *proceedings of the 12th annual ACM international conference on multimedia*, ACM, 48–55, https://doi.org/10.1145/1027527.1027537.

See also: DATA EXHAUST, DATA FLOW, DATA HARVESTING, DATA MINING, DATA STORAGE, PURPOSE LIMITATION, PURPOSE SPECIFICATION, SMART DEVICE

Data Centre

Data centres are used to host and manage an organisation's **application**s and **data**. They are designed to be reliable, **secure** and highly **available**, often using redundancy to support **resilience**, and can range in size from small corporate data centres to large **public** data centres that host **Internet** infrastructure and **cloud storage**. Where possible, cloud providers' data centres will be near clients to minimise latency of retrieval, but other factors come into play, including the large energy costs of running the centres. The regulation of data centres will come under the **jurisdiction** in which they are physically located, which has repercussions for how **privacy**, **data protection** and **data sovereignty** are managed.

See also: BIG DATA, CLOUD COMPUTING, INTERNET

Data Classification

A form of high-level **metadata** which categorises whole **dataset**s. Although the term is generic, the primary purpose of data classification for many organisations is to inform decision making about **data security**.

A common classification schema is a three-way split: **confidential**, internal and **public**. Some schemes split the confidential category into two subcategories, such as 'restricted' and 'highly restricted' (or similar).

See also: HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT, PRIMARY DATA

Data Controller

Under the EU's **GDPR**, the data controller is the actor (i.e., legal/**natural person**, public authority or any other kind of body) who determines the purposes and manner of personal **data processing**. This actor has the primary responsibility for data protection **compliance** and can be subjected to significant fines in the event of a personal **data breach**.

A common misconception is that all individuals using **personal data** are data controllers. In fact, employees are unlikely to be data controllers when acting in the course of their employment, and their employer would in fact be responsible for ensuring the processing of personal data in accordance with **data protection** law. Where a body dictates the terms of personal data use to a party who is not their employee, the latter is more likely to be a **data processor** and must be instructed in line with the GDPR's requirements.

Two bodies can use the same **information** as independent data controllers. However, where they jointly determine the purposes and manner of data processing, they become **joint data controller**s and must make **transparent** arrangements for their respective **compliance** organisations.

Further reading:

Information Commissioner's Office, 2022. *Controllers and processors*, https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/controllers-and-processors/controllers-and-processors-1-0.pdf.

Data Curation

For **data** to be reused, it needs to be understandable and usable in other contexts. However, data is often gathered from heterogeneous sources, of varying quality, reflecting different original intents of use. This poses potential problems of **data quality**, *consistency* and **data utility**.

Data curation consists of the methodological, technical and organisational processes of ensuring a relevant, high-quality, integrated **dataset**. Data curation processes can maximise **data utility** for a specific use or across a broader set of anticipated intended uses through the **data lifecycle**. This may include the critical selection of data sources, harmonising data/conceptual models and content/format, the addition of annotations or **metadata** to explain its **attribute**s, establishing **data provenance** mechanisms and supporting data discovery and accessibility.

Further reading:

Freitas, A. and Curry, E., 2016. Big data curation. *In*: Cavanillas, J.M., Curry, E. and Wahlster, W., eds, *New horizons for a data-driven economy*: *a roadmap for usage and exploitation of big data in Europe*. Cham: Springer, 87–118, https://doi.org/10.1007/978-3-319-21569-3_6.

See also: BIG DATA, DATA IN USE, DATA LIFECYCLE MANAGEMENT, DATA STEWARD, INFORMATION LIFECYCLE MANAGEMENT

Data Custodian

See: DATA STEWARD

Data Degaussing

Data destruction on magnetic media by exposing the device to strong magnetic forces. Degaussing can be difficult to verify.

Further reading:

Gutmann, P., 1996. Secure deletion of data from magnetic and solid-state memory. *Proceedings of the Sixth USENIX Security Symposium*, San Jose, CA, 77–90, www.usenix.org/legacy/publications/library/proceedings/sec96/full_papers/gutmann/.

See also: DELETION, SECURITY

Data Destruction

The complete elimination of all traces of some **data** from a system or device. The phrase emphasises that standard **deletion** is not sufficient, as traces of deleted data will remain recoverable on the storage medium. Actual destruction comes in multiple forms, including the physical destruction of the storage medium (shredding) the use of magnets to 'degauss' disks and the repeated overwriting of the storage medium with noise or arbitrary bits.

See also: CYBERSECURITY, DATA DEGAUSSING, DATA SANITISATION, DATA STORAGE

Data Divergence

A relationship between multiple **dataset**s whereby two pieces of **information** pertaining to the same underlying **attribute** for the same **population unit** are non-identical. There are several sources of divergence, including errors in datasets, differences in how information is coded and **data ageing**. Note that divergence is not synonymous with error as two pieces of **data** can be identically wrong and therefore convergent.

Elliot and Dale distinguish between data–data divergence (differences between two datasets) and data–world divergence (differences between a piece of data and the world). The former introduces errors into **record linkage** processes and the latter impacts on the validity of **inference**. Hence divergence can impact both *bona fide* analysis and adversarial **attack**s.

Further reading:

Elliot, M. and Dale, A., 1999. Scenarios of attack: the data intruder's perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring), 6–10, www.researchgate.net/publication/343963431_Scenarios_of_attack_the_data_intruder's_perspective_on_statistical_disclosure_risk.

See also: ADVERSARY, DATA LINKAGE, DATA UNIT

Data Dumping

The practice of extracting **data** from a system and saving it to another system. It is usually used for data migration and backup. To avoid **data breach**, it is important to store data dumps in **secure** locations and limit the access to authorised **user**s only, implementing **access control**.

Further reading:
Shu, X., Yao, D. and Bertino, E., 2015. Privacy-preserving detection of sensitive data exposure. *IEEE Transactions on Information Forensics and Security*, 10(5), 1092–1103, https://doi.org/10.1109/TIFS.2015.2398363.

See also: AUTHORISATION, DATA IN TRANSIT

Data Enclave

A **data environment** whereby **user**s (often researchers) are allowed **remote access** to **data** so that analyses can be run without the data themselves being transferred. This reduces **disclosure risk** by specifically preventing certain scenarios (that require specific linkage of **database**s). It has the advantage over on-site data labs of greater accessibility for analysts (as they do not need to travel to the lab's location) whilst still providing most of the control that on-site labs provide and so for many use cases it represents a sweet spot for organisations that wish to enable access to data for **secondary use**.

In common with on-site facilities, data enclave systems require some form of **output checking**.

See also: ACCESS CONTROL, DATA IN TRANSIT, DATA IN USE, DATA LINKAGE, DATA SAFE HAVEN, SAFE OUTPUT, SAFE SETTINGS

Data Environment

A core concept of **functional anonymisation** which can be best understood as a context for some **data**. Mackey and Elliot define a data environment as a set of formal and informal structures, processes, mechanisms and agents that either act on data, provide interpretable context for those data or define, control and/or interact with those data. Within the **Anonymisation Decision-Making Framework**, a data environment is deemed to comprise four components: agents, infrastructure, governance and other data. Data may be held in a number of environments/contexts at the same time, or as they transit through an organisation.

Further reading:
Mackey, E. and Elliot, M., 2013. Understanding the data environment. *XRDS: Crossroads, The ACM Magazine for Students*, 20(1), 36–9, https://doi.org/10.1145/2508973.

See also: DATA AT REST, DATA GOVERNANCE, DATA IN TRANSIT, DATA IN USE, DATA SITUATION

Data Environment Analysis

In general, a set of processes for identifying and analysing the components of a **data environment**. Elliot et al define a more specific meaning whereby the **data** that are available to an **adversary** within a given scenario are identified through an analysis of data collection instruments and then coded into **key variable**s though a process of key variable mapping.

Further reading:

Elliot, M., Lomax, S., Mackey, E. and Purdam, K., 2010. Data environment analysis and the key variable mapping system. In *International Conference on Privacy in Statistical Databases*, Berlin: Springer, 138–47, https://doi.org/10.1007/978-3-642-15838-4_13.

Smith, D. and Elliot, M., 2014. A graph-based approach to key variable mapping. *Journal of Privacy and Confidentiality*, 6(2), https://doi.org/10.29012/jpc.v6i2.641.

See also: MOTIVATED INTRUDER TEST, SCENARIO ANALYSIS

Data Ethics

Data ethics is often used as a synonym for **information ethics**. However, it may also refer in a more specific way to the ethics of **data** and **big data** analysis, especially where the data is **personal data** and **privacy** is particularly impacted. Principles of data ethics typically mirror those found in **data protection** law, and include the **transparency** of **data processing**, the means of obtaining and maintaining **consent** of **data subject**s and the nature of the rights over the data of both the **data controller** and the data subjects and the **harm**s arising from a particular use of data.

Rather than simply duplicating each other, however, data ethics and data law can be mutually informative, with ethical reflection providing depth and rigour to the interpretation of a controller’s legal obligations. The precise relationship between data ethics (e.g., research ethics) and data protection laws can be controversial, particularly where the former’s emphasis on the individual’s **informed consent** can be modulated by the latter’s more discretionary framework.

Further reading:

Gefenas, E., Lekstutiene, J., Lukaseviciene, V., Hartlev, M., Mourby, M. and Cathaoir, K.Ó., 2022. Controversies between regulations of research ethics and protection of personal data: informed consent at a cross-road. *Medicine, Health Care, and Philosophy*, 25(1), 23–30, https://doi.org/10.1007/s11019-02110060-1.

Rochel, J., 2021. Ethics in the GDPR: a blueprint for applied legal theory. *International Data Privacy Law*, 11(2), 209–23, https://doi.org/10.1093/idpl/ipab007.

Zwitter, A., 2014. Big data ethics. *Big Data and Society*, 1(2), 1–6, https://doi.org/10.1177/2053951714559253.

See also: CODE OF ETHICS, DATA PROTECTION PRINCIPLES, LAWFULNESS

Data Exhaust

Interaction online requires a computer as intermediary. As all interactions involve exchanges of digital **information**, the ambient **data** representing this information is a necessary by-product of the interaction. The metaphor of the exhaust gases of a chemical process indicates that the data exhaust is passively produced, is not essential to the process from the point of view of the interactors and may not be independently useful. However, since it can be captured and timestamped, it can be used to develop a more coherent model of the interactors. For instance, the data exhaust of a particular individual can be aggregated, to produce a **digital footprint**.

Further reading:

George, G., Haas, M.R. and Pentland, A., 2014. Big data and management. *Academy of Management Journal*, 57(2), 321–6, https://doi.org/10.5465/amj.2014.4002.

See also: BIG DATA, RECORD, SURVEILLANCE, SURVEILLANCE CAPITALISM

Datafication

Datafication is the practice of representing aspects of life, particularly social life, as **data**. It is generally held that datafication is accelerating, so that not only the amount of data, but also the spread of things represented, is increasing dramatically. Many aspects of human behaviour and psychology, as well as spaces such as cities, transport networks and infrastructure, have been *datafied* by technology. Once something has been represented, it can be reasoned about and subjected to **predictive analytics**.

Further reading:
Koopman, C., 2019. *How we became our data*: *a genealogy of the informational person*. Chicago: University of Chicago Press.

See also: BIG DATA, DIGITAL TWIN, RECORD

Data Flow

In technical use, data flow refers to the movement of **data** within a computer system or **network** or between networks. In a computer system, data flow can include the movement of data between hardware and **software**, or between different hardware components. Data flow can be controlled by network **protocol**s, which define the rules for **data transfer**. A record of a data flow is called **data provenance**.

More generally, data flows can describe the broader movement of **information** about people between places, and across contexts. Nissenbaum is particularly interested in movement of data from one context to another, meaning not only a change in location, but also changes in who has access to the information and for what purposes. She argues that norms of information flow govern every arena of human life, and **right**s **to privacy** ultimately require compatibility with these expectations as to how and when information will travel between contexts.

Further reading:
Nissenbaum, H., 2010. *Privacy in context: technology, policy, and the integrity of social life*. Stanford: Stanford University Press.

See also: CONTEXTUAL INTEGRITY, DATA IN TRANSIT, PURPOSE LIMITATION

Data Flow Diagram

A data flow diagram is a visual representation of how **data** is used within a system, which data is input to each process, and which other processes use its output. Typically, a data flow diagram does not include control **information**, about what triggers which processes and when, but this is a natural overlay on top of **data flow**. Such diagrams are valuable in enabling organisations to understand how their **data lifecycle**. For example, they should include representations of **data provenance**, where **personal data** is being used, which agents are involved in **data processing**, when **firewall**s prevent access to data, where **remote access** to data is possible, and whether data is being held on devices outside the direct control of the organisation. With a comprehensive focus on all aspects of data use, they will visualise the organisation's **data situation**.

Data flow diagrams facilitate the auditing of data processing within organisations, and to assess **vulnerabiliti**es to **data breach**es. For instance, the **Anonymisation Decision-Making Framework (ADF)** recommends the visualisation of the data situation, to determine exactly what **data controller**s are responsible for and whether data processing is **GDPR-compliant**.

Further reading:

Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners' guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

Jilani, A.A.A., Nadeem, A., Kim, T.H. and Cho, E.S., 2008. Formal representations of the data flow diagram: a survey. *In*: *Advanced software engineering and its applications*, IEEE, 153–8, https://doi.org/10.1109/ASEA.2008.34.

See also: ACCESS CONTROL, ANONYMISATION, DATA ENVIRONMENT, DATA GOVERNANCE, DATA IN USE, DATA LIFECYCLE MANAGEMENT, DATA SITUATION AUDIT

Data Governance

Data governance can be a synonym for **information governance**, or it can carry the implication of governance specifically of **data** held in digital format. It refers to a set of processes, policies and procedures ensuring the quality and timely **availability** of data for analysis, while keeping it **secure** and **compliant** with regulation.

Further reading:

Ladley, J., 2020. *Data governance*: *how to design, deploy, and sustain an effective data governance program*, 2nd edition. London: Academic Press.

See also: DATA LIFECYCLE MANAGEMENT, DATA QUALITY, DATA STEWARD

Data Harmonisation

The processing of multiple **dataset**s so that the coding schemes/variable constructions are identical. This enables **data linkage** and comparisons between datasets. Lack of harmonisation between different data sources creates a large overhead for an analyst.

It is easier to carry out **linkage attack**s or **mash attack**s on harmonised datasets, so an **adversary** may well also engage in harmonisation before conducting their attack.

Further reading:

Fichtinger, A., Rix, J., Schäffler, U., Michi, I., Gone, M. and Reitz, T., 2011. Data harmonisation put into practice by the HUMBOLDT project. *International Journal of Spatial Data and Infrastructures Research*, 6, https://ijsdir.sadl.kuleuven.be/index.php/ijsdir/article/view/191.

Nan, Y., Del Ser, J., Walsh, S., et al., 2022. Data harmonisation for information fusion in digital healthcare: a state-of-the-art systematic review, meta-analysis and future research directions. *Information Fusion*, 82, 99–122, https://doi.org/10.1016/j.inffus.2022.01.001.

Data Harvesting

Data harvesting is the practice of taking **data** from an **information** source, usually legally, but often without the explicit cooperation of the source's managers. One example would be **scraping** a website for the information contained on it, such as contact details, news stories, survey results or inventory. A second method would be to use a site's **application programming interface (API)**, which is designed to allow access to its data. Scraping allows the harvester to get at any data on display; the API may give access to **database**s below the surface, but equally may restrict access to **sensitive** data (APIs may also restrict the amount of data harvested).

The purpose may be to discover business data, for commercial or political intelligence or to conduct research. This can of course be done by hand, but at scale tools called *crawlers* are required, which parse websites, copy the information that is likely to be valuable, and process it into a structured format, such as a spreadsheet or data frame.

Further reading:

Glez-Peña, D., Lourenço, A., López-Fernández, H., Reboiro-Jato, M. and Fdez-Riverola, F., 2014. Web scraping technologies in an API world. *Briefings in Bioinformatics*, 15(5), 788–97, https://doi.org/10.1093/bib/bbt026.

Mitchell, R., 2018. *Web scraping with Python*: *collecting more data from the modern Web*, 2nd edition. Sebastopol, CA: O'Reilly Media.

See also: DATA CAPTURE, PUBLIC DOMAIN

Data in Motion

See: DATA IN TRANSIT

Data Intermediary

Data intermediaries are organisations, institutions or platforms that stand between **user**s and providers of data, providing support and **data stewardship** services, absorbing some of the costs and **risk**s associated with in-house **data processing**, and benefiting from economies of scale. Examples of existing data intermediaries include **data trust**s, data exchanges (platforms facilitating data discovery and **data sharing**), **Personal Information Management System**s **(PIMS)**, data cooperatives (shared data spaces), **data safe haven**s, **trusted research environment**s and other trusted third parties.

Further reading:

Centre for Data Ethics and Innovation, 2021. *Unlocking the value of data*: *exploring the role of data intermediaries*. London: Centre for Data Ethics and Innovation, www.gov.uk/government/publications/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries.

See also: DATA CENTRE, DATA STEWARDSHIP ORGANISATION, THIRD PARTY, TRUST, TRUSTED THIRD PARTY

Data In Transit

One component of a tripartite scheme encapsulating the states of digital **data**. When data is in transit it is currently moving between two locations. This movement may be within an organisation on its internal network(s) or moving between two organisations, or at large in the global **data environment** (e.g., on the **Internet**). The concept of a **dynamic data situation** implies data in transit.

See also: DATA AT REST, DATA IN USE, DATA SITUATION

Data Intruder

See: INTRUDER

Data Intrusion Simulation

A method of **disclosure risk** assessment for samples of **microdata** developed by Skinner and Elliot which calculates the probability of a correct **match** given a **unique** match on a given set of variables. The method has been shown to produce accurate estimates of the underlying prevalence of a combination of characteristics averaged across the whole file, and this gives a good estimate of the average **risk** posed by an **adversary** who attempts to **identify** a specific individual within a sample dataset. The method is useful in determining overall levels of risk, and therefore can feed into file specifications, but does not allow for variations in risk across a file, particularly **special unique**s.

Further reading:
Skinner, C.J. and Elliot, M.J., 2002. A measure of disclosure risk for microdata. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 64(4), 855–67, https://doi.org/10.1111/1467-9868.00365.

See also: DISCLOSURE, INTRUDER, RISK ASSESSMENT, STATISTICAL DISCLOSURE

Data in Use

One component of a tripartite scheme encapsulating the states of digital **data**. When data is in use it is currently being processed and will typically reside in RAM or a CPU cache or other temporary location.

See also: DATA AT REST, DATA IN TRANSIT

Data Lake

See: DATA WAREHOUSE

Data Lifecycle

While there is no overarching model of the **data** lifecycle, **data processing** can be seen as a number of (privacy-relevant) stages. Such stages might include: the *creation* of data or **data capture**; **data storage**; its *use*; its *archiving*; its **deletion**; or **data destruction**. Where the data is **personal data**, all of these stages are captured by the **GDPR** definition of data processing. **Data lifecycle management**, and planning for each of its stages, can increase the **value of data**, and facilitate reuse.

Creation includes the capture of **information** from devices, its purchase or its being volunteered by **data subject**s (e.g., **respondent**s in a survey), and should also include the development of **metadata** to facilitate its future use. Fair practice demands that there be a legitimate ground for its capture and **lawful basis** for its processing.

Storage, and, later, archiving each require suitable storage media and **security** measures to be in place, as well as making it possible for data subjects to retrieve and evaluate the data and have it changed if inaccurate. Funders may also wish access to data at a time of their choosing, and the data may also come under the scope of **Freedom of Information** legislation.

The use of data, which could include **data sharing** or **publication**, must be for the purposes specified upon creation. New **inference**s may increase the amount of information held about a data subject, while **anonymisation** or **pseudonymisation** techniques might support subjects' **privacy**. Data is likely to have to be 'cleaned up' before use. Future reuse of data is also possible, which may involve having to trace data subjects to gain fresh consent.

Deletion may have to occur by a particular deadline; alternatively, there may be regulations demanding that data be kept available for law enforcement agencies for a specific period of time (e.g., by Internet Service Providers), in which case it must be archived securely.

Further reading:

Van den Eynden, V., 2020. The research data lifecycle. *In*: Corti, L., Van den Eynden, V., Bishop, L. and Woollard, M., eds, *Managing and sharing research data*: *a guide to good practice*, 2nd edition. London: Sage, 33–43.

See also: CONSENT, DATA AT REST, DATA CURATION, DATA ENVIRONMENT, DATA FLOW, DATA GOVERNANCE, DATA IN TRANSIT, DATA IN USE, DATA PROTECTION, DATA SITUATION AUDIT, DATA STEWARD, DECLARED DATA, FAIRNESS, INFERRED DATA, INFORMATION SECURITY

Data Lifecycle Management

Data lifecycle management involves managing **data** by using the stages in the **data lifecycle** as a structuring principle. It is intended to help deliver the resources and equipment needed to gather, store and use the data, and to enable an organisation to continue to use data through time, either reusing it for future projects, **data sharing** with external partners or alternatively enabling different or new people in the organisation to use the data even if they were not directly involved in its original creation. Lifecycle management plans should be in place early, with organisational roles and responsibilities assigned, and sufficient resources and support services made available.

From the **privacy** perspective, most important management tasks include ensuring the privacy of **data subject**s, if the data is **personal data**; **compliance** with **data protection** law (e.g., **GDPR**); planning for **data breach**es or other potential future problems; maintaining effective **security**; and ensuring ethical practice. Where the data is personal this is usually referred to as **records management**.

Further reading:

Van den Eynden, V., 2020. The research data lifecycle. *In*: Corti, L., Van den Eynden, V., Bishop, L. and Woollard, M., eds, *Managing and sharing research data*: *a guide to good practice*, 2nd edition. London: Sage, 33–43.

See also: DATA AT REST, DATA IN TRANSIT, DATA IN USE

Data Linkage

Any process by which **data** (usually relating to the same **population**) might be joined or connected. This is generally carried out to enhance a **dataset** and to increase the range and complexity of analyses and the quality of **inference**s that can be made. A specific form of data linkage known as **record linkage** involves records in different **database**s corresponding to the same **population unit**s being joined to create a new dataset.

The **privacy concern**s arising from data linkage are multiple. **Data subject**s may not have **consented** to such linkage or, if they had, they may not have understood the consequences. Also, data that was previously anonymous could, when joined to other data, become **identifiable**. For instance, linking the data that X is 2m tall with the data that X is Scottish tells us that X is a 2m tall Scot, which will reduce the possibilities for X's identity dramatically, even in a large population.

Further reading:

Bohensky, M.A., Jolley, D., Sundararajan, V., Evans, S., Pilcher, D.V., Scott, I. and Brand, C.A., 2010. Data linkage: a powerful research tool with potential problems. *BMC Health Services Research*, 10(1), 1–7, https://doi.org/10.1186/1472-6963-10-346.

Christen, P., 2012. *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*. New York: Springer, https://doi.org/10.1007/978-3-642-31164-2_2.

See also: IDENTIFIABLE DATA, MATCHING

Data Map

A **data harmonisation** tool providing a **record** of the links between two or more data models which in turn enables the corresponding **database**s to be linked.

See also: DATA LINKAGE, REIDENTIFICATION

Data Minimisation

The intentional processing of the minimum amount of **data** required for the specified (lawful) purpose. The term derives from a longstanding principle within **data protection**. In 1973, the Council of Ministers (CoM, a body of the **Council of Europe**) adopted a resolution recommending that member states take action to preserve individual **privacy** in the context of automated **data processing** in the **private sector**. Many of the principles set out in this resolution can be found in contemporary form: obsolete data should be deleted, and the **information** retained should be appropriate and relevant for its purpose.

The **GDPR**'s expression of this principle is that **personal data** should be 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed'. This is essentially a replication of the principle as found in the previous **Data Protection Directive** 95/46 EC. A subtle shift, however, is that the Directive required that the data held should be 'not excessive' for their purpose (this being a replication of the principle as articulated in the Council of Europe's Convention 108 on automated data processing, which followed the CoM resolution eight years later). The GDPR tightens this requirement to 'limited to what is necessary' for a processing purpose, making it clear that anything beyond the minimum is excessive.

As a general rule, therefore, data processing under the laws of the EU and the Council of Europe should be limited to the minimum necessary for a legitimate purpose. The aim is to ensure the least amount of **intrusion** into the rights and freedoms of the **data subject** (Recital 156, GDPR). Although the data minimisation principle assumes that privacy is best served by keeping **identifiable data** processing to a minimum, the view can be taken that the maintenance of the connection between the individual and their data also serves their **legitimate interest** in controlling how it is used downstream.

See also: DATA PROTECTION PRINCIPLES

Data Minimisation Principle

See: DATA MINIMISATION

Data Mining

The processing of **data** – usually large scale – to discover new patterns using a battery of techniques from statistics and **machine learning** including cluster analysis, sequential pattern mining and anomaly detection. Since it was first coined in the 1990s the term has been popularised and broadened in usage to mean any large-scale processing of data typically for business intelligence purposes.

Because it involves extraction of patterns that may not be immediately visible in the data (sometimes called *latent constructs*), which might then be used to make predictions or for **personalisation** and **targeted adevertising**, data mining often raises **privacy concern**s. This in turn has led to the development of so-called **privacy preserving data mining** techniques.

Further reading:

Larose, D.T. and Larose, C.D., 2014. *Discovering knowledge in data*: *an introduction to data mining*, 2nd edition. Hoboken, NJ: John Wiley & Sons.

Xu, L., Jiang, C., Wang, J., Yuan, J. and Ren, Y., 2014. Information security in big data: privacy and data mining. *IEEE Access*, *2*, 1149–76, https://doi.org/10.1109/ACCESS.2014.2362522.

See also: BIG DATA, DATA PROCESSING, DATA WAREHOUSE

Data Ownership

Data ownership is a term that is polysemous and somewhat contested. As Fadler and Legner observe, the related debates in practice and research view the concept from different, often contrasting disciplinary perspectives. This point reflects the complex array of disciplines that are conceptual stakeholders: law, management science, **information** systems, ethics, economics and even psychology all have something to say on the topic. This complexity has been muddied further by the rise of **big data** for which 'ownership' is difficult to determine in practice, even if its definition were clear.

The related term 'data owner' is often used to connote a person or organisation with custody or control of information, without any firm conclusions about this entity's rights and responsibilities under relevant information/property law. Ownership might be assumed psychologically by an individual who has some stake in the development of some data resource or to whom some of the rights and responsibilities for a data **asset** have been delegated. In economics use of 'ownership' might be tied to contributions to the value chain. Ethicists on the other hand theorise who *ought* to be considered owners of data – irrespective of who is in practice or law.

A common legal conception of ownership is that it encompasses a 'bundle' of rights and responsibilities, with the precise composition of the bundle differing according to the nature of the property, and the circumstances in which it is held. These rights may include those of possession, control, exclusion, exploitation and destruction. However, answers to the question of whether **data** themselves are 'property' which can be 'owned' vary between different legal **jurisdiction**s.

Under English law, for example, the term 'data owner' may be used in contracts to signify that one party holds property rights (such as copyright) in the relevant data. In such a case, it is technically the rights in the data which are the owned property, not the (content of the) data themselves. Where access to the data is granted under a **licence agreement**, this does not transfer ownership of the rights in such data; such a transfer of ownership would require an assignment of **intellectual property** rights.

Where the data in question are **personal data** – that is, where they identify living **natural people** – the main concern of **privacy** regulators is not who owns the data in any sense, but who controls the data. This may be an entity with commercial rights in the data, but equally they may be a licensee, or perhaps even the data are not subject to any commercial or property rights and are not 'owned' by anyone.

Further reading:

Ballantyne, A., 2020. How should we think about clinical data ownership? *Journal of Medical Ethics*, 46(5), 289–94, https://10.1136/medethics-2018-105340.

Fadler, M. and Legner, C., 2021. Data ownership revisited: clarifying data accountabilities in times of big data and analytics, *Journal of Business Analytics*, 5(1), 123–39, https://doi.org/10.1080/2573234X.2021.1945961.

Grover, V., Chiang, R.H.L., Liang, T.-P. and Zhang, D., 2018. Creating strategic business value from big data analytics: a research framework. *Journal of Management Information Systems*, 35(2), 388–423, https://doi.org/10.1080/07421222.2018.1451951.

Ritter, J. and Mayer, A., 2017. Regulating data as property: a new construct for moving forward. *Duke Law and Technology Review*, 16(1), 220–77.

Royal Society, 2018. *Data ownership, rights and controls*: *reaching a common understanding*, https://royalsociety.org/-/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf.

See also: BIG DATA, DATA CONTROLLER, DATA CUSTODIAN, DATA ETHICS, DATA GOVERNANCE, DATA LIFECYCLE, DATA LIFECYCLE MANAGEMENT, DATA STEWARD, INFORMATION ETHICS, VALUE OF DATA

Data Portability

The EU’s **GDPR** introduced a right to receive **information** about oneself in a structured, transferable form. This enables individuals to exercise **privacy as control**; not only to access their **personal information**, but to withdraw their **consent** and entrust a new service provider with their **personal data**. However, the right only applies when personal data are processed on the basis of consent (or performance of a contract) and are processed using automated means, and if it can be applied without adversely affecting the rights and freedoms of others. This shows the limitations of the GDPR’s enactment of individual privacy as control; it may be an appropriate approach within consumer markets, but less so within (e.g.) public security or public health.

Further reading:

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L. and Sanchez, I., 2018. The right to data portability in the GDPR: towards user-centric interoperability of digital services. *Computer Law and Security Review*, 34(2), 193–203, https://doi.org/10.1016/j.clsr.2017.10.003.

See also: INTEROPERABILITY, RIGHT OF ACCESS, RIGHT TO DATA PORTABILITY

Data Privacy

See: INFORMATIONAL PRIVACY

Data Processing

The EU's **GDPR** defines data processing expansively, as any operation or set of operations performed upon **personal data**. This can include activity or inactivity: all uses of **personal information**, including retention and even **deletion**.

This broad concept of data processing should not be confused with that performed by a **data processor**. The latter is a more specific role, undertaken by an agent processing personal data on behalf of a **data controller**. Any person or organisation can engage in data processing, whether the GDPR would term them data processors or not.

See also: DATA IN USE, DATA RETENTION

Data Processor

The EU's **GDPR** defines a data processor as a legal or **natural person** who processes **personal data** on behalf of a **data controller**. This means that the data controller is the person or organisation responsible for determining the purposes and manner of the processing. The data processor, by contrast, can only use the personal data in question in accordance with the written instructions of the data controller.

Data Protection

The laws, policies, systems, **standard**s and measures variously designed to protect **information** which identifies people can be collectively termed 'data protection'. The extent to which data protection is distinct from **privacy**, and **informational privacy** in particular, has been widely debated. As data protection law and related **standard**s have grown in scope and complexity, so has their distinct value to regulate the processing of **personal information** (and not just the narrower subset of private information) with a more intricate balancing of the rights and obligations of precisely defined actors (**data subject**s, **data controller**s, **data processor**s and third parties).

Further reading:

Kokott, J. and Sobotta, C., 2013. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. *International Data Privacy Law*, 3(4), 222–8, https://doi.org/10.1093/idpl/ipt017.

Lynskey, O., 2014. Deconstructing data protection: the ‘added-value’ of a right to data protection in the EU legal order. *The International and Comparative Law Quarterly*, 63(3) 569–97, https://doi.org/10.1017/S0020589314000244.

See also: DATA PROCESSING, PERSONAL DATA, RIGHT TO DATA PROTECTION, THIRD PARTY

Data Protection Authority

All countries with **data protection** laws have public **regulators** charged with policy design and regulatory enforcement. Under the EU **GDPR**, these national regulators are termed Supervisory Authorities, and where more than one member state body might have **jurisdiction** over **data processing**, the GDPR makes provision for Lead and Competent Authorities to take charge of coordinated responses.

Further reading:

European Commission, 2023. *What are Data Protection Authorities?* https://commission.europa.eu/law/law-topic/data-protection/reform/what-are-data-protection-authorities-dpas_en.

See also: CROSS-BORDER PROCESSING, DATA PROTECTION POLICY, SUPERVISORY AUTHORITY

Data-Protection-By-Default

Article 25 of the EU **GDPR** requires **data controller**s to implement data-protection-by-design-and-by-default. This requirement is, for short, variously referred to as ‘data-protection-by-design’ and ‘data-protection-by-default’. The main distinction between these two terms is that data-protection-by-design is the process to be undertaken, with data-protection-by-default the desired outcome.

In essence, the requirement is for safeguards and **proportionate security** measures to be integrated into regular **data processing** activities, so they apply by default within an organisation.

Further reading:

European Commission, 2023. *What does data protection 'by design' and 'by default' mean*? https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_en.

See also: DATA PROTECTION POLICY

Data-Protection-By-Design

See: DATA-PROTECTION-BY-DEFAULT

Data Protection Directive

Before the European Union introduced the **GDPR** in 2018, many of the same provisions were brought into law via the Data Protection Directive (EC 95/46). The Directive was introduced in 1995 and based significantly on the **OECD Guideline**s on cross-border processing (1980), as well as the **Council of Europe**'s **Convention 108** (1981). As a Directive, the 1995 instrument did not have direct effect across the European Community (as it was then termed). Instead, member states were required to implement their own legislation to bring the Directive's provisions into force. In the UK, for example, the Directive was implemented by the Data Protection Act 1998.

Ellis and Oppenheim have argued that the Directive was inspired by concerns about **informational privacy**, but as the EC Treaties did not include **rights to privacy** within the Community's remit, the Commission placed emphasis on the importance of cross-border **data flows** for the functioning of the internal market, thus bringing **data protection** within the scope of its powers. Arguably, the EU **Charter on Fundamental Rights** has since developed the EU's powers to regulate human rights within its member states, and this is reflected in the expanded scope of the GDPR (e.g., in relation to the **risks** to individuals from **profiling**, and the **right to be forgotten**).

Further reading:

Ellis, S., and Oppenheim, C., 1993. Legal issues for information professionals, Part III: Data protection and the media – background to the Data Protection Act 1984 and the EC Draft Directive on Data Protection. *Journal of Information Science*, 19(2), 85–97, https://doi.org/10.1177/016555159301900201.

See also: CHARTER RIGHTS, DATA PROCESSING, RIGHT TO DATA PROTECTION

Data Protection Impact Assessment (DPIA)

Under the EU **GDPR**, the **data controller** must carry out a Data Protection Impact Assessment (DPIA) before processing data they consider likely to pose a high **risk** to the rights and freedoms of **natural person**s. The DPIA should identify risks, and mitigation measures to minimise them. If the identified risk cannot be mitigated, the controller should not proceed with the processing in question.

On the face of the Regulation, the scope of risks that should trigger a DPIA is not exhaustively defined. The most obvious category of natural persons the data controller should consider are the subjects of the **personal data** in question. Indeed, Article 35(7) states that the DPIA itself should contain ‘an assessment of the risks to the rights and freedoms of data subjects’. The **European Data Protection Board (EDPB)** has endorsed guidelines from the **Article 29 Working Party**, which state that the risks in question ‘primarily concern the rights to data protection and privacy but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion’.

However, it is not impossible that the natural persons at risk could in fact be people not represented in the data, but who will nonetheless be affected by downstream implications of the processing (e.g., where the processing is used to develop an **algorithm** used to make decisions about future **data subject**s, such as patients in a healthcare system). Unlike the US Federal Common Rule governing human research, there is nothing in the GDPR which specifically prohibits consideration of longer-term effects of the processing. That said, much of the available EU and national guidance focuses on **harm** to data subjects, which has the potential to limit the scope of harm contemplated.

National Supervisory Authorities in the EU have a significant say in when and how DPIAs are conducted. Two influential methodological frameworks for completing DPIAs have been published by the French and UK authorities and adapted in other jurisdictions. Friedewald and colleagues characterise the French approach as a software-supported checklist review, with the UK methodology requiring a more discursive, text-based practice seen in **Privacy Impact Assessment**s in the English-speaking world since the 1990s.

Supervisory Authorities also have a say in scope, as well as methodology, having the power to pass lists of processing which do and do not require a DPIA in their jurisdiction. The EDPB’s published opinions on these national DPIA exemptions refer mostly to risk of harm to data subjects, suggesting it is the narrow and more immediate scope of risks which is commonly contemplated within the EU. As such, the DPIA is an example of an individual-centred data protection provision not necessarily geared towards the **privacy risk**s posed by **big data** processing to groups across society.

Further reading:

Article 29 Working Party, 2017. *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is 'likely to result in a high risk' for the purposes of Regulation 2016/679.* Brussels: European Commission, https://ec.europa.eu/newsroom/article29/items/611236/en.

European Data Protection Board, 2019. *Opinion 13/2019 on the draft list of the competent supervisory authority of France regarding the processing operations exempt from the requirement of a data protection impact assessment (Article 35(5) GDPR)*, https://edpb.europa.eu/sites/default/files/files/file1/edpb_opinion_201913_fr_35.5_dpia_list_en.pdf.

Friedewald, M., Schiering, I., Martin, N. and Hallinan, D., 2022. Data Protection Impact Assessments in practice. *Computer Security*, 13106, http://dx.doi.org/10.1007/978-3-030-95484-0_25.

See also: ACCOUNTABILITY, DATA ETHICS, DATA PROTECTION, DATA PROTECTION AUTHORITY, GROUP HARMS, GROUP PRIVACY, RIGHT TO DATA PROTECTION, RISK ASSESSMENT

Data Protection Officer (DPO)

Article 37 of the EU **GDPR** requires **data controller**s to appoint a Data Protection Officer in certain privacy-relevant situations; for example, where **personal data** is processed by a public authority, or by any type of organisation using large volumes of **special category** (e.g., health-related) data, which poses greater risks to the rights and freedoms of **data subject**s.

The DPO should be sufficiently senior and independent to ensure the integrity of personal **data processing**. They may be employed by the data controller, or an independent professional providing external services. There is no qualification requirement for the role, and the DPO does not need to be a practising lawyer.

The DPO takes responsibility for **compliance** with the GDPR and other relevant legislation (e.g., national **data protection** legislation), and for ensuring that the controller can demonstrate compliance, in line with the **accountability** principle. They are, to some extent, the public face of the organisation's data protection practices, and where they are appointed their contact details must be made available to data subjects under the **transparency** principle.

Further reading:

Nissim, J., 2018. Accountability and the role of the Data Protection Officer. *In*: Carey, P., ed. *Data protection*: *a practical guide to UK and EU Law*, 5th edition. Oxford: Oxford University Press, 223–39, https://dl.acm.org/doi/abs/10.5555/3265270.

Data Protection Policy

A policy may refer either to a specific written document, or more generally to the understood stance taken on **information governance** within an organisation. Formal, written data protection policies are generally created at the discretion of an organisation, although some national legislation (e.g., the UK Data Protection Act 2018) does require policies relating to **data protection**.

Although not explicitly mandated at the EU level, data protection policies are often a tacit requirement. The **GDPR** requires **data controller**s to demonstrate **compliance** with its provisions and implement **appropriate technical and organisational measures** to safeguard **personal data**. The introduction of a written data protection policy, potentially by a **Data Protection Officer**, is one way to demonstrate data-protection-by-design-and-default. Larger organisations may have multiple sub-policies – for example, for marketing, **data breach notification**, **employee information** and so on.

In practice, there is often little meaningful distinction between ‘data protection policy’ and ‘**privacy policy**’ as the two terms are used interchangeably. Conceptually, however, data protection and **privacy** are distinct in law, with data protection comprising legislative codes governing operational concerns around the use of **personal data**. Privacy, on the other hand, stems from multiple legal sources, relates only to *private* personal **information** and is more amorphous, being determined according to broader legal principles without clear operational requirements.

Further reading:

Nissim, J., 2018. Creating a data protection compliance programme. *In*: Carey, P., ed. *Data protection*: *a practical guide to UK and EU Law*, 5th edition. Oxford: Oxford University Press, 240–9, https://dl.acm.org/doi/abs/10.5555/3265270.

See also: ACCOUNTABILITY, DATA GOVERNANCE, DATA PROTECTION, DATA-PROTECTION-BY-DEFAULT, IDENTIFIABLE DATA, INFORMATION GOVERNANCE

Data Protection Principles

Data protection principles are the principles embedded in **data protection** legislation, which form a way to organise their more specific requirements. For example, the EU **GDPR** has an overarching principle of **transparency**, within which are located more specific obligations to share **information** with **data subject**s. The other principles include **lawfulness**, **fairness**, **purpose limitation**, **storage limitation**, data **accuracy**, **data minimisation**, **confidentiality** and **accountability**.

Further reading:

Carey, P., 2018. Data protection principles. *In*: Carey, P., ed. *Data protection*: *a practical guide to UK and EU Law*, 5th edition. Oxford: Oxford University Press, 32–41, https://dl.acm.org/doi/abs/10.5555/3265270.

Kirby, M., 2010. The history, achievement and future of the 1980 OECD guidelines on privacy. *International Data Privacy Law*, 1(1), 6–14, https://doi.org/10.1093/idpl/ipq002.

See also: DATA MINIMISATION PRINCIPLE, HISTORY OF PRIVACY

Data Provenance

A record of where **data** have come from, what processes they have been through and who has interacted with them from collection up to their current state. As **data flow**s become more complex, recording provenance has become more important. A good provenance record promotes **trust** in the data and enables reusability and reproducibility.

Provenance is also important in **data governance**; for instance, without good provenance records, one may be unaware of the relationship between **data subject**s and the data one holds, and of any responsibilities that one may have to other **data controller**s upstream.

Data provenance **standard**s have been developed, most notably the W3C standard PROV which enables the formal representation of provenance and its visualisation as a graph. The **Anonymisation Decision Making Framework** highlights the importance of provenance in the **anonymisation** process. Recent work has explored the relationship in more detail.

Further reading:

Jarwar, M.A., Chapman, A., Elliot, M. and Raji, F., 2021. Provenance, anonymisation and data environments: a unifying construction. *arXiv*, https://doi.org/10.48550/arXiv.2107.09966.

Moreau, L. and Groth, P., 2013. *Provenance*: *an introduction to PROV*, Cham: Springer.
Simmhan, Y.L., Plale, B. and Gannon, D., 2005. A survey of data provenance techniques. Computer Science Dept., Indiana University, Technical Report IUB-CS-TR618, https://legacy.cs.indiana.edu/ftp/techreports/TR618.pdf.

Data Quality

There is no universally agreed set of dimensions for measuring data quality and different uses will focus on different dimensions. Some of the considerations are **availability**, correctness, completeness, consistency, flexibility, relevance, timeliness and validity. However, many of the concerns boil down to two questions: is it useful for my purposes, and does it represent the thing it's supposed to represent?

Data quality is a challenge for analysts and those responsible for **data governance** alike. Most **dataset**s are imperfect representations of the **population** or phenomena that they are supposed to represent, with issues of both **analytical completeness** and **analytical validity**. Errors in both **data capture** and downstream processing, problems with data specification and **data ageing** are just some of the issues.

The application of **statistical disclosure control**s or privacy models (such as **differential privacy**) to **data** invariably leads to reductions in quality, characterised by Purdam and Elliot as the loss of analytical validity and/or completeness. This in turn leads to considerations about the **risk/utility trade off**.

The concept has gained prominence in the context of **artificial intelligence** and other complex **algorithmic** processing, with the EU AI Act being one of the first pieces of major legislation to refer to 'high data quality' (Recital 44).The AI Act regulates the quality of training, validation and testing data for high-risk AI systems (e.g., systems used within medical devices, or otherwise posing a risk to the health and safety of a **natural person**). The proposed American Data Privacy and Protection Act contains similar provisions for algorithm design evaluations, including scrutiny of training data. Within **data protection** law, the EU's **GDPR** requires that personal data be accurate and adequate for their purpose, and so sets broad expectations of data quality.

Further reading:

Cichy, C. and Rass, S., 2019. An overview of data quality frameworks. *IEEE Access*, 7, 24634–48, https://doi.org/10.1109/ACCESS.2019.2899751.
Purdam, K. and Elliot, M., 2007. A case study of the impact of statistical disclosure control on data quality in the individual UK samples of anonymised records. *Environment and Planning A*, 39(5), 1101–18, https://doi.org/10.1068/a38335.

See also: DATA ENVIRONMENT, DATA PROCESSING, DATA UTILITY, SCRUTINY, US PRIVACY LAWS

Data Recipient

The EU’s **GDPR** defines a recipient as someone to whom **personal data** are **disclosed**. The Regulation does not specify whether the legal or **natural person** in question is the **data controller**, **data processor** or **data subject** of the data in question.

Data recipient is thus a broad term, which could describe a party to a **Data Sharing Agreement** or a Data Processing Agreement.

See also: DATA IN TRANSIT, DATA SHARING

Data Release

The act of making **data** more widely accessible, whether to a particular **user** group or to the general **public**, is referred to as data release. Data that was previously unavailable or only accessible to a select group of **authorised** people will be **published** as part of this process. Data releases can take place to support scholarly research, encourage **accountability** and **transparency**, make new services and **application**s possible, or because of statutory requirements. To support research and guide policy decisions, a government agency might, for instance, publish data on crime, transportation or health.

For users to effectively understand and use the data, data releases typically involve careful preparation, guaranteeing **data quality** and completeness, and the provision of appropriate documentation and **metadata**. This may entail choosing the best file formats, **de-identifying personal information** and setting up appropriate **access control**s and user agreements.

Data releases may also give rise to moral and legal questions about **confidentiality**, **privacy** and **intellectual property** rights. The privacy and confidentiality of the people or organisations represented in the data should be carefully considered, and appropriate measures should be taken to ensure that intellectual property rights are guaranteed.

See also: DATA IN TRANSIT, DATA USER, OPEN DATA, PERSONAL DATA, RIGHT OF ACCESS

Data Retention

Data retention is the preservation of **data** by an organisation after its initial use. This may be for business reasons, such as for **auditing**, operational reuse, business process management, and the like. Or it may be required by regulation.

Data retention is a complex matter, demanding the balancing of several requirements. First, it needs to comply with **privacy** and **confidentiality** regulation – hence, if **personal data**, with the **GDPR** or other relevant **data protection** regimes. This requires the provision of **security**, ensuring **consent** or other **lawful base**s for retention and **compliant data processing** when necessary. The **value of data** to the organisation may be offset against the costs of compliance.

But second, data retention may be mandated by regulation, so organisations may be legally obliged to store it. Examples of such mandates are the retention of telephone and **Internet traffic data** by telecommunications firms, to be made available to government under specified conditions for traffic analysis and **surveillance**, and the retention of banking data for a period of time to facilitate the investigation of money laundering. Only certain bodies, specified in law (typically law enforcement, intelligence or taxation agencies), will be allowed access to the retained data.

Third, there are practical matters to be managed, including not only the cost of secure retention but also issues such as ensuring formats remain up to date, policing **access control**s and managing **encryption**.

The arguments for data retention for commercial or operational reasons boil down to the cost/benefit analysis of secure, compliant **data storage** versus the operational gain. The arguments for compulsory retention are usually pitched around the social goods of security, law enforcement or effective tax collection, balanced against the rights to privacy for individuals, as well as the question as to how powerful the state becomes when potentially armed with the retained data. The costs of data retention are usually borne by businesses, while the social benefits are distributed across society, creating an issue of equity. Given that the argument for compulsory data retention has been won, there is still the political question to be resolved of how long the retention period should be.

The legal situation surrounding data retention has been controversial. The EU’s Data Retention Directive (2006) covered fixed and mobile telephones and the Internet, including email and **Voice over Internet Protocol communication**s, and required telecoms providers to retain data sufficient to allow communications to be traced in terms of senders, receivers, times, durations and devices used. Despite claims of value for security and criminal justice purposes, the Directive was criticised for breaching fundamental privacy rights, and was ultimately declared invalid by the European Court of Justice in 2014. The requirement to keep personal data no longer than necessary has been retained in the GDPR’s **storage limitation** principle.

Further reading:
Thierse, S. and Badanjak, S., eds, 2020. *Opposition in the EU multi-level polity*: *legal mobilization against the Data Retention Directive*. Cham: Palgrave Pivot, https://doi.org/10.1007/978-3-030-47162-0.

See also: DATA LIFECYCLE MANAGEMENT, DATA SITUATION, INFORMATION GOVERNANCE

Data Safe Haven

While there is no definitive characterisation, a data safe haven is a repository of **data** that ensures access for **certified** researchers, with a high level of **security**. Typically, access to the data is very strictly controlled, in terms of the people with access permissions, the **auxiliary data** with which it can be brought into juxtaposition, the technologies used upon it, the queries made of it, and the **publication** or further use of the query responses. It may be designed with security **standard**s such as **ISO27001** in mind or may go even further. Usually, data will be processed within the **bound**s of the safe haven. Data safe havens are most used for medical data, or highly **sensitive personal data**, where **trust** in the **data governance** regime is essential for the functioning of systems.

Further reading:
Lea, N.C., Nicholls, J., Dobbs, C., Sethi, N., Cunningham, J., Ainsworth, J., Heaven, M., Peacock, T., Peacock, A., Jones, K., Laurie, G. and Kalra, D., 2016. Data safe havens and trust: toward a common understanding of trusted research platforms for governing secure and ethical health research. *JMIR Medical Informatics*, 4(2), e22, https://doi.org/10.2196/medinform.5571.

See also: ACCESS CONTROL, DATA ENCLAVE, DATA IN USE, DATA TRUST, FIVE SAFES, SAFE PEOPLE, SAFE SETTING

Data Sanitisation

A high grade of **data destruction** where the complete **erasure** of the **data** in question is verified and usually evidenced using a tamper-proof certificate.

See also: DELETION

Data Schema

A data schema is a design that specifies how a **database** or **dataset** will be organised. It provides a framework for **data storage**, retrieval and manipulation of data elements, and their relations to one another. Definitions of tables, columns, keys and relationships between tables are frequently found in data schemas. Each field or **attribute**'s data types and formats are specified, and it may also include restrictions on data entry such as minimum and maximum values, or guidelines for data validation. There are many ways to represent data schema, including entity-relationship diagrams, UML diagrams and XML schemas, and they are used in database design, **software** development and other data management applications.

Data integration and **interoperability** between systems or **application**s can also be facilitated by the use of data schema. For improved data accessibility and usability, **data quality**, **accuracy** and consistency, and lowering the possibility of errors, data duplication and inconsistent **data processing** and analysis, a well-designed data schema plays a crucial role.

Further reading:
Scannapieco, M., Figotin, I., Bertino, E. and Elmagarmid, A.K., 2007. Privacy preserving schema and data matching. *In: Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, New York: ACM, 653–64, https://doi.org/10.1145/1247480.1247553.

Dataset

A dataset is a collection of **data** which is treated as a single unit. This often means it has a single manager (the **data steward** or **data custodian**, or in the case of a dataset of **personal data**, the **data controller**), and the pieces of data within it relate to each other. For example, they commonly are about the same things (the **population**) and have the same structure. However, datasets may be merged to create a larger aggregate dataset, usually about the same types of things; but the merged dataset may lack a common structure, **data schema** or ontology of terms.

High-level descriptions of such properties and **attribute**s of the dataset (called **metadata)** may be provided to enable its use by a **third party**. Datasets typically contain relatively fine-grained data (**microdata**), so that data units are distinguishable, rather than simple summary statistics. If the dataset is of this type, and the data is personal data, then the dataset could be used to compromise privacy. The use and analysis of very large datasets have given rise to the term **big data**.

Some datasets are specifically created for training **machine learning** algorithms. A *training dataset* is used for supervised learning, where the **algorithm** is given feedback on its classifications. A *validation dataset* is used to tune the algorithm's predictions. A *test dataset* is used to evaluate the algorithm's final model.

Further reading:
Kelleher, J.D. and Tierney, B., 2018. *Data science*. Cambridge, MA: MIT Press.

See also: DATABASE, DATA CURATION, DATA HARMONISATION

Data Sharing

Data sharing takes place when those in control of **data** allow outside agents to process it, usually to increase the value extracted from it. This can impact **privacy** when **personal data** is involved. Sharing personal data can only be done in **compliance** with regulation, and may involve the use of **anonymisation** techniques, including **access control**s and **de-identification** of the data.

The **risk**s of data sharing are exacerbated by the fact that the **data controller** will not have direct managerial influence over the outside **data recipient**s, and instead must set terms and conditions (a **data sharing agreement**) sufficient to ensure compliant behaviour. The data share may serve the interests of the data controller, or of a consortium of which the data controller is a member, or only of the recipients of the data, in which case they will typically pay for access. Data sharing may also be a requirement for holders of academic research data, and a necessity where the **data processing** requirements outstrip the capabilities of an individual organisation.

Further reading:
Joly, Y., Dyke, S.O.M., Knoppers, B.M. and Pastinen, T., 2016. Are data sharing and privacy protection mutually exclusive? *Cell*, 167(5), 1150–4, https://doi.org/10.1016/j.cell.2016.11.004.

See also: CONDITIONS FOR PROCESSING, DATA IN TRANSIT, DATA IN USE, DATA PROTECTION, DATA QUALITY, VALUE OF DATA

Data Sharing Agreement (DSA)

Data sharing agreements do not have a specific legal definition, but the term is typically used to refer to a legally binding, written contract governing the transfer (or general sharing) of **personal data** between two or more bodies. DSAs should be understood as distinct from Data Processing Agreements, which usually refer to the written instructions a **data controller** must provide to a **data processor** under EU **data protection** law. A data processor will receive personal data, but only to be used for the purposes of providing a service to the data controller. Under a DSA, the parties can each share and use personal data for their own purposes.

DSAs may or may not also be Data Transfer Agreements. While the latter are also not legally defined, they commonly refer to contracts governing the transfer of personal data outside its **jurisdiction** of origin – outside the European Economic Area, for example.

The EU's **GDPR** does not use the terms 'Data Sharing Agreement,' 'Data Processing Agreement' or 'Data Transfer Agreement'; 'agreements' referred to in the Regulation are generally instruments of public international law. The Regulation does, however, require **joint data controller**s to determine their respective responsibilities for data protection **compliance** in a **transparent** manner. In practice, given the **risk**s and potential liability involved for the parties, a contract under the private law of the relevant jurisdiction – such as a Data Sharing Agreement – is often the most practical arrangement.

See also: DATA IN TRANSIT, DATA IN USE, DATA RECIPIENT, DATA SHARING, DATA TRANSFER, INTERNATIONAL TRANSFER

Data Situation

The aggregate set of interactions between the **data** and the environment(s) in which they appear; the fundamental concept underpinning the **Anonymisation Decision-Making Framework** and **functional anonymisation**. The basic principle of functional anonymisation is that **risk** lies in the data situation rather than in the data themselves.

Further reading:
Elliot, M., Mackey, E. and O’Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners’ guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

See also: ANONYMISATION, DATA ENVIRONMENT, DATA SITUATION AUDIT

Data Situation Audit

The first stage of the **Anonymisation Decision-Making Framework**, the audit consists of six steps:

1. Capture the presenting problem
2. Sketch the **data flow**s and determine data holder’s responsibilities
3. Map the properties of the **data environment**(s)
4. Describe and map the **data**
5. Engage with **stakeholder**s
6. Evaluate the **data situation**

The intended outcome of the audit at step six is an evaluation of whether a full **risk assessment** is required.

Further reading:
Elliot, M., Mackey, E. and O’Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners’ guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

See also: DISCLOSURE RISK ASSESSMENT, ANONYMISATION, RISK, DISCLOSURE RISK

Data Sovereignty

Data sovereignty is the idea that **data** about individuals of a particular nationality should be held on servers in that nation’s **jurisdiction**.

This has been represented as being a **privacy** protection by governments of their citizens, on the ground that they are concerned with protecting their citizens’ privacy to a greater extent than foreign companies and governments. This may be the case, but governments may also be interested in having data about their citizens close at hand, and available if the law allows access. Some supporters of data sovereignty are authoritarian states.

There are other arguments for data sovereignty. It may be an aspect of mercantilist industrial policy, promoting the development of **data centre**s and complex data infrastructure within the country's jurisdiction. Furthermore, many countries now classify data about their citizens as a national **asset**, and sovereignty laws are drawn up to protect and foster it. Conversely, such data may be valuable to enemy states, and sovereignty may prevent its misuse, as part of a **national security** strategy. Finally, indigenous peoples have claimed that use of their data is another aspect of their historical exploitation.

Data sovereignty has also been argued to be in line with the **GDPR**'s restrictions on **international transfer**s of **personal data** to third countries. However, this argument is spurious. While GDPR does allow **data storage** relating to EU citizens anywhere within the EU, this is based not on jurisdiction but on **data protection** standards. If a third country can demonstrate equivalent **standard**s, then data can be exported there.

Further reading:
Hummel, P., Braun, M., Tretter, M. and Dabrock, P., 2021. Data sovereignty: a review. *Big Data and Society*, 8(1), 1–17, https://doi.org/10.1177/2053951720982012.

See also: ADEQUACY, SAFE HARBOR

Data Steward

Data curation is the process of managing high-quality, usable datasets. Data steward is the role tasked with curation of data, of preserving its utility and protecting any rights to privacy therein. *Data stewardship*, and the role of data steward, have not been formally defined, but can imply the idea that the steward has taken on informal duties to look after the data in trust, to ensure **fair information processing** and that all legitimate **stakeholder**s can benefit from the data's use. The steward's role is therefore wide and may include ensuring the data is available for academic research in the public good. There can therefore be an ethical connotation to data stewardship, absent from the role of a **data controller**, which may include exploitation of the data if that is **compliant** with **data protection** law. While 'stewardship' as respectful preservation is a common use of the term, some authors understand data stewardship more pragmatically, for example, as the day-to-day operation of **data governance**.

Further reading:
Plotkin, D., 2020. *Data stewardship*. San Diego: Elsevier Science & Technology.
Rosenbaum, S., 2010. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Services Research*, 45(5p2), 1442–55, https://doi.org/10.1111/j.1475-6773.2010.01140.x.

See also: COMMON LAW, DATA ETHICS, DATA INTERMEDIARY, DATA TRUST, DATA UTILITY, FAIRNESS, FIDUCIARY DUTY, INFORMATION ETHICS, INFORMATION GOVERNANCE

Data Stewardship Organisation

A term coined by Duncan et al to capture the multi-faceted nature of responsibilities for **data** and particularly those of **data curation** and/or providing access to useful data. The paradigm examples are the national statistical agencies who have responsibilities to **publish** statistics on their country's population (e.g., through national **census**es) while also protecting **confidentiality** of individual citizens.

Further reading:
Duncan, G.T., Elliot, M. and Salazar-González, J.J., 2011. *Statistical confidentiality*. New York: Springer, https://doi.org/10.1007/978-1-4419-7802-8.
Rosenbaum, S., 2010. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Services Research*, 45(5p2), 1442–55, https://doi.org/10.1111/j.1475-6773.2010.01140.x.

See also: DATA CONTROLLER, DATA GOVERNANCE, DATA STEWARD

Data Storage

Data is an ordered collection of symbols and can be created by various practices and mechanisms that generate symbols as means of representing events, propositions, measurements, observations or other reflections of an environment or model. For the data to remain useful in future, it must be stored after collection, in such a way as to be accessible to systems that can process the symbols. Such storage might use media such as clay, paper, punched cards, photographic film, vacuum tubes, magnetic tape or solid-state devices (semiconductors) and will also require a standardised *format* for its representation, to allow for straightforward interpretation after retrieval. Analogue storage involves continuous variation (as with a vinyl record), while digital storage only allows combinations of discrete values. Analogue storage is more prone to include noise but is less likely to suffer a catastrophic loss of meaning when the signal degrades.

The 21st century has seen a dramatic increase in the amount of data created, stored and (at least in principle) open for retrieval, which has had two effects on **privacy**. First, there is a greater quantity of data about individuals stored, and so a much larger chance (a) that some of this data is pertinent to an individual, and (b) that **dataset**s may be combined to be even more **disclosive**. This is the world of **big data**, very large datasets that can be analysed using **machine learning** techniques to produce greater statistical power (note that the requirement to bring datasets together highlights the need for common data formats or **data schemata**, to reduce the amount of processing necessary). Particular data storage formats have been developed which facilitate retrieval and large-scale analysis, such as Hadoop, which includes software for *distributed storage* (i.e., data stored on multiple servers), allowing *parallel processing*.

The second privacy-related issue is that of the **security** of the storage system. If valuable data is stored, then it may be the target of **hacker**s, and under various regulations and **code**s **of conduct**. Those with responsibilities for storing **personal data** must ensure that it is securely held. Distributed storage and local processing of data are one way to avoid storage with a single point of failure.

Further reading:

Computer History Museum, 2022. *The storage engine*: *a timeline of milestones in storage technology*, www.computerhistory.org/storageengine/timeline/.

Reis, J. and Housley, M., 2022. *Fundamentals of data engineering: plan and build robust data systems*. Sebastopol, CA: O'Reilly Media.

See also: CYBERSECURITY, DATA AT REST, DATA ENVIRONMENT, DATA LIFECYCLE, DATA PROTECTION, DISCRETE DATA

Data Subject

Within its definition of **personal data**, the EU's **GDPR** sub-defines a data subject as an identified or **identifiable natural person**. These people are thus living individuals whose unique **identity** can be determined, either directly from the personal data in question, or from the personal data combined with other **information**. They are the people who can exercise **data subject** rights under the GDPR, and their rights and freedoms are the main objects of its protection.

See also: DATA CONTROLLER, DATA PROTECTION, IDENTIFIABLE DATA, IDENTIFIED DATA, NATURAL PERSON, PERSONAL INFORMATION

Data Subject Access Request

Under the EU **GDPR**, and similar legislation such as that within the **UK GDPR**, individuals have a right to access a copy of the **information** which a **data controller** processes that identifies and relates to them (i.e., a copy of their **personal data**). They also have a right to some information about how their personal data is used by the data controller.

The underlying rationale (per Recital 63, GDPR) is to enable individuals to check the **accuracy** of **information** held about them, and the **lawfulness** of the way it is used. As with all **data subject** rights, some exemptions apply, such as for **national security**, or where **compliance** would infringe the rights and freedoms of others. For example, **redaction**s of information may be permitted to protect the **privacy** of other data subjects.

Further reading:
Lloyd-Jones, H. and Carey, P., 2018. The rights of individuals. *In*: Carey, P., ed. *Data protection*: *a practical guide to UK and EU Law*, 5th edition. Oxford: Oxford University Press, 122–54, https://dl.acm.org/doi/abs/10.5555/3265270.

See also: DATA PROTECTION, PERSONAL INFORMATION, PRIVACY AS CONTROL, RECTIFICATION, RIGHT OF ACCESS

Data Synthesis

The creation of artificial **dataset**s usually using a model of a real dataset (referred to as the *original data*). The initial research was done by Rubin using an extension of **multiple imputation** – effectively treating all **data** as **missing data**. CART (classification and regression tree) models have subsequently been popular. Most recently, researchers have started to explore **machine learning** techniques, particularly generative adversarial models (GANs).

In principle, synthetic data has much lower **disclosure risk** than the original data because the direct link between the **data subject** and the data has been broken but the actual **risk** depends on the data, the model used, and the context. Specifically, although some have argued that **reidentification** is meaningless in synthetic data, **attribution** is still a possibility as is **membership inference**.

Further reading:

Rubin, D.B., 1993. Statistical disclosure limitation. *Journal of Official Statistics*, 9(2), 461–8.

Drechsler, J., 2011. *Synthetic datasets for statistical disclosure control: theory and implementation*. New York: Springer.

Little, C., Elliot, M. and Allmendinger, R., 2022. Comparing the utility and disclosure risk of synthetic data with samples of microdata. *In: International Conference on Privacy in Statistical Databases*, Cham: Springer, 234–49, https://doi.org/10.1007/978-3-031-13945-1_17.

See also: DATA UTILITY, DISCLOSURE RISK, INFERENCE, PUBLISHING

Data Transfer

The simple meaning of ‘data transfer’ would be the act of making **data** available to another party, usually involving the data physically moving (or being copied to) a new location. However, the term is often used in a more specific sense to refer to the transfer of **personal data** to another party, be they another **data controller**, a **data processor** or a **third party**. The **data recipient** would thus acquire legal responsibility for the data in accordance with **data protection** law.

Even more specifically, data transfers are often discussed in the context of transfers out of a legal **jurisdiction**, such as outside the European Economic Area. These are commonly referred to as **international transfer**s or international data transfers. The **GDPR** has specific requirements for the lawful transfer of personal data outside the EEA, referring to these as transfers of personal data to third countries or international organisations. If the European Commission has not formally approved the recipient’s country as having adequate data protection laws, an EU data controller must select an additional means of protecting personal data, such as contractual safeguards.

Further reading:

Ustaran, E., 2018. International data transfers *In*: Carey, P., ed. *Data protection*: *a practical guide to UK and EU Law*, 5th edition. Oxford: Oxford University Press, 105–21, https://dl.acm.org/doi/abs/10.5555/3265270.

See also: CROSS-BORDER DATA PROCESSING, SAFE HARBOR, SCHREMS

Data Trust

A data trust is a **data stewardship organisation** which exploits the notion of **fiduciary duti**es to manage **data**, thereby, it is hoped, removing conflicts of interest. A trustee administers a trust not in their own interests, but to benefit the named *beneficiaries* of the trust.

There are two major senses of ‘data trust’ in circulation, although neither is at the time of writing being implemented very widely. The first has a literal interpretation in terms of *trust law*, where individuals would join and be the beneficiaries (hence these have been called ‘bottom up’ data trusts). The beneficiaries would deposit **personal data** about themselves, including data they are able to access via **data protection** regulations respecting **data portability**. Trustees would administer the data to maximise beneficiaries’ benefits. If a data trust could develop sufficient scale, it would have economies of scale and bargaining power to extract favourable terms from **data user**s. However, trust law is complex, and derives from equity in **common law jurisdiction**s, fitting less well with civil law systems. Given the responsibilities of the trustees, it is also not clear what their business model might be.

The second sense of ‘data trust’ is a means to increase **data sharing** to benefit the **artificial intelligence** industry. Hall and Pesenti argued in a UK government report that the lack of data sharing (whether personal or non-personal) across European companies was leading to a failure to exploit **big data** at scale. They proposed structures called data trusts to manage the perceived **risk**s to **data owner**s and **data controller**s of data sharing, although did not spell out in any detail what these might be.

Further reading:

Delacroix, S. and Lawrence, N.D., 2019. Bottom-up data trusts: disturbing the ‘one size fits all’ approach to data governance. *International Data Privacy Law*, 9(4), 236–52, https://doi.org/10.1093/idpl/ipz014.

Hall, W. and Pesenti, J., 2017. *Growing the artificial intelligence industry in the UK* [online]. London: Department for Digital, Culture, Media and Sport/ Department for Business, Energy and Industrial Strategy, www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.

O’Hara, K., 2020. Data trusts. *European Data Protection Law Review*, 6(4), 484–91, https://doi.org/10.21552/edpl/2020/4/4.

See also: DATA GOVERNANCE

Data Unit

Information within a **dataset** relating to a single **population unit**. Most often these are **record**s or cases within the dataset.

See also: MICRODATA

Data User

A data user or *data consumer* is a person or organisation that processes **data** as part of its business model or purpose. In the *knowledge* or **digital economy**, data users seek data to add value, either for their own purposes, or to monetise (e.g., to sell enhanced versions to other organisations, or to publish reports, summaries or aggregated statistics). *Academic* data users use *data science* to produce original and publishable empirical conclusions. *Data journalists* seek data to help create or support newsworthy narratives. Although the term technically covers people or organisations that generate their own data (e.g., from transactions they are involved in), it is used more frequently in relation to data marketplaces and/or **data sharing**, where data users source data from external organisations.

Further reading:
Kelleher, J.D. and Tierney, B., 2018. *Data science*. Cambridge, MA: MIT Press.
Schweidel, D.A., 2015. *Profiting from the data economy: understanding the roles of consumers, innovators, and regulators in a data-driven world*. Upper Saddle River, NJ: Pearson Education.

See also: DATA INTERMEDIARY, DATA IN USE, VALUE OF DATA

Data Utility

The value of a given **dataset** as an analytical resource. The key issue is whether, and how well, the **data** represent whatever it is they are supposed to represent. **Disclosure control method**s can have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximise data utility while minimising **disclosure risk**, and in practice disclosure control decisions trade off these two parameters. This then intersects with two components of **trust** – can data users trust the data, and can **data subject**s trust the **data steward**s or **data controller**s to protect their **confidentiality**?

Utility metrics come in different types: broad and narrow. Broad measures attempt to capture the difference in utility between an original dataset and the one that has been treated, in terms of **information loss**. Narrow measures attempt to capture the capacity of datasets to produce a specific piece of analysis. Some authors, such as Taub et al., advocate using a basket of narrow measures.

Purdam and Elliot distinguish between **analytical completeness** and **analytical validity** as capturing distinct impacts of utility.

Further reading

Li, T. and Li, N., 2009. On the tradeoff between privacy and utility in data publishing. *In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, 517–26, https://doi.org/10.1145/1557019.1557079.

Taub, J., Elliot, M. and Sakshaug, J.W., 2020. The impact of synthetic data generation on data utility with application to the 1991 UK samples of anonymised records. *Transactions on Data Privacy*, 13(1), 1–23, www.tdp.cat/issues16/tdp.a306a18.pdf.

Woo, M.J., Reiter, J.P., Oganian, A. and Karr, A.F., 2009. Global measures of data utility for microdata masked for disclosure limitation. *Journal of Privacy and Confidentiality*, 1(1), https://doi.org/10.29012/jpc.v1i1.568.

See also: DATA QUALITY, DISCLOSURE, RISK–UTILITY TRADE-OFF, UTILITY FIRST, VALUE OF DATA

Dataveillance

The monitoring of individuals or groups through their **digital footprint**. Data utilised for dataveillance can be any digital traces, including **social media** accounts, credit card transactions, **browsing history**, email, GPS coordinates from mobile phones, and so on. The purpose of dataveillance is to create a representation of a person or a group and their activity (both online and offline). Dataveillance has uses in crime detection and prevention, and in counterterrorism. On the other hand, it raises numerous **privacy concern**s. The representations created from the data gathered are essentially used for **profiling**. Individuals that are surveilled will certainly not have **consented** to this activity and dataveillance seems to run contrary to **GDPR**'s **right to be forgotten**.

Note that dataveillance is distinct from electronic **surveillance**, which generally is used to refer to the direct real-time monitoring of a person's oral and video **communication**s.

Further reading:
Clarke, Roger A., 1988. Information technology and dataveillance. *Communications of the ACM*, 31(5): 498–511, https://doi.org/10.1145/42411.42413.

See also: DATABASE OF RUIN, DATA EXHAUST, DIGITAL BREADCRUMBS, GROUP PRIVACY, SOCIAL PROFILING

Data Warehouse

A data warehouse is a sizable, centralised repository of **data** created to support data analysis. It is a system that collects and organises historical and current data from various sources, including **database**s, **customer relationship management** systems, and other data sources, in a way that facilitates analysis and the extraction of insights.

Data warehouses are made to meet the needs of decision makers who need access to their data in a simple to use format, such as business analysts, executives, and other **stakeholder**s within the enterprise.

Implementing and maintaining data warehouses can be difficult and expensive, and they call for specialised knowledge and skills in **database** technologies, data management, and data analysis. However, they are essential to extracting value from all the data, including historical data, within a complex enterprise, and realising the potential of **big data**.

Data warehouses are distinct from *data lakes*. A data warehouse typically includes a set of tools and techniques for extracting, transforming, and loading data from source systems. Additionally, features such as **data mining**, data modelling and data visualisation are frequently built into their operation. Data lakes on the other hand are simply repositories of raw data in whatever format is native, without any of the functionality of a data warehouse.

Further reading:
Inmon, W.H., 1995. What is a data warehouse? *Prism Tech Topic*, 1(1), 1–5, www2.cs.sfu.ca/CourseCentral/741/jpei/slides/Data%20warehousing%202.pdf.

See also: DATA AT REST, DATA GOVERNANCE, DATA IN USE, DATA UTILITY, VALUE OF DATA

DDOS

See: DENIAL OF SERVICE

Deanonymisation

See: REIDENTIFICATION

Decentralisation of the Web

The **World Wide Web** was originally designed as a permissionless, decentralised **information** space unified by the **uniform resource identifier** naming scheme. As it grew popular, however, some websites and platforms became very large, thanks to network effects by which the size of the **network**s they fostered (sometimes in the billions of people) added immensely to their value for their **user**s. This enabled them to create *walled gardens* – centralised areas such as social networking sites, which provided many valuable services but which were not easily linked to other parts of the Web, and from which users' **data** could be harvested to create even more services, at the cost of their **privacy**. Filter bubbles, **personalisation** and **targeted advertising** threatened to undermine user **autonomy** and **decisional privacy**. Some platforms became default **identity provider**s. **Personal data**, being held in centralised stores owned by the platforms, also created a **security** problem, with the platform being a single point of failure in the event of a hack.

This led to pressure to decentralise, or *re-decentralise*, the Web, to restore its original libertarian vision. The basic idea behind this is to separate **data storage** from services, suggesting either the possibility of the use of **personal data store**s, so that individuals have control over their data, or the use of distributed ledger technology, where data is stored on a decentralised **blockchain**. The **Social Linked Data (SOLID)** project is an example of the former approach. The latter approach has been termed **Web 3.0**. In either case, service providers would need to ask for access to data to deliver their services. Users' incentives to provide access would depend on the services they required, but they could also protect their privacy if they preferred.

Further reading:
Verborgh, R., 2019. Re-decentralizing the Web, for good this time, https://ruben.verborgh.org/articles/redecentralizing-the-web/.

See also: NETWORK, SOCIAL NETWORK, SOCIAL PROFILING

Decisional Privacy

Decisional privacy refers to the **integrity** of an individual’s decisions, actions, plans and choices with respect to their **private** affairs. This quality of ‘cognitive liberty’ is distinct from the external influence exercised by (for example) a manager, whose bounds of action are circumscribed by terms of employment. In a **breach** of decisional privacy, an intended action of an individual is prevented by an **other**, or the individual is coerced into one course of action or an artificially narrow choice. Influence does not have to involve force: it could include ridicule or hostility, and Rössler has argued that even praise may be an undue kind of influence. Attempts to influence behaviour, especially covert ones such as nudging, are often seen as interfering with decisional privacy, possibly in a paternalistic way justified by some greater good for the manipulated individual. Some, such as Thomson and Gavison, have argued that decisional privacy is really a form of freedom or **autonomy**, not privacy.

Decisional privacy is not a legal term, but it is a concept which some scholars have read into the privacy jurisprudence of the American and European courts. Van der Sloot has argued that the emergence of decisional privacy has implications not only for conventional ‘private’ decisions (e.g., relating to healthcare), but also for **profiling** and automated nudging.

Further reading:

Gavison, R., 1980. Privacy and the limits of law. *Yale Law Journal*, 89(3), 421–71, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2060957.

O’Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Rössler, B., 2005. *The value of privacy*. Cambridge: Polity Press.

Thomson, J.J., 1975. The right to privacy. *Philosophy and Public Affairs*, 4(4), 295–314, www.jstor.org/stable/2265075.

Van der Sloot, B., 2017. Decisional privacy 2.0: the procedural requirements implicit in Article 8 ECHR and its potential impact on profiling. *International Data Privacy Law*, 7(3), 190–201, https://doi.org/10.1093/idpl/ipx011.

See also: ABORTION, RECOMMENDATION SYSTEM

Declared Data

Declared data is **data** received from a **data subject**, usually a **user** or consumer of a good or service, or a **respondent** to a survey, that is voluntarily and intentionally declared, as opposed to **inferred data**, which is the result of an **inference** by the researcher. Data is generally declared in response to specific questions placed on a form, which may be asking for **identifying** or demographic data, such as name, address or age, or for less precise **attribute**s, such as motivations, preferences or interests. Declared data is regarded as relatively authoritative because its production is unmediated. However, for some classes of data where declarations are subject to biases such as social conformity, declared data may be less reliable than inferred data.

Further reading:
Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. and Rao, V., 1994. Combining revealed and stated preferences data. *Marketing Letters*, 5(4), 335–49, https://doi.org/10.1007/BF00999209.

Decryption

Decryption is the process of converting **encrypted** data back into its original, readable **plaintext** form. A decryption **algorithm** must have access to the correct decryption key to be able to decrypt it, providing an extra layer of **security**.

Depending on the type of encryption used, different decryption algorithms may be used. In **asymmetric cryptography**, different keys are used for encryption and decryption, while **symmetric key encryption** is less secure (and less complex) as it uses the same key for encryption and decryption.

Further reading:
Zhou, X. and Tang, X., 2011. Research and implementation of RSA algorithm for encryption and decryption. *In: Proceedings of 2011 6th international forum on strategic technology,* 1118–21, https://doi.org/10.1109/IFOST.2011.6021216.

See also: ENCRYPTION ALGORITHM, ENCRYPTION KEY, RSA ENCRYPTION

Deepfake

Deepfake is the use of **artificial intelligence (AI)** and **machine learning** methods to produce modified films or images that are hard to distinguish from reality. Deepfakes may be used to produce fake news, political smear campaigns and other types of misinformation that are hard to rebut and can be swiftly and extensively disseminated online.

Deepfakes pose issues such as the possibility of unauthorised use of **personal information** and the danger of **reputational** harm from libellous or detrimental manipulation of a person's voice or image. They may even be used for **harassment** or extortion.

They may also be used to construct plausible false **identiti**es or **fake profile**s that can be exploited for fraud or other illegal actions. This may be a serious **risk** to people and companies, as well as to the general **security** of **network**s and online platforms.

Further reading:

Westerlund, M., 2019. The emergence of deepfake technology: a review. *Technology Innovation Management Review*, 9(11), 40–53, http://doi.org/10.22215/timreview/1282.

See also: DEFAMATION, HARM

Deep Learning

Deep learning is a branch of **machine learning** and **Artificial Intelligence** characterised by the use of neural networks (NNs) built with multiple layers of nodes, or neurons, connected hierarchically. Each layer processes input from the layer above it and its output then forms an input to the layer below. Basic features or patterns are detected by the input layers, while more complex abstraction and analysis are carried out by the deeper layers. The depth of the **network**, which gives the technique its name, increases the power of NNs.

Images, video, audio and text are among the types of media that deep learning has been used to analyse and learn from. **Natural language processing**, **speech recognition**, object and image recognition and **predictive analytics** have all benefited from its extra power. The capacity to learn automatically from **data** and improve without the aid of explicit programming or human intervention is one of its main advantages. However, the danger is that it becomes a black box, with the systems themselves unable to explain their complex and abstract reasoning. The discipline of **explainable AI** is intended to open the black boxes to human understanding, but without definitive success at the time of writing.

Further reading:

Boulemtafes, A., Derhab, A. and Challal, Y., 2020. A review of privacy-preserving techniques for deep learning. *Neurocomputing*, 384, 21–45, https://doi.org/10.1016/j.neucom.2019.11.041.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 436–44, https://doi.org/10.1038/nature1453.

Deep Packet Inspection (DPI)

The action of examining and classifying **network** traffic. DPI introduces intelligence into the network, against the spirit of the **Internet**'s end-to-end principle that **data** should flow freely across it by a simple packet-forwarding mechanism, with the main processing of **data** occurring at the end **users**' devices. DPI effectively allows real-time network management functionality, at the cost of the openness of the network. DPI can be part of **firewall** defences, enabling the detection of **intruder**s and other unwanted incoming traffic (e.g., **virus**es or **spam**) and can also potentially reduce the **risk** of data leaks by identifying and blocking outgoing confidential **information**.

DPI can also be used a mechanism for systematic **surveillance**. For example, the Chinese government uses advance DPI for **censorship** of traffic (including blocking services such as Facebook and Google) and in the United States it is employed by the National Security Agency for increasing the efficiency of intelligence gathering.

DPI are partially restricted to what can be identified in a single packet, so for example some **worm**s which are spread over multiple packets may evade detection by DPI systems.

Further reading:
Bendrath, R. and Mueller, M., 2011. The end of the net as we know it? Deep packet inspection and Internet governance. *New Media and Society*, 13(7), 1142–60, https://doi.org/10.1177/1461444811398031.

See also: DATA FLOW, INTRUSION DETECTION SYSTEM, NATIONAL SECURITY, TRAFFIC DATA

Deep Web

See: DARK WEB

Defamation

A collective term for both libel (written) or slander (spoken), defamation covers false statements which give rise to a civil cause of action to protect **harm** to an individual's **reputation**.

Some accounts of **rights to privacy** encompass the value of personal reputation. Park argues that privacy rights in Europe encompass a broader right to control one's image as perceived by others, and as such overlap with causes of action (such as defamation) which protect reputation. He suggests that this more expansive understanding of privacy can be seen in the development of the **right to be forgotten** by the Court of Justice of the European Union, which regulates **information** already made **public** and thus in one sense not **private**.

This interpretation of privacy rights as encompassing a right to defend one’s reputation is borne out by the **false light privacy tort** in the US, and the **Council of Europe**’s 1970 Declaration on mass communication media and human rights, which defines privacy in expansive terms: ‘The right to privacy consists essentially in the right to live one’s own life with a minimum of interference. It concerns private, family and home life, physical and moral integrity, honour and reputation, avoidance of being placed in a false light, non-revelation of irrelevant and embarrassing facts.’

Further reading:

Park, K.S., 2020. Do we need to separate privacy and reputation? USA, Europe and Korea compared. *In*: Koltay, A. and Wragg, P., eds, *Comparative Privacy and Defamation*. Northampton: Edward Elgar Publishing, 130–46.

Santolaya, P., 2012. The right to private life (notably extended right to privacy) (art 8 ECHR). *In*: Roca, J. G., and Pablo, S. eds, *Europe of rights*: *a compendium on the European Convention of Human Rights*. Leiden: Brill, 337–51, https://doi.org/10.1163/9789004219915_019.

See also: DIGNITY, EUROPEAN CONVENTION ON HUMAN RIGHTS, HARM, INFORMATIONAL SELF-DETERMINATION, INTEGRITY, INTERFERENCE, PRIVACY AS CONTROL, PUBLISHING

Default Settings

Many **privacy** and **data protection** regimes rely on **consent**, and empowering **data subject**s. Consistent with this, Web services, such as **search engine**s, **social media**/networking sites and **e-commerce** sites, provide **privacy settings**, which allow **user**s to specify the amount and quality of **information** about them that the sites can gather. One can set privacy high, perhaps accepting a lower level of service, or low, to receive more or higher quality services.

Setting privacy to an optimal level can be a chore (and can be made even more of a chore by a badly designed interface). Privacy settings may cover a large number of activities, such as (on a **social networking** site) covering which audiences have access to which bits of a user’s **profile**, how they may be **tagged**, who may **search** for them, and so on. Forcing users to fix their privacy settings in advance may put them off, so most sites have default settings (indeed, if the system is to be used even by recalcitrant users, defaults are inevitable). Users are free to change these settings at any time in the future.

However, few do. Hence the default setting, which may not be protective of privacy, has heightened influence. It may be suspected that some sites, particularly commercial ones, set their defaults at a lower level of protection. Furthermore, defaults may be changed without notice, and so privacy settings that the user agreed to when signing up to a **privacy policy** may be subject to alteration in the future.

Further reading:

Liu, Y, Gummadi, K.P., Krishnamurthy, B. and Mislove, A., 2011. Analyzing facebook privacy settings: user expectations vs. reality. In: *IMC '11*: *proceedings of the 2011 ACM SIGCOMM conference on Internet measurement*, ACM, 61–70, https://doi.org/10.1145/2068816.2068823.

See also: CONSENT, DARK PATTERN, PRIVACY NOTICE

De-Identification

De-identification broadly refers to the process of removing direct **identifier**s from **personal data** or **personally identifying information**. In some **data protection** jurisdictions, 'de-identification' has a specific legal meaning. In the United States and Australia, de-identified **data** constitute **information** which no longer identifies individuals, and is no longer covered by the law. The equivalent term within the EU is '**anonymisation**'.

In **jurisdiction**s where 'de-identification' is not legally defined (such as within EU **data protection** law), de-identification may simply constitute **pseudonymisation**. This prevents the direct identification of individuals but nonetheless leaves sufficient **risk** of **reidentification** that the information should still be treated as personal data and covered by law.

Deletion

The removal of files from a computer's **filing system**. In most cases this does not equate to complete **erasure** or **data destruction**. Most operating systems simply remove the link to the **data** from the file system so that the data is no longer visible to the **user** and the file space that the data is in is available for other uses, but in principle the data can be recovered and indeed in some operating systems have built in undelete features.

In the context of **cloud computing**, deletion refers to the process of permanently removing data from a cloud-based service. Users should refer to the policies provided by their cloud service provider to understand how deletion works for that service and if it is compliant with relevant laws and regulations.

Further reading:

Kopo, M.R., Awais, R., and Jose, M., 2016. Assured deletion in the cloud: requirements, challenges and future directions. *In Proceedings of the 2016 ACM on Cloud Computing Security Workshop (CCSW '16)*, Association for Computing Machinery, 97–108, https://doi.org/10.1145/2996429.2996434.

See also: DATA DEGAUSSING, DATA SANITISATION, DATA STORAGE, FILING SYSTEM, CLOUD COMPUTING, SECURITY

Delta

In **differential privacy**, delta is a parameter representing the theoretical probability of **information** being leaked (from a query on a **database**). In standard differential privacy, it is set to zero, so any use of the delta parameter at a higher level can be seen as a relaxation of the strict standard.

See also: PRIVACY METRIC

Demographic Advertising

Also known as demographic targeting, the term refers to the use of demographic data – either **declared data** or inferred from behavioural **information** – to segment an audience to enable the delivery of **targeted advertisement**s (that their demographic profile indicates they are most likely to respond positively to).

See also: BEHAVIOURAL ADVERTISING, INFERRED DATA

Demonstration Attack

A demonstration attack aims to demonstrate the viability of a specific **attack** or technique to highlight a **vulnerability** and motivate developers or administrators to take **remedial** action. Depending on the type of vulnerability or exploit being used for the demonstration, demonstration attacks can take many different shapes.

Security researchers, ethical hackers and malicious adversaries can all conduct demonstration attacks. When carried out by security experts, they can be a helpful tool for locating and fixing vulnerabilities before they can be used by an **adversary**.

See also: BLACK HAT ATTACK, ETHICAL HACKING, GREY HAT ATTACK, HACKING, PENETRATION TESTING, WHITE HAT ATTACK

Denial of Service (DoS)

The goal of a Denial of Service (DoS) attack is to disable a **network**, server, website or **application** from performing its regular tasks by flooding it with excess traffic. The **adversary** sends additional packets, taking advantage of **software** flaws to use up all of the target's bandwidth and resources. The effects of this kind of attack can be lost sales, **reputational harm** and other consequences.

There are mitigation techniques to protect against DoS attacks, such as network traffic filtering, traffic splitting over several servers, **intrusion detection system**s and **intrusion prevention system**s, cloud-based mitigation services, routine testing and **risk assessment**.

DoS attacks can happen at any layer of the network stack (i.e., network, application, etc.). A DoS attack coordinated from multiple locations is called a DDoS (Distributed Denial of Service) attack. In this case the adversary usually takes control over various devices and launches the attack at the same time from all of them. The network of devices which has been compromised in this way for a coordinated DDoS attack is often referred to as a **botnet**.

Further reading:

Carl, G., Kesidis, G., Brooks, R.R. and Rai, S., 2006. Denial-of-service attack-detection techniques. *IEEE Internet Computing*, 10(1), 82–9, https://doi.org/10.1109/MIC.2006.5.

See also: APPLICATION LAYER ATTACK, INTERNET, MIRAI, NETWORK LAYER ATTACK, SECURITY

Deterministic Record Linkage

A form of **record linkage** where only exact matches are allowed to be linked. It is much faster to run than **probabilistic record linkage** as no complex comparisons are needed. It sacrifices recall for precision, accepting a higher rate of **false negative**s to minimise **false positive**s. Hybrid approaches can carry out deterministic record linkage first, followed by probabilistic record linkage.

Further reading:
Christen, P., 2012. *Data matching*. Berlin: Springer, https://doi.org/10.1007/978-3-642-31164-2.

See also: RECORD

Device Fingerprinting

Using specific characteristics and configurations, such as the operating system, browser type and version, hardware configuration and **network** settings, device fingerprinting is a technique for **identifying** and **tracking** devices. This **data** can be gathered and analysed to create a basic **identity** for a specific device, and to track its use across the **Internet**.

Online advertising, fraud detection and **security** monitoring frequently use device fingerprinting. For instance, it allows online marketers to follow **user**s across multiple websites and show them relevant ads based on their **browsing history**. It can also be used by fraud detection systems to spot suspicious activity, such as attempts to open multiple accounts on the same device. Security **surveillance** systems may be used.

Common precautions against it include using **ad blocker**s, clearing browser caches and **cookie**s, disabling JavaScript, using **virtual private network**s (VPNs) or the **TOR** network and avoiding online activities that could reveal sensitive **information**. However, since some techniques can be extremely challenging to detect and counter, it is hard to completely prevent device fingerprinting.

Further reading:
Xu, Q., Zheng, R., Saad, W. and Han, Z., 2015. Device fingerprinting in wireless networks: challenges and opportunities. *IEEE Communications Surveys & Tutorials*, 18(1), 94–104, https://doi.org/10.1109/COMST.2015.2476338.

See also: BROWSER FINGERPRINTING, DIGITAL FINGERPRINTING, IP ADDRESS, TARGETED ADVERTISING, TRACKER BLOCKER

DICOM Standard (Digital Imaging and Communications in Medicine)

DICOM (Digital Imaging and Communications in Medicine) is a **standard** for the management (and sharing) of medical images and related **metadata**. It enables the effective integration of different medical imaging devices from multiple manufacturers.

DICOM files can be exchanged between any organisations that are capable of processing DICOM format data. Devices come with DICOM Conformance Statements which state which DICOM classes they support. The standard includes a file format definition and a network communications **protocol**.

In principle, standards such as DICOM increase **data portability** and therefore **compliance** with article 20 of **GDPR**. They also make it easier to develop general solutions to **anonymising** the data. Recently, however, concerns have been raised that the DICOM standard may leave medical images vulnerable to **malware** insertion attacks.

Further reading:

Monteiro, E., Costa, C., and Oliveira, J.L., 2017. A de-identification pipeline for ultrasound medical images in DICOM format. *Journal of Medical Systems*, 41(5), 1–16, https://doi.org/10.1007/s10916-017-0736-1.

Mustra, M., Delac, K., and Grgic, M., 2008. Overview of the DICOM standard. *In*: *50th International Symposium ELMAR*, 1, 39–44, IEEE, https://ieeexplore.ieee.org/abstract/document/4747434.

Ortiz, M.O., 2019. HIPAA-protected malware? Exploiting DICOM flaw to embed malware in CT/MRI imagery. *Cylera Labs*, https://researchcylera.wpcomstaging.com/2019/04/16/pe-dicom-medical-malware/.

See also: ELECTRONIC HEALTH RECORD, HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT

Differencing

A **reidentification** attack whereby two different but overlapping codings for a variable are overlain leading to cross-classified categories which might contain small numbers of cases. Geographical coding is often considered the most like source of such cross-classified **information**, as many possible geographies exist, but in principle any variable could be subject to differencing. The issue is most likely to arise with flexible query systems in which a **user** can request their own classifications by querying a detailed **database**. This is one of the rationales for **differential privacy** where such attacks are – in principle, at least – prevented by the **privacy budget**.

See also: QUERY OVERLAP, REIDENTIFICATION ATTACK

Differential Identifiability

A variant on **differential privacy** that focuses on the **identifiability** of **data subject**s in some statistical output rather than how much an individual **respondent** affects the output. This has the advantage of being closer to the requirements of regulations such as **GDPR**.

Further reading:
Lee, J. and Clifton, C., 2012. Differential identifiability. *In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining*, 1041–9, https://doi.org/10.1145/2339530.2339695.

See also: OUTPUT PRIVACY

Differential Privacy

Differential privacy is a data analysis framework that aims to safeguard the **privacy** of people whose **data** is being examined by reducing the risk of **disclosing** sensitive **information** about specific users, while allowing the release of aggregate statistics and other information. The key principle of differential privacy is to allow the querying of a **database** so that the answer betrays no information about an individual, including whether they are in the database or not. The required level of protection will therefore depend upon the query.

The fundamental principle of differential privacy is to **mask** any information that could be used to specifically identify individuals, by **noise addition** or introducing randomness to the data being analysed. To maintain **accuracy** and **data utility** and to safeguard privacy, this noise is carefully calibrated and controlled so that wider statistics are unchanged. A key parameter is **epsilon**, or ε, which is a measure of the privacy loss that is incurred by making a query to the dataset.

Differential privacy techniques can be applied in a variety of ways, such as subsampling, **Laplace noise** and **randomised response**. Implementing differential privacy can be difficult, though, as it necessitates carefully weighing the **privacy risk**s and the data being analysed.

Further reading:
Dwork, C., and Roth, A., 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407, http://dx.doi.org/10.1561/0400000042.

See also: BIG DATA, DATA RELEASE, DELTA, PRIVACY METRIC

Digital Assistant

A digital assistant is a type of device that can offer users a variety of services on request, designed to be responsive to simple and intuitive commands, such as by voice. **Artificial Intelligence** and **natural language processing** are therefore vital components. Answering queries, providing **information**, setting reminders, placing calls or sending messages, setting appointments, playing music, offering navigation and controlling smart home devices are some of the many tasks that digital assistants can handle. Widely used examples include Siri from Apple, Alexa from Amazon, Google Assistant and Microsoft’s Cortana.

These virtual helpers can be found on a range of gadgets, such as smartphones, smart speakers and smart home appliances. A growing number of people are using digital assistants because of their convenience, as well as the ease with which they can be integrated with a variety of other services and gadgets. However, because they may gather and store **personal information** to provide their services, their use has also brought up questions about **privacy** and **security**. The result is a trade-off between the **privacy risk** and the value of the services. Furthermore, there is a question – for instance with a **recommendation** – as to whether the ultimate beneficiary of the service is the **user**, or the service provider.

Further reading:
Dubois, D.J., Kolcun, R., Mandalari, A.M., Paracha, M.T., Choffnes, D. and Haddadi, H., 2020. When speakers are all ears: characterizing misactivations of IOT smart speakers. *Proceedings on Privacy Enhancing Technologies*, 2020(4), 255–76, https://doi.org/10.2478/popets-2020-0072.

See also: DATA EXHAUST, DATAFICATION, DIGITAL FOOTPRINT, EDGE COMPUTING, INFORMATION SECURITY, INTERNET OF THINGS, SMART DEVICE

Digital Breadcrumbs

Digital breadcrumbs are the pieces of **information** that the user of an electronic device leaves behind in use. The metaphor implies that the 'breadcrumb trail' can be 'followed' until the follower 'reaches' the user, by **identifying** them or **inferring** key, perhaps monetisable, information about them.

Further reading:
George, G., Haas, M.R. and Pentland, A., 2014. Big data and management. *Academy of Management Journal*, 57(2), 321–6, https://doi.org/10.5465/amj.2014.4002.

See also: DATA EXHAUST, DIGITAL FOOTPRINT, JIGSAW IDENTIFICATION

Digital Certificate

A digital certificate is a form of **identification** that can be used to confirm the legitimacy of a person, business or device. A **Certification Authority (CA)**, a dependable **third party**, confirms the identity of the certificate holder and issues the certificates. Online banking, **secure communication**s and **e-commerce**, where the **security** of a remote transaction needs to be ensured, are common users of digital certificates. A browser will verify that the certificate of a website is valid and issued by a reputable CA when the **user** accesses it. It will then create a safe, **encrypted** connection between the user's computer and the website only if the certificate is legitimate, and recommend the connection be closed otherwise.

Digital certificates can be used to **authenticate** the identities of the parties involved in an online transaction or communication, but the need for a third party may introduce a **vulnerability**, for example if certificates are issued by a non-reliable CA or if the **private key** that accompanies the certificate is compromised.

Further reading:
Leavitt, N., 2011. Internet security under attack: the undermining of digital certificates. *Computer*, 44(12), 17–20, https://doi.org/10.1109/MC.2011.367.

See also: CERTIFICATION, CRYPTOGRAPHIC KEY, DIGITAL IDENTITY, ENCRYPTION KEY

Digital Divide

See: DIGITAL INEQUALITY

Digital Economy

The digital economy is a loose term describing the evolution of exchange and economic activity facilitated by the **Internet**, the **World Wide Web**, smartphones and other digital technologies and **smart device**s. Such technologies may facilitate traditional economic activities of buying and selling, or digital resources may also be exchanged. **Application**s can manage interactions, allowing otherwise redundant resources to be rented in small quantities (the ‘gig economy’). Digital payments systems, micropayments and digital and cryptocurrencies may be used to make purchases, while **data** generated by individuals may also be used in exchange for ‘free’ services.

Further reading:

Tapscott, D., 2015. *The digital economy*: *rethinking promise and peril in the age of networked intelligence*, 20th anniversary edition. New York: McGraw Hill Education.

See also: COMMODIFICATION, CRYPTOCURRENCY, E-COMMERCE, ECONOMICS OF PRIVACY, NEGATIVE EXTERNALITIES OF DISCLOSED DATA, SURVEILLANCE, SURVEILLANCE CAPITALISM, VALUE OF DATA, VALUE OF PRIVACY

Digital Fingerprinting

The process of gathering different **data** about a **user**’s activity to create a digital **profile** or ‘fingerprint’ of that user is known as digital fingerprinting. Digital fingerprinting aims to **track** and **identify** specific users via their devices as they navigate various websites or online services.

Digital fingerprinting employs several techniques to gather details about a computer or device, including screen size and resolution, **IP address**, installed fonts and plugins, browser and device settings, and other distinctive system characteristics. The device's digital profile or fingerprint is created using this data, and it can be used to track the device or its user across various websites or online services.

Digital fingerprinting is frequently utilised in online marketing and **targeted advertising** because it enables advertisers to target users or devices with relevant ads. Digital fingerprinting, however, can also be used for more nefarious objectives, such as monitoring user behaviour or gathering private data without the user's knowledge or **consent**. As digital fingerprinting technology advances, worries about its effects on **security** and **privacy** are becoming more widespread. To safeguard user privacy and stop the misuse of this technology, some experts have called for more regulation and **transparency** surrounding digital fingerprinting practices.

Further reading:

Laperdrix, P., Bielova, N., Baudry, B. and Avoine, G., 2020. Browser fingerprinting: a survey. *ACM Transactions on the Web*, 14(2), article no.8, https://doi.org/10.1145/3386040.

See also: BROWSER FINGERPRINTING, CROSS-DEVICE TRACKING, CUSTOMER TRACKING, DEVICE FINGERPRINTING, USER MODELLING, WEB PROFILING

Digital Footprint

Interaction online requires a computer as intermediary. As all interactions involve exchanges of digital **information**, it is possible to store all such interactions together with some means of indexing using perhaps timestamps and/or **IP address**es. Broadly speaking, the aggregate of the exchanges involving an individual that are stored can be thought of as a **digital footprint** of that individual. An individual may have different footprints in different computer systems.

Other phrases for the same phenomenon exploit different aspects of the metaphor; a **digital fingerprint** also focuses on the link between the **record** and the **natural person**, although a fingerprint is more strongly connected with a **unique** individual than the more generic footprint, while **data exhaust** emphasises the creation of data as a by-product, without suggesting that it makes up a coherent picture of an individual.

There are various ways of understanding what relationship individuals bear to their digital footprints. Different people, organisations, corporations and the state may each hold a digital footprint of an individual: a retailer will hold a set of transactions; a **social network**ing site will hold a set of interactions, friendships, likes and postings; a **search engine** may remember time-stamped **search**es; an Internet Service Provider will hold downloads from IP addresses, and so on. Collectively, this would be a rich and informative data source, so regulations usually make it difficult to merge these footprints into a single picture. However, regulations often demand that these footprints are preserved for a period of time, in case they might be useful, e.g., **data retention** for law enforcement purposes. Footprints can also be used for **inference**s about individuals, for example using addresses and purchases to cluster them into different types of purchasers.

In their lexicographical analysis of the literature, Parkinson et al suggest that a digital footprint (whether provided by the individual’s **declared data** or by a **third party**) best refers to the data by-products of interactions. A *digital mosaic* is created from aggregating a range of digital footprints of an individual. Analysis of digital footprints of an individual, possibly alongside the digital footprints of others, produces a model of the individual called a *digital persona*. Finally, we might consider the sum total of all such representations (which may be impossible to achieve) as a *digitally extended self*.

Further reading:

Parkinson, B., Millard, D.E., O’Hara, K. and Giordano, R., 2018. The digitally extended self: a lexicological analysis of personal data. *Journal of Information Science*, 44(4), 552–65, https://doi.org/10.1177/0165551517706233.

See also: CLICKSTREAM DATA, COMMODIFICATION, DATABASE OF RUIN, DIGITAL TWIN, INTERNET OF PEOPLE, PERSONAL DATA, SELF

Digital Footprint Eraser

A system or service which facilitates the **deletion** of **information** about an individual across the **Internet**, or on specific systems such as those owned by **data broker**s.

See also: DIGITAL FOOTPRINT, INFORMATIONAL SELF-DETERMINATION, PRIVACY AS CONTROL, RIGHT TO BE FORGOTTEN

Digital Hygiene

A set of basic **cybersecurity** practices for individuals. Different cybersecurity experts have a different list about what this involves but examples include regularly updating and cleaning all electronic devices, keeping **anti-virus software** up-to-date and live scan on, using strong **password**s, a **password manager** and **two-factor authentication**, identifying **phishing** attempts, being cautious in sharing identity **information** online and keeping operating systems up to date.

Further reading:
Gelbstein, E., 2013. *Good digital hygiene: a guide to staying secure in cyberspace*. Bookboon, https://bookboon.com/en/good-digital-hygiene-ebook.

See also: DIGITAL LITERACY

Digital Imaging and Communications in Medicine

See: DICOM STANDARD

Digital Identity

An electronic representation of a person used to **identify** and authenticate them in digital interactions is called a digital identity. Digital identities frequently include signifiers such as names, email addresses, **username**s, and **password**s along with other identifying **information** like **biometric data**, **security token**s, **digital signature**s or **digital certificate**s. These identities can be created and managed by both individuals and organisations.

Digital identities are becoming more and more crucial for establishing **trust** and safeguarding **personal data** as users’ activities move online; in fact, without digital identities the **digital economy** could hardly function. However, the use of digital identities also raises **data protection**, **security** and **privacy** issues. In particular by creating a single digital entity, **identity theft** leading to unauthorised access to **personal information** or other resources becomes more serious.

Further reading:
Camp, J., 2004. Digital identity. *IEEE Technology and Society Magazine*, 23(3), 34–41, https://doi.org/10.1109/MTAS.2004.1337889.

See also: AUTHENTICATION, DIGITAL FINGERPRINTING, DIGITAL FOOTPRINT, IDENTITY, INFORMATION SECURITY, PERSON, SELF-SOVEREIGN IDENTITY, SINGLE OUT

Digital Inequality

Digital inequality refers to the idea that people are unequally able to take advantages of the opportunities, and to protect themselves from the threats, of the digital world. This view was initially labelled the *digital divide*, with the idea that people who were already relatively privileged (particularly in terms of race, gender, class and education) received greater access to digital technology. As access evened out, digital divides were reconceptualised to mean that, even where those on the wrong side of the divide had access to the technology, they were less able to gain advantage from it, owing to disadvantages in skills and education.

Digital inequalities affect **privacy** in that protection against **intrusion** often requires money (to pay for legal or technological defences), support (such as high-quality IT support, perhaps from an employer) and education (to understand the threat and master **digital hygiene**). It seems to follow that those in low-paid employment and of lower levels of education will, all things being equal, be less able to defend themselves against intrusion. Furthermore, Coles-Kemp argues that **cybersecurity** is currently largely an elite concept, which focuses on the protection of technology rather than the **security** of **user**s and excludes the views of users in favour of addressing highly technical threats with similarly technical solutions. The cybersecurity element of a social security system, for instance, is likely to be directed against the possibility of users receiving more benefits than they are entitled – requiring intrusive questioning of those users and **interference** in their ways of life.

Further reading:

Coles-Kemp, L., 2020. Inclusive security: digital security meets Web Science. *Foundations and Trends in Web Science*, 7(2), 88–241, http://dx.doi.org/10.1561/1800000030.

Eubanks, V., 2019. *Automating inequality*: *how high-tech tools profile, police, and punish the poor*. New York: Picador.

See also: BENEFITS OF PRIVACY, DIGITAL LITERACY, GROUP HARMS

Digital Inheritance

When someone dies, they nowadays tend to leave behind digital **asset**s which may be part of the inheritance of their beneficiaries. This is their digital inheritance, and may include **social media** accounts, playlists, photographs, email accounts, calendars and medical records. Access to these, and even knowledge of their existence, may be hard for executors and beneficiaries to secure. As these resources move into the **cloud** and are increasingly co-created by service providers who have some **intellectual property** rights in the assets, the legal and technological issues raised by digital inheritance are complex. Furthermore, the posthumous **privacy** of the deceased, and the sensitivities of the bereaved, may be compromised where the digital inheritance contains evidence of concealed activity, such as previously unknown relationships, or quantities of pornography.

Further reading:
Nemeth, K. and Carvalho, J.M., 2017. Digital inheritance in the European Union. *Journal of European Consumer and Market Law*, 6(6), 253 (and subsequent country reports in issues 6(6) and 7(1)). https://jorgemoraiscarvalho.com/wp-content/uploads/2018/02/Digital-Inheritance-in-the-European-Union.pdf.

See also: DIGITAL FOOTPRINT, INTELLECTUAL PRIVACY, LIFELOGGING

Digital Literacy

Digital literacy refers to people’s understanding of the digital tools, practices and industries that impact their lives. This includes both their ability to use digital tools effectively for their own purposes, and their **awareness** of how **information** has an impact on their own behaviour, preferences and exposure. Hence it involves an understanding not only of the tools themselves, but also of the business practices including them, and the cultures in which they are embedded.

It is often asserted that a digitally literate citizenry will be empowered relative to the use of digital tools, will be able to use the technology in their own interests and to pursue their goals, will be able to produce **informed consent** and will be better able to resist others’ attempts to exploit their digital personae (including attempts to **breach informational privacy**). Many assume that *digital natives* (those brought up in a world characterised by digital technology) are, on average, more digitally literate than *digital immigrants* (those who learned these technologies as adults), although the evidence is equivocal.

Further reading:
Bawden, D., 2008. Origins and concepts of digital literacy. *In*: Lankshear, C. and Knobel, M., eds, *Digital literacies: concepts, policies and practices*. New York: Peter Lang, 17–32, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b6883b3ab0d8c172ccc2f4ac668de5d004a85da5.

Digital Persona

See: DIGITAL FOOTPRINT

Digital Rights Management (DRM)

It is common practice within digital commerce for content to be shared in a way that prevents infringement of digital rights; for example, for software to be sold in a form that does not permit **third-party** modification contrary to the **user**'s **intellectual property** licence. Digital rights management (DRM) is less discussed in the context of preserving **right**s **to privacy** in **information**, but the same tools and techniques often applied in DRM (**security** and **integrity** features, **encryption** etc.) can equally be used to protect the information of **identifiable individual**s as **data** are made available to a downstream chain of **data user**s.

As Feigenbaum and colleagues note, DRM systems can create a tension between the **intellectual property** rights of owners/distributors and the privacy rights of end-users, particularly when user **tracking** or **network** control is used as an enforcement tool by copyright holders.

Brownsword also discusses, more broadly, the loss of human **autonomy** stemming from technological management, that is, the use of technology to make certain behaviour impossible, rather than appealing to personal morality through legal or social norms. The alleged corrosion (or marginalisation) of human decision-making can be seen as engaging our need for cognitive integrity, and thus for **decisional privacy**.

Further reading:
Brownsword, R., 2019. *Law, technology and society*: *re-imagining the regulatory environment.* Abingdon: Taylor & Francis.
Feigenbaum, J., Freedman, M.J., Sander, T. and Shostack, A., 2002. Privacy engineering for digital rights management systems. *In*: Sander, T., ed. *Security and privacy in digital rights management.* Heidelberg: Springer, 76–105, https://doi.org/10.1007/3-540-47870-1_6.

See also: CUSTOMER RELATIONSHIP MANAGEMENT, PRIVACY ENGINEERING, INFORMATION LIFECYCLE MANAGEMENT

Digital Self-Determination

Digital self-determination conveys the extrapolation of the social construct of human self-determination into the digital sphere. It is unclear whether the digital manifestation of self-determination brings something qualitatively different or simply a change in scale and volume of already existing processes.

On the one hand, the digital sphere has manifested an unprecedented degree of self-expression and **communication** for individuals (across the globe). However, it has also been shaped by and exacerbated existing inequalities and power structures, sometimes referred to as **digital inequality**. Moreover, this transformation has enabled **breach**es of **privacy** at scale through the mass collection, and analysis, of **personal data** to infer individuals' preferences and to influence their behaviour and attitudes. What may appear to the individual as a means of self-expression may in fact be subtly influenced by the tools, structures and platforms which are used, or a dialectical process such as Giddens described in his work on structuration.

Further reading:
Giddens, A., 1984. *The constitution of society: outline of the theory of structuration*. Cambridge: Polity Press.

See also: AUTONOMY, COMMUNICATION PRIVACY, DIGITAL FOOTPRINT, INFORMATIONAL SELF-DETERMINATION, SURVEILLANCE, SURVEILLANCE CAPITALISM

Digital Signature

Public-key cryptography underpins digital signatures. With such a system, Alice creates an unencrypted message to send to Bob, and adds her signature **encrypted** with her **private key**. When Bob receives the message, he can verify that the signed document was signed by Alice, and has not been amended or edited, by **decrypting** the signature with Alice's **public key**.

This is different from an electronic signature, which is simply a digital version of a physical (wet ink) signature. In EU law the digital signature is known as an *advanced electronic signature*.

Further reading:
Merkle, R.C., 1989. A certified digital signature. *In: Conference on the Theory and Application of Cryptology*, 218–38, https://doi.org/10.1007/0-387-34805-0_21.

See also: CRYPTOGRAPHY, DIGITAL CERTIFICATE

Digital Twin

A virtual representation of a physical entity or system, often combined with a range of means of visualisation. Digital twins were initially developed in an engineering context where they have value in allowing systems to be simulated before they are built to identify problems in advance. But increasingly they are being used to run alongside physical systems for early warning of faults, pre-testing system stresses, and the like.

Datafication of humans and their activity has given rise to the possibility of human digital twins with significant implications for **identity** and **privacy**.

Further reading:

El Saddik, A., 2018. Digital twins: the convergence of multimedia technologies. *IEEE Multimedia*, 25(2), 87–92, https://doi.org/10.1109/MMUL.2018.023121167.

Far, S.B. and Rad, A.I., 2022. Applying digital twins in metaverse: user interface, security and privacy challenges. *Journal of Metaverse*, 2(1), 8–15, https://dergipark.org.tr/en/pub/jmv/issue/67967/1072189.

Wang, Y., Su, Z., Guo, S., Dai, M., Luan, T.H. and Liu, Y., 2023. A survey on digital twins: architecture, enabling technologies, security and privacy, and future prospects. *IEEE Internet of Things Journal*, 10(17), 14965–87, https://doi.org/10.1109/JIOT.2023.3263909.

See also: AUGMENTED REALITY, DIGITAL IDENTITY, INTERNET OF THINGS

Digital Wallet

A digital wallet is a software resource for making electronic payments. It will take note of payments into and out of an account and verify that a payment is covered by the funds in the account; paying into the account is described as paying into the wallet. The account may be held separately, or the wallet itself may be the account. Hence it is analogous to a physical wallet (if a non-metaphorical wallet does not contain sufficient cash, then the transaction cannot take place). Association with an account means that a wallet may be used to **identify** the owner.

In a **cryptocurrency** such as Bitcoin, the wallet contains the owner’s **private key** which can be used to verify ownership of the cryptocurrency on the **blockchain**. Because it only stores the key, not the coin, the wallet need not identify the owner. If the wallet owner uses no other means to **authenticate** their ownership, they are effectively **anonymou**s.

Further reading:

Antonopoulos, A.M., 2015. *Mastering Bitcoin*: *unlocking digital cryptocurrencies*. Sebastopol CA: O'Reilly Media.

Hassan, M.A. and Shukur, Z., 2019. Review of digital wallet requirements. *In*: *2019 International Conference on Cybersecurity (ICoCSec)*, IEEE, https://doi.org/10.1109/ICoCSec47621.2019.8970996.

See also: CRYPTOGRAPHY, PUBLIC-KEY CRYPTOGRAPHY

Dignity

Dignity is an ethical conception of innate human value, which came to prominence in the moral philosophy of the eighteenth century. Immanuel Kant, in particular, crystallised the concept within his categorical imperative that people should be treated as ends and not means.

The assumption of inalienable value of human dignity continues to shape human rights law, particularly within the jurisprudence of the European Court of Human Rights. This has implications for the right to private life under Article 8 of the **European Convention on Human Rights** and the legal duty to respect **privacy** as a way of upholding the dignity of individuals from humiliation, degradation and unauthorised exposure. While dignity may be most obviously compromised through intervention with **bodily privacy**, the large-scale **commodification** of **personal data** increasingly raises dignity concerns in an **information**al context.

Further reading:

Düwell, M., Braavig, J., Brownsword, R. and Meith, D., eds, 2014. *The Cambridge handbook of human dignity*: *interdisciplinary perspectives*. Cambridge: Cambridge University Press.

See also: AUTONOMY, INFORMATIONAL PRIVACY, VALUE OF PRIVACY

Direct Access Attack

Gaining physical access to the computing system, or some part thereof. The implication is that the **adversary** will then perform various malicious actions, such as installing devices to compromise **security**, **virus**es or **Trojan horse**s, or will simply download important **data** using portable media.

See also: SECURITY

Direct Identifier

A direct identifier is a symbol representing an **attribute** of individuals that identifies them within a **dataset**, ideally uniquely. This may be the individual’s name, although this may not be unique (it can be made so within the dataset by using additional characters, as ‘john-smith-32’). It could be an **identifier** that singles out the individual against a schema designed for uniqueness – for example, a **personal identification number**, customer reference number, social security number, NHS number, email address, mobile number or passport number. Not all direct identifiers are equally reliable; Elliot et al define five classes of direct identifier, driven largely by the relationship between the identifier and **the data subject**.

Formal anonymisation or **de-identification** is achieved by the removal or replacement of direct identifiers in a dataset. However, this may still leave open the possibility of *indirect identification* of individuals.

Further reading:

Elliot, M., Mackey, E. and O’Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners’ guide*, 2nd edition. United Kingdom Anonymisation Network, https://ukanon.net/framework/.

International Organization for Standardization, 2018. *Privacy enhancing data de-identification terminology and classification of techniques*, definition 3.10, https://www.iso.org/obp/ui/#iso:std:iso-iec:20889:ed-1:v1:en.

See also: ANONYMISATION, FUNCTIONAL ANONYMISATION, FUNCTIONAL UNIQUE IDENTIFIER, INDIRECT IDENTIFIER, IDENTIFIABLE INDIVIDUAL, UNIQUE IDENTIFIER

Direct Marketing

An approach to advertising and sales where sellers communicate directly with consumers to promote their products or services. This minimally requires possessing contact **information** for consumers and ideally some profile information to allow targeting of marketing.

Direct marketing therefore raises **data protection** and other **privacy** concerns.

Further reading:

Palmer, A. and Koenig-Lewis, N., 2009. An experiential, social network-based approach to direct marketing. *Direct Marketing*: *An International Journal*, 3(3), 162–76, https://doi.org/10.1108/17505930910985116.

See also: CUSTOMER TRACKING, TARGETED ADVERTISING

Directory Indexing

Directory indexing is the process of creating and maintaining a list of the files and subdirectories contained within a directory on a computer. A software program typically creates the index, which is then saved in a file or **database** for quick and simple access.

In the context of Web servers, by default the directory indexing function lists the contents of a directory accessed by a **user** if the default webpage is not present. This was a feature that was originally intended to allow **Internet** users to **search** sites in the same way as folder structures on their own computers, but it can lead to **security** vulnerabilities, as it may expose configuration, temporary and backup files. An **adversary** could exploit this to gain access to the contents of a directory. To prevent these issues, a Web server's configuration should be adjusted to disable the listing of directory contents.

Disassociability

The minimisation of the connection between **data unit**s and **population units**. **Anonymisation** and data **security** processes can all be seen as driven by the disassociability principle.

The minimisation of connections between different elements in a system. Flows of **data**, **people** and **communication**s are kept down to the lowest level that meets operational requirements. This increases system **security**, **resilience** and trustworthiness.

See also: CYBERSECURITY, DATA FLOW, DATA MINIMISATION, TRUST

Disclosive Data

Data that allow **data subject**s to be **identified** (either directly or indirectly) and/or reveal **information** about data subjects. Data can be disclosive without any actual disclosures having (yet) happened.

Structured **categorical data** can be technically disclosive if it contains empirical zeroes in the underlying table of counts. However, it is debated whether such a strict definition is practically useful.

See also: DISCLOSURE, DISCLOSURE RISK, TABULAR DATA

Disclosure

The revelation of **information** that was previously **secret**, hidden, **obfuscated**, **obscured** or simply not known. Disclosure may be deliberate, where the entity holding the **data** chooses to release or share it (e.g., I publish my CV online), it may be **consented** (e.g., I agree to tell you about my work history to apply for a job) or it might occur through a **breach** (e.g., after a **reidentification attack**, the **adversary** sells my employment history).

See also: DATA BREACH, DATA RELEASE, DATA SHARING, DISCLOSIVE DATA, DISCLOSURE AND BARRING, DISCLOSURE CONTROL METHODS, NON-DISCLOSURE AGREEMENTS, PUBLICATION, STATISTICAL DISCLOSURE, STATISTICAL DISCLOSURE CONTROL

Disclosure and Barring (Check)

Under UK law, prospective employers can check applicants' **record**s on the Police National Computer if the role in question falls within certain categories (in healthcare, finance, childcare, etc.). Whether this check should include spent convictions, or whether an individual's right to rehabilitation should override any ongoing threat they may pose in a professional role, is a contentious issue.

Further reading:
Zalnieriute, M., 2013. Blanket criminal record data disclosure system incompatible with privacy rights. *International Data Privacy Law*, 3(3), 197–201, https://doi.org/10.1093/idpl/ipt012.

See also: CONFLICT OF RIGHTS, DEFAMATION, RIGHT TO BE FORGOTTEN

Disclosure Control Methods

Data-focused methods for reducing statistical **disclosure risk**, usually based on restricting the amount of, or modifying, the data released or shared.

Further reading:
Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K. and De Wolf, P.P., 2012. *Statistical disclosure control*. New York: Wiley.

See also: DATA, DATA RELEASE, DATA SHARING, DISCLOSURE, STATISTICAL DISCLOSURE CONTROL

Disclosure Risk

The **risk** – present in all useful data – that some entity (often an individual person) will be **identified** in the data even when the data have been subjected to **de-identification**. More specifically, the term is used to refer to the probability that an **adversary** can identify and/or reveal new **information** about at least one **data subject** in disseminated data. That probability is notoriously difficult to estimate as it is subject to a large range of practical uncertainties. Most measures of risk make estimations based on the structure of the data themselves, but, as Elliot et al have argued, the risk resides in the relationship between the data and their **data environment** and **data situation**.

Further reading:

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K., & De Wolf, P.P., 2012. *Statistical disclosure control*. New York: Wiley.

Elliot, M., O'Hara, K., Raab, C., O'Keefe, C.M., Mackey, E., Dibben, C., Gowans, H., Purdam, K. and McCullagh, K., 2018. Functional anonymisation: personal data and the data environment. *Computer Law and Security Review*, 34(2), 204–21, https://doi.org/10.1016/j.clsr.2018.02.001.

See also: ANONYMISATION, ATTRIBUTION DISCLOSURE, DISCLOSURE, FUNCTIONAL ANONYMISATION, IDENTIFIED DATA, REIDENTIFICATION, STATISTICAL DISCLOSURE, STATISTICAL DISCLOSURE CONTROL, RISK/UTILITY TRADE-OFF

Discrete Data

Also known as *ordinal data*, discrete data is **information** is in the form of ordered variables, such as level of education, Likert scale data and shoe sizes. Discrete data may or may not relate to an underlying continuous dimension.

See also: CATEGORICAL DATA, CONTINUOUS DATA

Discretionary Access Control

See: ACCESS CONTROL

Discretisation

A type of **global recoding** where continuous variables are converted to discrete ones.

See also: CONTINUOUS DATA, DISCRETE DATA, GLOBAL RECODING

Disguise

A disguise changes the appearance of something to conceal its **identity**. Most commonly, disguises are for the human face (such as wigs, false facial hair or makeup) or body (such as costume or prosthetics to mislead about body shape). The *practice* of disguise also includes changing behaviour patterns, gait, voice and accent, and so on. A *master* (or *mistress*) *of disguise* is someone who specialises in adopting convincing false personae. Those who change identity, such as transitioning to another gender, are not usually thought of as disguising themselves when changing their appearance, because their new identity is not intended to mislead.

Camouflage is a kind of disguise, which is often found in the animal kingdom, and also with objects such as military gear. Its aim is to blend in with the background and not be noticed, rather than to mislead about identity.

Further reading:

Mazzuki, A., Siljander, R. and Mitchell, S., 2015. *Undercover disguise methods for investigators: quick-change techniques for both men and women*. Springfield: Charles C. Thomas.

See also: IDENTITY, FACIAL RECOGNITION TECHNOLOGY, GAIT RECOGNITION, MASK, VEIL

Distributed Denial of Service

See: DENIAL OF SERVICE

Distributed Ledger

See: BLOCKCHAIN

DNS Server

A computer server that converts domain names into **IP addresses** is known as a Domain Name System (DNS) server. The IP address of the server hosting the website associated with a domain name is requested from a DNS server whenever a **user** enters that domain name into their **Web** browser. The browser can then connect to the Web server and access the desired content after the DNS server returns the IP address.

Internet Service Providers, Web hosts or large organisations that need their own private DNS infrastructure typically run DNS servers. Additionally, there are open DNS servers that anyone can use, including those run by Google and Cloudflare. DNS servers are essential to the operation of the **Internet**, and the user experience can be significantly impacted by their performance. As a result, there is ongoing research and development in the area of DNS with the goal of enhancing DNS resolution’s speed, security and effectiveness.

The **privacy**-relevance of this infrastructure is that the handling of the browser’s request for an IP address could, if not secure, reveal sensitive **information** about which sites the browser is being asked to connect to. To maintain privacy, this should not be inferable from whatever information is made **public**.

Further reading:

Zhao, F., Hori, Y. and Sakurai, K., 2007. Analysis of privacy disclosure in DNS query. *In: 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE 07)*, 952–7, https://doi.org/10.1109/MUE.2007.84.

See also: BROWSING HISTORY, DIGITAL FOOTPRINT, TRACKING

Domain Generalisation

See: GLOBAL RECODING

Dominance Rule

A form of **output statistical disclosure control** which is applied to tables of summary statistics (typically volumes or means), and which aims to prevent **information** about specific contributors being **disclosed**. A cell in such a table is triggered by the rule that the *n* largest units cannot contribute more than *k*% to the cell total. For example, if n=2 and k=70, a cell is deemed to be vulnerable to confidentiality breaches if the two largest units contribute more than 70 per cent to the cell total. The setting of the parameters *n* and *k* is a matter of judgment that will usually be made by the **data stewardship organisation** itself. But the basic principle is to prevent one of the larger contributors being able to make tightly bounded estimates of the size of other contributors' contributions to the cell.

Also known as the *concentration rule* and the *(n,k) rule*.

See also: OUTPUT STATISTICAL DISCLOSURE CONTROL, DATA, DATA STEWARDSHIP ORGANISATION

Do Not Track (Protocol)

Do Not Track was an addition to the Hypertext Transfer Protocol (HTTP) that is the basis for **information** transfer on the **World Wide Web**. It was designed to enable **user**s to opt out of being **tracked** by websites that they had visited. It was created in 2009, and incorporated in several Web browsers, including Mozilla Firefox, Google Chrome and Internet Explorer (one version of which even made it a default setting, so that users actively had to opt into tracking, to the chagrin of advertisers).

However, there was no means of enforcement – it merely allowed users to express **privacy preference**s that could be ignored by websites (and users would be unaware of which sites respected their preferences). This lack of legal mandate, together with concerns about user interest and the usability of the **protocol**, meant that it was barely used, and the World Wide Web Consortium disbanded the Do Not Track Working Group in 2019.

Further reading:
Bott, E., 2012. Why Do Not Track is worse than a miserable failure. *ZDNet*, 21 Sep 2012, www.zdnet.com/article/why-do-not-track-is-worse-than-a-miserable-failure/.

See also: PLATFORM FOR PRIVACY PREFERENCES, PROFILING

DOS

See: DENIAL OF SERVICE

Doxxing

Doxxing, sometimes spelt doxing or d0xing, is the practice of releasing someone’s **personal information**, often their real name or address, onto the **Internet** in the clear. This is not only a **breach** of **informational privacy**, but also removes the **anonymity** of people with an online presence, opening them up to real-world retaliatory action, intimidation, or **harassment**.

Further reading:
Douglas, D.M., 2016. Doxing: a conceptual analysis. *Ethics and Information Technology*, 18(3), 199–210, https://doi.org/10.1007/s10676-016-9406-0.

See also: INFORMATION ETHICS, OUTING

DPI

See: DEEP PACKET INSPECTION

DPIA

See: DATA PROTECTION IMPACT ASSESSMENT

DPO

See: DATA PROTECTION OFFICER

D-Privacy

An extension of **differential privacy** to domains other than statistical **databases**, with different metrics of distance.

Further reading:

Chatzikokolakis, K., Andrés, M.E., Bordenabe, N.E. and Palamidessi, C., 2013. Broadening the scope of differential privacy using metrics. *In*: *Proceedings of Privacy Enhancing Technologies: 13th International Symposium, PETS 2013*, Berlin: Springer-Verlag, 82–102, https://doi.org/10.1007/978-3-642-39077-7_5.

DRM

See: DIGITAL RIGHTS MANAGEMENT

DSA

See: DATA SHARING AGREEMENT

Duty of Confidence

In **common law** jurisdictions, a duty of confidence is a legal obligation to respect the **confidentiality** of an individual's **identifiable information**. In English law, the duty began as a kind of **fiduciary duty** under the law of equity. Since the advent of human rights law in the 20th century, however, a duty of confidence is more likely to be determined by the jurisprudence surrounding the right to **privacy** than by reference to equitable principles. As such, it is an obligation which has largely outgrown its equitable origins but retained the original equitable term of 'duty'.

See also: BREACH OF CONFIDENCE, CONFIDENCE, HISTORY OF PRIVACY

Duty to Protect

See: DUTY TO WARN

Duty to Warn

It is a tenet of **common law confidentiality** that **personal information** can be disclosed without **consent** in circumstances of overriding **public interest**. In the United States, the psychiatric profession discusses a specific duty to protect or duty to warn, in which a patient's **privacy** may be outweighed by the **risk** they pose to others to the extent that **disclosure** (e.g., alerting the authorities) is warranted. Whether the risk in question needs to constitute an immediate threat to an **identifiable person**, or can be broader in scope, is a matter of controversy.

Further reading:
Leeman, C.P., 2004. Confidentiality and the duty to warn of possible harm. *The American Journal of Psychiatry*, 161(3), 583, https://doi.org/10.1176/appi.ajp.161.3.583.

See also: CONFLICT OF RIGHTS, HARM

Dyad

The fundamental unit of a **social network**; a pair of **population units** or **data unit**s that exist in some sort of relation with one another. In social network graphs, a dyad is usually represented by a pair of nodes connected by an edge.

In **privacy** terms, dyadic relationships are potentially quite **disclosive** as who an individual knows or spends time with can, in aggregate, say much about them and their preferences.

Dynamic Consent

The scope of a valid **consent** to the reuse of **personal information** is a key debate within **privacy** literature. Some favour broad consent, in which **data subject**s entrust their **information** for multiple potential purposes (e.g., for research in general), on the understanding that other appropriate safeguards will be in place. The EU **GDPR**, however, requires narrow consent, in which each use of the information is clearly and unambiguously specified at the time of giving consent.

Dynamic consent is an **information governance** model which attempts to strike a balance between broad and narrow consent. As new uses of information emerge over time, data subjects are updated and retain the ability to modify their consent preferences. An accessible platform for data subjects to use for granular opt-outs, and a population sufficiently engaged to deliberate the potential uses of their **data**, are key prerequisites for the model to support participants' informational **autonomy**.

Further reading:

Budin-Ljøsne, I., Teare, H.J.A., Kaye, J., et al., 2017. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics* 18(4), article no.4, https://doi.org/10.1186/s12910-016-0162-9.

See also: DATA GOVERNANCE, DATA IN USE, PERSONAL DATA, PRIVACY AS CONTROL

Dynamic Data Situation

A **data situation** where data is being moved from one **data environment** to another. Note that this is distinct from dynamic **data** which implies a constant flow of data (sometimes called a data stream). Although dynamic data will create a dynamic data situation, so will static data which is being moved by some agent.

See also: DATA FLOW, DATA IN TRANSIT

E

E2EE

See: END-TO-END ENCRYPTION

E3

See: ENCRYPT-EVERYTHING-EVERYWHERE

Eavesdropping

See: EAVESDROPPING ATTACK

Eavesdropping Attack

An attack which occurs when an **adversary** intercepts and keeps track of **data** or **communication**s between two parties without either party’s **awareness** or **consent**. The term ‘eavesdropper’ entered the English language in the seventeenth century to denote a person who hung from the eaves of a house to listen to the conversations of others.

Current-day eavesdropping attacks may entail intercepting wireless communications over Wi-Fi or **Bluetooth**. **Packet sniffer**s, **key logger**s and wireless scanners are just a few of the tools and methods that can be used to carry out such attacks. These tools give an **adversary** the ability to record keystrokes or other **user** inputs as they are being entered, as well as to intercept and analyse data packets as they are being transmitted over a **network**. Computer screens also emit radio frequency radiation that sophisticated eavesdroppers might use to reconstruct what a user is looking at. Physical listening devices are also still very much in use; video camera and audio recording devices provide simple, effective ways to eavesdrop. The **Internet of Things**, where **sensor**s are built into material objects, affords the possibility of hybrid digital and non-digital approaches.

An eavesdropping attack typically aims to disclose or access **confidential** or sensitive **information**. **Secure communication** protocols such as HTTPS or TLS, which **encrypt data in transit** and reduce the risk of interception by an adversary, provide some protection against some possible attack vectors, but user vigilance is still essential.

Further reading:

Balakrishnan, S., Wang, P., Bhuyan, A. and Sun, Z., 2019. Modeling and analysis of eavesdropping attacks in 802.11 ad mmWave wireless networks. *IEEE Access*, 7, 70355–70, https://doi.org/10.1109/ACCESS.2019.2919674.

Yu, J., Lu, L., Chen, Y., Zhu, Y. and Kong, L., 2019. An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing. *IEEE Transactions on Mobile Computing*, 20(2), 337–51, https://doi.org/10.1109/TMC.2019.2947468.

See also: SPYWARE, TRACKER, TRANSPORT LAYER SECURITY WIRETAPPING

ECHR

See: EUROPEAN CONVENTION ON HUMAN RIGHTS

E-Commerce

E-commerce is the practice of commercial activity, buying and selling of goods and services, through a digital medium. The **World Wide Web** is the most likely venue for such activity, and special sites have been set up for the purpose. Individual companies have specific websites from which one can view their inventory and purchase products. E-commerce platforms create **searchable** marketplaces, where buyers and sellers are brought together for a commission. E-commerce may be business sales to individual consumer purchasers (business-to-consumer, *B2C*), business-to-business (*B2B*) or, as with large marketplaces such as eBay or Alibaba, consumer-to-consumer (*C2C*). Marketplaces can provide other services, such as recommendations, rating systems and price comparison services.

E-commerce tries to replicate many of the processes used in ordinary offline commerce. Goods are chosen and placed in a *shopping basket*; this means that the **software** uses **cookie**s to **track** which goods have been selected by consumers and stores the list until they are ready to pay. Electronic payment systems are used. Non-electronic goods and services need to be delivered, which has led to an expansion of the freight transport industry, and the growth of distribution warehouses as important real estate and employers.

The result of this activity is that a large amount of **data** can be gathered from **customer tracking**, not only about what they buy and what they are (un)satisfied with, but even which goods they have examined, and for how long. This data can be monetised, to support product optimisation,

personalised services or **targeted advertising**, at some cost to consumers' privacy.

Further reading:

Bandara, R., Fernando, M. and Akter, S., 2020. Privacy concerns in e-commerce: a taxonomy and a future research agenda. *Electronic Markets*, 30(3), 629–47, https://doi.org/10.1007/s12525-019-00375-6.

Laudon, K.C. and Traver, C.G., 2022. *E-commerce 2021–2022: business*, *technology, society*, 17th edition. Harlow: Pearson Education.

See also: SURVEILLANCE CAPITALISM

Economics of Privacy

Privacy has economic effects, and so creates benefits and costs that some have argued should be paid for by individuals. Alternatively, since the practice of privacy involves complex trade-offs, the tools of economics can be used to value the different alternatives, and so enable rational decision-making about the **value of privacy**.

Privacy itself, however, is valued differently by people (and differently across contexts), sometimes benefits individuals and sometimes provides societal benefits, and so we should not expect simple equilibria to emerge. Furthermore, markets for **personal information** are complex and rarely open to all. In particular, **data subject**s are often excluded, making it hard to factor in the value they place on their **personal data** being protected. Because of this, we lack a reference for the value of privacy: should it be the price a data subject would demand to allow access, or the potential **harm** caused by misuse, or an actuarial value based on a **privacy insurance** market, or the price a data subject would pay for privacy protection services?

The most influential argument is that of Posner, based on **information asymmetry**. If privacy is seen as the concealment of **information**, then, since the efficiency of markets is affected by the amount of information that buyers and sellers possess, privacy will tend to give an advantage to the private person. For instance, an applicant for a job is advantaged relative to the employer and to other job applicants if their previous egregious indiscretions are kept private. The privacy of the applicant creates *negative externalitie*s with respect to the employer and other participants in the labour marketplace.

Further reading:

Acquisti, A., Taylor, C. and Wagman, L., 2016. The economics of privacy. *Journal of Economic Literature*, 54(2), 442–92, https://doi.org/10.1257/jel.54.2.442.

Posner, R.A., 1981. The economics of privacy. *American Economic Review*, 71(2), 405–9, www.jstor.org/stable/1815754.

See also: BENEFITS OF PRIVACY, COMMODIFICATION, DATA UTILITY, DIGITAL ECONOMY, NEGATIVE EXTERNALITIES OF DISCLOSED DATA, NEGATIVE EXTERNALITIES OF PRIVACY, VALUE OF DATA

Edge Computing

A computing infrastructure where **data processing** happens closer to the data source, such as a device (i.e., home router) or **user**s, rather than in the **cloud** or on a central server. This allows computation to be performed locally.

Edge computing is used for applications that require real time processing (for example, **Internet of Things**) and in principle might enhance **security** as there is reduced need to transmit sensitive **information** to the cloud, decreasing the risk of interception during transmission reducing data **exposure** to **eavesdropping attack**s. However, with large numbers of users, edge computing might be riskier as **data** is being processed on multiple devices with increased security **risk** from weaknesses in multiple edge systems.

Further reading:

Ranaweera, P., Jurcut, A.D. and Liyanage, M., 2021. Survey on multi-access edge computing security and privacy. *IEEE Communications Survey*s *& Tutorials*, 23(2), 1078–1124, https://doi.org/10.1109/COMST.2021.3062546.

Varghese, B., Wang, N., Barbhuiya, S., Kilpatrick, P. and Nikolopoulos, D.S., 2016. Challenges and opportunities in edge computing. In: *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, 20–6, https://doi.org/10.1109/SmartCloud.2016.18.

See also: DATA IN TRANSIT

EDPB

See: EUROPEAN DATA PROTECTION BOARD

EDPS

See: EUROPEAN DATA PROTECTION SUPERVISOR

EHR

See: ELECTRONIC HEALTH RECORD

Electronic Health Record (EHR)

Electronic health records are a collation of a patient’s medical **information** into a single **longitudinal record**. In principle, the coverage of this record could contain all tests, diagnoses, treatments, medications, medical practitioners’ notes, demographic and lifestyle data, and so on, from the cradle to the grave. In practice, the scope of EHRs tends to be constrained by issues in **data sharing** between various healthcare providers and between the patient and the holder of the EHR.

However, EHRs do represent a more significant innovation than merely permitting the storage of medical information in digital form. Under the common use of the term, electronic health records should provide **data portability** and **interoperability** between healthcare providers. As such, they should form part of an infrastructure in which patients can change providers and have their medical records travel with them. Their interoperable nature also means the records can be combined, scaled and built into a powerful resource for research and analytics into human health conditions.

Further reading:
Keshta, I. and Odeh, A., 2021. Security and privacy of electronic health records: concerns and challenges. *Egyptian Informatics Journal*, 22(2), 177–83, https://doi.org/10.1016/j.eij.2020.07.003.

See also: BIG DATA, HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT, SPECIAL CATEGORY DATA

EM Algorithm

In the context of probabilistic models, the Expectation-Maximization (EM) technique is a statistical tool used to estimate model parameters for incomplete **data**. It can be used for **imputation**, latent variable estimation, **data linkage** and **data synthesis**.

The **algorithm** consists of the E-step (Expectation step) and the M-step (Maximization step). Using the observed data and the most recent estimated model parameters, the algorithm generates expected values for the **missing data** in the E-step. To increase the likelihood of the combined observed and predicted missing data (from the previous step), the algorithm modifies the model parameters in the M-step. It then iterates between the two steps until it reaches an optimum for both data and model estimates.

The EM algorithm is an example of the sophisticated tools available to analysts that allow them to make inferences beyond the data they have available. Although this is useful for legitimate purposes, it also increases the vulnerability of **data subject**s to **disclosure** via **inference attack**s.

Further reading:

Ng, S.K., Krishnan, T. and McLachlan, G.J., 2012. The EM algorithm. *In:* Gentle, J., Härdle, W. and Mori, Y., eds, *Handbook of computational statistics: concepts and methods*, Berlin: Springer, 139–72, https://doi.org/10.1007/978-3-642-21551-3.

See also: MULTIPLE IMPUTATION

Emotion Recognition

The detection and interpretation of someone's emotional state using their facial expressions, speech and other behavioural indicators.

AI vision and **machine learning** algorithms are used to automate emotion recognition. To identify emotional states, the technology uses algorithms to detect and classify changes in facial features. Similarly, speech analysis technology examines vocal tonality and pitch.

Emotion recognition is then used for marketing and advertising to gain **consumer preference information**. There are live discussions over the ethical implications of emotion recognition technologies and concerns about **data security**, privacy and exploitation. As Cowie et al. observe, in human interaction, emotion recognition is part of an implicit secondary **communication** channel, the negotiation of which is part of how we manage our relationships, and also an element of **communication privacy** management and **psychological privacy**. Automation of emotion recognition threatens to subvert these most human of processes.

Further reading:

Dzedzickis, A., Kaklauskas, A. and Bucinskas, V., 2020. Human emotion recognition: review of sensors and methods. *Sensors*, 20(3), 592, https://doi.org/10.3390/s20030592.

Hernandez, J., Lovejoy, J., McDuff, D., Suh, J., O'Brien, T., Sethumadhavan, A., Greene, G., Picard, R. and Czerwinski, M., 2021. Guidelines for assessing and minimizing risks of emotion recognition applications. *In: 2021 9th International conference on affective computing and intelligent interaction,* IEEE, 1–8, https://doi.org/10.1109/ACII52823.2021.9597452.

See also: ARTIFICIAL INTELLIGENCE, FACIAL RECOGNITION TECHNOLOGY, SPEECH RECOGNITION, TARGETED ADVERTISING

Employee Information

An organisation will process the **personal data** of its employees, as well as its customers. When **data subject**s are discussed, concern is usually expressed about consumers, patients or third parties using an organisation's services. However, privacy *within* organisations can be just as important.

The COVID-19 pandemic was an instructive example of the complexities involved, as employers were required to collect more health-related **information** (e.g., lateral flow test results) from their employees to ensure the safety of on-site workers. The legitimacy of this data collection was important to establish, particularly for employers falling under the EU **GDPR**, which frames the employer–employee relationship as involving too great a power imbalance for employee **consent** to be said to be freely obtained.

Further reading:

Suder, S., 2021. Processing employees' personal data during the covid-19 pandemic. *European Labour Law Journal*, 12(3), 322–37, https://doi.org/10.1177/2031952520978994.

See also: DATA CONTROLLER

Encrypt-Everything-Everywhere (E3)

This principle establishes that, to reduce the risk of unauthorised access to a minimum, all **data** should always be encrypted – whether **data at rest**, **data in use** or **data in transit**. Data **security** and **confidentiality** are prioritised over **data utility** in E3.

Further reading:

Chielle, E., Tsoutsos, N.G., Mazonka, O. and Maniatakos, M., 2020. E3X: Encrypt-Everything-Everywhere ISA eXtensions for private computation. *IEEE Transactions on Dependable and Secure Computing*, 19(2), 848–61, https://doi.or/10.1109/TDSC.2020.3007066.

See also: RISK, RISK–UTILITY TRADE OFF

Encryption

Encryption is a process of encoding **data** by converting the original data, referred to as **plaintext**, into an encoded form known as **ciphertext**. Encryption is used to protect data from unauthorised users; the data might be files saved on a storage device (**data at rest**), being transferred over a **network** or over the **Internet** (**data in transit**) or even while the **user** is interacting with the data (**data in use**).

Encrypted data is converted before being stored, sent or processed: the method of carrying out this conversion is referred to as the encryption **algorithm** and the calculations performed for the transformation are called a **cipher**.

Further reading:
Bhanot, R. and Hans, R., 2015. A review and comparative analysis of various encryption algorithms. *International Journal of Security and Its Applications*, 9(4), 289–306, https://doi.org/10.14257/ijsia.2015.9.4.27.

See also: CRYPTOGRAPHIC KEY, CRYPTOGRAPHIC PROTOCOL, DATA STORAGE, DATA USER, NETWORK ENCRYPTION, TRANSPORT LAYER SECURITY

Encryption Algorithm

See: ENCRYPTION

Encryption Key

Most **encryption** uses pseudo-random encryption keys. An authorised recipient of the encrypted **data** will be able decrypt it by simply using the key (provided by the originator), but unauthorised users will not – for any sufficiently robust encryption scheme – be able to convert the encrypted **ciphertext** into **plaintext**, without considerable computational resources and skill.

See also: CRYPTOGRAPHY

Endpoint Security

The practice of safeguarding endpoints, such as laptops, servers, mobile devices, **sensor**s and other devices that connect to a **network**. These endpoints tend to be points of **vulnerability** in systems and are growing in number, particularly with the proliferation of **Internet of Thing**s devices. Endpoint security covers a range of different types of activity, from policies and staff training to systems-based solutions.

Systems-based solutions consist of two distinct approaches: (i) endpoint protection platforms (EPP), systems deployed at the endpoints themselves to prevent **malware** attacks and detect malicious activity; and (ii) endpoint detection and response (EDR) systems, which tend to be server-based and continually monitor endpoints for threat detection, often supported by some automated response capabilities.

Further reading:

Karantzas, G. and Patsakis, C., 2021. An empirical assessment of endpoint detection and response systems against advanced persistent threats attack vectors. *Journal of Cybersecurity and Privacy*, 1(3), 387–421, https://doi.org/10.3390/jcp1030021.

See also: EDGE COMPUTING, NETWORK SECURITY

End-to-End Encryption (E2EE)

End-to-end encryption (E2EE) is a method for encrypting **communication**s to keep them private to the originator and receiver. As such, to qualify as E2EE, no one, including the communication system provider, telecom providers, **Internet** providers or malicious adversaries, can access the **cryptographic key**s needed to be able to decrypt the communication.

As no **third parti**es can decipher the **data** being communicated, companies providing E2EE-based services cannot, for example, hand over **plaintext**s of **user**s' messages to criminal investigations, which has resulted in some concerns being expressed by law enforcement agencies. E2EE does not provide an absolute **security** guarantee. Specifically, it does not secure the end points of the system (the sender and receiver of the data), which are still prone to standard security vulnerabilities.

See also: CRYPTO WARS, ENDPOINT SECURITY

End-User Licence Agreement (EULA)

A standard document specifying the rights and restrictions regarding the use of some digital artefact, often a piece of **software** but sometimes a **dataset**. If the **data** in such datasets include **information** which identifies living **natural person**s, the end-user licence will usually include a clause whereby the **user** agrees to respect the confidentiality of the **data subject**s; for example, from the UK Data Service's EULA: 'To comply with all obligations to preserve the confidentiality of, and not attempt to identify, individuals, households or organisations in the data.'

Although the amount of **confidentiality** protection that such agreements provide is limited, Elliot et al demonstrate that they reduce **risk** compared to releasing data as **open data**.

Further reading:

Elliot, M., Mackey, E., O'Shea, S., Tudor, C. and Spicer, K., 2016. End user licence to open government data? A simulated penetration attack on two social survey datasets. *Journal of Official Statistics*, 32(2), 29–348, https://doi.org/10.1515/jos-2016-0019.

UK Data Service, 2023. *End User Licence Agreement version 11*, https://dam.ukdataservice.ac.uk/media/455131/cd137-enduserlicence.pdf.

See also: DATA IN USE, DATA SHARING AGREEMENT, DATA USER, LICENCE AGREEMENT, SERVICE USER AGREEMENT

Engineering Ethics

The aim of engineering ethics is to ensure that technology is developed and used in a way that promotes common goods, preserves human rights and **dignity** and minimises **harm** and adverse effects on society and the environment. This needs a proactive approach to technical development that considers the wider social and ethical consequences of technological progress. Since technology continues to have a significant influence on **user**s' lives, both personally and collectively, engineering ethics is becoming increasingly important.

Privacy-by-design is an integral part of engineering ethics, requiring engineers to embed privacy principles into the design of new products, systems and technological artefacts.

Further reading:

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E., 2018. AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707, https://doi.org/10.1007/s11023-018-9482-5.

See also: DATA ETHICS, INFORMATION ETHICS, PRIVACY ENGINEERING

Enhanced Privacy ID (EPID)

Digital certificates and remote attestation employ a **security** technology known as the Enhanced Privacy ID (EPID). It is intended to offer a safe and private **authentication** of the **identity** of a platform or device without disclosing any private data. Each device or platform is given a distinct **private key** in addition to a common **public key** shared by the group of devices and platforms. When a device needs to authenticate itself, it generates a **digital signature** using its private key, which the shared public key may then be used to verify.

Multiple applications use EPID, such as **secure communication protocol**s, **digital rights management** systems and remote attestation for **trust**ed computing platforms, for privacy-preserving authentication.

Further reading:
Brickell, E. and Li, J., 2007. Enhanced privacy ID: a direct anonymous attestation scheme with enhanced revocation capabilities. *In: Proceedings of the 2007 ACM workshop on privacy in electronic society*, 21–30. https://doi.org/10.1109/TDSC.2011.63.

See also: DIGITAL IDENTITY

EPID

See: ENHANCED PRIVACY ID

Eprivacy Directive

See: EPRIVACY REGULATION

Eprivacy Regulation

The EU’s ePrivacy Regulation is set to replace the ePrivacy Directive, commonly known as the ‘Cookie Directive’. Adopted by the European Commission in January 2017, it was originally intended to take effect in May 2018, to coincide with the **GDPR** (which also updated a Directive into a Regulation, with uniform effect across the EU).

While the GDPR applies to all processing of **personal data**, this will be supplemented by the ePrivacy Regulation in the context of electronic communications. The updates of the ePrivacy Directive include a broader scope of regulated technologies (including email and text messaging, not just traditional telecoms networks), streamlined **cookie consent** processes and stricter controls for **metadata**, which have become more personally **disclosive** since the Cookie Directive was introduced in 2002.

Further reading:

European Commission, 2023. *Proposal for an e-Privacy Regulation*, https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation.

González, E.G., De Hert, P. and Papakonstantinou, V., 2020. The proposed ePrivacy Regulation: the Commission’s and the Parliament’s drafts at a crossroads? *In:* Hallinan, D., Leenes, R., Gutwirth, S. and De Hert, P. eds, *Data protection and privacy: data protection and democracy.* Oxford: Hart Publishing, 267–98.

See also: DATA PROCESSING, DATA PROTECTION

Epsilon

Epsilon or ε is the key parameter of **differential privacy**. It is a measure of the maximum distance between a query on a **database** and the same query on another database which differs from the first by a single entry (for **data subject** *a*). Epsilon is a measure of the maximum amount of **information** about *a* that is leaked by the contribution of data subject *a* to the database. Epsilon is also referred to by proponents of the approach as ‘privacy loss’.

Further reading:

Dwork, C., Kohli, N. and Mulligan, D., 2019. Differential privacy in practice: expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), https://doi.org/10.29012/jpc.689.

Dwork, C. and Roth, A., 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407, http://dx.doi.org/10.1561/0400000042.

See also: PRIVACY BUDGET

Equivalence Class

A set of **record**s in a **dataset** which contain identical values for a given set of variables. In ***k*-anonymity**, *k* is the minimum class size for a **key variable** set.

Equivalence Class Structure

A frequency table of the counts of **equivalence class** sizes for a given set of **key variable**s. This table is used in some **disclosure risk** assessment measures. For example, in the **Data Intrusion Simulation** method, Skinner and Elliot use the counts of uniques and pairs to calculate a risk measure based on the provability of successful **data linkage**. Elamir and Skinner extend this to include triples to account for measurement error.

Further reading:

Elamir, E.A. and Skinner, C.J., 2003. Modelling the re-identification risk per record in microdata. *In: 54th session of the International Statistical Institute*, 13–20. www.researchgate.net/publication/2944660_Modeling_the_Re-identification_Risk_per_Record_in_Microdata.

Skinner, C.J. and Elliot, M.J., 2002. A measure of disclosure risk for microdata. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 855–67, https://doi.org/10.1111/1467-9868.00365.

See also: DATA IN USE, DISCLOSURE, FREQUENCY DATA, RISK ASSESSMENT, STATISTICAL DISCLOSURE

Erasure

The EU **GDPR** gives **data subject**s a right to the erasure of their **personal data** (subject to some exceptions). This is otherwise known as a **right to be forgotten**, following the judgment of the Court of Justice of the European Union in the *Google Spain* case. In this case, ‘erasure’ meant de-referencing from **Internet search engine** results, but erasure generally means rendering the **information** unusable by the **data controller** or any other entity through physical **data destruction** or assured data deletion. The UK Information Commissioner’s Office specifies that erased data should be removed from live systems and rendered beyond use on backup systems even if it cannot be immediately overwritten.

A right to erasure of personal data was first articulated in the **Council of Europe’**s 1973 *Resolution on the Protection of Privacy of Individuals vis-à-vis Electronic Data Banks in the Private Sector*, although this was limited to obsolete or unlawfully obtained data. Article 17 GDPR sets out a list of circumstances in which personal data which should be erased – including when **consent** is withdrawn, or when there is no overriding **legitimate interest** to justify **retention**.

In many cases erasure requires that **data** go through a *data destruction* process. Specifications of data destruction processes are often written into **data sharing agreement**s.

Further reading:

Aidinlis, S., 2020. The right to be forgotten as a fundamental right in the UK after Brexit. *Communications Law*, 25(2), 67–78, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3554625.

Ausloos, J., 2020. *The right to erasure in EU data protection law.* Oxford: Oxford University Press.

Information Commissioner's Office, n.d. *Right to erasure*, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/.

Sharma, S., 2019. Data subjects' rights. *In:* Sharma, S., ed., *Data privacy and GDPR handbook.* Hoboken: John Wiley & Sons Inc, 193–232.

See also: DATA PROTECTION, DATA RETENTION, DELETION, INFORMATIONAL SELF-DETERMINATION, INFORMATIONAL PRIVACY, PRIVACY AS CONTROL, RIGHT TO DELETION

Escrow

Escrow is the holding of an item involved in a transaction by a **trusted third party**, until the transaction is completed. For instance, a purchaser may place the payment in escrow; once the seller has verified that the payment has been made, they can release the goods to the purchaser. Once the goods have been received, the third party is notified and hands the payment to the seller.

Escrow facilitates **trust** in many types of interaction. Perhaps the most important in the privacy field is *key escrow*, where the **private key** needed to decrypt a piece of **information** is held in escrow for third parties (including government and law enforcement). However, many experts are concerned that key escrow creates a **vulnerability** in otherwise **secure** systems.

Further reading:

Abelson, H., Anderson, R., Bellovin, S.M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Green, M., Landau, S., Neumann, P.G., Rivest, R.L., Schiller, J.I., Schneier, B., Specter, M.A. and Weitzner, D.J., 2015. Keys under doormats: mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity*, 1(1), 69–79, https://doi.org/10.1093/cybsec/tyv009.

See also: CRYPTO WARS, MANDATORY DECRYPTION, PUBLIC-KEY INFRASTRUCTURE

Ethical Hacking

Ethical hacking is the practice of using the techniques of malicious hackers to test the **security** of computer systems, by exposing vulnerabilities. Ethical hackers usually have the permission of the system managers, but even if not, they will inform managers, rather than making the vulnerability **public** or exploiting it for gain.

Further reading:
Palmer, C.C., 2001. Ethical hacking. *IBM Systems Journal*, 40(3), 769–80, https://doi.org/10.1147/sj.403.0769.

See also: MOTIVATED INTRUDER TEST, PENETRATION TEST, RED TEAM, VULNERABILITY, WHITE HAT ATTACK

Ethics

See: CODE OF ETHICS, DATA ETHICS, ENGINEERING ETHICS, ETHICAL HACKING, ETHICS COMMITTEE, INFORMATION ETHICS, NEUROETHICS

Ethics Committee

Ethics committees as a formal mechanism for ethical oversight of research started to appear in the 1960s, with the first often cited as being located at the UK's Porton Down military research facility. The scope of ethics committees' work is usually the wellbeing of human and animal research participants and the handling of **personal data** (including justification for **data processing**, **security**, **data minimisation** and **anonymisation**). They may also cover issues of **public interest**.

The EU **GDPR** requires **data controller**s to evidence **appropriate technical and organisational measures** to protect personal data. In the context of research use of personal data, oversight by an ethics committee is a common example of an organisational measure.

However, their connection to **data protection** law does not always materialise into a smooth institutional interface with **information governance** teams, which can have a different focus.

Further reading:

Gefenas, E., Lekstutiene, J., Lukaseviciene, V., Hartlev, M., Mourby, M. and Cathaoir, K.Ó., 2022. Controversies between regulations of research ethics and protection of personal data: informed consent at a crossroad. *Medicine, Health Care, and Philosophy*, 25(1), 23–30, https://doi.org/10.1007/s11019-021-10060-1.

Schmidt, U., 2019. Creating a 'Father Confessor': the origins of research ethics committees in UK military medical research, 1950–1970. Part I, context and causes. *BMJ Military Health*, 165, 284–90, http://dx.doi.org/10.1136/jramc-2019-001206.

See also: CONSENT, DATA ETHICS, DATA GOVERNANCE, INFORMED CONSENT

EULA

See: END-USER LICENCE AGREEMENT

European Convention on Human Rights (ECHR)

A key touchstone for the right to privacy in Europe and beyond is the Council of Europe's European Convention on Human Rights, of which Article 8 expresses a right to private and family life. Distinct from the European Union, the Council of Europe contains most countries on the European continent, including Turkey, and some post-Soviet states, such as Armenia and Azerbaijan (Russia was expelled in 2022). Its articles apply in all signatory countries.

Citizens in signatory states can challenge their governments in the European Court of Human Rights (ECtHR) in Strasbourg, when they believe their human rights (as defined in ECHR) have been breached. At the time of writing, the European Union is currently negotiating accession to the ECHR; if successful, this will mean that EU bodies (such as the European Commission) can also be sued in the ECtHR for **breach**es of human rights.

The ECHR must be understood in its original context: the international reaffirmation of human rights following the atrocities of the Second World War, and the consequent desire to regulate state power over individuals. Following the 1948 United Nations Declaration on human rights, the

ECHR was signed in 1950 as a European equivalent, with the authority of the ECtHR to enforce its provisions.

The right to privacy under Article 8 ECHR was further articulated by the Council of Europe's 1970 Declaration on Mass Communication Media and Human Rights. Its influence was extended by the 1981 **Convention 108** for the Protection of Individuals regarding Automatic Processing of Personal Data, which defines **data protection** as the 'right to privacy with regard to automatic processing or personal data'. The ECHR can therefore be seen as having initiated the creation of legally binding international instruments in data protection law.

Further reading:

Greenleaf, G., 2012. The influence of European data privacy standards outside Europe: implications for globalization of Convention 108. *International Data Privacy Law*, 2(2), 68–92, https://doi.org/10.1093/idpl/ips006.

Santolaya, P., 2012. The right to private life (notably extended right to privacy) (art 8 ECHR). *In*: Roca, J.G. and Pablo, S. eds, *Europe of rights: a compendium on the European Convention of Human Rights*. Leiden: Brill, 337–51.

See also: CHARTER RIGHTS, RIGHT TO DATA PROTECTION

European Data Protection Board (EDPB)

When the EU **GDPR** entered into force in May 2018, the **Article 29 Working Party (A29WP)** of **data protection** regulators was replaced with the European Data Protection Board. The EDPB is similar to its predecessor but has been given a greater role in ensuring **consistency** and **cooperation** within GDPR enforcement across the EU.

Like the A29WP, the EDPB is made up of representatives from the national data protection regulators (Supervisory Authorities) in each member state. It issues guidance on specific topics relating to GDPR **compliance** and coordinates regulatory responses between Supervisory Authorities when **cross-border data processing** affects multiple countries within the EU.

Further reading:

Janciute, L., 2020. European Data Protection Board: a nascent EU agency or an 'intergovernmental club'? *International Data Privacy Law*, 10(1), 57–75, https://doi.org/10.1093/idpl/ipz021.

See also: SUPERVISORY AUTHORITY

European Data Protection Supervisor (EDPS)

The role of the European Data Protection Supervisor was created by the EU Regulation 45/2001 on the processing of **personal data** by Community Institutions. The EDPS supervises the **GDPR compliance** of EU bodies, which are not regulated by national authorities and thus require separate means of oversight. The office of the EDPS was established for this purpose.

The EDPS also issues guidance on **data protection** law matters, which can supplement those of the **European Data Protection Board** (EDPB). As the EDPS is a single agency, rather than a committee of multiple national **regulator**s (like the EDPB), its guidance can be produced more expeditiously – the EDPB can take years to negotiate its draft guidance. The EDPS has also collaborated with the Spanish data protection agency to publish guidance on **biometric** identification, indicating a political freedom to join forces where perspectives align.

Further reading:
Busch, C., Czajka, A., Deravi, F., et al., 2022. A response to the European Data Protection Supervisor ‘misunderstandings in biometrics’ by the European Association for Biometrics. *IET Biometrics* 11(1), 79–86, https://doi.org/10.1049/bme2.12057.

See also: SUPERVISORY AUTHORITY

Exfiltration

The unauthorised extraction of **data** from a computer system or **network**.

See also: HACKING, NETWORK

Expectation-Maximization Algorithm

See: EM ALGORITHM

Explainable AI (XAI)

Explainable Artificial Intelligence is a branch of **Artificial Intelligence** (AI) that focuses on creating artificial intelligence systems that can explain their decisions and reasoning processes in human terms. The goal of XAI is to make AI **algorithms** more **transparent** and understandable to humans, so they can be used more **safely** and ethically. This has become more salient in the context of powerful **machine learning** methods such as **deep learning** and **generative AI**, which present as black boxes. XAI is in demand in many fields, such as medicine, law and finance, where it is important that the humans who use AI algorithms understand their outputs.

Further reading:

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G. and Ranjan, R., 2023. Explainable AI (XAI): core ideas, techniques, and solutions. *ACM Computing Survey*s, 55(9), 1–33, https://doi.org/10.1145/3561048.

O'Hara, K., 2020. Explainable AI and the philosophy and practice of explanation. *Computer Law and Security Review*, 39, 105474, https://doi.org/10.1016/j.clsr.2020.105474.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D. and Zhu, J., 2019. Explainable AI: a brief survey on history, research areas, approaches and challenges. *In: Natural Language Processing and Chinese Computing: 8th CCF International Conference, Proceedings*, *Part II 8*, Cham: Springer, 563–74, https://doi.org/10.1007/978-3-030-32236-6_51.

Explanatory Variable

In statistical models a variable that is associated with the outcome of interest. Explanatory variables are also known as 'predictor variables' or 'X-variables', and in experimental contexts 'independent variables'.

If the statistical model is good enough (usually if enough explanatory variables are available and the **dataset** is large and of high quality) the value of the **response variable** might be disclosed about specific **population unit**s. In such circumstances the explanatory variables have effectively become model-based **key variable**s.

See *also*: PREDICTIVE MODELLING

Explicit Consent

See: EXPRESS CONSENT

Exposure

In his 2006 taxonomy of privacy, Daniel Solove draws a distinction between **disclosure** and exposure. He argues that both involve the dissemination of true **information**, but disclosed information is sufficiently novel to inform an assessment about an individual. Exposure, within his taxonomy, does not reveal new information about a **person**, but instead uncovers the **sensitive**, primordial aspects of their existence: for example, their nudity, sexuality, urination, defecation, trauma or injury. While it is not usually a surprise that an individual has a body, or defecates, the revelation of this aspect of self can nonetheless be shameful and humiliating for the person in question, as a collapse of the boundaries which permit dignified, civilised selves to participate in **public** life.

Moreham has endorsed this characterisation but lent it additional nuance through considerations of **consent** (e.g., the contrast between images of a streaker vs an unconscious patient), as well as the cultural contingency of what is considered exposure. Varying attitudes about the need to cover head hair, as well as the male/female torso, highlight the fluid boundaries of our secret, primordial selves.

Further reading:

Moreham, N.A., 2018. Unpacking the reasonable expectation of privacy test. *Law Quarterly Review*, 134, 651–74, https://search.informit.org/doi/abs/10.3316/agispt.20190213006682.

Solove, D.J., 2006. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477–564, https://heinonline.org/HOL/LandingPage?handle=hein.journals/pnlr154&div=20&id=&page=.

See also: BODILY PRIVACY, CULTURAL VARIATION OF PRIVACY, REASONABLE EXPECTATION OF PRIVACY, REVENGE PORN, SECRET, VEIL, VOYEURISM

Express Consent

A form of **consent** where an individual explicitly gives permission for an action by another. Express consent can either be verbal or written. Express consent often goes hand in hand with unambiguous and **informed consent** (requirements for valid consent under the EU **GDPR**) but is distinct from any legal term and can be used in a more informal sense to connote explicitness. The GDPR, on the other hand, refers to explicit consent as a condition for processing **special category personal data**; while this could also be given verbally, it must be documented for

accountability purposes (e.g., through an audio recording, or attested in writing by a witness).

See also: CONSENT FORM, OPT-IN, UNAMBIGUOUS CONSENT

Extranet

An extranet is a private **network** that makes use of **Internet** technology to give authorised external **user**s safe and regulated access. It expands an organisation's **Intranet** beyond its walls so that outside users may access internal resources otherwise only accessible by internal system users. Extranets usually use **authentication** checks before allowing users access, and they may also employ **encryption** and other **security** measures to safeguard **data** and resources. They can provide a variety of advantages, including enhanced cooperation, coordination and effectiveness between organisations and their outside partners, as well as more control over and visibility of **information** and resources.

See also: ACCESS CONTROL, NETWORK ENCRYPTION, NETWORK SECURITY

Extrinsic Privacy

See: OBTRUSION

F

Face Recognition

See: FACIAL RECOGNITION TECHNOLOGY

Facial Recognition Technology

Various types of **software** exist which can assess whether two facial images correspond to the same person, and in doing so **identify** individuals whose faces have been captured by (for example) **CCTV**. The **privacy concern**s associated with this technology have intensified as it has become more accurate and accessible.

Facial recognition poses significant **privacy risk**s, including:

- **Surveillance**: facial recognition can be used to **track** and monitor people's movements and activities, both in **public** and in private spaces.
- **Biometric data** collection: facial recognition relies on the collection and storage of biometric data, such as images of faces. This data is **sensitive** and could be used for **identity theft** or other malicious purposes if it falls into the wrong hands.
- False identifications: facial recognition **algorithm**s are not perfect and can produce **false positive**s.
- Bias: facial recognition algorithms have been found to be less accurate for certain groups of people, such as people with darker skin tones or those with certain facial features. This can lead to discrimination and further marginalisation of already marginalised groups.

Given the above risks, there have been (successful) calls to ban the use of facial recognition by law enforcement agencies pending stricter regulation. While many **jurisdiction**s do not have laws specifically governing facial recognition technologies, **data processing** of facial data is covered by **data protection** laws in the UK and EU. Images of faces have been termed 'facial biometric data' by the English High Court in *Bridges v South Wales Police* and treated as **information** which constitutes a direct identification (i.e., a **direct identifier**). This means facial images are treated as **personal data**, and protected by data protection law, even if they are not linked to any other information.

Mitigations can include regulation, **transparency**, and user education about how the technology works and how **data** is being used.

Further reading:

Mourby, M. and Mackey, E., 2023. Pseudonyms, profiles and identity in the digital environment. *In*: van Der Sloot, B. and van Schendel, S., eds, *The boundaries of data: technical, practical and regulatory perspectives*. Amsterdam: Amsterdam University Press.

Roussi, A., 2020. Resisting the rise of facial recognition. *Nature*, 587(7834), 350–4. https://doi.org/10.1038/d41586-020-03188-2.

See also: CHILLING EFFECT, DATA STORAGE

FAIR

The FAIR principles for scientific **data** management require **information** to be Findable, Accessible, **Interoperable** and Reusable. They were first published by Wilkinson and colleagues in 2016, with an emphasis on automating the discovery and use of data, to facilitate researcher access to larger volumes of previously disparate data.

The FAIR principles are thought to be a new cornerstone for more systematic **data sharing** for research. Boeckhout and colleagues argue that they offer a middle ground between **open data** for scientific research and privacy protections for **data subject**s. Applied responsibly, they can enable researchers to find data and ascertain the nature of its content, with access to individually **identifying** information still regulated by the **data controller**.

Further reading:

Boeckhout, M., Zielhuis, G.A. and Bredenoord, A.L., 2018. The FAIR guiding principles for data stewardship: fair enough? *European Journal of Human Genetic*s, 26(7), 931–6, https://doi.org/10.1038/s41431-018-0160-0.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, https://doi.org/10.1038/sdata.2016.18.

See also: BIG DATA, DATA IN USE, IDENTIFIABLE INDIVIDUAL, OPEN ACCESS

Fair Information Practice Principles (FIPPS)

Fair **Information** Practice was brought into the political arena via a 1973 report, *Records, Computers and the Rights of Citizens*, by Willis Ware, and elaborated in a series of reports about **data protection** and **fair** practice, for example by the OECD in 1980 and the **Council of Europe** in 1981. In 1998, the US **Federal Trade Commission (FTC)** enumerated the Fair Information Practice Principles (FIPPs) of *notice*, *choice*, *access*, *security* and *enforcement*.

- Notice/**awareness**: customers should be given notice of a company’s information practices before **personal information** is collected, including potential recipients of the information and the uses to which it will be put.
- Choice/**consent**: customers should have an option to opt into or opt out of (consent to) the information use.
- Access/participation: customers must be able to access information held about them, verify it and contest its **accuracy**.
- **Integrity/security**: those holding information should ensure it is securely held and accurate.
- Enforcement/redress: potential enforcement mechanisms for the FIPPs include self-regulation by those holding information, private **remedies** from civil actions in the event of **harm**s, and undetermined civil and criminal penalties to be levied by the US government.

Further reading:

Landesberg, M.K., Levin, T.M., Curtin, C.G. and Lev, O., 1998. *Privacy online: a report to Congress*. Federal Trade Commission, www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf.

Ware, W.H., 1973. *Records, computers, and the rights of citizens: report of the Secretary’s Advisory Committee on Automated Personal Data Systems*, US Department of Health, Education and Welfare Publication no.(OS)73-94, www.justice.gov/opcl/docs/rec-com-rights.pdf.

See also: DATA PROTECTION PRINCIPLES, INFORMATION SECURITY

Fairness

Fairness has been a longstanding consideration in the **common law** of **confidence**, where a possessor of **information** is bound by **confidentiality** if it should fairly and reasonably apply. More recently, fairness has become part of the first principle of the EU **GDPR**: that **personal data** should be processed lawfully, fairly and in a **transparent** manner. Some of the Recitals (e.g., 60 and 71) indicate that a particular concern from the perspective of fairness is the use of potentially inaccurate information for **profiling** individuals, emphasising the connection between accurate representation and issues of justice.

This is not entirely new. Before **big data**-generated profiling began to open individuals up to automated injustice (e.g., ill-founded elimination from early rounds of recruitment, or poor credit scores), unjust detriment through **reputational** damage was a core concern of the law of **defamation**. Fair comment has thus long been a defence in **libel** or **slander** suits. Fairness can relate to the veracity of any stated facts or the soundness of any **inference** drawn from them. Fairness now needs to regulate not only the publicly stated comments of an individual, but a whole ecosystem of **data broker**age and automated profiling. In a digital context, assessing the soundness of inferences about a person is not necessarily an evaluation of human reasoning; it can also involve **scrutiny** of the statistical methodology behind the probabilistic connection between a data point and an automated prediction.

Such technocratic evaluation may seem far removed from fairness as an ethical principle, or the anthropologically observed human (and sometimes non-human) need for equitable treatment within a social group. The principle, however, is fundamentally the same: that detriment should not be inflicted without justification. It is rather how we secure this outcome in a digital environment that must change – particularly as the **algorithm**s used to make decisions about people become increasingly complex. This is another reason why regulation of automated decision-making is a subject of sufficient anxiety for privacy scholars and practitioners.

Further reading:

Gil González, E. and de Hert, P., 2019. Understanding the legal provisions that allow processing and profiling of personal data – an analysis of GDPR provisions and principles. *ERA-Forum*, 19(4), 597–621, https://doi.org/10.1007/s12027-018-0546-z.

Fair Processing Notice

See: PRIVACY NOTICE

Fake Profile

A fake profile on a **social media** account is one that misrepresents the owner of the account, either **impersonating** another person or creating an entirely false **identity**. While some fake profiles are intended to protect the

privacy of an otherwise genuine person (for example, on a dating site or a political chatroom), they are more likely to be created to dupe others into believing the agent they are dealing with (a person or a **bot**) is real (and then are sometimes called **sock puppet**s). This can facilitate several types of misleading behaviour, such as spreading misinformation, **phishing**, trolling and **harassment**, extortion or damaging (or improving) the **reputation**s of individuals and brands by leaving false reviews. It is estimated that fake profiles account for a large minority of online profiles, and **social networking** platforms conduct a good deal of research into methods for discovering and banning them.

Further reading:
Ramalingam, D. and Chinnaiah, V., 2018. Fake profile detection techniques in large-scale online social networks: a comprehensive review. *Computers and Electrical Engineering*, 65, 165–77, https://doi.org/10.1016/j.compeleceng.2017.05.020.

See also: DEFAMATION, IDENTITY THEFT, OBFUSCATION, REPUTATION MANAGEMENT

False Light

Publicity which presents a plaintiff to the **public** in a false light is the third of William Prosser's four **privacy tort**s. In an influential paper of 1960, Prosser argued against the salience of the **right to be let alone**, traced in US law by Warren and Brandeis. He claimed instead that the privacy torts that existed in law did not furnish a broader principle of integrated coverage of a right to be let alone but were instead a set of four discrete and discontinuous protections.

False light is the public declaration of falsehoods or misleading truths about the plaintiff. The defendant must have known the claims were false, and a reasonable person of ordinary sensibilities would find the plaintiff's association with the claims offensive. Where the plaintiff is a **public figure**, the tort also requires actual malice on the part of the defendant to be proven. The tort provides compensation for hurt feelings, and so only applies to people, and not to corporations. It is close to **defamation**, and some US courts have refused to recognise false light as a separate tort. However, one difference between it and defamation is that false light can be found even when the statements were true, if they were offensive and misleading (for example, using someone's undoctored photograph without **consent** to illustrate an article discussing bad behaviour, implying that the person in the photograph was behaving in that way).

Further reading:
Prosser, W.L., 1960. Privacy. *California Law Review*, 48, 383–423.
Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220.

See also: LIBEL, SLANDER, RIGHT TO PRIVACY

False Negative

A false negative is an error that occurs when a test or classification system incorrectly identifies an entity as not belonging to a class or as not present, when in fact it does belong/is present. For example, in medicine, a false negative might occur when a test for a disease is negative, but the patient does have the disease. In **machine learning**, a false negative is an outcome where the model incorrectly predicts the non-membership of a class to which the instance does in fact belong.

In a privacy context, false negatives can affect both sides of the **security** divide. For example, in **reidentification attack**s an **adversary** may decide that a **record** does not refer to a particular individual when in fact it does. Conversely, in a security system, a false negative might occur when the system does not detect a genuine **intrusion**.

See also: ACCURACY, FALSE POSITIVE

False Positive

A false positive is an error that occurs when a test or classification system incorrectly identifies an entity as belonging to a class or as present, when in fact it does not belong or is not present. For example, in medicine, a false positive might occur when a test for a disease is positive, but the patient does not actually have the disease. False positives can have negative consequences, as they can lead to unnecessary medical treatment or unnecessary safety interventions. In **machine learning**, a false positive is an outcome where the model incorrectly predicts membership of a class.

In a privacy context, false positives can affect both sides of the **security** divide. For example, in **reidentification attack**s an **adversary** may decide that a record refers to a particular individual when in fact it does not. Conversely, in a security system, a false positive might occur when the system detects a hazard that does not exist, such as a security alarm that goes off for no reason.

See also: ACCURACY, FALSE NEGATIVE

Family Resemblance Theory of Meaning

The family resemblance theory of meaning was developed by Ludwig Wittgenstein. According to the theory, while we might assume that all things that we refer to with the same term must have something in common, in many (perhaps most, or even all) cases this is not so. Rather than any common feature, the things may be linked by overlapping similarities, and connected usages (including metaphorical ones) of the term. In particular, the term does not refer to any underlying abstract concept, but rather is marked by recognised similarity of use.

Daniel Solove proposed that ‘**privacy**’ is such a family resemblance term, to explain its many different usages and disagreements over its definition. O’Hara, while agreeing that it is a family resemblance term, argued that such a diagnosis failed to explain the disagreement. Family resemblance explains agreement in the absence of an obviously common feature but, O’Hara points out, it cannot explain disagreement.

Further reading:

O’Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Solove, D.J., 2008. *Understanding privacy*. Cambridge, MA: Harvard University Press.

Wittgenstein, L., 1953. *Philosophical investigation*s. Oxford: Blackwells.

See also: PRIVACY, CULTURAL VARIATION OF

FARAS

See: FULLY AUTOMATED REMOTE ANALYSIS SYSTEM

Feature

Also referred to as an **attribute**, a feature is a measure or classification that is used as input to a **machine learning algorithm** or other model. Features have a numerical or symbolic value and may be **continuous** or **categorical**. A machine learning model is trained using features to classify based on the values of its attributes.

In machine learning, an essential step is to select the appropriate features for a given context. Poorly chosen features can result in models that are unnecessarily complicated or underfit to the **data**, whereas well-chosen features can result in more accurate and effective ML models.

Federal Trade Commission (FTC)

Originally created in 1914 as an antitrust agency, the Federal Trade Commission has become the principal enforcer of the multiple **privacy** and consumer protection laws in the United States of America.

The FTC brought some of the first **Internet** privacy enforcement actions in the mid 1990s and is able to fill some of the gaps between the various sectoral **information** laws in the US. However, its main **regulatory** focus and experience is on enforcing privacy promises made to consumers via a **privacy policy**, and the resulting expectations created. For instance, it championed a **Do Not Track** initiative in 2011, which failed to become law despite an attempted revival in 2019. As such, it does not operate across the same breadth of **data protection** frameworks as the EU Supervisory Authorities.

Further reading:
Craig, T. and Ludloff, M., 2011. *Privacy and big data*. Sebastopol: O'Reilly Media.
Hoofnagle, C.J., 2016. *Federal Trade Commission privacy law and policy.* Cambridge: Cambridge University Press.

See also: US PRIVACY LAWS, DATA PROTECTION AUTHORITY, SUPERVISORY AUTHORITY

Federated Identity

This term describes the use of a single **digital identity** to log into different **application**s and online services owned by different organisations. Using a federated identity, **user**s can access services from various providers without having to log in over and again or create additional accounts, by first **authenticating** with their first **identity provider**. To facilitate **identity** and **attribute** sharing across several domains, federated identity depends on industry-**standard protocol**s and technologies like the **Security Assertion Markup Language (SAML)** and OpenID Connect. It provides some advantages, including a better user experience, easier **identity management** and **security** and privacy features.

There is a trade-off also in terms of **privacy risk**; a single compromised set of federated **credentials** can grant an **adversary** access to multiple applications, risking more significant **data breach**es and significantly heightening the risk of **identity theft**. It also requires that the user **trust**s the organisation providing the federated identity service. Against the risk, the user no longer requires multiple identities across different services, while reusing the same credentials for different services helps with the psychological load of managing multiple accounts.

Further reading:
Shim, S., Geetanjali, B. and Vishnu, P., 2005. Federated identity management. *Computer*, 38(12), 120–2, https://doi.org/10.1109/MC.2005.408.

Federated Learning

A **machine learning** methodology that enables numerous devices jointly to train a machine learning model without sharing their **data**. In federated learning, the model is trained locally on each device and the modified parameters are then transmitted to a central server, where they are combined to form a new model that captures the **information** distributed across all the devices.

By minimising the amount of **data sharing**, federated learning reduces **privacy risk** and risks of unauthorised access to data, since the raw data remains distributed and is never transferred to the central server. Some derived data such as model parameters and possibly descriptive statistics will be transferred both to the server and in some use cases to each client as well. This transferred **information** does still carry some **disclosure risk** and so to preserve **confidentiality** during the federated learning process, secondary privacy and security measures may be necessary. For example, some federated learning systems employ **differential privacy** on transferred data and **encryption** for protecting the **data in transit** between the client devices and the server.

Further reading:
Li, T., Sahu, A.K., Talwalkar, A., & Smith, V., 2020. Federated learning: challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60, https://doi.org/10.1109/MSP.2020.2975749.

See also: INFORMATION SECURITY, INTEGRITY

Feminist Critique of Privacy

Within the liberal tradition, **privacy** is seen as an essential protection for individual **autonomy**. However, historically, the autonomy of some individuals was seen as more valuable than that of others. In the 19th-century liberalism of John Stuart Mill, gendered assumptions meant that privacy was conceived as protecting households, with an unstated implication that the (usually male) head of household remained sovereign within it (an understanding of the household economy that dated back to Aristotle). As liberalism moved toward a more individualistic framework post-Mill, the role of individual autonomy within the private space of the household was not resolved. For instance, in Article 8 of the **European Convention on Human Rights**, privacy of family life is explicitly protected, even though the right applies to individuals and not to family groups.

The feminist critique therefore framed privacy as a means of obscuring domestic inequalities. MacKinnon argued that this made the household an unscrutinised space within which the model of the free, autonomous individual was illusory. Rössler demanded a post-Millian liberalism by identifying the Millian account in particular as fundamentally contradictory, because its moral component leaves existing norms of privacy (specifically in the household) untouched, despite their being discriminatory against women and detrimental to their autonomy.

Further reading:

MacKinnon, C.A., 1987. Privacy v. equality: beyond Roe v. Wade. *In:* MacKinnon, C.A., *Feminism unmodified: discourses on life and law*. Cambridge, MA: Harvard University Press, 93–102.

Rössler, B., 2005. *The value of privacy*. Cambridge: Polity Press.

See also: HISTORY OF PRIVACY, INTIMACY, PRIVATE SPHERE

FHE

See: FULLY HOMOMORPHIC ENCRYPTION

Fiduciary Duty

A *fiduciary* is a person placed in a position of **trust** and **confidence** towards someone else (a beneficiary) which gives rise to duties under the law of equity. The overarching duty of a fiduciary is to act in the best interests of the beneficiary, without pursuing their own interests even as a secondary matter. The duties of a fiduciary to serve the best interests of the beneficiaries are collectively termed fiduciary duties. Trustees and company directors are common examples of fiduciaries.

While a fiduciary *can* be a **data controller**, if they process **personal data**, most data controllers are *not* fiduciaries, and perform **data processing** for purposes other than the best interests of the data subjects. However, **data**

trusts have been proposed as a means of introducing fiduciary duties into personal data management. These would be formally constituted legal trusts, managing the personal data of the beneficiaries (i.e., data subjects) to serve their best interests. The controllers' fiduciary duties would include a duty of **confidentiality**, a duty of care to keep the data secure, and a duty of loyalty, for example avoiding conflicts of interest and respecting terms of **consent**.

The potential to introduce fiduciary duties as a safeguard of **informational privacy** has its proponents but also its sceptics, with little evidence of implementation in practice.

Further reading:

Balkin, J.M., 2020–1. The fiduciary model of privacy. *Harvard Law Review Forum*, 134(1), 11–33, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3700087.

Delacroix, S. and Lawrence, N.D., 2019. Bottom-up data trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, 9(4), 236–52, https://doi.org/10.1093/idpl/ipz014.

See also: DATA GOVERNANCE, DUTY OF CONFIDENCE

Filing System

The **GDPR** defines a filing system as a structured set of **personal data**. **Data protection** legislation, such as the GDPR, mostly applies to digital **information**. However, non-digital **data** is covered by the GDPR if it is contained in a filing system. Unstructured handwritten notes do not fall within the scope of the legislation, but if they are (or are intended to be) organised according to a specific criterion (such as name), then the information is in the **material scope** of the law.

Financial Privacy

Financial privacy refers to the privacy that attaches to the details of the financial affairs of individuals, for example the restriction of access of outsiders to **information** about an individual's income, their bank account details or their financial arrangements with respect to capital, savings, payments to others or dispositions of inheritance. The opposite of financial privacy is sometimes called financial **transparency**.

Financial information is informally regarded as sensitive by many. In the US, financial information is protected by the 1999 Gramm–Leach–Bliley

Act and others (including laws in some states as well). Banks, insurance firms and securities firms must follow the Financial Privacy Rule and the Safeguards Rule, which together govern how financial information is collected, stored and disclosed. However, in the EU, financial data is not defined as **special category data** in **GDPR**, unlike other **data** perceived to be **sensitive**.

Financial privacy is aided or restricted by the forms of **currency**. Cash is totally **anonymous** (although high-denomination paper notes can be laboriously traced), while most **cryptocurrency** affords strong measures of anonymity. Bank and credit accounts are traceable, and non-cryptocurrency digital payments leave data trails. Many central banks are experimenting with digital currencies, and these will, in the absence of privacy-preserving measures, centralise information about all transactions.

Further reading:

Berg, C., 2018. *The classical liberal case for privacy in a world of surveillance and technological change*. Cham: Palgrave Macmillan.

Johannesen, N. and Zucman, G., 2014. The end of bank secrecy? An evaluation of the G20 tax haven crackdown. *American Economic Journal: Economic Policy*, 6(1), 65–91, https://doi.org/10.1257/pol.6.1.65.

Meier, W., 1973. Banking secrecy in Swiss and international taxation. *The International Lawyer*, 7(1), 16–45, https://scholar.smu.edu/cgi/viewcontent.cgi?article=3915&context=til.

See also: PRIVATE PROPERTY, SECRECY

FIPPS

See: FAIR INFORMATION PRACTICE PRINCIPLES

Firewall

A firewall is a type of **security** system which regulates both incoming and outgoing network traffic in accordance with pre-established security rules. It serves as a barrier between a **trust**ed internal **network** and an untrusted external network, such as the **Internet**, by analysing all network packets and deciding whether to allow or block them in accordance with the rules. Firewalls are frequently used in companies and residential network settings to guard against unauthorised access, **malware** and other security **risk**s. They can be implemented as hardware, **software** or a mixture of the two.

Further reading:

Voronkov, A., Iwaya, L.H., Martucci, L.A. and Lindskog, S., 2017. Systematic literature review on usability of firewall configuration. *ACM Computing Surveys (CSUR)*, 50(6), 1–35, https://doi.org/10.1145/3130876.

See also: ACCESS CONTROL, DEEP PACKET INSPECTION

Firmware

Software that is integrated into a piece of hardware, such as a computer, cell phone, router or printer. It is a form of software made to regulate the fundamental operations and actions of the device. The firmware supplies low-level control over the device and is often stored in non-volatile memory. Initialising the hardware components, controlling input and output processes and providing a hardware-to-operating system or other software application interface are all part of firmware.

To address faults, enhance performance or add new features to the physical device, firmware upgrades are frequently required. Firmware is more difficult for the end user to modify or update than other kinds of software. Consequently, the manufacture of the device usually remotely updates the firmware (via software updates) instead of replacing the physical memory chip.

Just as firmware is difficult for **user**s to directly access, it is more difficult for **adversari**es to access too. However, there is a class of attacks called *firmware exploits*, most of which involve reverse engineering firmware to identify vulnerabilities. Although they are more difficult to deliver than orthodox software-based attacks, they can also be more difficult to detect because of the hidden nature of firmware.

Further reading:

Bettayeb, M., Nasir, Q. and Talib, M.A., 2019. Firmware update attacks and security for IoT devices: survey. *In: Proceedings of the ArabWIC 6th Annual International Conference Research Track*, 1–6, https://doi.org/10.1145/3333165.3333169.

Shah, Y. and Sengupta, S., 2020. A survey on classification of cyber-attacks on IoT and IIoT devices. *In: 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 406–13, https://doi.org/10.1109/UEMCON51285.2020.9298138.

Fishing Attack

A form of **reidentification attack** whereby the **adversary** examines a **de-identified** dataset looking for **data unit**s that appear unusual (and therefore may be **population unique**) – they then attempt to find a **population unit** that matches the data unit on the assumption that because the unit is unusual such a match is likely to be correct. The notion of **special unique**ness is an attempt to capture the **risk** of a fishing attack at the data unit level.

See also: REIDENTIFICATION

Five Safes

A system of **confidentiality** risk management which focuses on organisational controls. The five safes are Settings, People, Projects, Data and Outputs. The system was specifically designed by Felix Ritchie to provide a system of thinking about the core elements of **data safe haven**s. Although a useful framework for this purpose, having the benefit of intuitive simplicity, the five safes system lacks a companion architecture of tools to support rigorous **risk assessment**, so it might be regarded as a context-specific framework rather than a general **risk** management tool applicable to all **data situation**s. Although some recent work has moved to remedy this gap, the framework has also come under some criticism primarily because of this lack of rigour.

Further reading:

Arbuckle, L. and Ritchie, F., 2019. The Five Safes of risk-based anonymization. *IEEE Security & Privacy*, 17(5), 84–9, https://doi.org/10.1109/MSEC.2019.2929282.

Culnane, C., Rubinstein, B.I. and Watts, D., 2020. Not fit for purpose: a critical analysis of the 'Five Safes'. *arXiv preprint,* https://doi.org/10.48550/arXiv.2011.02142.

See also: INFORMATION GOVERNANCE, DATA ENVIRONMENT, FUNCTIONAL ANONYMISATION

Flexible Output

A system (digital or procedural) by which the holder of some **data** may allow **user**s to request statistical extracts rather than having to use a fixed set of statistical outputs. This approach has been explored particularly by **census** agencies where the tradition of publishing large books of fixed tables is being replaced by virtual on-demand systems.

Although this flexibility is beneficial for users, the big challenge with such systems is the assessment of **disclosure risk**, which needs to be done automatically in real time.

Further reading:

Chipperfield, J., Gow, D. and Loong, B., 2016. The Australian Bureau of Statistics and releasing frequency tables via a remote server. *Statistical Journal of the IAOS*, 32(1), 53–64, https://doi.org/10.3233/SJI-160969.

Shlomo, N., Antal, L. and Elliot, M., 2015. Measuring disclosure risk and data utility for flexible table generators. *Journal of Official Statistic*s, 31(2), 305–24, https://doi.org/10.1515/jos-2015-0019.

See also: DATA UTILITY, DISCLOSURE, PUBLISHING

Formal Anonymisation

Formal anonymisation is the practice of removing or **masking** the **direct identifier**s from a **dataset**. This means that individuals are not **identifiable** from within the dataset, but formal anonymisation fails to protect against the possibility that an **adversary** will use **auxiliary knowledge** from outside the dataset to identify individuals using **indirect identifier**s. It is therefore not usually sufficient protection against **disclosure risk** without other measures also being employed.

Further reading:

Elliot, M., O’Hara, K., Raab, C., O’Keeffe, C.M., Mackey, E., Dibben, C., Gowans, H., Purdam, K. and McCullagh, K., 2018. Functional anonymisation: personal data and the data environment, *Computer Law and Security Review*, 34(2), 204–21, https://doi.org/10.1016/j.clsr.2018.02.001.

See also: ANONYMISATION, DATA ENVIRONMENT, FUNCTIONAL ANONYMISATION, PERSONAL DATA

Formal Privacy

Used to denote a method which provides some provable guarantee regarding the **security** of **information**. The term is something of a misnomer as the guarantees provided are almost always within the realm of **confidentiality** (they concern the leakage of information) rather than **privacy** *per se*.

The best-known formal privacy technique is **differential privacy**, other examples are **secure multi-party computation** and **homomorphic encryption**.

Further reading:

Acar, A., Aksu, H., Uluagac, A.S. and Conti, M., 2018. A survey on homomorphic encryption schemes: theory and implementation. *ACM Computing Surveys*, 51(4), 1–35, https://doi.org/10.1145/3214303.

Dwork, C., 2008. Differential privacy: a survey of results. *In*: *International conference on theory and applications of models of computation.* Berlin: Springer, 1–19, https://doi.org/10.1007/978-3-540-79228-4_1.

See also: PRIVACY ENGINEERING, PRIVACY GUARANTEE

Format Preserving Encryption (FPE)

A form of secure **encryption** that maintains the format (and length) of the original **plaintext** which makes it simpler to integrate the encrypted **data** with current systems that take specific data formats. FPE employs mathematical techniques to modify the data in a way that maintains the format and data type, in contrast to conventional encryption approaches that transform data into an unreadable form.

FPE can be used to encrypt potentially **sensitive** and or **identifying** data such as credit card numbers or social security numbers.

Further reading:

Bellare, M., Ristenpart, T., Rogaway, P. and Stegers, T., 2009. Format-preserving encryption. *In: Selected areas in cryptography: 16th annual international workshop*, 295–312, https://doi.org/10.1007/978-3-642-05445-7_19.

Foundation Model

See: GENERATIVE AI

FPE

See: FORMAT PRESERVING ENCRYPTION

Freedom of Expression

The idea of freedom of speech has a long history, but, the expressions of human rights after the Second World War frame the current context. The United Nations’ 1948 Declaration of Human Rights included a right to freedom of expression and opinion, including ‘freedom to hold opinions without interference and to seek, receive and impart **information** and ideas through any media and regardless of frontiers’. A similar right was then articulated in the **European Convention on Human Rights** in 1950. This has equally been carried forward into EU law through the **Charter of Fundamental Rights** in 2012.

Personal free expression requires, and contributes to, a free flow of information to nurture the public discourse within which individuals can express themselves. This principle can come into conflict with the (qualified) right of other individuals to control the information about them which remains in the **public sphere** under the **right to be forgotten**.

Rights to privacy and freedom of expression are both qualified rights that may have to be balanced against each other in practice. Arguably, the introduction of the right to free expression into EU law through the Charter has strengthened its role in European **data protection**. In 2018, the **GDPR** expanded the derogations previously available for artistic and literary expression, to create the possibility for new exemptions for academic expression.

Further reading:

Ausloos, J., 2020. *The right to erasure in EU data protection law*. Oxford: Oxford University Press.

Mourby, M., Gowans, H., Aidinlis, S., Smith, H. and Kaye, J., 2019. Governance of academic research data under the GDPR – lessons from the UK. *International Data Privacy Law*, 9(3), 192–206, https://doi.org/10.1093/idpl/ipz010.

O’Connor, N., 2015. International trends in freedom of information. *In*: Adshead, M. and Felle, T., eds, *Ireland and the Freedom of Information Act: FOI@15*. Manchester: Manchester University Press, 6–31.

See also: CHARTER RIGHTS, CONFLICT OF RIGHTS, FREEDOM OF INFORMATION

Freedom of Information

Under EU and European human rights law, freedom of **information** is closely associated with **freedom of expression**. In the **GDPR**, for example, they are referred to collectively as the ‘right of freedom and information’ (e.g., in Article 17 governing **erasure** of **personal information**).

Aside from exemptions in **data protection** law, freedom of information has its own statutory regime in many countries, albeit often focused on the public sector. The United States, the United Kingdom, Ireland, Finland and Sweden are among the 90-plus countries with some form of public access to official information granted by national legislation. While this can appear to represent an opposing value to that of **privacy**, there can in fact be overlap between privacy as a form of **informational self-determination** and public access to information. For example, the US Privacy Act of 1974 gave citizens the ability to see the information collected about them by federal agencies, and thus exercise **privacy as control** over personal information.

Further reading:
O'Connor, N., 2015. International trends in freedom of information. *In*: Adshead, M. and Felle, T., eds. *Ireland and the Freedom of Information Act: FOI@15.* Manchester: Manchester University Press, 6–31.

See also: CONFLICT OF RIGHTS

Freely Given Consent

The EU **GDPR** requires **consent** to be freely given for it to constitute a valid **lawful basis** for processing **personal data**. The **Article 29 Working Party** has interpreted this to mean that the **data subject** must not fear any detriment from declining to give consent, either from an imbalance of power with the **data controller**, or because consent is an essential precondition to receiving a service. It thus requires consent to serve as an expression of individual **autonomy** and attempts to protect **decisional privacy**.

Further reading:
Article 29 Working Party, 2017. *Guidelines on consent under Regulation 2016/679*, https://ec.europa.eu/newsroom/article29/items/623051.

See also: DATA PROCESSING, DATA PROTECTION

Frequency Data

A form of data representation that captures how frequently a specific event or observation occurs within a given **dataset** (possibly conditioning on the values of other **attribute**s, such as timeframe). In statistical analysis, this kind of **data** is frequently used to identify distributions, trends and patterns.

Frequency data may be disclosive, if not adequately anonymised. This can be of particular concern as frequency tables are a commonly used format for **publication**. In this context, rare (combinations of) values may be a cause for **privacy concern**; this is known as the **special unique** problem.

Techniques to reduce **risk**s in frequency tables include using a formal model such as **differential privacy**, aggregating the data to higher levels, **rounding** values and using statistical methods for **noise addition**.

FTC

See: FEDERAL TRADE COMMISSION

Fully Automated Remote Analysis System (FARAS)

A computer-based system to enable **remote analysis** of data to be carried out without direct human intervention.

This kind of system is intended to automate data collection, analysis and reporting, which can decrease mistakes and save time when compared to manually processing data. FARAS generally combines hardware and **software** components. The software components may include **machine learning** algorithms, **artificial intelligence** and other kinds of **data processing** tools, while the hardware components may include **sensors**, cameras and other kinds of data collecting equipment. Applications for FARAS include traffic analysis, environmental monitoring and security **surveillance**.

Further reading:

O’Keefe, C.M. and Chipperfield, J.O., 2013. A summary of attack methods and confidentiality protection measures for fully automated remote analysis systems. *International Statistical Review*, 81(3), 426–55, https://doi.org/10.1111/insr.12021.

Fully Homomorphic Encryption (FHE)

Fully homomorphic encryption (FHE) enables computation on **encrypted** material without the requirement to first decrypt it. Put another way, it makes it possible to execute computations on **encrypted** data without disclosing the **data** to anyone, not even the person doing the computation. This is accomplished by permitting arithmetic operations to be carried out directly on the **ciphertext**, producing a new ciphertext that, when decoded, produces the correct answer.

FHE has potential uses in **cloud computing**, where data may be handled and analysed by outside service providers while still being encrypted, giving the **data owner** more protection and protecting privacy. Moreover, it may be utilised in **secure multi-party computing**, which allows several participants to work with their individual encrypted data without requiring a **data share**.

Further reading:

Acar, A., Aksu, H., Uluagac, A.S. and Conti, M., 2018. A survey on homomorphic encryption schemes: theory and implementation. *ACM Computing Surveys*, 51(4), 1–35, https://doi.org/10.1145/3214303.

Fan, J. and Vercauteren, F., 2012. Somewhat practical fully homomorphic encryption. *Cryptology ePrint Archive*, https://eprint.iacr.org/2012/144.

Functional Anonymisation

Anonymisation has often disappointed critics, partly because it cannot be perfect, and partly because its focus on the **dataset**s to be anonymised has missed key aspects of their context. Functional anonymisation is a holistic conception of anonymisation that rests on the insight that anonymity of **data** is not a property of data alone, but rather is a property of data in the environment in which it is stored and used. Given the properties of data in the situation in which it is held, the aim of functional anonymisation is to reduce the **risk** of **reidentification** (*deanonymisation*) to an acceptably low level. Furthermore, the level of risk needs to be **proportional** to the **sensitivity** of the data, *and* the **data utility** of the functionally anonymised dataset. A dataset with very limited utility might present a small reidentification risk which is still disproportionately high compared to the benefits of **data-sharing**.

A process for functionally anonymising data is described in the **Anonymisation Decision-Making Framework (ADF)**. In accordance with functional anonymisation’s holistic view, the anonymisation of data involves not only the manipulation of the data, but also of its context or **data environment**. This latter is characterised by other datasets that may be used by an **adversary**, the set of **data user**s who could access the data, the **data governance** and the nature of the **data storage** infrastructure. Manipulating the environment may involve such aspects as **access control**s, query controls or other data governance methods, rather than data manipulation such as high-level **encryption** methods or **noise addition**. The precise focus of the anonymisation methods will need to be proportionate to the risk, but also sensitive to the purpose of sharing and the requirements of data users.

It also follows from the characterisation of functional anonymisation that it does not end with **data release**. The data environment will always be changing, even after sharing and release (for example, as other related datasets become public), and so will need to be monitored. Where possible, there may need to be further measures taken as a result. Furthermore, as functional anonymisation is a risk-based method, there will always be the possibility of a **data breach**. For that eventuality, functional anonymisation of data should include impact management, including means to contact **stakeholder**s, and (if possible) means to suppress any further dissemination of data.

Further reading:

Elliot, M., O’Hara, K., Raab, C., O’Keeffe, C.M., Mackey, E., Dibben, C., Gowans, H., Purdam, K. and McCullagh, K., 2018. Functional anonymisation: personal data and the data environment, *Computer Law and Security Review*, 34(2), 204–21, https://doi.org/10.1016/j.clsr.2018.02.001.

Elliot, M., Mackey, E. and O’Hara, K., 2020. *The anonymisation decision-making framework 2nd edition: European practitioners’ guide*. Manchester: UKAN Publications, https://ukanon.net/framework.

See also: ANONYMITY, DATA AT REST, DATA IN USE, DATA SITUATION, DATA SITUATION AUDIT, PERSONAL DATA

Functional Unique Identifier

A borderline type of **direct identifier** where a combination of two or more **attribute**s will single out most individual **population unit**s. The paradigm example is the combination of name and address. This combination is not 100 per cent reliable as a **unique identifier** as there remains the possibility of statistical twins (it might be that there are two people called ‘John Henry Smith’ living at address X), but these will be rare enough that we can in practice treat functional unique identifiers as if they are unique.

Further reading:

Mackey, E., Elliot, M., O’Hara, K. and Tudor, C., 2016. *The anonymisation decision-making framework*. Manchester; UKAN Publications.

See also: POPULATION UNIQUE, SINGLE OUT, UNIQUENESS

Function Creep

Function creep is the use of something beyond its originally designed purpose. In the context of **informational privacy**, it is often discovered that **information** gathered for one purpose will be useful for other purposes. There is then pressure to co-opt that information for the new purpose. **Purpose limitation** restrictions are intended to guard against function creep, but it is difficult, in advance, to prevent bureaucracies or parliaments changing the rules in the future.

Further reading:
Koops, B.-J., 2021. The concept of function creep. *Law, Innovation and Technology*, 13(1), 29–56, https://doi.org/10.1080/17579961.2021.1898299.

See also: MISSION CREEP, PURPOSE SPECIFICATION

Fuzzing

A technique used to discover vulnerabilities or defects in **software** by inputting random, unexpected, and/or malformed **data**. It is primarily designed to test the **security** and **resilience** of cybersystems.

Further reading:
Zeller, A., Gopinath, R., Böhme, M., Fraser, G. and Holler, C., 2019. *The fuzzing book*. Saarbrücken: CISPA + Saarland University, https://publications.cispa.saarland/3120/.

See also: VULNERABILITY

G

Gait Recognition

A person’s gait is their manner of walking, running or other personal locomotion. While the gait of animals was studied by Aristotle, it became a more scientific enterprise with the invention of photography and cinematography. Further instrumentation now allows quantification of many aspects of biomechanics, often for medical and osteopathic uses.

It has also transpired that precise measurement of the location of body parts such as ankle, knee and hip produce unique **biometric** data for an individual. This biometric can be studied remotely, at a distance, with relatively low resolution, without the subject’s knowledge or cooperation. It is also hard to spoof. Hence, while it is not as accurate a recognition technique as some other biometrics, it has advantages, particularly for covert **surveillance**.

Further reading:
Bouchrika, I., Goffredo, M., Carter, J. and Nixon, M., 2011. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4), 882–9, https://doi.org/10.1111/j.1556-4029.2011.01793.x.

See also: BODILY PRIVACY

Game Theory

Game theory is based on a mathematical formulation representing the interactions between entities that make decisions that have an impact on their individual or collective outcomes. It is frequently used to represent and evaluate the strategic interactions between individuals or organisations in economics, political science, psychology and other social sciences.

A game consists of multiple players, each with a variety of potential moves or approaches, and multiple outcomes that rely on the decisions made by all participants. Game theory examines the incentives and limitations that affect each player’s choices, and forecasts both the likely consequences of those choices and the logical outcome of the interaction of the players’ preferences for specific outcomes.

Game theory can aid in designing privacy-preserving **recommendation systems** by incentivising **users** to share their preferences while protecting their sensitive **information**. It can also enhance privacy in data-driven decision-making and be used to model the interaction between **adversary** and defender in **cybersecurity**. It has also been applied to represent the interaction between adversaries and data systems in **statistical disclosure** scenarios.

Further reading:

Mackey, E. and Elliot, M.J., 2009. An application of game theory to understanding statistical disclosure events. *In: Joint UNECE/Eurostat work session on statistical data confidentiality*, 1–12, UNECE, www.researchgate.net/publication/375837804_An_Application_of_Game_Theory_to_Understanding_Statistical_Disclosure_Events.

Manshaei, M.H., Zhu, Q., Alpcan, T., Baccsar, T. and Hubaux, J., 2013. Game theory meets network security and privacy. *ACM Computing Surveys*, 45(3), 1–39, https://doi.org/10.1145/2480741.2480742.

See also: BOUNDED RATIONALITY, RATIONAL CONSUMER

Gatekeeper

Where there are **access** **control**s to a **dataset**, a gatekeeper has the responsibility to ensure that only those who are **authorised** gain access, by **authenticating** their **credentials**. The gatekeeping function covers both the operational process of managing authorisation, and the responsibility for authorising specific access requests. These processes may rest with different individuals.

Further reading:

Sandhu, R.S. and Samarati, P., 1994. Access control: principle and practice. *IEEE Communications Magazine*, 32(9), 40–8, https://doi.org/10.1109/35.312842.

See: RESTRICTED ACCESS

GDPR (General Data Protection Regulation)

The General Data Protection Regulation is commonly abbreviated as the GDPR. Its full title is:

> *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*

The Regulation contains many of the same requirements of the **Data Protection Directive** it replaced. It governs the use of **personal data**;

information which identifies a living **natural person**. As a Regulation, the GDPR has direct effect across EU member states, meaning that further implementing legislation is not required at national level for its provisions to have legal effect in those **jurisdiction**s.

The GDPR does, however, contain several provisions in which member states have scope to create derogations, such as on the age of **consent** for **children** using **Internet** services, or the safeguards over **data processing** for research. Despite its aim to achieve a single market for personal data in the EU, therefore, the GDPR cannot create complete uniformity across national implementations.

The relative prominence of the GDPR stems in part from its territorial reach – it applies whenever the personal data of people in the EU is processed to monitor them or offer them services. Its expanded agenda in emphasising **transparency**, **accountability** and **explicit consent** has affected most people's lives the most in the proliferation of **cookie** consent notifications. The greater level of maximum fines (whichever is the higher of €20 million or 4 per cent of a **data controller**'s global turnover) makes non-**compliance** a greater commercial risk.

Many years in the making, the GDPR gained political momentum within the EU legislative bodies after the Edward Snowden intelligence leak in 2013 highlighted the global vulnerability to **surveillance** of **personal information**. It is has since been hugely influential on other jurisdictions, an effect that has been called the **Brussels effect**. Its extraterritorial reach means that many multinational organisations have found it simpler to comply with its requirements across all personal data processing, rather than attempting to differentiate according to the location of the **data subject**s in question. It has also inspired the California Consumer Privacy Act & Privacy Rights Act, as well as attempts at large-scale harmonisation reflected in measures such as the proposed American Data Protection and Privacy Act in the United States, and the African Union's Malabo Convention.

Further reading:
Kuner, C., Bygrave, L.A. and Docksey, C., 2020. *The EU General Data Protection Regulation (GDPR): a commentary*. Oxford: Oxford University Press.

See also: DATA HARMONISATION, DATA PROTECTION, US PRIVACY LAWS

Gendered Spaces

Gendered spaces are physical or virtual spaces which those of a particular gender can occupy or meet in, and where the presence of others is forbidden, inappropriate or unwelcome. Such a space may be created by a formal definition, such as with public bathrooms, changing rooms, dormitories, single-sex schools, gentlemen's clubs and women-only gyms. Spaces may also become effectively gendered despite being theoretically open to all, such as clothing shops, certain sporting events or areas of the workplace. Certain rooms in a household may be seen by some in gendered terms.

Gendered spaces, particularly women's spaces, may be constructed for **safety** reasons, as for example refuges for victims of domestic or sexual violence, or women's prisons. However, most gendered spaces primarily protect privacy, allowing members of the admitted gender to interact in the absence of others. The division between gendered spaces is culturally variable, depending on convention. Feminists have long argued that symbolic barriers between the genders have worked to enhance men's status relative to that of women.

In recent years gendered spaces have become more controversial, as some have argued that all spaces should be degendered, while others have framed gender as more expansive and less binary than traditionally understood.

Further reading:
Spain, D., 1992. *Gendered spaces*. Chapel Hill: University of North Carolina Press.

See also: FEMINIST CRITIQUE OF PRIVACY, PRIVACY, CULTURAL VARIATION OF

General Data Protection Regulation

See: GDPR

Generative AI

A type of **Artificial Intelligence** that can automatically create text, images, videos, audio and other content after being trained on very large volumes of **data**. Examples of generative AI are Bidirectional Encoder Representation from Transformers (BERT) and Generative Pre-trained Transformers (GPT).

Generative AI models have great potential, but they also raise **privacy concern**s, as they are trained on vast quantities of data, often scraped from the **Internet** but also from other restrictive resources which might have been linked together either directly or through the generative modelling itself. Such data may well be **personal data**, or at the very least hard to verify as non-personal. There is broad **regulatory** consensus that **information** should be made available to individuals about how their data is collected and processed, so it is intrinsically in tension with the principle of **data-protection-by-design**. The ingestion of such vast amounts of data also makes privacy through **obscurity** less meaningful.

A further risk is *catastrophic forgetting*, a phenomenon where the AI system loses information from previous tasks while learning new ones, which breaks the **audit trail** so that **scrutiny** of what data has been used and how becomes difficult if not impossible. Relatedly, generative AI **algorithm**s are not **transparent** and lack **explainability**. Generative AI can be used to generate **deepfake** images and videos which, while not real, can still be invasive.

Major issues concerning **trust** in these systems, fairness and bias in **machine learning** need still to be addressed. The rise of generative AI makes this urgent.

Further reading:

Hacker, P., Engel, A. and Mauer, M., 2023. Regulating ChatGPT and other large generative AI models. *In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, New York, ACM, 1112–23. https://doi.org/10.1145/3593013.3594067.

Veselovsky, V., Ribeiro, M.H. and West, R., 2023. Artificial Artificial Artificial Intelligence: crowd workers widely use large language models for text production tasks. *arXiv*, a2306.07899, https://doi.org/10.48550/arXiv.2306.07899.

See also: ACCOUNTABILITY, DEEP LEARNING

Genetic Data

See: GENOMIC DATA

Genetic Fingerprint

A reference to the capacity of **genomic data** to **single out** individuals.

Genetic Privacy

Genetic privacy essentially refers to the **privacy concern**s and personal interests in **information** derived from human DNA. As Taylor has pointed out, the term ‘genetic data’ is broad, encompassing different types of technologically generated information, as well as a breadth of **inference**s about people which this information can yield. He argues that **genomic data** are not unique in the concerns they raise for human rights and freedoms, but that the (perceived or actual) stakes are particularly high when this type of information is generated. Genetic data are unusually rich in interpretative potential, and as such are a form of **personal data** which constitutes a particularly fluid regulatory object.

All personal data to some degree present varying risks to **data subject**s depending on the context in which they are interpreted, but genetic data can be an especially (if not quite uniquely) heightened example of this contextuality. One particular concern that is often raised is *heritability*. If someone chooses to make their genetic information available, then they also to some extent reveal information about those related to them. The case of Joseph James DeAngelo, known as the Golden State Killer, who was captured after his relatives were cross-referenced with crime scene DNA on the commercial genomics website 23andMe highlights the potential uses, benefits and **risk**s.

However, the broad consensus among commentators is that *genetic exceptionalism* – and the corresponding idea that genetic privacy represents a unique set of interests and concerns – should be treated with caution. Clayton and colleagues note that the United States has created federal protection for genetic privacy but suggest this is a matter of legislative pragmatism, as the task of passing more a more general privacy law was not politically attainable at the time. Other **jurisdiction**s – such as the European Union – are less piecemeal in their approach. The **GDPR** treats genetic data the same as any other health-related information, as **special category data**, requiring additional safeguards.

Further reading:

Clayton, E.W., Evans, B.J., Hazel, J.W., Rothstein, M.A., 2019. The law of genetic privacy: applications, implications, and limitations. *Journal of Law and the Biosciences*, 6(1), 1–36, https://doi.org/10.1093/jlb/lsz007.

Paradis, M.A., 2018. The Golden State Killer case shows how swiftly we’re losing genetic privacy. *Vox*, 5 May 2018, www.vox.com/the-big-idea/2018/5/3/17313796/genetic-privacy-killer-golden-state-serial-killer-genealogy-genome.

Taylor, M., 2012. *Genetic data and the law: a critical perspective on privacy protection*. Cambridge: Cambridge University Press.

See also: BIOMETRIC DATA, BODILY PRIVACY, INFORMATIONAL PRIVACY, US PRIVACY LAWS

Genomic Data

The term ‘genomic data’ can be broad, encompassing both the genetic code immediately gleaned from analysis of human DNA and also subsequent **inference**s made about an individual through a genomic expertise framework. The slight differentiation between ‘genetics’ and ‘genomics’ is that the latter perhaps captures more of the knowledge accumulated through the study of human genes and their associated phenotypes, but the two terms are often used interchangeably.

The EU **GDPR** defines genetic data as ‘personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique **information** about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question’. Genetic data are treated as a **special category** of data under the GDPR – requiring additional safeguards and justification for their use – but this simply places genomic data on a level with other health-related information, which are also special category data.

Further reading:
Alser, M., Bingöl, Z., Cali, D.S., Kim, J., Ghose, S., Alkan, C. and Mutlu, O., 2020. Accelerating genome analysis: a primer on an ongoing journey. *IEEE Micro*, 40(5), 65–75, https://doi.org/10.1109/MM.2020.3013728.
Taylor, M., 2012. *Genetic data and the law: a critical perspective on privacy protection*. Cambridge: Cambridge University Press.

See also: GENETIC PRIVACY, NATURAL PERSON, PERSONAL DATA

Geographical Resolution

The granularity of a piece of geographical data (within a **dataset**).

Geographical data is regarded among of the most problematic in terms of **statistical disclosure risk**; the higher the resolution, the more **disclosive** the geocoding is. Consequently, **record**-level **census** and survey data, for example, is only routinely shared by statistical agencies with regional level geocodes.

See also: GEOPRIVACY, LOCATIONAL PRIVACY, LOCATION DATA

Geoprivacy

Technically a form of **confidentiality** rather than **privacy**, the term refers to the protection of **data** about an individual's location, place of residence and place of employment and any geographical data tied to a person.

As with all such **privacy concern**s, the tension arises from in enabling analysts to build useful spatial models while still protecting individual identities. It is probably true that geographical data presents this trade-off more acutely than other data. Key to resolving this is the determination of the appropriate level of **geographical resolution** or granularity. Detailed **location data** pins down individuals in a manner that other **information** does not. On the other hand, coarse-grained geographical data can quickly become less useful for drawing valid inferences.

Further reading:

Kounadi, O. and Leitner, M., 2014. Why does geoprivacy matter? The scientific publication of confidential data presented on maps. *Journal of Empirical Research on Human Research Ethics*, 9(4), 34–45, https://doi.org/10.1177/1556264614544103.

Keßler, C. and McKenzie, G., 2018. A geoprivacy manifesto. *Transactions in GIS*, 22(1), 3–19, https://doi.org/10.1111/tgis.12305.

See also: LOCATIONAL PRIVACY, RISK–UTILITY TRADE-OFF

Geo-Social Data

Geo-coded **social media** data (i.e., data including the **location data** of the poster).

Analysts use this type of **data** for geographical analysis of sentiment in social media posts for example. Sophisticated statistical models such as multilevel **network** models allow the combined analysis of geographical and social **information**; Kuchler and colleagues highlighted how useful this could be in predicting the geographical spread of a virus.

A common **privacy concern** arises from this type of analysis being non-**transparent**; although social media users will have agreed to **term**s **of service** that include provision for research by **third parti**es, it is doubtful that this is fully understood by most **user**s.

Further reading:

Kuchler, T., Russel, D. and Stroebel, J., 2022. JUE Insight: the geographic spread of COVID-19 correlates with the structure of social networks as measured by

Facebook. *Journal of Urban Economics*, 127, 103314, https://doi.org/10.1016/j.jue.2020.103314.

See also: GEOPRIVACY, SOCIAL NETWORK

Geotagging

Geotagging is the practice of adding location **metadata** to digital content. The location may be specified by place names, by significant features, or more technically through latitude and longitude coordinates derived from the global positioning system (GPS). These latter will probably be derivable by the device upon which the content is created (smartphones automatically embed GPS coordinates in content by default).

In combination with a timestamp, the geotag effectively places a device at a place and time and therefore can disclose patterns of behaviour. The content may also disclose that people or activities are (taking place) at the geotagged location: an **identifiable individual** may appear in a photograph on **social media**; an illustrated advert on Craigslist may place high-value goods at an address (or a person in the personal column); a celebrity Twitter feed may reveal a home address.

Further reading:
Friedland, G. and Sommer, R., 2010. Cybercasing the joint: on the privacy implications of geo-tagging. *In: HotSec'10: Proceedings of the 5th USENIX Conference on Hot Topics in Security*. New York: ACM, 1–8, www.usenix.org/legacy/events/hotsec10/tech/full_papers/Friedland.pdf.

See also: LOCATIONAL PRIVACY, LOCATION DATA

Globally Unique Identifier (GUID)

A file, document or **record** can be uniquely identified by a Globally Unique Identifier (GUID), a 128-bit **identifier** created by **software**. As GUIDs are intended to be distinct across both space and time, two GUIDs should never be the same, even when used on different computers. The most common way to display GUIDs is as a series of hexadecimal numbers divided into groups. There are many different approaches to creating GUIDs, including random number generators, timestamp-based techniques and **network** address-based techniques.

Global Privacy Control (GPC)

Users may inform websites and online services of their **privacy** choices via the Global Privacy Control (GPC) feature on their Web browser. The GPC project, a non-profit organisation, created the GPC to give users a uniform, accessible way to exercise their rights to privacy and **privacy preference**s. The **GDPR** and the California Consumer Privacy Act are two examples of current privacy laws and regulations that the GPC is meant to complement. Users can exercise their right to object to data collection and **data sharing** by activating the GPC setting in their Web browser, which sends a signal to the websites and online services they visit.

The GPC is implemented by a browser extension or plugin that transmits a standardised 'Do Not Sell' signal to websites and online services. This will prevent the gathering and sharing of users' **personal information** by those websites and online services that have agreed to honour it.

Further reading:

Fisher, D., 2020. Global Privacy Control Protocol aims to pick up where Do Not Track left off. *Decipher*, 7 Oct 2020, https://duo.com/decipher/global-privacy-control-protocol-aims-to-pick-up-where-do-not-track-left-off.

See also: DO NOT TRACK (PROTOCOL), TRACKING

Global Recoding

A form of **statistical disclosure control** whereby categories of a nominal or ordinal variable are aggregated together (e.g., age might be indicated by ranges rather than single integers, or geography might be changed from postcodes to regions). This reduces the detail of the data and thus is detrimental to **analytical completeness** but unlike **perturbation** does not impact **analytical validity**.

Global recoding is also referred to as domain generalisation.

Further reading:

Duncan, G.T., Elliot, M. and Salazar-González, J.J., 2011. *Statistical confidentiality: principles and practice.* Cham: Springer.

See also: GEOGRAPHICAL RESOLUTION

Global Suppression

A method of **statistical disclosure control** for **microdata** where whole variables are removed from the **dataset** before **data sharing** or **publication**.

Further reading:
Willenborg, L. and De Waal, T., 2012. *Elements of statistical disclosure control.* Cham: Springer.

See also: GLOBAL RECODING, LOCAL SUPPRESSION

Gossip

Gossip refers to both the practice of discussing the private affairs of someone who is not present, and the topics of such discussion. Gossip is usually superficial and trivial, often malicious and sometimes true. It may exaggerate the truth, leading to a general opinion that ‘there is no smoke without fire’ which harms the subject. Because it is unattributable, it is hard to counter or to hold people to **account** for it (in an Agatha Christie story, Hercule Poirot compared it to the Lernean Hydra, which had many heads and every time one was cut off, two more would grow back). Many social psychologists see it as an important bonding mechanism. It is a breach of **attentional privacy** but there is some cultural ambiguity, aptly conveyed by Oscar Wilde’s famous quote: ‘there is only one thing worse than being talked about and that is not being talked about.’

Further reading:
Peters, K., Jetten, J., Radova, D. and Austin, K., 2017. Gossiping about deviance: evidence that deviance spurs the gossip that builds bonds. *Psychological Science*, 28(11), 1610–19, https://doi.org/10.1177/0956797617716918.

See also: CULTURAL VARIATION OF PRIVACY

GPC

See: GLOBAL PRIVACY CONTROL

Graduated Security

A **security** strategy whereby security measures are stronger or weaker depending on the (perceived) amount of **risk** or **threat** that an organisation is facing.

This is intended to create a balance between the demand for security and the need for usability of assets and information. Differing levels of security might also be given to various assets or pieces of information depending on their **value** or **sensitivity**. For instance, more sensitive data could be subject to more stringent security controls, such as **encryption**, **restricted access** and monitoring, but less sensitive data would simply need simple **password** protection or **access control**s.

See also: BUSINESS IMPACT LEVEL, RISK–UTILITY TRADE-OFF

Grey Hat Attack

An attack on an organisation's **system**s and/or **data** designed to demonstrate – to the organisation or possibly to the wider world – that its systems are unsafe. It is distinguished from a **white hat attack** because it has not been pre-authorised by the organisation, and from a **black hat attack** because a grey hat **adversary** does not have malicious intent.

Further reading:

Kirsch, C., 2014. The grey hat hacker: reconciling cyberspace reality and the law. *Northern Kentucky Law Review*, 41(3), 383–403, https://heinonline.org/HOL/LandingPage?handle=hein.journals/nkenlr41&div=24&id=&page=.

Morgan, G. and Gordijn, B., 2020. A care-based stakeholder approach to ethics of cybersecurity in business. *In*: Christen, M., Gordijn, B. and Loi, M., eds, *The ethics of cybersecurity*, Cham: Springer, 119–38, https://doi.org/10.1007/978-3-030-29053-5_6.

See also: ADVERSARY, CYBERSECURITY, ETHICAL HACKING, HACKING, PENETRATION TEST

Group Harms

Groups as understood for the purposes of **informational privacy** could be thought of as united by a common susceptibility to a specific **harm** or set of harms. For example, if people with fast eye movements are routinely screened out of recruitment processes by an **algorithm** that correlates this pattern in the footage with poor employee performance, these 'fast-retinal-movement-people' are unlikely to realise they have this **risk** of detriment in common, but they nonetheless have a shared interest in how eye-movement profiles are used in an employment context.

The recognition of groups as potential victims of harm has filtered out of academia and found recognition in policy circles. The **European Data Protection Supervisor**, for example, has recommended 'safeguards and rights to be provided to individuals, and groups of individuals, that may be impacted by the use of AI systems'. It may be some time, however, before the dualistic model of the individual and the **public** embedded in many legal systems gives way to a more nuanced recognition of groups and communities who may also require legal protection. In the United States, reform of the (application of) the Common Rule which apparently prohibits consideration of long-term policy effects in ethical review of research would be a potential step forward for the recognition of group harms.

Further reading:

Doerr, M., and Meeder, S., 2022. Big health data research and group harm: the scope of IRB review. *Ethics and Human Research*, 44(4), 34–8, https://doi.org/10.1002/eahr.500130.

Wachter, S., 2022. The theory of artificial immutability: protecting algorithmic groups under anti-discrimination law. *Tulane Law Review*, 97(2), 149–204, https://heinonline.org/HOL/LandingPage?handle=hein.journals/tulr97&div=11&id=&page=.

See also: ARTIFICIAL INTELLIGENCE, ATTRIBUTION, BIG DATA, DATA ETHICS, GDPR, GROUP PRIVACY, PRIVACY RISK

Group Privacy

In law, privacy has typically been cast as an individual human right. However, outside rights discourse, groups also have privacy interests and are open to **group harms**. Most obviously, the family is a locus of privacy, and **intimacy**, as a type of privacy, seems to require at least two people together. The household is often seen as a private space or part of the **private sphere**, by writers ranging from Aristotle to John Stuart Mill, in which state or social interference would be inappropriate. More recently, the development of data analytics has led to questions as to whether clustered **data subject**s are appropriately seen as requiring group privacy protections.

Opponents of group privacy argue that all group privacy ultimately can be expressed as the sum of the privacy interests of all the members of the group. However, O'Hara draws two important distinctions. First is that between privacy in a group setting, where the privacy interests of others in a group are equivalent to their individual privacy interests (for example, those attending a banned event), and between group privacy itself, where the group's privacy interests are over and above the sum of its individual members' interests. This is most clear with respect to a nuclear family, where the privacy interest seems to go beyond those of the individual interests of parents and children. Note also that the privacy of a group may entail the reduction or loss of privacy of the members of the group with respect to other members of the group.

O'Hara's second distinction is between associations which are formed consciously and voluntarily, and clusters which are formed without the knowledge of their members (such as a classification of people by a **machine learning** system). He argues that the salience of group privacy is easier to defend in instances of the first type; with the second type, group members probably do not perceive themselves to have joint interests and are more likely to pursue their privacy interests as individuals.

Further reading:

O'Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Taylor, L., Floridi, L. and van der Sloot, B., eds, 2017. *Group privacy: new challenges of data technologie*s. Cham: Springer.

See also: RIGHT TO PRIVACY

GUID

See: GLOBALLY UNIQUE IDENTIFIER

H

Hacking

Hacking is the act of gaining unauthorised access to a computer system, **network** or **software** program, sometimes with the intention of stealing, altering or destroying sensitive **data** or interfering with the system or network's regular operation. Hacking may be done maliciously by people or organisations, as well as by **security** experts and researchers looking to find weak points in systems and strengthen their security. Hackers may enter a system or network via various methods, such as **social engineering**, **password** cracking and taking advantage of any **vulnerability** in software or hardware. Hackers may employ malware or other harmful software to infiltrate the system.

By undermining standard security measures and ignoring norms about property and boundaries, hacking may have harmful effects on people, businesses and society at large. Systems that have been compromised might be used to transmit **virus**es or **malware**, steal financial or **personal data** or launch **attack**s on other networks or systems. In addition to financial losses, **reputation**al **harm** and legal liabilities, hacking can result in **data breach**es and other security problems.

Ethical hacking uses the techniques of hacking to discover vulnerabilities before external agents do; **penetration testing** is one ethical hacking approach that security experts may employ to find vulnerabilities in systems and networks and strengthen their security posture.

Further reading:

Erickson, J., 2010. *Hacking: the art of exploitation.* Seoul, Acorn Pub. https://digtvbg.com/files/books-for-hacking/Hacking%20-%20The%20Art%20of%20Exploitation,%202nd%20Edition%20by%20Jon%20Erickson.pdf.

See also: BLACK HAT ATTACK, CYBERSECURITY, GREY HAT ATTACK, NETWORK SECURITY, SECURITY POSTURE, WHITE HAT ATTACK

Harassment

Harassment encompasses a range of unwanted **intrusion**s from **public** or **private** actors that are humiliating or threatening and targeted at a specific victim or victims.

In the digital era, harassment takes on new forms, which engage other aspects of privacy. In **doxxing**, an individual's **personal information** – often their address or contact details – is revealed without their **consent** to facilitate further harassment from other **social media** users. This is a privacy violation in both its intrusion into the **right to be let alone** (by facilitating unwanted contact via private channels), and in the sense of a violation of informational **autonomy**, as **information** is revealed against the will and wishes of the victim. A harasser may also create a **fake profile**, purporting to be their victim, and behave badly to attack their **reputation**. Even if no one directly contacts or disturbs the victim, and they thus remain 'let alone', this is still a privacy violation in its co-opting of the victim's **informational self-determination**, and right to control their image.

Further reading:

Eckert, S. and Metzger-Riftkin, J., 2020. Doxxing, privacy and gendered harassment: the shock and normalization of veillance cultures. *Medien & Kommunikationswissenschaft,* 68(3), 273–87, https://doi.org/10.5771/1615-634X-2020-3-273.

See also: DEFAMATION, FALSE LIGHT, INFORMATIONAL PRIVACY, PRIVACY AS CONTROL, RIGHT TO PRIVACY

Harm

The damage done by a violation of **privacy** can be difficult to characterise, let alone quantify. However, the law requires evidence of loss or injury before a court can provide compensation for **breach** of rights to privacy.

Van der Sloot has shown that the awards from the European Court of Human Rights for breaches of the **European Convention on Human Rights** Article 8 privacy rights are mostly made up of 'non-pecuniary' damages. These constitute financial compensation for non-material, intangible harms such as distress or loss of **reputation**. Damage to tangible property (or loss of quantifiable income, such as from employment) is also included within the Court's awards. Nevertheless, the Court clearly places significant emphasis on the subjective harm experienced by individuals whose privacy rights have been breached.

Privacy and **data protection** laws have been criticised for their emphasis on harm to the rights and freedoms of individuals, and a relative disregard for the downstream, societal consequences of personal **data processing**, such as **group harm**s and **chilling effect**s. McMahon and colleagues have therefore advocated 'Harm Mitigation Bodies' to adjudicate more systemic impacts of **big data**.

Further reading:

McMahon, A., Buyx, A. and Prainsack, B., 2020. Big data governance needs more collective responsibility: the role of harm mitigation in the governance of data use in medicine and beyond. *Medical Law Review*, 28(1), 155–82, https://doi.org/10.1093/medlaw/fwz016.

van der Sloot, B., 2017. Where is the harm in a privacy violation? Calculating the damages afforded in privacy cases by the European Court of Human Rights. *JIPITEC*, 8, 322–51, https://heinonline.org/HOL/Page?handle=hein.journals/jipitec8&div=38&g_sent=1&casa_token=&collection=journals.

See also: DATA GOVERNANCE, DATA PROTECTION IMPACT ASSESSMENT, FINANCIAL PRIVACY, HUMAN RIGHTS IMPACT ASSESSMENT, OBJECTIVE HARM, ONTOLOGICAL SECURITY, PERSONAL DATA, SUBJECTIVE HARM

Hashing

A hash function is a deterministic function that takes an arbitrary piece of **data** (of arbitrary size) and maps it onto a data value of fixed size and structure (a *hash*). The relationship between input data and hash can be stored in a *hash table*. Hash functions should always be easy to compute and should minimise the (inevitable) duplication of output clashes where the hashes of two different inputs are identical. Hashing in general facilitates **data storage** and retrieval and is an important tool in **cryptography**.

Further reading:

Knott, G.D., 1975. Hashing functions. *Computer Journal*, 18(3), 265–78, https://doi.org/10.1093/comjnl/18.3.265.

See also: CRYPTOGRAPHIC HASH FUNCTION

Header Information

Header information is the **metadata** that is present at the start of a data packet or file that contains details about the content and format of the **data**. The header of each packet normally comprises the source and destination addresses, the packet sequence number, the date and time of transmission, the type of data being transferred, the format or encoding employed and any error detection or correction codes.

Header information is used to describe the structure and content of documents in file formats such as HTML. In general, header **information** is crucial for guaranteeing accurate data transmission, interpretation and **processing** between various systems and **application**s. It is essential for routing the data packets through a network to their destination, and for assembling the packets in the right sequence into a complete document when they arrive. In **end-to-end encryption**, therefore, even though the packet's contents are encrypted, the header information remains in the clear, because the nodes in the **network** need to know how to treat the packet.

See also: CONTENT DATA, HYPERTEXT TRANSFER PROTOCOL SECURE, INTERNET, LINK ENCRYPTION

Health Information Exchange (HIE)

Within a healthcare system, there are likely to be multiple actors, ranging from hospitals, doctors and specialists, nurses, researchers, policymakers, public health officials, billing systems, insurance services, drug prescribers, laboratory test providers and so on, as well as individual patients themselves, and sometimes their carers, parents or guardians. There is therefore a strong interest in exchanging information between these actors, but this must be done securely and with patients' consent: the process is called Health Information Exchange (HIE).

The issues involved in HIE vary, and regularly require legal, administrative and technological measures, as well as often substantial funding. In the United States, with its fragmented and largely privately owned healthcare system, HIE has to be mandated by federal or state regulation. The Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009 provided for billions of dollars of funding to set up HIE systems within entities covered by the **Health Insurance Portability and Accountability Act**, and required privacy provisions, especially **data breach notification**s. At the other end of the scale, in the highly centralised British National Health Service, many providers will be internally based, but even there exchanging information is non-trivial.

Further reading:

Kuperman, G.J., 2011. Health-information exchange: why are we doing it, and what are we doing? *Journal of the American Medical Informatics Association*, 18(5), 678–82, https://doi.org/10.1136/amiajnl-2010-000021.

See also: CONSENT, DATA SHARING, ELECTRONIC HEALTH RECORD, RECORD

Health Insurance Portability and Accountability Act (HIPAA)

Commonly abbreviated to ‘HIPAA’, this US federal legislation was passed in 1996 with the primary aim of regulating the health insurance market. The HIPAA Privacy Rule was subsequently drafted by the Department of Health. Together with HIPAA, it provides some privacy safeguards for patients, but only applies to ‘covered entities’ such as healthcare providers and insurers.

The United States has historically struggled to pass privacy or **data protection** legislation of a more general nature. As a result, HIPAA and GINA (the Genetic Information and Nondiscrimination Act of 2008) respectively regulate the use of health and genetic **data**, even though the **privacy concern**s associated with these types of **personal information** may be equally pertinent for data relating to race, ethnic origin, sexuality, religious beliefs and so on. Rothstein has argued that the piecemeal nature of the American privacy framework reflects the legislative pragmatism of lawmakers in passing the laws they can, and not holding out for wider protection.

Both HIPAA and the subsequent Privacy Rule have been heavily criticised by academic commentators for being out of date and excessively reliant on individualistic **informed consent** as a regulatory tool, as well as for being unable to guarantee patient **confidentiality** following the Supreme Court’s reversal of *Roe v Wade*.

Further reading:

Cohen, I.G. and Mello, M.M., 2019. Big data, big tech, and protecting patient privacy. *JAMA*, 322(12), 1141–2, https://doi.org/10.1001/jama.2019.11365.

Shachar, C., 2022. HIPAA, privacy, and reproductive rights in a post-Roe era. *JAMA*, 328(5), 417–18. https://doi.org/10.1001/jama.2022.12510.

See also: ACCOUNTABILITY, CONSENT, GENETIC PRIVACY, HEALTH INFORMATION EXCHANGE, PERSONAL DATA, US PRIVACY LAWS

Hellinger Distance

A metric of the similarity of two probability distributions that is frequently used to measure the residual utility of **data** after **disclosure control method**s are applied.

Further reading:

Nikulin, M.S., 2001. Hellinger distance. *In*: *Encyclopedia of mathematics*, http://encyclopediaofmath.org/index.php?title=Hellinger_distance&oldid=47206.

See also: DATA UTILITY, DISCLOSURE, STATISTICAL DISCLOSURE CONTROL

Hidden Service

An online service which is anonymous to its users. Collectively, such services are often referred to as the **dark web**.

Further reading:
Owen, G. and Savage, N., 2016. Empirical analysis of Tor hidden services. *IET Information Security*, 10(3), 113–18, https://doi.org/10.1049/iet-ifs.2015.0121.

See also: TOR

HIE

See: HEALTH INFORMATION EXCHANGE

Hierarchical Data

A form of **microdata** where groups of **data unit**s are explicitly represented in the **data** structure. Examples are people within households, employees within companies, pupils within schools. Taking account of such hierarchical structure in multi-level models leads to better analyses. However, the presence of such structure in a **dataset** also increases the **disclosure risk**. For example, simply knowing the age-sex structure of a large household provides a significant **attack** vector for an **adversary**.

HIPAA

See: HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT

History of Privacy

Although **privacy** has been discussed for millennia in multiple cultural contexts, the word 'privacy' was formed within the English language in the 16th century, derived from the medieval term ‘privitee’, meaning a hidden aspect of God or nature. The early modern, humanist emphasis of ‘privacy’ places the individual at the centre of the concept, acknowledging some social right of voluntary **seclusion** of the **self**, or aspect of the self, from wider view or **interference**.

By the 19th century, the idea of the ‘private’ had become a key part of the architecture of legal and political discourse. Society was understood to be organised along the lines of **public sphere** and **private sphere**, and the law within many Western **jurisdiction**s was understood as either ‘public law’ (governing the relationships between the individual and the state) or ‘private law’ (governing the relationships between citizens).

The genesis of privacy as a freestanding legal right, and potential cause of action in itself, is attributed to Warren and Brandeis in their 1890 *Harvard Law Review* article. The impact of this **right to be let alone** spilled over from academia into the American courts, particularly as photography became an increasingly accessible consumer technology, and individuals tried to defend non-consensual uses of their image. The right to privacy was famously the basis on which the Supreme Court ruled that **abortion** could be lawfully provided to women in *Roe v Wade*, an interpretation of privacy which has since been interpreted as at odds with the US constitution in *Dobbs v Jackson Women's Health Organization* (2022).

In the UK, multiple attempts were made in the 20th century to introduce privacy as a distinct statutory right, but ultimately the Human Rights Act 1998 gave the **European Convention on Human Rights (ECHR)** the status of domestic legislation. The ECHR right to respect for private and family life has reinforced **common law** causes of action, and in essence created a right to privacy in **jurisdiction**s party to the ECHR.

The binary, neoliberal concept of privacy formed in the 19th century came under fire in the 20th. Feminist scholars criticised what they saw as patriarchal oppression in the defence of a private sphere in which the state should not interfere, exemplified in the view that no sexual acts within marriage could constitute rape. Technology further undermined the neat distinction between public and private, with the **Internet** bringing global connections within the domestic setting.

Further reading:

Gavison, R., 1992. Feminism and the public/private distinction. *Stanford Law Review*, 45(1), https://doi.org/10.2307/1228984.

O’Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220.

See also: CELEBRITY PRIVACY, FEMINIST CRITIQUE OF PRIVACY, INTIMACY, PRIVACY, CULTURAL VARIATION OF

Homomorphic Encryption

A form of **encryption** that allows the performance of computation without **decryption** being necessary first. For example, if a **user** adds two encrypted numbers together, they get the same outcome as if they added the **plaintext** numbers together and then encrypted the sum. In theory, this enables **security** to be maintained for **data in use** and opens the possibility of privacy-preserving **predictive analytics**. Use cases include secure **data processing** in the cloud, enhancing **privacy-preserving machine learning**, secure financial transactions and encryption of **database** queries.

However, the computational resources required to enable homomorphic encryption means that it is still inefficient for many applications and so the potential of this technology is yet to be fully realised.

Further reading:
Acar, A., Aksu, H., Uluagac, A.S. and Conti, M., 2018. A survey on homomorphic encryption schemes: theory and implementation. *ACM Computing Surveys*, 51(4), 1–35, https://doi.org/10.1145/3214303.

See also: FORMAT-PRESERVING ENCRYPTION, FULLY HOMOMORPHIC ENCRYPTION, ORDER-PRESERVING ENCRYPTION, PRIVACY-PRESERVING DATA ANALYTICS

Honeypot

A honeypot is a **cybersecurity** instrument or strategy that entails setting up a computer or network system to detect **attack**s. To trap as many adversaries as possible, a honeypot is made to seem weak or alluring, but is instrumented to detect unauthorised activity. There are many different ways to construct honeypots. The honeypot system might be set up to mimic a flaw or **vulnerability** or to pose as a genuine system holding important **information** or resources. This variability is a strength as there is not a single honeypot signature that adversaries could themselves detect (and therefore avoid).

Once an **adversary** starts interacting with the honeypot system, the **security** team may keep track of and document their actions, including how they gained access, what tools and methods they used, and whether they tried to steal **data** or exploit vulnerabilities. This information may be utilised to find existing vulnerabilities, enhance security procedures and create fresh defences and remedies.

Moreover, honeypots may be used to obtain information about adversaries' intentions, objectives and affiliations, and may even be used to identify them. Organisations may utilise this information to better understand the **risk**s and threats they face by using it to inform threat intelligence and **threat modelling** initiatives.

Further reading:
Niels, P., 2004. A virtual honeypot framework. *In*: *13th USENIX Security Symposium (USENIX Security 04)*, San Diego, CA: USENIX Association, www.usenix.org/conference/13th-usenix-security-symposium/virtual-honeypot-framework.

HRIA

See: HUMAN RIGHTS IMPACT ASSESSMENT

HTTP

See: HYPERTEXT TRANSFER PROTOCOL

HTTPS

See: HYPERTEXT TRANSFER PROTOCOL SECURE

Human-Centred Cybersecurity

A discipline that addresses the human factor in **information security**. It is recognised that humans play an important role in **cybersecurity** decisions via their interaction with **security** technology. Therefore, sound technology cannot solve all cybersecurity problems, because cybersecurity itself exists within a human context that will affect its reliability. It highlights the importance of usability, user experience, awareness and human adaptation in the context of cybersecurity.

Further reading:
Grobler, M., Gaire, R. and Nepal, S., 2021. User, usage and usability: redefining human centric cyber security. *Frontiers in Big Data*, 4, https://doi.org/10.3389/fdata.2021.583723.

See also: USER

Human Rights Impact Assessment (HRIA)

In policy-making circles, impact assessments are a well-established means of investigating, ahead of time, the potential implications of a particular programme or change in the law. The Human Rights Impact Assessment (HRIA) has been championed as a way of considering the social impacts of a proposal through the lens of international human rights law (including the **right to privacy**).

Mantelero and Esposito have advocated the use of HRIAs to evaluate the impacts of **Artificial Intelligence**, to help stabilise the values debated in a field often dominated by culturally contingent 'ethical guidelines'. The right to privacy would be one such stabilising force: a legal requirement with a long-established international jurisprudence, lending any assessment greater weight and generalisability than an evaluation based on ethical principles alone.

Further reading:
Mantelero, A. and Esposito, M.S., 2021. An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law and Security Review*, 41, 105561. https://doi.org/10.1016/j.clsr.2021.105561.

See also: DATA PROTECTION IMPACT ASSESSMENT, EUROPEAN CONVENTION ON HUMAN RIGHTS, IMPACT MANAGEMENT, PRIVACY IMPACT ASSESSMENT, TRUST

Hypertext Transfer Protocol (HTTP)

A fundamental **protocol** used for **communication** on the **World Wide Web**. It enables the transfer of data between a client and a server, allowing the display of web pages. The client sends a HTTP request, the server then processes it and returns the requested resource to the client.

HTTP transmits data in **plaintext**, meaning that any **information** exchanged between the user and the web server can be intercepted by a **third party**. This lack of **encryption** makes users vulnerable to **eavesdropping** and unauthorised access. Adversaries can capture and misuse session **cookie**s, and in general the lack of encryption in HTTP makes **user**s' data more susceptible to interception. To address these concerns, HTTPS has been introduced.

Further reading:

Naylor, D., Finamore, A., Leontiadis, I., Grunenberger, Y., Mellia, M., Munafò, M., Papagiannaki, K. and Steenkiste, P., 2014, December. The cost of the 's' in https. *In: Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies*, 133–40.

Kristol, D.M., 2001. HTTP cookies: standards, privacy, and politics. *ACM Transactions on Internet Technology*, 1(2), 151–98.

See: COOKIE, HYPERTEXT TRANSFER PROTOCOL SECURE, INTERNET PROTOCOL

Hypertext Transfer Protocol Secure (HTTPS)

A protocol called Hypertext Transfer Protocol Secure (HTTPS) is used to offer secure **Internet communication**. It is an expansion of the HTTP protocol, which enables a Web server and browser use to exchange **data**. To prevent unwanted access or tampering, HTTPS operates by encrypting the data passed between server and browser. This is done by creating a secure connection between them using the **Transport Layer Security (TLS)** or **Secure Socket**s **Layer (SSL)** encryption protocols.

When a **user** connects to a website using HTTPS, the browser checks the server's **digital certificate**, which is issued by a reputable **certificate authority**, to confirm its **identity**. This guarantees that the user is corresponding with the appropriate server, that their data is encrypted, and that it is safe from interception or alteration. For critical transactions like online banking and **e-commerce**, and other delicate operations involving the transmission of **personal data** or financial **information**, HTTPS is frequently utilised. Many websites employ it to guard against **attack**, including **man-in-the-middle attack**s and **data breach**es.

Further reading:

Felt, A.P., Barnes, R., King, A., Palmer, C., Bentzel, C. and Tabriz, P., 2017. Measuring HTTPS adoption on the Web. In: *Proceedings of the 26th USENIX Conference on Security Symposium*, USENIX Association, 1323–38, www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/felt.

See also: COMMUNICATION PRIVACY, DATA TRANSFER, FINANCIAL PRIVACY, INTERNET PROTOCOL, SECURE COMMUNICATION, SECURITY

I

I2P

An anonymous network layer called I2P (Invisible **Internet** Project) enables private and secure online communication, offering a decentralised, autonomous **network** that is impervious to monitoring, **censorship** and other types of intervention. At each layer, I2P adds a new level of **encryption** and **anonymity** to the **data** as it is routed across the network. The network is made up of endpoints, which are nodes that send or receive traffic, and routers, which relay **communication** between nodes.

The use of **hidden service**s or websites and services that are only available through the I2P network is one of its main characteristics. These covert services serve as an essential tool for safe, **secure communication** and free expression since they are designed to be anonymous and resistant to censorship. I2P also has other uses, including email, instant chat and file sharing.

Further reading:
Hoang, N.P., Kintis, P., Antonakakis, M. and Polychronakis, M., 2018. An empirical study of the I2P anonymity network and its censorship resistance. *In*: *Proceedings of the Internet Measurement Conference 2018*, ACM, 379–92, https://doi.org/10.1145/3278532.3278565.

See also: DARK WEB, NETWORK ENCRYPTION, TOR, TRAFFIC DATA

ID Card

See also: IDENTIFICATION CARD

Idem-Identity

In his 1990 volume *Oneself as Another*, French philosopher Paul Ricoeur criticised simplistic views of **identity**, which ignored identity's self-referential nature, and introduced a distinction between **ipse-identity** and *idem-identity*. Idem-identity is the self as understood from outside, an external, objective, third person attribution of sameness, a social judgment that *this* person is numerically identical to *that* one.

Further reading:
Ricoeur, P., 1994. *Oneself as another*. Chicago: University of Chicago Press.

Identifiability

Identifiability refers to a property of **data** by which identifying a **natural person** within it is possible. It is not a legal term, but as **data protection** laws, such as the EU's **GDPR**, cover **information** which relates to identified and identifiable individuals, data will be within scope if they have the capacity to identify someone through means reasonably likely to be used.

See also: IDENTIFIABLE DATA, IDENTIFIABLE NATURAL PERSON, PERSONAL DATA

Identifiable Data

Data in which it is possible to identify a **natural person** are sometimes referred to as identifiable data. The term is more likely to be used within **jurisdiction**s that apply the EU's **GDPR**, which governs **personal data** which it defines as data relating to **identified** and **identifiable individual**s.

Personal data can therefore be seen as comprising two subtypes: data relating to identified people, and data relating to identifiable people. To determine whether a person is identifiable, the GDPR requires reference to all *means reasonably likely to be used* to identify someone.

See also: DATA SUBJECT, IDENTIFIABLE NATURAL PERSON, DATA PROTECTION, PERSONHOOD

Identifiable Individual

See also: IDENTIFIABLE NATURAL PERSON

Identifiable Natural Person

The **material scope** (i.e., subject matter) of the EU's **GDPR** is **personal data**. Personal **data** are defined as **information** relating to an identified or an identifiable **natural person**. An individual is *identifiable* if they can be identified by a means reasonably likely to be used. Whether or not a means is reasonably likely to be used depends on the informational context, and the available technology according to the state of the art. It is this circumstantial contingency in the quality of *identifiability* that has led Elliot, O’Hara and colleagues to discuss **anonymisation** as a functional (rather than intrinsic) process.

Further reading:
Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S.E., Bell, J., Smith, H., Aidinlis, S. and Kaye, J., 2018. Are pseudonymised data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law and Security Review*, 34(2), 222–33, https://doi.org/10.1016/j.clsr.2018.01.002.

See also: DATA PROTECTION, DATA SUBJECT, FUNCTIONAL ANONYMISATION, IDENTIFIABLE DATA, IDENTIFIED DATA PERSONHOOD

Identification Card

An identification card or *ID card* is a card, usually mandated by a government, through which a person can authenticate their **identity** either to the government itself, to commercial organisations such as banks or to service providers such as healthcare companies. Although they have a long history, ID cards were rare before the First World War in 1914. More abstractly, an ID card is a member of the class of identity documents, which also includes passports or written documents (*identity papers*). Some identity documents, such as driving licences, are sometimes used as *de facto* ID cards.

Typically, the ID card is linked with a **database** that contains the **information** about the **person** that may officially be known, which includes items such as birthdate, birthplace, nationality or full name. It may also include more intrusive elements, such as the person’s home address or profession. Often, gender is also included, although this has become more controversial as its meaning and significance has become increasingly contested. Race or religion may also be recorded, which historically has led to injustice and discrimination in some cases; it has been claimed that the efficient pre-war Dutch ID system, which recorded religion, allowed the Nazis to discover Jews more easily there than elsewhere in Europe.

Governments use such cards for assorted reasons: for keeping track of citizens, for purposes of taxation, conscription, or determining eligibility for government services; for keeping track of foreigners and controlling borders; for controlling populations, restricting travel and movement and abetting **surveillance**; for other policy purposes that require knowledge of **population** behaviour. Establishing identity in the first place may be difficult and may depend on the citizen being able to authenticate their identity with some other document.

ID cards are a boon for authoritarian systems. In some **jurisdiction**s, failure to gain an ID card may mean exclusion from certain government services or freedoms, but conversely a comprehensive ID register may be helpful for governments to ensure they respond to their citizens’ needs (for example, ID systems were valuable in supporting economic relief programmes during the COVID pandemic). As part of its administrative modernisation, for instance, India has unrolled a large-scale identity programme called Aadhaar.

ID cards are often contested. In those jurisdictions where a citizen has a specific identity assigned by the government, the ID card works as the proof of that identity. However, in **common law** jurisdictions such as the United States or the United Kingdom, citizens do not have fixed identities, and so the ID card represents an increase in government powers vis-à-vis the citizen. Attempts to introduce ID cards in Britain, or to extend their use beyond wartime, have always been unpopular, and successfully resisted up to the time of writing.

Further reading:

Aiyar, S., 2017. *Aadhaar*: *a biometric history of India’s 12-digit revolution*. Chennai: Westland Publications.

Lyon, D., 2009. *Identifying citizens*: *ID card*s *as surveillance*. Cambridge: Polity Press.

See also: IDENTITY DOCUMENTS, PERSONHOOD, PUBLIC RECORDS

Identification File

In disclosure **risk assessment**, the identification file is a representation of the auxiliary information that the **adversary** holds (or is considered to). This **information** could be in the form of a **database**, paper records, or it could simply reside in the adversary’s memory. Whatever its form, the file contains **direct identifier**s and other information on one or more **data subject**s, and it may therefore be useful for the adversary to carry out **linkage attack**s against de-identified **target file** if the other information overlaps with the **data** contained within the target file and hence provide **key variable**s for the purposes of linkage.

Further reading:

Elliot, M., and Dale, A., 1999. Scenarios of attack: the data intruder's perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring), 6–10, http://tinyurl.com/SCEN-ATTACK.

Duncan, G.T., Elliot, M., and Salazar-González, J.J., 2011. *Statistical Confidentiality*. New York: Springer, https://doi.org/10.1007/978-1-4419-7802-8.

See also: RECORD, RECORD LINKAGE, STATISTICAL DISCLOSURE RISK, SCENARIO ANALYSIS

Identified Data

Under the EU's **GDPR**, there are, in essence, two forms of **personal data**: **information** relating to *identified* and to **identifiable natural person**s. The term 'identified data' is sometimes used to refer to information falling within the first category of personal **data**: information relating to *identified* natural people.

A person is identified by data if it reveals their identity without the need to refer to any further information. Little has been written on what constitutes an **identity** in the context of digital information. Mourby and Mackey have argued that the benchmark for when information can be considered an identity lies in its capacity to impact an individual. Rather than suggesting that **privacy** is engaged when an individual is identified, they suggest that identity should be understood when some aspect of privacy – **intrusion**, monitoring, **profiling**, **reputation**, **autonomy** and so on – is likely to be affected by the information. As such, identity in information (and thus identified data) lies in the capacity to engage the values and interests captured by the idea of privacy.

Further reading:

Mourby, M., and Mackey, E., 2023. Pseudonyms, profiles and identity in the digital environment. *In*: van der Sloot, B. and van Schendel, S. eds, *The boundaries of data: technical, practical and regulatory perspectives*, Amsterdam: Amsterdam University Press.

See also: DATA PROTECTION, DIGITAL IDENTITY, IDENTIFIABLE DATA

Identified Natural Person

See also: IDENTIFIED DATA

Identifier

A piece, or combination of pieces, of **information** which make it possible to link some data to a real-world entity. Identifiers can either be direct (indicating that there is reliable one-to-one mapping between the identifier and the entity) or indirect (indicating that the mapping is likely but contingent on empirical facts).

The EU’s **GDPR** refers to a list of identifiers within its definition of **personal data** within Article 4(1). These include name, identification number, **location data**, online identifier, or various aspects of, for example, the physical, physiological or genetic **identity** of that **person**. This list is not intended to be an exhaustive list of the types of information that could be used to identify an individual. Instead, its contents are examples of the signifiers that could – alone or in combination with other pieces of information – enable an identification of an individual.

An important distinction under EU law is that the presence of an identifier is not the same thing as an identification. While – for example – a name or a genetic variant *might* point to the identity of a specific individual, if millions of people share that name, or that variant, this piece of information will not be sufficient to identify a **natural person**. Therefore, to determine whether the information containing the identifier is personal **data**, and covered by **data protection** law, it is necessary to consider the **uniqueness** of the identifier, and/or the likelihood of it being combined with other information.

Other **jurisdiction**s take a more data-centric and less contextual approach to the concept of identifiers. For example, the **HIPAA** Privacy Rule in the United States lists 18 identifiers which are taken to constitute **direct identifier**s of an individual, regardless of circumstances.

Further reading:
Information Commissioners Office, 2022. *What is personal data?* https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/personal-information-what-is-it/what-is-personal-data-1-0.pdf.

See also: UNIQUE IDENTIFIER, GENETIC DATA, DATA PORTABILITY

Identity

Identity is a complex nexus of conceptions with numerous implications for **privacy**. Broadly speaking, such conceptions can be classified into three groups: the metaphysical, the **other**-directed and the other-generated.

A metaphysical idea of identity relates to essential matters pertaining to an individual. That which distinguishes the individual from everything else is its identity, the relation that it bears only to itself, otherwise called *numerical identity*. This often persists through time, creating philosophical problems about how numerical identity is established (how do we know that this person is the same person as the child in this photograph taken 50 years ago?). A human individual's numerical identity over time is often referred to as their *personal identity*. There are many deep questions as to whether personal identity resides in a **self** or soul, or spatiotemporal continuity of the body or mind, and there are many paradoxes which are often the subject of science fiction counterfactuals.

An other-directed idea of identity aims to present an individual to others in a certain way. This may help the other to distinguish the individual – for example, certain physical characteristics are helpful for this purpose, such as the face, as are labels, such as the name. Such conceptions may also help individuals to assimilate into social groups, as when a person identifies as a particular gender, nationality or religion. One can have a cultural, political (in a class, or a party), national, sexual, racial or ethnic, professional or generational identity, among others. Such identities are often signalled by individuals in the ways they speak, dress or present themselves.

An other-generated idea of identity is created by a nation, civic society, group, institution or computer system for it to be able to distinguish and **single out** those individuals it deals with. Such identities typically generate identifiers or **credentials**, including passports, ID cards, social security numbers and other labels, non-obvious biometrics such as fingerprints, behaviours such as purchase histories, associations with devices signalled by cookies and passwords, all of which serve to confirm that the individual is indeed the correct individual from the point of view of the institution. When another gains access to such an **identifier**, they can present themselves falsely as the original individual, a process known as **identity theft**.

Identity affects privacy in many ways, principally as the means of singling out an individual and providing a route of access to that individual from others. Furthermore, if two identifiers can be linked to the same identity, an individual can be traced across systems. The absence of such a means entails that the identity of the individual is concealed, resulting in the individual being *anonymous*. Other-generated identities are a particular issue, as individuals may not input on or control the means used to distinguish them – a point which may be seen as an assault on their **dignity**. To address this latter problem, the idea of **self-sovereign identity** has emerged, where individuals manage computational resources to generate their own unique identifiers which will suffice to identify them to others.

Further reading:

Kerr, I., Steeves, V. and Lucock, C., eds, 2009. *Lessons from the identity trail*: *anonymity, privacy and identity in a networked society*. New York: Oxford University Press.

Martin, R. and Barresi, J., 2006. *The rise and fall of soul and self*: *an intellectual history of personal identity*. New York: Columbia University Press.

Sullivan, C., 2018. Digital identity – from emergent legal concept to new reality. *Computer Law and Security Review*, 34(4), 723–31, https://doi.org/10.1016/j.clsr.2018.05.015.

See also: ANONYMITY, AUTONOMY, DIGITAL IDENTITY, DIGNITY, IDENTITY ASSURANCE, IDENTITY DISCLOSURE, IDENTITY MANAGEMENT, PERSONHOOD, PRIVACY AS CONTROL, SELF-DISCLOSURE

Identity Assurance

In **identity management**, identity assurance is the key task of *authenticating* those wishing to access a system. Typically, a user presents a credential to a gatekeeper, and the quality of the credential is assessed by the assurance system. Access to the system (virtual or physical) must only be possible at specific points where the gatekeepers can be located. **Credentials** may take the form of **password**s, biometrics, digitally signed certificates or devices such as smart cards, and the assurance process may be in several stages. It could also involve a dialogue with the agent, for example sending a one-time passcode to their phone, as in **two factor authentication**.

Identity assurance for individuals requires holding some **personal data** since, by definition, **information** about their credentials must be sufficient to identify them. Centralised provision of identity assurance therefore poses **security** and **privacy risk**s. Identity assurance may be outsourced to external specialist providers, which raises the question of how such a service should be funded. It may also be *federated*, so that several organisations participate in the process (perhaps some holding the **data**, others performing the authentication).

Depending on the security requirements of the system, the credentials must be reliable. This means that, except in low-risk environments, the agent should not provide their own credentials (although they might have control over which credentials are presented to whom, as in a **self-sovereign identity** system). Credentials must be supplied either by outside providers, such as a trusted **certification authority**, or by a secure in-house process, such as the generation by the identity management system of a smart card with a biometric such as a photograph, possibly verified via a passport.

It may be that the credentials of an agent could be portable, so that they can be used across identity management systems.

Further reading:
Chapple, M., 2021. *Access control and identity management*, 3rd edition. Burlington, MA: Jones & Bartlett Learning.

See also: ACCESS CONTROL, AUTHORISATION, AUTHENTICATION, CERTIFICATION, FEDERATED IDENTITY, INFORMATION SECURITY

Identity Cloning

A form of **identity theft** where the thief is not motivated by financial gain but rather seeks to create a false **identity** to conceal their true one. This may also involve synthesising new elements of an identity rather than simply copying an existing one. Cloning may be more effective if the stolen identity corresponds to a person who is no longer alive (and therefore the sole digital footprint being created is that of the cloners).

Identity Disclosure

In **statistical disclosure control**, an **identity** disclosure is the association of a **data unit** with a specific person. Identity disclosure is the result of a successful **reidentification attack**.

The **self-disclosure** of an attribute that is central to a person's sense of self. Often used as a formalisation of 'coming out' in LGBT+ communities but also with respect to neurotypes and, in principle, any significant **attribute**.

It is noteworthy that these two definitions are converse in terms of who is doing the disclosing and what is disclosed.

Further reading:
Hunter, S., 2007. *Coming out and disclosures*: *LGBT persons across the life span*. London: Psychology Press.
Skinner, C., 2009. Statistical disclosure control for survey data. *In*: Pfeffermann, D. and Rao C.R., eds, *Handbook of statistics 29A Sample Surveys*: *Design, Methods and Applications*, Amsterdam: Elsevier, 381–96.

See also: DISCLOSURE, PERSONHOOD, REIDENTIFICATION

Identity Documents

See also: IDENTIFICATION CARD

Identity Management

In any organisation or company, some **information** will be sensitive or confidential, and so access to it must be restricted. However, such information will also be important for timely decision-making, so that, for instance, everyone in a discussion about a sensitive issue is sufficiently briefed. The principle of an identity management system is to determine which persons or occupiers of which roles within an organisation should have access to information, when, and under what circumstances.

As well as information, access may govern entry into a space, as for instance when an **identification card** is used to unlock a door. A **social network** user should be the only person able to access and change their profile and should be able to determine which other **user**s have access to their uploaded content. Access may also be to delivery of services – for instance, one may need to identify oneself to receive a welfare payment.

Because the focus of identity management is the persons who may require access, an identity management system must be able to *register* individuals on the system (and de-register them when they leave), *identify* registered individuals, *authenticate* that they are who they say they are and *enforce* restrictions on access appropriately. It should also be able to *update* profiles and *repair* problems, such as managing the refreshment of forgotten **password**s, and keep the system *secure*, so that identities cannot be stolen. Some people will be employed by or otherwise associated with the organisation, and so identity management can be conducted as a part of their job, but sometimes external agents may be involved (for instance, a private sector provider of an outsourced function for a welfare agency may need access to confidential medical records, or an e-commerce company may need to manage the identities of its customers), which demands a more robust system. The ISO has developed several standards for identity management systems. Identities may also be imported from **identity provider**s or managed by users (**self-sovereign identity**); this has the advantage for the user that they can use the same **identifier**, such as a password, across different sites (**single sign-on**), and for the manager that the key functions of identification, **authentication**, and so on are outsourced to a specialist. A **federated identity** system distributes the necessary identification across various providers, so that no one holds the complete record.

Privacy can be an issue in several ways. For example, the system **database** is likely to associate **personal data** with the means of identification. A management system may also provide a trace of the activities of an **identified individual** (where they logged on, which floors of the building they entered and when), which may be used to evaluate their performance. A social network depends on the management of the identities of all its members, and of their relationships. The network graphs are themselves valuable commodities, but also the flow of information between users depends on effective management, so that a private post is not seen more widely than set out in the **privacy policy**.

Further reading:

Cao, Y. and Yang, L., 2010. A survey of identity management technology. *In*: *2010 IEEE international conference on information theory and information security*, IEEE, https://doi.org/10.1109/ICITIS.2010.5689468.

Wilson, Y. and Hingnikar, A., 2019. *Solving identity management in modern applications*: *demystifying OAuth 2.0, OpenID Connect, and SAML 2.0*. New York: Apress.

See also: ACCESS CONTROL, AUTHORISATION, DIGITAL IDENTITY, IDENTITY ASSURANCE, PASSWORD MANAGER, RESTRICTED ACCESS

Identity Provider

An **identity** provider is a service that manages **user** identification and **authentication**. It oversees confirming users' identities and authenticating them for applications, giving them access to resources and services in accordance with their status. It is, therefore, a **third-party** service provider mediating between **user** and **application**. An identity provider is frequently used in a **single sign-on** (SSO) system, which allows users to access multiple services with a single authentication. Other services and apps receive the appropriate **credentials** from the identity provider, who also manages the authentication and **authorisation** procedure.

Identity providers utilise a variety of authentication methods, including SAML (**Security Assertion Markup Language**), OpenID Connect, and OAuth, to confirm a user's identity. These protocols enable users to log in to the identity provider using a variety of credentials, including biometric **information**, security tokens, usernames and **password**s.

Identity Theft

The appropriation or use of an individual’s **personal data**, usually for financial gain, or less commonly to cause difficulty to the victim. Typical examples of such misuse are opening bank accounts, obtaining loans or procuring goods or services, while posing as the victim; the victim’s *bona fides* authorise the transaction, while the rewards accrue to the **identity** thief. As Armstrong’s testimony demonstrates, the impact on a victim’s life can be significant and long lasting.

Further reading:

Armstrong, D., 2017. My three years in identity theft hell. Bloomberg.com, https://web.archive.org/web/20170919142519/https://www.bloomberg.com/news/articles/2017-09-13/my-three-years-in-identity-theft-hell.

Williams, M.L., 2016. Guardians upon high: an application of routine activities theory to online identity theft in Europe at the country and individual level. *British Journal of Criminology*, *56*(1), 21–48, https://doi.org/10.1093/bjc/azv011.

See also: APPROPRIATION OF LIKENESS

Ideological Privacy

O’Hara defines ideological privacy as a type of **privacy** that does not involve **secrecy** or ignorance. An individual has ideological privacy when their political or religious belief system is not held against them or does not influence others’ behaviour towards them, even if their beliefs are widely known. It is therefore a type of freedom of thought and action, and as such is related to **decisional privacy**. It is mainly discussed when it is under threat, for example where **data** is used to infer ideological preferences, or when the display of overt symbols of belief is penalised in the workplace.

Further reading:

O’Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

See also: BIG BROTHER, PRIVACY THREAT

IDS

See also: INTRUSION DETECTION SYSTEM

Impact Management

A process which, acknowledging that **confidentiality** risks (or other **cybersecurity** risks) cannot be reduced to zero, puts in place strategies to mitigate the negative impact of a **breach** should one happen.

The term is particularly used in the **Anonymisation Decision Making Framework** to denote the third managerial activity, which is designed to steer **information** managers away from a **release-and-forget** mentality.

See ANONYMISATION

Impersonation

Impersonation is the act of representing oneself as another existing individual. This can be done with one's physical presence, adopting the mannerisms and patterns of speech of the **other**. Or, in an administrative context, one can use the means that the impersonated use to identify themselves (**password**s, **identity** documents, forged biometrics, **credential**s, etc.). The appropriation of such means is called **identity theft**. Impersonation is a comedic art form, but in a criminal context is a type of fraud, enabling the impersonator to commit actions for which the impersonated person has permission or authority (e.g., voting, withdrawing money from an account, accessing confidential **information** or making decisions).

Further reading:
Campobasso, M. and Allodi, L., 2020. Impersonation-as-a-service: characterizing the emerging criminal infrastructure for user impersonation at scale. *In*: *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, 1665–80, https://doi.org/10.1145/3372297.3417892.

See also: DEEPFAKE, IDENTITY CLONING

Implicit Consent

A form of **consent** where an individual's permission for an action by another is inferred from that individual's other behaviour and/or from the context. Implicit consent can be adequate where the contextual cues are clear – Taylor and Wilson give the example of a patient rolling up their sleeve in a doctor's office, having been offered an injection. However, they warn against the 'overextension' of the concept to justify disclosure of confidential **information**, particularly when there is no active behaviour from which the assent of the **data subject** can be inferred.

Further reading:
Greenwald, A.G. and Krieger, L.H., 2006. Implicit bias: scientific foundations. *California Law Review*, 94(4), 945–67. https://doi.org/10.2307/20439056.
Taylor, M.J. and Wilson, J., 2019. Reasonable expectations of privacy and disclosure of health data. *Medical Law Review,* 27(3), 432–60, https://doi.org/10.1093/medlaw/fwz009.

See also: CONTEXTUAL INTEGRITY, EXPRESS CONSENT, INFORMED CONSENT, OPT OUT

Imputation

A technique used both as a remedy for **missing data** and as a method for **data synthesis** where *m* draws from the posterior distribution derived from a model of the available/original **data**. This leads to the creation of *m*, **dataset**s requiring an analytical approach which accounts for the additional variance. Imputed data are less risky than the original data but may still carry **risk** of **attribute disclosure**. Where m = 1 the term *single imputation* is sometimes used; where m > 1 then multiple **imputation** is common.

Further reading:
Schafer, J.L., 1999. Multiple imputation: a primer. *Statistical method*s *in medical research*, 8(1), 3–15, https://doi.org/10.1177/096228029900800102.

Inadvertent Disclosure

Inadvertent **disclosure** involves the accidental release of private or confidential **information** to one or more unauthorised people. It is a particular ethical issue in legal cases where one litigant accidentally discloses damaging confidential information to an opponent. It may also be that even well-anonymised data may be inadvertently disclosive if seen by someone who happens to be able to piece together an **identity** from personal knowledge.

Further reading:
Stewart, C.E., 2017. Ethics corner: inadvertent disclosure – traps await the unwary. *Business Law Today*, 27 April 2017, www.americanbar.org/groups/business_law/publications/blt/2017/04/ethics_corner/.

See also: CONFIDENTIALITY, ETHICS, IDENTITY DISCLOSURE, SPONTANEOUS RECOGNITION

Incognito Mode

Several web browsers provide an option called incognito mode, or *private browsing mode*, which enables users to access the **internet** anonymously. The browser does not save any temporary files, cookies or browsing history on the device when in incognito mode. Those who do not want their surfing activities to be monitored or recorded can utilise incognito mode to increase their **privacy** and **security**. It can be helpful for a variety of tasks, including sensitive research, anonymous gift-buying and utilising public computers or **network**s where security and privacy may be an issue.

Incognito mode does not completely guarantee security or **anonymity**. Internet service providers (ISPs) and websites may still monitor a user’s activities and location even when their **browsing history** is not kept locally via their **IP address**.

See also: BROWSER FINGERPRINTING

Incremental Authorisation

See also: JUST-IN-TIME CONSENT, PERSONAL INFORMATION MANAGEMENT SYSTEM

Indirect Identifier

An indirect **identifier** is an **attribute** of an individual that would not usually by themselves **single out** the individual (unlike a **direct identifier**), but which, by combination with other attributes, could create a **unique identifier** for at least some individuals. Indirect identifiers are also sometimes called quasi-identifiers. When combined into an **attack** vector for a specific **linkage attack**, they are **key variable**s.

Sime combinations of mundane attributes can be particularly revealing. For instance, neither age, biological sex nor marital status will normally be directly identifying, but for some combinations may be disclosive (consider for example a 16-year-old widower). Sweeney took a publicly available hospital **dataset** from Washington State in 2012, which contained no direct identifiers but was complete (i.e., it contained all the hospitalisations in the state during that year). She used **information** from news reports and was able to match some of the information in the reports uniquely against the hospital data and thus identify some people, including politicians and sports stars. The medical data provided further information about these people that was not public, such as medical history and drug and alcohol use. The attributes that enabled the matches were therefore, in this context, indirect identifiers.

Further reading:

International Organization for Standardization, 2018. *Privacy enhancing data de-identification terminology and classification of techniques, definition 3.10*. www.iso.org/obp/ui/#iso:std:iso-iec:20889:ed-1:v1:en.

Sweeney, L., 2015. Only you, your doctor, and many others may know. *Technology Science*, 2015092903, https://techscience.org/a/2015092903/.

See also: ANONYMISATION, FUNCTIONAL UNIQUE IDENTIFIER, JIGSAW IDENTIFICATION

Inference

Inference is the capacity to derive a piece of **information** from one or more other pieces of information. This is one *modus operandi* of much scientific research and in particular of statistical approaches to the analysis of **data**.

One **privacy** issue that arises from this is that strong inference may disclose information about a specific individual at a probability sufficiently close to certainty. Consequently, the EU's **Article 29 Working Party** listed inference as one of the risks that effective **anonymisation** must counter.

The underlying problem is the strong relationship between *bona fide* statistical inference and statistical disclosure risk. Specifically, the property that is of most interest to an analyst – variance within a **population** – is precisely the property that an **adversary** can exploit by for **reidentification** or **attribution disclosure** attacks. Conversely, attempts to reduce the **disclosure risk** can ruin the usability of the data for *bona fide* analysts, so a trade-off is often necessary.

A related issue concerns the inferential capabilities of **machine learning**, **artificial intelligence** and in particular **generative AI**, which are also able to use probabilistic techniques to produce surprising results from data analysis. Their ability to find even weak signals in noisy data (especially **big data**) means that their performance outstrips that of purely human or bureaucratic techniques. Privacy tends to be compromised, simply because novel inferences will be made, bringing many relationships within the data out from **obscurity**.

Further reading:
Article 29 Working Party, 2014. *Opinion 05/2014 on Anonymisation Techniques*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

See also: STATISTICAL DISCLOSURE, RISK–UTILITY TRADE-OFF, PROFILING

Inference Attack

Any **attack** through which an **adversary** can infer a characteristic of a **population unit** of which they were not previously aware. This has been paradigmatically considered in the context of single statistical **database**s where it is often referred to as **attribute disclosure** and for which **risk assessment** approaches exist. However, in the context of **big data**, inference attacks can be far more complex and difficult to measure the risk of.

To be considered successful from the adversary's point of view, an inference attack does not need to be 100 per cent correct, and even marginal improvements in **accuracy** may be sufficient (depending on the purposes of the attack). Relatedly, as **reidentification** is not required for an inference attack to be successful, the population unit in question need not even be in the data for an inference attack to be feasible. In fact, a well-formed sample of the **population** may well be sufficient (as sample design specifically aims to allow valid **inference**s). This does partially undermine **sampling** as a form of disclosure control.

See also: DISCLOSURE RISK, STATISTICAL DISCLOSURE CONTROL, TARGETED ADVERTISING

Inferential Disclosure

See: ATTRIBUTE DISCLOSURE

Inferred Data

Inferred data is **data** that is attributed to a **data subject** without being directly captured, usually during an interaction with them. The data upon which inferences are made may be directly provided by the subject (**declared data**) or created through their actions on a platform or channel (for example, their purchases or browsing behaviour on a company’s website). Inferred data stands in contrast to declared data, in that the data subject does not provide it intentionally, but only as a by-product of their other interactions and/or as a derivative of their declared data.

Further reading:
Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. and Rao, V., 1994. Combining revealed and stated preferences data. *Marketing Letters*, 5(4), 335–49, https://doi.org/10.1007/BF00999209.

See also: DATA EXHAUST, PERSONALISATION

Inforgs

See: INFOSPHERE

Information

An abstract concept with overlapping definitions in physics, maths, philosophy and computer science. A core idea underpinning these definitions is the interpretation of **data**. As a result of this interpretative process, information may be considered to have meaning whereas data does not. One can also reverse this relation and consider that data are signals for the underlying information.

Consequently, it is more coherent to refer to agents as ‘informational beings’ but less so ‘data beings’. For this reason, **informational privacy** is preferred as a term to data privacy, though the latter is used in some literatures.

Informational Privacy

Informational **privacy** is the prevention of access to **information** about an individual or a group and is typically breached when information about them flows freely or is transferred to third parties, particularly against their wishes. It is the form of privacy that receives the most attention in legal and academic studies, owing to its being the chief type of privacy breached by and within computer systems. In a **breach**, information might be revealed directly, or may be inferred by an **adversary**. One serious consequence of informational privacy breaches is that they can be hard to undo – once a piece of information is published, it is extremely hard to suppress.

The question as to what constitutes information about someone, and how that might be quantified, is quite complex, and explored, for example, by Gavison. The most used definition, prominent in legal regulation, is the idea of **personal data** (in European **data protection** law), or **personally identifying information** (in US law), in which the information is sufficient to allow the subject to be identified, or to be the basis of action concerning them.

The reach of informational privacy norms is a matter of dispute. Some information about an individual is seen as ‘public’. In that case, there is a question about whether informational privacy applies to that type of information: the individual is not entitled to keep the information private, but they may still have a privacy interest in suppressing it in some contexts. Informational privacy also has links with **confidentiality** (e.g., the protection of information about someone by another person in a professional role, such as a doctor or accountant) and **secrecy** (where information might circulate around an in-group, while being protected from outsiders).

Informational privacy is an important and prominent type of privacy, so much so that some have gone as far as to define privacy in entirely informational terms; a common type of definition of privacy (see for example Westin or Inness) focuses on the amount of control one has over the flow of information about oneself. Solove, cataloguing those privacy breaches that could be remedied under US law, devoted three of his four major classes of breach to informational privacy breaches. Others (such as Allen), however, have criticised such approaches as ignoring many other important types of privacy that have nothing to do with information.

Further reading:

Allen, A.L., 1988. *Uneasy access*: *privacy for women in a free society*. Totowa, NJ: Rowman & Littlefield.

Gavison, R., 1980. Privacy and the limits of law. *Yale Law Journal*, 89(3), 421–71, https://doi.org/10.2307/795891.

Inness, J., 1992. *Privacy, intimacy and isolation*. New York: Oxford University Press.

O’Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Solove, D.J., 2008. *Understanding privacy*. Cambridge, MA: MIT Press.

Westin, A., 1967. *Privacy and freedom*. New York: Ig Publishing.

Informational Self-Determination

Also referred to as a 'right of personality', the idea of a right to informational self-determination was coined by the German Constitutional Court when considering the lawfulness of the 1983 **Census** Act. In principle, the right confers upon the individual the opportunity to make decisions about the **disclosure** and use of their **personal data**. However, even in its inception, the right is qualified and can be overridden in the **public interest** (subject to safeguards). More broadly, the informal influence of this jurisprudence has led some to see the rights of **data subject**s under the EU's **GDPR** as mechanisms of informational self-determination, even if the specific legal right is not technically available outside German law.

Technology that supports informational self-determination includes **personal data store**s and **digital footprint eraser**s, although the technical and institutional infrastructure to support these tools is not yet available at the scale needed for widespread adoption.

Further reading:
Ausloos, J., 2020. *The right to erasure in EU data protection law.* Oxford: Oxford University Press.
Asikis, T. and Pournaras, E., 2020. Optimization of privacy–utility trade-offs under informational self-determination. *Future Generation Computer System*s, 109, 488–99, https://doi.org/10.1016/j.future.2018.07.018.

See also: ERASURE, PRIVACY AS CONTROL, RIGHT TO DATA PROTECTION, RIGHT TO BE FORGOTTEN, RIGHT OF ACCESS

Information Broker

See also: DATA BROKER

Information Classification Table

A document that lists the various forms of **information** categories inside an organisation. It is used to recognise and categorise information according to its level of **secrecy** and **sensitivity**, and to establish the proper type of **security** and handling techniques it needs. An information classification often includes classification categories or levels such as public, internal, confidential and top secret.

See also: CIA TRIAD, CLASSIFIED INFORMATION, INFORMATION SECURITY, PUBLIC, SECRET

Information Ethics

Information **ethic**s is a branch of the **philosophy of information** that deals with the moral challenges posed by information technology. This, as well as the ethical issues surrounding **privacy**, covers several other areas, including the moral status of artificial agents, the value of truth and definitions of mis**information**, the status of the information space as an environment, the nature of rights of access to information (and rights to withhold it) and digital divides.

Floridi’s *ontological theory of privacy* is related to his information ethics; privacy is a function of the forces that hinder or promote the flow of information. As Floridi also argues that people are at least partly constituted by their information, so **informational privacy** has a direct effect on personal **identity**.

Further reading:
Floridi, L., 2013. *The ethics of information*. Oxford: Oxford University Press.

See also: PERSONHOOD, VALUE OF PRIVACY

Information Governance

Information governance is the administration by an organisation of the **information** it holds, covering among other things discovery, storage, **compliance**, **security**, **privacy**, management and quality assurance. The aim is to increase the business value of information held by the organisation, while minimising the **risk**s it poses (which include privacy **breach**es for which the organisation could be held responsible).

Further reading:
Smallwood, R.F., 2020. *Information governance*: *concepts, strategies, and best practices*. Hoboken: John Wiley & Sons.

See also: DATA LIFECYCLE MANAGEMENT, DATA CURATION, INFORMATION SECURITY, PRIVACY RISK

Information Lifecycle Management

The purpose of information lifecycle management is analogous to that of **data lifecycle management**, that is, managing **information** within an organisation using the stages in its lifecycle as a structuring principle, so that the information is available when needed in a timely fashion. The major difference between **data lifecycle** management and information lifecycle management is that the former carries an implication of digital resources, while information management may include other media, including paper, microfilm, photographs and video, and so on – although as more records are held digitally, the distinction is becoming less salient.

Further reading:
Stephens, D.O., 1998. Megatrends in records management. *ARMA Records Management Quarterly*, 32(1), 3–9.

See also: RECORDS MANAGEMENT

Information Loss

A reduction in the **information** provided by some data.

Data protection processes such as **statistical disclosure control** or **differential privacy** invariably result in some information loss. How these losses affect **data utility** can be difficult to estimate as the impact will depend on the specific uses of the **data**.

Further reading:
Domingo-Ferrer, J., 2009. Information Loss Measures. *In*: Liu, L. and Özsu, M.T. eds, *Encyclopedia of Database Systems*. Springer: Boston, https://doi.org/10.1007/978-0-387-39940-9_1505.

See also: DATA QUALITY, ANALYTICAL VALIDITY, ANALYTICAL COMPLETENESS

Information Ownership

See also: DATA OWNERSHIP

Information Security

The methods and steps used to safeguard **information** from unauthorised access, use, **disclosure**, interruption or alteration. Protecting sensitive **data** and assets, such as private financial and personal information, **intellectual property** and proprietary corporate data, requires careful attention to information security, especially when they are held online.

To protect the **privacy**, **accuracy** and accessibility of information, a variety of technologies, rules and practices are used. Access restrictions, **encryption**, **firewall**s and **network security** and security awareness are examples of standard **information security** procedures.

Further reading:
Whitman, M.E. and Mattord, H.J., 2021. *Principles of information security.* London: Cengage Learning.

Informed Consent

Just as **consent** takes on different meanings according to context, informed consent can also refer to multiple terms of art. Faden and Beauchamp presented a thorough historical and conceptual overview of informed consent in 1986, arguing that the disciplines of law and moral philosophy have been the most influential in shaping the term in recent years. This characterisation remains apposite, with informed consent continuing to refer to:

- An ethical tenet for (some) scientific research on human subjects.
- A legal requirement for human-subject research in many jurisdictions – particularly for clinical trials.
- A broader legal doctrine whereby an individual's cognisant acceptance can convert what would otherwise have been a violation of civil or criminal law into a justified **interference** with their physical and/or moral **integrity**.

Physical integrity in this instance refers to **bodily privacy**, whereas moral integrity might be more commonly violated by a breach of **informational privacy** (e.g., a **breach of confidence**). As such, informed consent can legally and/or ethically justify interferences as varied as medical treatment, contact sports, tattoos and use of **personal data** for research.

More recently, informed consent has emerged as a basis for processing personal **information** under data protection law. The proposed American Data Protection and Privacy Act defines affirmative **express consent** as being informed; the **GDPR** defines consent as *inter alia* an informed indication of a **data subject**'s wishes. The debate within **data protection** circles as to when informed consent is an appropriate basis for **data processing** reflects wider controversies about the use of consent as a model of regulation based on individual behaviour, rather than systemic **scrutiny**.

Elliot et al argue that informed consent can be best understood as the intersection of two more fundamental psychological processes: **awareness** and **agreement**, which can occur independently of one another. They further posit that breaking the concept down into these two components both makes it easy to understand the nuance (of, e.g., what has been consented to) and to map on to other higher order concepts such as **transparency**.

Further reading:

Elliot, M., Mackey, E. and O'Hara, K., 2020. *The anonymisation decision-making framework, 2nd Edition*: *European practitioners' guide*. Manchester: UKAN Publications, https://ukanon.net/framework.

Faden, R. and Beauchamp, T., 1986. *A history and theory of informed consent*. New York: Oxford University Press.

Brownsword, R., 2004. The cult of consent: fixation and fallacy. *King's College Law Journal*, 15, 223–9, https://doi.org/10.1080/09615768.2004.11427572.

See also: PHYSICAL PRIVACY, DATA BREACH, US PRIVACY LAWS

Infosphere

The infosphere is a name for the realm of **information** and its operation in the world, for example in **communication**, in governing natural and artificial processes and in learning, memory and knowledge; the term was developed as an analogy to other spheres, such as the biosphere (the realm of living things), the noosphere (the realm of reason) and the hydrosphere (the realm of water). The infosphere is larger than the digital realm or *cyberspace*, and indeed the term was coined in a world of analogue telephone, radio and television.

If the infosphere is taken as a fundamental aspect of reality, then those operating within it need to be information-based too – so-called information organisms or **inforgs**. Humans are inforgs, with their identity formed by the information that exists about them. It follows that breaches of **informational privacy** are serious threats to personal **identity**, which led philosopher Luciano Floridi to advance his ontological interpretation of informational privacy.

Further reading:

Floridi, L., 2005. The ontological interpretation of informational privacy. *Ethics and Information Technology*, 4(4), 287–304, https://doi.org/10.1007/s10676-006-0001-7.

See also: PHILOSOPHY OF INFORMATION, ETHICS, INFORMATION ETHICS

Inherence

A property of a piece of evidence for **authentication** whereby it is intrinsically tied to the person's (unique) physical identity (i.e., the evidence *inheres in* their physical identity). The more specific term **biometrics** is often used synonymously. Examples are fingerprints, iris scans, gait, DNA and hand geometry.

See also: MULTI-FACTOR AUTHENTICATION, IDENTITY, IRIS SCANNING, GAIT RECOGNITION

Input Privacy

Input **privacy** is the enabling of **machine learning** to infer **information** from the input **data** without revealing the input to the **algorithm**. Where there is input privacy, complex computations can be carried out without requiring access to the input data.

Input privacy also applies to **secure multi-party computation**, where different parties contribute their own data to a machine learning effort but wish to keep their data confidential from their partners. Input privacy here means that the output of the learning, distributed to all the partners, is performed without direct access to partners' data.

Techniques include secure multi-party computation, where data remains in the custody of its owner or **data controller**, and the **data mining** algorithm either sends queries to the owners or takes only encrypted versions of the data as input. The algorithm may be run by the data owner over the data, and then passed to the analyst.

Further reading:

Ricciato, F., Bujnowska, A., Wirthmann, A., Hahn, M. and Barredo-Capelot, E., 2019. A reflection on privacy and data confidentiality in official statistics. *Presented at 62nd ISI World Statistics Conference*, https://ec.europa.eu/eurostat/cros/content/reflection-privacy-and-data-confidentiality-official-statistics-0_en.

See also: PRIVACY-PRESERVING MACHINE LEARNING, OUTPUT PRIVACY, INPUT STATISTICAL DISCLOURE CONTROL

Input Statistical Discloure Control

Where analysts have controlled access to data in a **safe setting**, **statistical disclosure** is controlled though both the inputs (the data that is stored in the safe setting) and the outputs (the results of the analysts' processing that might then be published).

See also: OUTPUT STATISTICAL DISCLOURE CONTROL, INPUT PRIVACY

Integrity

Integrity is a complex concept, referring to the properties of wholeness and consistency. Moral integrity applies to those who adhere to their, or society's, ethical principles, especially trustworthiness, honesty and openness.

Integrity often features in the privacy literature as a beneficial moral property supported by privacy. Edward Bloustein described the legal protection of privacy as part of the protection of a person's dignified personality, including their independence and integrity. The American Fifth Amendment, which protects people against incriminating themselves, was connected directly with people's religious integrity by Robert Gerstein, by conferring on them the right to set their own consciences in order.

In oppressive regimes – whether governmental or social – family life, associations and confidential relationships are undermined to leave no space for the individual to flourish. However, the so-called Nicodemite option of outward conformity still leaves space for the private individual's inner integrity. Named after the Biblical character Nicodemus, Nicodemites were originally those who professed one religion while practising another during the strife between Protestants and Catholics in Reformation Europe in the 16th and 17th centuries.

Further reading:

Bloustein, E.J., 1964. Privacy as an aspect of human dignity: an answer to Dean Prosser. *New York University Law Review*, 39, 962–1007, https://heinonline.org/HOL/LandingPage?handle=hein.journals/nylr39&div=71&id=&page=.

Gerstein, R.S., 1970. Privacy and self-incrimination. *Ethics*, 80(2), 87–101, https://doi.org/10.1086/291757.

See also: SELF, AUTONOMY, CONTEXTUAL INTEGRITY, TRUST, PRIVACY, RIGHT TO PRIVACY, DIGNITY, ETHICS

Intellectual Privacy

Intellectual privacy is a term that has been given to several types of **privacy**. Neil Richards, while arguing that free speech should usually trump privacy in an open society, makes an exception for the work of generating knowledge and ideas by speaking, reading and conversing with colleagues and opponents. The space for this vital activity he called intellectual privacy, although Anita Allen argued that this was merely a hybrid of **associational privacy** and **informational privacy**. Richards was supported by Koops et al., although they conceded that it was protected by rights that were distinct from traditional privacy rights. Allen herself used the term to mean something like **psychological privacy**.

Further reading:

Allen, A.L., 2011. *Unpopular privacy*: *what must we hide?* New York: Oxford University Press.

Koops, B.-J., Newell, B.C., Timan, T., Škorvánek, I., Chokrevski, T. and Galič, M., 2017. A typology of privacy. *University of Pennsylvania Journal of International Law*, 38(2), 483–575, https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1938&context=jil.

Richards, N., 2015. *Intellectual privacy*: *rethinking civil liberties in the digital age*. New York: Oxford University Press.

See also: IDEOLOGICAL PRIVACY, FREEDOM OF EXPRESSION, MENTAL PRIVACY

Intellectual Property (IP)

Intellectual property (IP) is the class of protected intangible products of creative activity. IP may be used to create material goods, but IP itself denotes the product of the intellectual work of design, invention or creation (it may be material, as with a sculpture, but more usually a distinction is made between an object and its associated IP, such as its design). IP is protected to provide incentives for innovation, so that it cannot be copied or used without permission of the IP owner (often under licence, for a fee), unless there is a **public interest** (for instance, a critic might quote short passages from a book). Protection generally lasts a period of time, often decades-long, after which it lapses. At this point, the IP becomes part of the **public domain**, and can be used or copied by others. Standard types of IP are *copyright* (the right to copy, distribute, perform or display a creative or artistic work), *patents* (the right to make or use an invention), *trade marks* (signs that denote products from a particular company or creator), and *trade **secrets*** (confidential **information** that is kept from the **public**, and which is used in a process).

The connections between IP law and **privacy** law are complex and convoluted. In their seminal paper of 1890, Warren and Brandeis identified many of the **common law** instruments that made up the **right to be let alone**, from IP doctrines, such as an author’s right to be the first publisher of unpublished material (copyright), and **breach of confidence** (a protection of trade secrets). Such doctrine, they argued, gave ordinary people some tools for preventing exposure of their **private life**. In general, both IP law and privacy law share the basic aim of creating rights of exclusion restricting the flow of information, to benefit creators, in the former, and subjects, in the latter. Some indeed have suggested that giving data subjects property rights over the **data** about them might be a means of addressing privacy problems, although this remains controversial and has rarely been tried. As property can easily be exchanged, it may result in a reduction of privacy, albeit alongside a profit for the **data subject**s.

However, the two may also clash, for example over a representation of one person by another (such as a photograph, a biography, **social media** content, a profile or a **dataset**). In the creative digital world, the intersection of the two areas of law has become ever more entangled, as **personal data** increasingly underlies innovative services and processes.

Further reading:

Liebenau, D., 2016. What intellectual property can learn from informational privacy, and vice versa. *Harvard Journal of Law and Technology*, 30(1), 285–307, https://jolt.law.harvard.edu/assets/articlePDFs/v30/30HarvJLTech285.pdf.

Samuelson, P., 2000. Privacy as intellectual property? *Stanford Law Review*, 52(5), 1125–73, https://doi.org/10.2307/1229511.

Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220, https://doi.org/10.2307/1321160.

See also: CONFIDENTIALITY, PRIVATE PROPERTY, DATA OWNERSHIP, INTELLECTUAL PRIVACY

Intentional Data

Data which are generated explicitly to be used as data, and not as a by-product of some other activity. Typical examples are **census**es and surveys.

Further reading:
Purdam, K. and Elliot, M., 2015. The changing social data landscape. In: Halfpenny, P. and Proctor, R., eds. Innovations in digital social research methods. London: Sage, 25–58. https://doi.org/10.4135/9781473920651.

See also: CONSEQUENTIAL DATA, DECLARED DATA

Intention–Behaviour Gap

See: ATTITUDE–BEHAVIOUR GAP

Interference

The intervention by a person, organisation or authority in another's body, image, actions, business or personal life. When the object of interference is a **natural person**, their **privacy** will be impacted.

In law, an interference with the right to privacy should be understood as distinct from a **breach** of privacy rights. An interference with the right means only that the right is engaged – the action in question touches on an individual's privacy to a degree recognised in law. For example, disclosing an individual's identifiable **information** will normally constitute an interference with their privacy.

An interference is only deemed a breach of privacy rights, however, if it is not justified. An interference justified on – for example – **public interest** grounds, and subject to safeguards ensuring legitimacy and **proportionality** (e.g., review by an **ethics committee**, or **compliance** with a privacy policy), will be legally justified, and thus not a breach of privacy or **confidentiality** rights.

See also: IDENTIFIABLE NATURAL PERSON, RIGHT TO PRIVACY

Internal Security Testing

Internal security testing is the practice of examining an organisation's networks, applications and **information** systems for flaws and vulnerabilities from within its **network**. It is often carried out by an internal **security** team or a security company that has been given permission by the company to work with a **third party**. **Penetration testing**, **vulnerability** scanning and **security audit**s are examples of procedures that are frequently used in conjunction with human and automated testing techniques.

Internal security testing is a crucial part of an organisation's overall security strategy, since it may identify and fix security flaws before an **adversary** can take advantage of them.

See also: SOCIAL ENGINEERING, INFORMATION SECURITY, NETWORK SECURITY

International Transfer

See also: DATA TRANSFER

Internet

The Internet (originally called the Internetwork) is a **network** of computer networks that uses the TCP/IP **Internet Protocol** suite to pass data across the network. Created in the 1960s by Vinton Cerf, Robert Kahn and others, originally as an academic and military tool, the Internet has now become a global network crucial to almost every walk of life, from commerce to healthcare, to entertainment, to warfare, to government, to scientific research. It is a decentralised permissionless network designed to scale easily as the number of users increases.

While the value of the Internet is incalculable, it also poses serious **privacy risk**s. The gathering and sharing of **personal data** online is a significant privacy issue. People leave digital traces while they use the Internet, which may be used to track their preferences, behaviours and online activities. As more activities move online, the great the prospect for their **datafication**. The shift online has been accelerated since the 1990s by the appearance of the **World Wide Web**, an application that uses the Internet and has adopted similar principles of decentralisation.

Furthermore, the Internet not only supports the creation and collection of data, but it also simultaneously provides an access point to it. This risk is increased by the fact that many websites and online services lack robust **security** mechanisms, leaving them open to **intrusion**, especially as the Internet was originally designed with **information** flow and good faith communication between like-minded people in mind, so that security was a secondary consideration from the outset.

Further reading:
Handley, M., 2006. Why the Internet only just works. *BT Technology Journal,* 24(3), 119–29, https://doi.org/10.1007/s10550-006-0084-z.
O'Hara, K. and Hall, W., 2021. *Four Internets*: *data, geopolitics*, *and the governance of cyberspace*. New York: Oxford University Press.

See also: BROWSER FINGERPRINTING, INTERNET OF PEOPLE, INTERNET OF THINGS, PROTOCOL

Internet of Humans

See: INTERNET OF PEOPLE

Internet of People

A developing idea known as the **Internet** of People (IoP) describes the blending of people, technologies and data to produce a more connected and customised experience. Although the IoP has numerous advantages, such as improved efficiency and convenience, it also poses serious **privacy** issues as the necessary gathering and sharing of **personal data** are intrinsic to an Internet of People.

The terms 'Internet of bodies' and 'Internet of humans' are also used.

Further reading:
Miranda, J., Mäkitalo, N., Garcia-Alonso, J., Berrocal, J., Mikkonen, T., Canal, C. and Murillo, J.M., 2015. From the Internet of Things to the Internet of People. *IEEE Internet Computing*, 19(2), 40–7, https://doi.org/10.1109/MIC.2015.24.

See also: INTERNET OF THINGS, PROFILING, DATA SHARING

Internet of Things

The **network** of actual physical items that are connected to the **Internet** and can gather and share data is known as the Internet of Things (IoT). The

IoT has a lot of advantages, including convenience and efficiency gains, but it also poses serious **privacy** issues. As more devices, which may include **sensor**s, smart home or **smart city** devices, wearable devices, **autonomous vehicle**s and their computer systems and networked medical devices, are linked, massive volumes of **data** on people's tastes, habits and activities are produced. Individuals and their behaviours may be thoroughly profiled using this **information**, which third parties may then utilise for **targeted advertising** or other uses.

Unauthorised access to **personal data** is another threat to privacy. With the IoT, there are more possible ports of entry for hackers as more devices and data are connected. The risk is increased by the fact that many IoT devices, which are often small and designed to operate with low power, lack robust **security** safeguards, leaving them open to **intrusion**. They often exist in networks of devices passing information between them, providing a broad attack surface. The security deficiencies of centralised **data storage** in such networks are now being countered by the promotion of **edge computing**, a more decentralised view putting security, storage and control nearer the devices themselves.

Further reading:

Ren, J., Dubois, D.J., Choffnes, D., Mandalari, A.M., Kolcun, R. and Haddadi, H., 2019. Information exposure from consumer IOT devices: A multidimensional, network-informed measurement approach. *In*: *Proceedings of the Internet Measurement Conference*, 267–79, https://doi.org/10.1145/3355369.3355577.

O'Hara, K., 2014. The fridge's brain sure ain't the icebox. *IEEE Internet Computing*, 18(6), 81–4, https://doi.org/10.1109/MIC.2014.122.

See also: SMART DEVICE, NETWORK SECURITY, WEARABLE TECH

Internet Protocol (IP)

The Internet Protocol (IP) is a **network** protocol belonging to the TCP/IP suite of Internet protocols on which the operation of the **Internet** is based. It is a network interconnection protocol (Inter-Networking Protocol), classified at the network layer of the ISO/OSI model, created to interconnect heterogeneous networks in terms of technology, performance, and management, above other link layer protocols, such as Ethernet or ATM.

It is a connectionless packet protocol of the best effort type in the sense that it does the maximum it can do without guaranteeing any form of **communication reliability** in terms of error control, flow control and congestion control. These functions therefore need to be provided by higher level transport protocols such as TCP for IP to work effectively.

A device or machine that is connected by IP to a network receives an **IP address**. These serve two main functions: they identify the host (e.g., a user's home router or a Web server for a particular website) and provide the location of the host in the network.

There are three main privacy concerns related to IP addresses. First, they can be used to track a user's online activity and location, because the IP address space is managed by a hierarchy of regional registries that associate an IP address with a small geographical area. Hence, a website or online service may log a user's IP address when they visit the site, enabling them to link that user's activity on the site to their physical location. Second, they can be used to identify a user's Internet Service Provider (ISP). In some cases, this information can be used to infer a person's **identity**. Third, IP addresses can be used to link **data** to build up a profile of a user's online behaviour by the owners of the servers they visit, facilitating **targeted advertising** or malicious purposes, via the cookie installation on the user's computer.

To mitigate these issues, some users install a **Virtual Private Network** to obscure their IP address and encrypt their Internet connection. Additionally, there are technologies (such as **TOR**) that can be used to further anonymise a person's online activity by routing their internet traffic through multiple nodes in a network. Privacy policies may also play a role, perhaps self-imposing rules on the collection of IP addresses. Businesses and website operators also must comply with regulations such as the **GDPR**, which regulates the use of IP addresses for **tracking** and **data sharing**, as they are classified as **personal data**, since users are identifiable from them.

Further reading:

International Standards Organisation, 1996. *ISO/IEC 7498-1*: *1994 information technology-open systems interconnection-basic reference model*: *the basic model. International Standard* ISOIEC, 74981, 59, www.iso.org/standard/20269.html.

See also: LOCATION TRACKING

Interoperability

Interoperability is a property of distinct systems that can work together. This can mean, for example, that they can transfer **data** between them, that they apply the same semantics to data or that their processing is coordinated. They may be designed for interoperability, for instance by using the same technical standards for their implementation.

Interoperability creates potential **privacy** issues when interoperable systems become able to share and link **information** about individuals, or to check whether information refers to the same individual. Interoperable stores of **big data** pose this problem acutely, particularly when **dataset**s are rendered interoperable after their creation. In such cases, the original security **protocol**s may not have treated **data linkage** as an issue.

Further reading:
National Institute of Standards and Technology, 2019. *Big data interoperability framework volume 4*: *security and privacy*. Gaithersburg: NIST, https://doi.org/10.6028/NIST.SP.1500-4r2.

See also: COMMUNICATION SECURITY, DATA TRANSFER, LINKABLE INFORMATION

Interval Publication

A method of statistical disclosure control whereby, instead of exact values, ranges of values are published with the real value falling somewhere within the range creating uncertainty for an **adversary**. This technique is most often used with tables of counts but in principle could be applied to any structured **data** format. Duncan et al observe that many other forms of **statistical disclosure control** (e.g., **rounding** or **cell suppression**) do in effect amount to interval **publication** as the bounds on cell value are a *de facto* interval.

Further reading:
Duncan, G.T., Elliot, M., and Salazar-González, J.J., 2011. *Statistical confidentiality*. New York: Springer, https://doi.org/10.1007/978-1-4419-7802-8.

Intimacy

Intimacy is a state of physical and/or mental closeness, and often emotional intensity, between people. **Privacy** is an important facilitator of intimate relationships, which are often seen as essential for human psychological well-being. Paradoxically, while intimate relationships require privacy for the intimates, they also entail reduced privacy *between* the intimates.

Julie Inness described intimacy as being based on care, liking and love and characterised by **consent**, **fairness** and mutuality, and suggested that the US Supreme Court’s privacy protections tended to focus on areas to do with the body and aspects of intimacy, including contraception, homosexual relations and **abortion** (writing in 1992, prior to the overturning of **Roe v Wade**). Dahrl Pedersen’s empirical investigation of people’s categorisations of privacy distinguished between intimacy between friends and intimacy within the family, although this seems to suggest a third category of sexual intimacy which is not properly covered by either of the two. Beate Rössler argued that intimacy was not a category of privacy in itself but instead involved a combination of **informational**, **decisional** and **spatial privacy**.

Further reading:

Inness, J., 1992. *Privacy, intimacy and isolation*. New York: Oxford University Press.
Pedersen, D.M., 1997. Psychological functions of privacy. *Journal of Environmental Psychology*, 17(2), 147–56, https://doi.org/10.1006/jevp.1997.0049.

See also: AUTONOMY, INVIOLATE PERSONALITY, MENTAL PRIVACY, PHYSICAL PRIVACY, PSYCHOLOGICAL PRIVACY, ISOLATION

Intranet

A private **network** used within an organisation, called an Intranet, promotes internal **communication**, teamwork and **information** exchange. It is like the Internet in operation, using World Wide Web and **Internet protocol**s such as TCP/IP, HTML and HTTP, but it is only available to authorised people within the company and may not be connected to the **Internet** at all. If it is connected, then it will be guarded by **firewall**s and other **security** measures to prevent unwanted access from outside the enterprise.

See also: INFORMATION SECURITY, NETWORK SECURITY

Intruder

An agent who seeks to invade some space determined to be private or confidential, thereby breaching **privacy** and/or **confidentiality**.

In a **data protection** context, an intruder is a type of **adversary** who attempts to disclose **information** either by identifying a **data subject**, or by **attribution**. Intruders may be motivated or inadvertent. An inadvertent intruder may simply stumble across information or accidentally recognise a **data** subject in a **dataset**. A **motivated intruder** is someone who is seeking information, usually for gain. Motivated intruders are also called opponents, enemies or **attacker**s.

A **motivated intruder test** assesses the risks of **identifiability** within a dataset. The **data controller** imagines (or models) an intruder with a reasonable set of skills and a motive for attacking the **database**: would such an adversary be able to achieve their goals?

Further reading:
Information Commissioner’s Office, 2012. *Anonymisation*: *managing data protection risk code of practice*, https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

See also: INFORMATIONAL PRIVACY, INTRUDER TESTING

Intruder Testing

A variant of **penetration testing** where the object of the simulated **attack** is one or more **dataset**s rather than an organisation’s cybersystems. Typically, the tester will be using an auxiliary dataset and/or publicly available **information** to attack a dataset which has been de-identified/anonymised, to establish if it is possible to reidentify individuals within that dataset. The attack will involve some form of **linkage** between the auxiliary information and the de-identified dataset. Generally, **intruder** testing should be tied to a well-formed **scenario analysis**.

Further reading:
Elliot, M., Mackey, E., O’Shea, S., Tudor, C., and Spicer, K., 2016. End user licence to open government data? A simulated penetration attack on two social survey datasets. *Journal of Official Statistic*s, 32(2), 329–48. https://doi.org/10.1515/jos-2016-0019.
Narayanan, A. and Shmatikov, V., 2009. De-anonymizing social networks. *In*: *Proceedings of the 2009 IEEE Symposium on Security and Privacy*, 173–87, http://dx.doi.org/10.1109/Sp.2009.22.

See also: RECORD LINKAGE, ATTACK, LINKAGE ATTACK

Intrusion

An act of entering a space (physical or virtual) or becoming involved in a situation against the wishes of owners or other stakeholders of that space or situation. Most intrusions involve a breach of privacy and/or security.

Intrusion Detection System (IDS)

A **security** tool called an **intrusion** detection system (IDS) is used to watch **network** traffic for unauthorised or suspicious behaviour posing a threat to a system. By examining network traffic and finding patterns or abnormalities that would signal a threat, it is intended to detect possible security **breach**es and take appropriate action. IDSs may be host-based or network-based. Installed on certain PCs or servers, host-based IDSs keep an eye on the activities affecting them directly. Network-based IDSs are placed on a network and track the traffic that moves through it. An IDS can sound an alarm or send a message to administrators or security staff when it discovers anomalous activity.

IDS systems can spot a variety of security risks, including **virus**es, **malware**, **hacking** attempts and unauthorised access. They are a crucial tool for defending against cyberattacks and safeguarding networks, and they are frequently used in conjunction with other security technologies such as **firewall**s and **anti-virus software**.

Further reading:

Ashoor, A.S. and Gore, S., 2011. Importance of intrusion detection system (IDS). *International Journal of Scientific and Engineering Research*, 2(1), 1–4, www.ijser.org/researchpaper/importance_of_intrusion_detection_system.pdf.

See also: ATTACK, NETWORK SECURITY

Intrusion Prevention System (IPS)

An **intrusion** prevention system (IPS) is a security tool used to proactively block and reduce unwanted traffic to address **network security** concerns. It performs real-time network traffic analysis to compare with a set of predetermined security rules or policies. If the system notices suspicious activity, it may respond right away to restrict the traffic.

In contrast to **intrusion detection system**s (IDS), which simply notify administrators of possible security concerns, IPS systems are more proactive, blocking traffic, isolating infected computers, and stopping malicious **software**, such as **malware**, **virus**es, **worm**s and **botnet**s, before it has a chance to do any damage. They frequently function in tandem with other security tools like **firewall**s and **anti-virus software** to offer a thorough and layered defence against network security threats.

Further reading:
Xinyou, Z., Chengzhong, L. and Wenbin, Z., 2004. Intrusion prevention system design. *In*: *The Fourth International Conference on Computer and Information Technology*, 386–90, https://doi.org/10.1109/CIT.2004.1357226.

See also: LAYERED SECURITY MODEL

Intrusion upon Seclusion

Intrusion upon **seclusion** is the first of William Prosser's four **privacy tort**s. In an influential paper of 1960, Prosser argued against the salience of the **right to be let alone**, traced in American law by Warren and Brandeis. He claimed instead that the **privacy** torts that actually existed in law did not furnish a broader principle of integrated coverage of a right to be let alone but were instead a set of four discrete and discontinuous protections.

Intrusion upon seclusion includes the examples concerning **publication** that Warren and Brandeis exhibited, as well as physical intrusions into private spaces, and those using technology such as microphones. He described the tort as protecting a mental interest that bridged the gaps between **trespass**, nuisance and infliction of mental distress, as well as constitutional protections against government **interference**. However, not all such intrusions would be protected: Prosser gave the example of police fingerprinting of suspected criminals.

Further reading:
Prosser, W.L., 1960. Privacy. *California Law Review*, 48, 383–23, https://doi.org/10.2307/3478805.
Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220, https://doi.org/10.2307/1321160.

See also: RIGHT TO PRIVACY

Invasive BCI

A form of Brain–Computer Interface requiring surgery. Invasive BCIs usually involve the implantation of electrodes and sensors into the brain to detect and stimulate brain activity.

Further reading:
Straw, I., Ashworth, C., and Radford, N., 2022. When brain devices go wrong: a patient with a malfunctioning deep brain stimulator (DBS) presents to the

emergency department. *BMJ Case Reports CP*, 15(12), e252305. https://doi.org/10.1136%2Fbcr-2022-252305.

See also: BRAIN–COMPUTER INTERFACE

Inversion Attack

A sort of cryptographic **attack** known as an inversion attack aims to undo the **encryption** process to recover the original **plaintext** from a given **ciphertext**. In other words, an inversion attack works backwards from the ciphertext to try to uncover the **decryption** key. The two basic categories of inversion attacks are cryptanalysis attacks and brute force attacks. In a **brute force attack**, every decryption key is tried until the right one is discovered. This can be time- and resource-consuming, particularly for longer keys or intricate encryption schemes.

On the other side, cryptanalysis attacks use mathematical methods to examine the encryption algorithm and identify any flaws or vulnerabilities that may be used to retrieve the plaintext. While requiring specific knowledge and experience in **cryptography**, **cryptanalysis** assaults have the potential to be more efficient and successful than brute-force operations. The **security** of encrypted **data** is seriously threatened by inversion attacks, which may also be used to undermine a system's integrity or steal important data. It is crucial to employ robust **encryption algorithm**s with suitably long keys.

Further reading:

Fredrikson, M., Jha, S. and Ristenpart, T., 2015. Model inversion attacks that exploit confidence information and basic countermeasures. *In*: *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 1322–33. https://doi.org/10.1145/2810103.2813677.

See also: CRYPTOGRAPHIC KEY, ENCRYPTION KEY, INTEGRITY

Inviolate Personality

Samuel Warren and Louis Brandeis' seminal paper on rights to **privacy** in American **common law** traced the evolution of a right to life from a basic set of rights to **security** and protections from physical harms to broader protections of what they called an 'inviolate personality', including the **right to be let alone**.

The roots of inviolate personality were in Romantic thought about individuals' struggles to assert themselves against social pressures, and, following Warren and Brandeis, evolved further in the legal literature. Edward Bloustein saw it as the link connecting the Prosser **privacy tort**s, Fourth Amendment constitutional protections against **unreasonable search**es, and Brandeis' dissent in *Olmstead* (a 1928 US Supreme Court judgment that affirmed the constitutionality of **wiretapping** without a **search** warrant, overturned in 1967). He argued that it was the moral basis for the law's support of the individual's independence, **dignity** and **integrity** as a unique and self-determining being, which would be undermined if they became a public spectacle. Hence, on this reading, a **breach** of privacy goes beyond contingent harms such as distress and mental suffering and is instead an **intrusion** into the dignity of the person.

Further reading:

Bloustein, E.J., 1964. Privacy as an aspect of human dignity: an answer to Dean Prosser. *New York University Law Review*, 39, 962–1007, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3537968.

Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220, https://doi.org/10.2307/1321160.

See also: DIGNITY, MENTAL PRIVACY, PERSONHOOD, RIGHT TO PRIVACY, TELEPHONE TAPPING

Invisible Computing

See also: UBIQUITOUS COMPUTING

Invisible Internet Project

See also: I2P

IP

Can stand for **intellectual property** or **Internet Protocol**.

IP Address

See also: INTERNET PROTOCOL

IPS

See also: INTRUSION PREVENTION SYSTEM

IPSE-Identity

In his 1990 volume *Oneself as Another*, French philosopher Paul Ricoeur criticised simplistic views of **identity**, which ignored identity’s self-referential nature, and introduced a distinction between *ipse-identity* and **idem-identity**. Ipse-identity is the self as a reflexive structure, which exists via reference to itself, a subjective, internal, first-person understanding of the self, including the values and norms to which the person adheres and professes fidelity.

Further reading:
Ricoeur, P., 1994. *Oneself as another*. Chicago: University of Chicago Press.

See also: PERSONHOOD

Iris Scanning

Iris scanning involves taking a high-resolution image of the iris with a specialised camera or scanner, analysing its patterns with **software** and comparing those findings to a **database** of recognised patterns, to identify an individual. Since iris patterns are so distinctive and challenging to copy or fake, iris scanning is regarded as a highly accurate method of biometric identification. It is also more hygienic and user-friendly than many other biometric identification methods since it is non-invasive and does not involve direct physical contact with the individual being identified.

Applications for iris scanning include **access control** systems, border and immigration management, and law enforcement. It is also utilised in a few mobile devices, including tablets and smartphones, enabling mobile payments and secure **authentication**. Despite its efficacy and **accuracy**, iris scanning raises questions regarding **security** and **privacy**, which are common to all types of handling and storage of **biometric data**. However, because it requires the explicit cooperation of the individual involved, it cannot be used for covert **surveillance**, and does ensure at least some kind of **consent**.

Further reading:
Sanderson, S. and Erbetta, J., 2000. Authentication for secure environments based on iris scanning technology. *In*: *IEEE Colloquium on Visual Biometrics*, https://doi.org/10.1049/ic:20000468.

See also: IDENTITY

Irreversibility

See also: REVERSIBILITY

ISO27001

An international **standard** for **information** security management systems. It provides a framework for organisations to manage sensitive **data** within their locus of responsibility.

ISO 27001 is part of the ISO/IEC 27000 family of information security standards, which have been developed to formalise best security practices. The standard provides a risk-based approach, requiring organisations to identify and assess **information security** risks, and implement appropriate measures to mitigate those risks.

ISO 27001 certification, which involves a formal assessment of the organisation by an accredited certification body is used by organisations that wish (or are required) to demonstrate the **trust**worthiness of their information systems.

Further reading:
International Standards Organisation, 2022. *ISO/IEC 27001 Information security management systems*, www.iso.org/standard/27001.

See also: CYBERSECURITY ISO27002, SECURITY INFORMATION MANAGEMENT

ISO27002

A companion to **ISO27001** that provides a comprehensive description of best **information security** management practices ranging from security policies and **asset** management to cryptographic practice.

Further reading:
International Standards Organisation, 2022. *ISO/IEC 27001 Information security management systems*, www.iso.org/standard/27002.

See also: CYBERSECURITY, CRYPTOGRAPHY, SECURITY INFORMATION MANAGEMENT

Isolation

Isolation is the act of ensuring no contact between a person or object and **other** people or things. While some may prefer to isolate themselves, the term carries an implication of unwillingness on the part of the isolatee, and the **privacy** literature tends to treat it as negative pathological (in contrast to **solitude**, to which it is structurally similar). For example, Altman described isolation as the state of having more privacy than is desired. Many argue that unwilling isolation (e.g., of Robinson Crusoe) is not privacy at all. However, others, such as Pedersen, have pointed out positive aspects of isolation: it aids recovery from adverse social experience, and facilitates desired behaviour that others disapprove of (e.g., smoking). Small groups may wish to isolate themselves to have confidential discussions.

The term has specific uses in several fields. In medicine, it means separating a patient from others to prevent contagion; the patient may be held in an *isolation ward*. Emotional isolation is a psychological condition where there is no one in whom one feels able to confide. In computing, resources are placed in isolation to prevent them interacting; for example, in a **database**, **data** held in isolation can only be accessed by one user at a time, to prevent conflicts occurring. In security context, **trusted execution environment**s and **sandbox**es are examples of approaches to keeping code isolated to avoid corruption of resources while allowing testing and sanitisation to take place.

Further reading:
Altman, I., 1975. *The environment and social behavior*: *privacy, personal space, territory, crowding*. Monterey: Brooks/Cole.

Pedersen, D.M., 1997. Psychological functions of privacy. *Journal of Environmental Psychology*, 17(2), 147–56, https://doi.org/10.1006/jevp.1997.0049.

See also: PERSONHOOD, PERSONAL SPACE, PSYCHOLOGICAL PRIVACY

J

Jensen–Shannon Divergence

A measure of dissimilarity between two probability distributions.

It is sometimes used in the context of **statistical disclosure control** to quantify the difference between two versions of a **dataset**: one before **disclosure control methods** have been applied and another afterwards. This is one measure of the **information loss** caused by the disclosure control.

See also: DATA UTILITY

Jigsaw Identification

Jigsaw identification, a term which originated in journalism, is the process of **identifying** individuals or attributing some previously unknown **attribute** to an individual, by bringing together two or more pieces of **information** that were previously kept separate (also known as *mosaic identification*). As an example, if a crime victim is referred to in one news report as being a Member of Parliament, and in another as living in a particular village, then together the two reports may well enable the victim's identity to be revealed via further **public** information (the registered address of the MP). In a visual example, a person's face may be pixelated in a photo, but they may be seen getting into the driver's seat of a car identifiable by its number plate; the person may then be identified via the car.

Further reading:

Narayanan, A. and Shmatikov, V., 2010. Myths and fallacies of 'Personally Identifiable Information'. *Communications of the ACM*, 53(6), 24–6, https://doi.org/10.1145/1743546.1743558.

Wilson, J., 1996. *Understanding journalism: a guide to issues*. London: Routledge.

See also: DEANONYMISATION, REIDENTIFICATION

Joint Data Controller

Under the **GDPR**, a person or organisation who determines why and how **personal data** are used is known as a **data controller**. Where they share responsibility for making these decisions with others, these two or more entities are known as joint data controllers. Under Article 26, they must make **transparent** arrangements to determine their respective obligations to collectively achieve GDPR **compliance**.

Further reading:
UK Information Commissioner's Office, n.d. *What does it mean if you are joint data controllers?* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/controllers-and-processors/what-does-it-mean-if-you-are-joint-controllers/.

Jurisdiction

The national state with power to regulate a particular entity or activity is said to have jurisdiction. It is a legal sphere of competence usually determined by geography, and where the activity in question takes place.

In the context of digital information, **data processing** can easily cross national boundaries. This can make it difficult to identify the appropriate jurisdiction to deal with legal matters of privacy and **data protection**. For this reason, the EU's **GDPR** has multiple rules to help determine its **territorial scope** (i.e., when the **personal data** of EU residents is processed for marketing or **tracking**, anywhere in the world), as well as how national regulators within the EU should cooperate when processing affects citizens in both their respective countries.

Further reading:
Hörnle, J., 2021. Data protection regulation and jurisdiction. *In*: *Internet Jurisdiction Law and Practice.* Oxford: Oxford University Press, 233–63.

See also: CONSISTENCY MECHANISM, CROSS-BORDER DATA PROCESSING, INTERNET, LEAD SUPERVISORY AUTHORITY, REGULATORS, SUPERVISORY AUTHORITY

Just-In-Time Consent

The collection of **consent** from a **data subject** at the point at which a new piece of **data processing** is about to happen. Previously, a logistic challenge, online connectivity has made this increasingly viable. An example is the **cookie** notice which allows individuals to opt out of some or all cookies.

In principle, a development of **personal data store**s could enable the possibility of semi-automated just in time consent systems which would operate using **AI** sub-systems acting as personalised **privacy avatar**s to manage each individual’s **data sharing** according to their **privacy preference**s.

Further reading:

Van Kleek, M. and O’Hara, K., 2014. The future of social is personal: the potential of the personal data store. *In*: Miorandi, D., Maltese, V., Rovatsos, M., Nijholt, A. and Stewart, J., eds. *Social collective intelligence: combining the powers of humans and machines to build a smarter society*, Cham: Springer, 125–58, https://doi.org/10.1007/978-3-319-08681-1_7.

Vickers, A.J., Young-Afat, D.A., Ehdaie, B. and Kim, S.Y., 2018. Just-in-time consent: the ethical case for an alternative to traditional informed consent in randomized trials comparing an experimental intervention with usual care. *Clinical Trials*, 15(1), 3–8, https://doi.org/10.1177/1740774517746610.

Just-In-Time Notice

A **privacy notice** which provides **data subject**s a brief message explaining how the **information** they are about to provide will be used at the point when it is collected. In principle, this allows for more nuanced decision-making by data subjects. In practice, the evidence is that most data subjects dismiss them without reading.

Further reading:

Information Commissioner’s Office, 2018. The right to be informed, https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed-1-0.pdf.

See also: RIGHT TO BE INFORMED

K

K-Anonymity

A **standard** developed by Samarati and Sweeney, that there be at least *k* **data unit**s within a **dataset** that have the same combination of (specified) **indirect identifiers**. Sometimes termed as using a threshold of *k* (typically *k* = 3 or *k* = 5). *k*-anonymity can be implemented using a variety of **disclosure control methods**, including **microaggregation** and **domain generalisation**.

k-anonymity on its own is now regarded as being too weak in the protection it provides. In particular, all *k* units within a given **equivalence class** may share the value of a **target variable** allowing **attribute disclosure**. This led to the development of a suite of companion measures such as ***l*-diversity** and ***t*-closeness**. In common with other **statistical disclosure control** methods, *k*-anonymity also assumes a complete understanding of the indirect identifiers available to an **adversary**.

Further reading:

Samarati, P. and Sweeney, L., 1998. Protecting privacy when disclosing information: *k*-anonymity and its enforcement through generalization and suppression, https://dataprivacylab.org/dataprivacy/projects/kanonymity/paper3.pdf.

Sweeney, L., 2002. k-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05), 557–70, https://doi.org/10.1142/S0218488502001648.

Key Disclosure

See: MANDATORY DECRYPTION

Key Logger

See: KEY LOGGING

Key Logging

The collection and recording of each keystroke made on a keyboard, (potentially including **password**s, credit card numbers, and other **confidential** data). Key loggers are sometimes used for legal purposes – such as a computer owner monitoring use – and also have a role to play in law enforcement (equivalent to **wiretapping**), but perhaps more common is their use as a form of **spyware** to capture **personal data**.

Keyloggers can be installed using various methods: as hardware that is attached to the keyboard or the computer, or **malware** that infects a computer carrying the key logging **software** as payload.

Further reading:
Sagiroglu, S. and Canbek, G., 2009. Keyloggers: increasing threats to computer security and privacy. *IEEE Technology and Society Magazine*, 28(3), 10–17, https://doi.org/10.1109/MTS.2009.934159.

See also: ATTACKER, DATA CAPTURE, EAVESDROPPING ATTACK, SPYWARE

Key Variable

An operationalisation of the concept of an **indirect identifier**; a variable common to two (or more) **dataset**s, which may therefore be used for **record linkage** between them. More generally, in **scenario analysis**, a variable that is likely to be captured in **data** available to the **adversary**.

Further reading:
Elliot, M. and Dale, A., 1999. Scenarios of attack: the data intruder’s perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring), 6–10.

See also: DATA LINKAGE, DISCLOSURE RISK, STATISTICAL DISCLOSURE

Knowledge Economy

See: DIGITAL ECONOMY

Kompromat

Kompromat is a style of governance based on **blackmail**, with the state or other political actors holding a threat of releasing compromising **information** about **public figure**s, to ensure compliant behaviour, or to remove political opponents. It is particularly suited to non-democratic and corrupt states with the capability for large-scale **surveillance**, especially of elites. To function effectively in such states, political actors need to take part in corrupt activities, but then evidence of those activities can be held in dossiers to ensure future **compliance**, and so the system tends to sustain itself.

Further reading:

Choy, J.P., 2020. *Kompromat*: a theory of blackmail as a system of governance. *Journal of Development Economics*, 147, 102535, https://doi.org/10.1016/j.jdeveco.2020.102535.

See also: PUBLIC DISCLOSURE OF PRIVATE FACTS, PUBLICITY, REPUTATION

L

Laplace Noise

The Laplace distribution is a probability distribution which is derived from the exponential distribution. Laplace noise is therefore Laplace-distributed noise added to a function or **data**. Its **privacy** relevance is that it is the most used **noise addition** method associated with **differential privacy** because of its natural relationship with the **epsilon** parameter.

Large Language Model

See: GENERATIVE AI

Lawful Basis

Under the EU **GDPR**, as in the preceding **Data Protection Directive**, the processing of **personal data** must be carried out with reliance on an identified legal basis. These bases for **data processing** are listed at Article 6 of the GDPR, and include **consent**, **public interest**, **legitimate interest** and contractual purposes. Ausloos credits the 1970 *Hessische*s *Datenschutzgesetz* with being the first regional **data protection** legislation and points out that it also contained the negative default rule, whereby data processing always constitutes **interference** with personal rights to privacy requiring legal legitimation.

The term 'lawful basis' can be used more generally for any act requiring legal justification. The **disclosure** of private **information** or confidential information, or the transfer of **personal information** to another **jurisdiction**, may also require a further basis in law, aside from the six legal bases specified in the GDPR.

Further reading:

Ausloos, J., 2020. *The right to erasure in EU data protection law*. Oxford: Oxford University Press.

See also: DATA TRANSFER, IDENTIFIED DATA

Lawfulness

Lawfulness is part of the first principle of the **GDPR**: that **personal data** should be processed lawfully, **fairly** and in a **transparent** manner. Essentially, it requires that the processing of personal data be carried out in accordance with all applicable laws, be they national or international. Lawful processing will therefore be ascertained according to different laws depending on where the processing takes place and the people involved.

The lawfulness requirement of the GDPR thus imports other laws into its provisions. If, for example, it is a requirement of national legislation to obtain **consent** for a research project, then this will be necessary for processing to comply with the GDPR principle of lawfulness (even if the legal basis for processing under the Regulation itself is not consent – a point which can cause significant confusion).

A further layer of complexity stems from the fact that EU Member States also signatory to the **European Convention on Human Rights (ECHR)** have a duty to interpret their domestic laws in light of the Convention, and the associated jurisprudence of the European Court of Human Rights. Therefore, the GDPR’s incorporation of domestic law also imports transnational principles of fundamental rights, which are at the same time mirrored in the GDPR’s reference to the EU **Charter on Fundamental Rights** in its Recitals.

To complete the cycle of circularity, the ECHR in turn requires any interference with Article 8 privacy rights to be ‘in accordance with the law’, which in turn refers to a basis for **interference** which could be rooted in national law.

Lawfulness thus requires **compliance** with a complex web of mutually influential legal requirements – a quest to find coherence in multiple intersections of law – and is not necessarily the straightforward question of obedience it might initially seem. In practice, many **data controller**s will rely on guidance from their **Supervisory Authority** rather than attempting to distil this legal matrix from first principles.

Further reading:
Bjorge, E., 2015. *Domestic application of the ECHR: courts as faithful trustees*. Oxford: Oxford University Press.

See also: CHARTER RIGHTS, LAWFUL BASIS

Layered Notice

EU **data protection** law, most notably the **GDPR**, requires a significant amount of technical **information** to be disclosed to **data subject**s about the processing of their information under Articles 13 and 14. Somewhat paradoxically, the GDPR also requires this information to be communicated in a way which is meaningful and accessible to data subjects under Article 12.

Layered notices are one way to reconcile these two aims of the GDPR: to convey detailed information (e.g., the **lawful basis** for processing, categories of recipients and the **Data Protection Officer**'s contact information) in a way that does not overwhelm the data subject. The top layer of information which is immediately visible to the reader gives the highlights, with additional detail available through hyperlinks and click-throughs. The layered approach has emerged as a best-practice model in the UK since the introduction of the GDPR in 2018.

Further reading:
Information Commissioner’s Office, 2023. *What methods can we use to provide privacy information?* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/what-methods-can-we-use-to-provide-privacy-information/.

See also: TRANSPARENCY

Layered Security Model

A **security** strategy that employs many levels of defence to thwart possible threats and **attack**s. Each new layer of security offers a different type of defence, making it more challenging for an **adversary** to breach the entire system. Threats are mitigated by the different levels, and the layering means that a single point of failure is avoided. Digital security measures like **firewall**s, **intrusion detection system**s and **anti-virus software** might be layers in a layered security model. The security system should be comprehensive and well coordinated, with each layer of protection intended to strengthen and supplement the others.

Layered security does not mean perfect security, as was aptly dramatised in the now classic film *Die Hard*, in which the **adversary** had a plan for each of the first six layers of security around a bank vault, with the seventh being circumvented by the FBI’s well-meaning decision to cut power to the building.

Further reading:

Hong, J.B. and Kim, D.S., 2016. Towards scalable security analysis using multi-layered security models. *Journal of Network and Computer Applications*, 75, 156–68, https://doi.org/10.1016/j.jnca.2016.08.024.

See also: ACCESS CONTROL, NETWORK SECURITY

LBS

See: LOCATION-BASED SERVICE

L-Diversity

One of the ***k*-anonymity** family of privacy models. Simple *l*-diversity requires that with each **equivalence class** of a set of **key variable**s there are at least *l* possible values of a **target variable**. More complex forms, such as entropy *l*-diversity, make more nuanced distributional assumptions about the target variable.

Further reading:

Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M., 2007. L-diversity: privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1), https://doi.org/10.1145/1217299.1217302.

Lead Supervisory Authority

The EU **GDPR** has a rule commonly known as the **one-stop shop** to determine which national **regulator** takes the lead in **cross-border data processing** cases. The country in which the **data controller** has its **main establishment** shall have **jurisdiction**, meaning that its **data protection** regulator becomes the lead supervisory authority. The regulators in other affected countries have a duty to cooperate as concerned supervisory authorities.

Further reading:

Chemlali, L., 2022. The competence of non-lead supervisory authority under the EU GDPR’s one-stop-shop mechanism: CJEU judgment in Facebook and others (C-645/19). *The Journal of Media Law*, 14(2), 208–17, https://doi.org/10.1080/17577632.2022.2109852.

See also: CONSISTENCY MECHANISM, DATA PROTECTION AUTHORITY, SUPERVISORY AUTHORITY

Least Privilege

The principle of least privilege in an **information** system is the idea that every **user**, process or agent should only be able to access the information that it needs to carry out its legitimate function. Exactly how this is assessed can be quite vague and may depend on how the system is defined. Nevertheless, it is accepted as an important principle in system design that has a range of efficiency and **security** benefits.

In **privacy** terms, the principle implies that no one in a system should be able to access more **data** about others than they need to do their legitimate job. For instance, to determine whether a user is over 18, it is not necessary to know their birthdate, only to receive a yes/no answer to the question 'Are you over 18?'

Further reading:

Saltzer, J.H., 1974. Protection and the control of information sharing in multics. *Communications of the ACM*, 17(7), 388–402, https://doi.org/10.1145/361011.361067.

See also: ACCESS CONTROL, CONFIDENTIALITY, DATA MINIMISATION, DATA MINIMISATION PRINCIPLE, MANAGEMENT INFORMATION SYSTEM, NEED TO KNOW

Legal Basis For Processing

See: LAWFUL BASIS

Legitimate Interest

Under Article 6 of the EU **GDPR**, the legitimate interest of the **data controller** is one of the six legal bases on which **personal data** may be processed. One of these bases must be satisfied for the personal **data processing** to be lawful.

There is no definition of a legitimate interest; a potentially infinite number of reasonable purposes could be considered legitimate purposes for processing personal data. It is therefore one of the more flexible bases for processing personal data, particularly in the **private sector**. It is a common basis for marketing uses of personal data, as well as for maintaining **network security**.

Despite the flexibility of the basis, the data controller is still required to weigh up their legitimate interests in performing the processing against the

rights **to privacy** and freedoms of the **data subject**. A Legitimate Interests Assessment is one template exercise that **data protection** practitioners have developed to weigh up any risks of processing against the importance of interest pursued. A completed LIA is one way of demonstrating **compliance** with the GDPR per the **accountability** principle.

Further reading:
Shervin, N., 2021. The principles of data protection. *In*: Room, S., ed., *Data protection and compliance*, 2nd edition. Swindon: BCS Learning and Development, 101–19.

See also: LAWFUL BASIS

Libel

See: DEFAMATION

Licence Agreement

A licence agreement is a contract between two parties, in which the holder of property rights (e.g., copyright, **database** rights or trade **secret**s) in an **asset** grants the licensee permission to use, subject to specified terms. The assets in question could be **software**, a trade mark or a **dataset**.

Information subject to property rights may also identify living **natural person**s, triggering the additional application of **privacy** and **data protection** laws. Although a licence agreement is fundamentally a contractual regulation of property law rights, the rights to privacy of a **third party** can also be considered within its terms. When **personal data** is licensed to another party, therefore, the licensor should also consider the rights of the **data subject**s within the agreement. The licence agreement may therefore include **remedies** for improper **disclosure** or use of the data. Licence agreements may also be referred to as *data use agreement*s and **data sharing agreement**s, although these latter names are more colloquial, and do not necessarily signify that any property rights are at stake.

Further reading:
Publications Office of the European Union, 2021. Commercialising intellectual property: licence agreements. https://op.europa.eu/en/publication-detail/-/publication/e510929d-f015-11eb-a71c-01aa75ed71a1/language-en.

See also: DATA GOVERNANCE, DATA IN USE, DATA OWNERSHIP, DATA SHARING, DISCLOSURE, INTELLECTUAL PROPERTY

Lifecasting

See: LIFESTREAMING

Lifelogging

Lifelogging is the practice of capturing, recording, storing and retrieving details of one's personal life as it unfolds using digital technology. It combines the retrospective glance of the diary, with the objective perspective of the device(s) used to capture the information. The **information** can be deliberately and consciously created, for example by storing digital photos, blogs, tweets, emails, texts or messages. Or devices can be used to gather information independently of the lifelogger's volition, such as **wearable** medical **sensor**s; cameras such as Microsoft's SenseCam or Google Glass, which can record scenes; or smartphones, which log **location data** and positional data as well as calls and downloads.

The whole collection of information, called the *lifelog*, presents a detailed and fine-grained record of someone's life while the devices are being used. This presents several potential **privacy** issues, both with respect to the **security** of the lifelog itself, which could compromise the lifelogger's privacy, and with respect to the **privacy** of others, some of whose activities may be recorded.

Further reading:

Bell, G. and Gemmell, J., 2009. *Total recall: how the e-memory revolution will change everything*. New York: Penguin.

Gurrin, C., Smeaton, A.F. and Doherty, A.R., 2014. LifeLogging: personal big data. *Foundations and Trends in Information Retrieval*, 8(1), 1–125, http://dx.doi.org/10.1561/1500000033.

O'Hara, K., Tuffield, M.M. and Shadbolt, N., 2008. Lifelogging: privacy and empowerment with memories for life. *Identity in the Information Society*, 1(2), 155–72, https://doi.org/10.1007/s12394-009-0008-4.

See also: AUGMENTED REALITY, BIG DATA, DATA CAPTURE, DIGITAL FOOTPRINT, DIGITAL IDENTITY, IDENTITY, LIFESTREAMING, PERSONAL DATA, PERSONAL DATA STORE, PERSONAL INFORMATION, PERSONAL INFORMATION MANAGEMENT SYSTEM, RECORD, SOUSVEILLANCE

Lifestreaming

The process of **lifelogging** in a single place. This is invariably an online activity but lifestreams could be constructed using physical media.

A particular form of lifestreaming is lifecasting, which is the continuous video recording of a person's daily activity using **wearable** audio-visual equipment.

Lifestreams could in principle be **private**, but an individual's decision to **publish** their lifestream turns it into a *de facto* **public record**. It also makes the **data** available for **data mining** and is effectively a form of deliberate creation of a **big data** representation of oneself.

Further reading:
Selke, S., ed., 2016. *Lifelogging: digital self-tracking and lifelogging: between disruptive technology and cultural transformation*. Cham: Springer.

See also: DATA CAPTURE, DIGITAL FOOTPRINT, DIGITAL IDENTITY, IDENTITY, SOUSVEILLANCE

Linkability

Defined by the EUs **Article 29 Working Party** as 'the ability to link, at least, two records concerning the same **data subject** or a group of data subjects (either in the same **database** or in two different databases)'. In terms of **record linkage** methodology this would correspond to a link between two **record**s that has a high probability of being a match.

It should be stressed that linkage is a *bona fide* analytical technique that can be used by analysts for benign statistical purposes. However, linkage is also the main tool available to **adversar**ies for **reidentification attack**s and therefore linkability implies **reidentification risk**.

Further reading:
EU Article 29 Data Protection Working Party, 2014. *Opinion 05/2014 on anonymisation techniques*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

See also: ANONYMISATION, DATA LINKAGE, DATA PROTECTION

Linkable Information

Some **information** about a **population unit** (e.g., a **natural person** or a group of people) which could in principle be linked to a **record** of the population units (within a **de-identified dataset**).

See also: AUXILIARY KNOWLEDGE, LINKABILITY, RECORD LINKAGE

Linkage

See: DATA LINKAGE

Linkage Attack

The paradigmatic *modus operandi* of a **reidentification attack** or **attribute disclosure** attack. The **adversary** possesses some auxiliary information about relevant **population unit**s (i.e., those which *could* have contributed to the **target dataset** whether they have done so or not) and that **information** overlaps with the **data** in the target dataset. The overlap – sometimes referred to as the **key variable**s – is then used to link the auxiliary information to the target dataset. A linkage attack is an adversarial form of **data linkage** or **record linkage** and uses a similar set of techniques.

Most **motivated intruder test**s take the form of **red team** linkage attacks.

Further reading:

Ahmed, N., Michelin, R.A., Xue, W., Ruj, S., Malaney, R., Kanhere, S.S., Seneviratne, A., Hu, W., Janicke, H. and Jha, S.K., 2020. A survey of COVID-19 contact tracing apps. *IEEE Access*, 8, 134577–601, https://doi.org/10.1109/ACCESS.2020.3010226.

Elliot, M., Mackey, E., O'Shea, S., Tudor, C. and Spicer, K., 2016. End user licence to open government data? A simulated penetration attack on two social survey datasets. *Journal of Official Statistics*, 32(2), 329–48, https://doi.org/10.1515/jos-2016-0019.

Merener, M.M., 2012. Theoretical results on de-anonymization via linkage attacks. *Transaction*s *on Data Privacy*, 5(2), 377–402, http://tdp.cat/issues11/tdp.a074a11.pdf.

See also: ATTRIBUTION, REIDENTIFICATION

Link Encryption

Link encryption is a means of protecting **communication**s by **encrypting** all packets of data between any two points on a **network**. As the packets are received at a node, they must be decrypted, to allow the **header information** to be read, so that the next destination of the packet can be determined, and the packet sent. After the header **metadata** is read, the packet is encrypted again and sent on.

This contrasts with **end-to-end encryption**, where only the packet contents are encrypted, and the header metadata sent in the clear, so it can be read by the intermediate network nodes. The advantage of link encryption is that traffic analysis is made far more difficult because the metadata cannot be **tracked**. The advantage of end-to-end encryption is that only the intended receiver of the message can read the actual **content data** in **plaintext**.

Local Shared Object (LSO)

Local shared objects (LSOs) are like HTTP **cookie**s but can retain more **data** and are saved in a separate place on the computer. LSOs are frequently used by websites to track user activity across several visits and to record **consumer preference information**, such as language or layout choices. They can, however, also be utilised for more malicious objectives, such as **secretly monitoring** user behaviour.

LSOs they may be removed or blocked by users. Some browsers have built-in options that let **user**s control LSOs; others require the **user** to install add-ons or **third-party software**.

Further reading:
McDonald, A.M. and Cranor, L.F., 2011. A survey of the use of Adobe Flash local shared objects to respawn http cookies. *Isjlp*, 7, 639, https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/isjlpsoc7&section=25.

See also: CUSTOMER TRACKING, TARGETED ADVERTISING, WEB PROFILING

Local Suppression

The **redaction** of a characteristic for some **data unit**s within a **data** file. This would normally be done in situations where the characteristic is in general not **disclosive**, but for those data units – perhaps in combination with other characteristics – it is unusual.

Further reading:

Chen, R., Fung, B.C., Mohammed, N., Desai, B.C. and Wang, K., 2013. Privacy-preserving trajectory data publishing by local suppression. *Information Sciences*, 231, 83–97, https://doi.org/10.1016/j.ins.2011.07.035.

See also: DISCLOSIVE DATA, OVERIMPUTATION, PUBLISHING, SPECIAL UNIQUE, SUPPRESSION

Locational Privacy

The location or whereabouts of an individual, especially when the time they occupied that point is also known, is very revealing. It not only suggests a means by which an **adversary** can gain physical access to the individual, but it also makes certain **inference**s about them possible (e.g., if they visit a particular church regularly, stay at a workplace on a near-daily basis, or appear to be present at a demonstration or political event). Preserving locational privacy involves protecting that information.

Location is relatively **disclosive**. A study of several months of **location data** for 1.5m people revealed that four spatiotemporal points will single out 95 per cent of individuals. Applying **disclosure control** to the **data** is difficult, because even reducing the specificity of the information to a coarser grain has little effect on the **uniqueness**. There is some dispute about this, however, with Sanchez et al arguing that standard techniques are sufficient. This type of debate highlights the importance of **functional anonymisation**, which takes account of such contextual matters.

Locational privacy is a kind of **attentional privacy**, but in a digitally enabled world where mobile devices collect and give out much location data, it is also often seen as a kind of **informational privacy**.

Further reading:

De Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. and Blondel, V.D., 2013. Unique in the crowd: the privacy bounds of human mobility. *Scientific Reports*, 3, 1376, https://doi.org/10.1038/srep01376.

Sánchez, D., Martínez, S. and Domingo-Ferrer, J., 2016. Comment on ‘Unique in the shopping mall: on the reidentifiability of credit card metadata’. *Science*, 351(6279), 1274. https://doi.org/10.1126/science.aad9295.

See also: ANONYMISATION, GEOPRIVACY, LOCATION-BASED SERVICE, LOCATION TRACKING, PHYSICAL PRIVACY, SINGLE OUT

Location-Based Service (LBS)

Location-based services are programs and services that use **location data** from a **user**’s device, such as a smartphone, including location-based marketing, mapping and navigation, **social networking** and emergency services. Location-based social networking **software**, for instance, enables users to interact with people nearby, while a mapping and navigation **application** uses location data to deliver turn-by-turn directions to a destination.

Most often, location **information** is gathered via a variety of technologies, including GPS, Wi-Fi and cellular network data. Applications and services utilise this data to deliver useful information and functionality to users based on their present location.

Further reading:
Junglas, I.A. and Watson, R.T., 2008. Location-based services. *Communications of the ACM*, 51(3), 65–9, https://doi.org/10.1145/1325555.1325568.

See also: LOCATION TRACKING

Location Data

Location data is **information** about someone’s whereabouts, especially, when timestamped, at a particular moment. Location data taken at a series of times will give a person’s movements. Information about places on the Earth is often referred to as *geospatial* **data**, and there exist many standards for expressing this, sometimes based on coordinate systems (e.g., latitude–longitude), or alternatively in terms of *points of interest* (such as towns, streets or landmarks).

One means of producing location data is via (sometimes manual or semi-automatic) **geotagging**, where geographical data is added to the **metadata** of an object. For instance, a photo might be tagged with the place it was taken, or the location coordinates might be recorded by the camera and added automatically to the metadata. The mobile devices carried by individuals necessarily produce location data, as they must at a minimum connect to local network providers and leave a **record** of the connection, enabling **location tracking**. More than that, many smartphone services are tailored to the movements and behaviour of their users, and such **location-based service**s have an active requirement for location data. The smartphone detects signals from external sources, such as the Global Positioning System (GPS), Wi-Fi networks or mobile network towers.

The location data thus produced is associated with the device, not the **user**, but there is still a strong connection between them. Hence the device's identifier in the location data is a **proxy** for the owner. The **locational privacy** implications are clear, especially if the user **consent**s to an **application** getting access to more location data than it needs to provide its services, and if the **data** can be delivered to a **third party** under the **privacy policy**.

Further reading:
Bettini, C., Wang, X.S. and Jajodia, S., 2005. Protecting privacy against location-based personal identification. *In:* Jonker, W. and Petković, M., eds, *Secure data management: second VLDB workshop, SDM 2005*, Berlin: Springer, 185–99, https://doi.org/10.1007/11552338_13.

See also: BEHAVIOURAL ADVERTISING, GEOPRIVACY, NETWORK

Location Tracking

Monitoring and logging a **user**'s current or past position using timestamped **location data**. It often entails locating a target using tools such as GPS, **RFID** tags or cellular **network**s, then sending that **information** to a centralised server or application for processing and analysis. **Asset** monitoring, fleet management, personal safety, **location-based service**s and geolocation-based marketing are just a few uses for location tracking, as well as the **surveillance** of the individual.

Further reading:
Bajaj, R., Ranaweera, S.L. and Agrawal, D.P., 2002. GPS: location-tracking technology. *Computer*, 35(4), 92–4, https://doi.org/10.1109/MC.2002.993780.

See also: GEOPRIVACY, LOCATIONAL PRIVACY, TRACKING

Logic Bomb

An example of **malware** that is intentionally placed into a computer **system** or **software** program. Typically, it is configured to run at a specific time or in response to a trigger event, such as a **user** action. A logic bomb might disable the system, delete files, corrupt or steal data and/or have other negative consequences when it 'detonates'. As they are frequently hidden in legal code and can be engineered to activate a considerable amount of time after being injected into the system, they can be challenging to detect.

Further reading:
Denning, D.E., 2017. Cyberterrorism: the logic bomb versus the truck bomb. *In*: Wall, D.S., ed., *Cyberspace crime*. Abingdon: Routledge, 217–25, https://doi.org/10.4324/9781315199627.

See also: CYBERTERRORISM, STEGANOGRAPHY, TIME BOMB

Longitudinal Data

Longitudinal data is **data** collected from the same sources, **sensor**s or **respondent**s, in comparable formats, sequentially through time. When applied to a fixed group of individual respondents surveyed through time, it is sometimes called *panel data*.

It often contains data about multiple phenomena, or a cross-section of a community, rather than about particular objects or individuals. It therefore allows trends to be mapped and has **privacy** implications because **auxiliary knowledge** about different **attribute**s of an individual through time can significantly increase their **identifiability** within a longitudinal **dataset**.

Further reading:
Frees, E.W., 2004. *Longitudinal and panel data: analysis and applications in the social sciences*. New York: Cambridge University Press.

See also: TIME SERIES

Loyalty Card

The cards issued to customers in the retail environment provided an important tool for the advent of **surveillance capitalism**. Early store credit cards – which predated universal credit cards – were originally designed to encourage repeat business. A similar function was then provided more cheaply with cards that were not used for payment, but rather registered a purchase in order to provide rewards for the customer. With the rise of **big data**, however, the spending patterns collected on the cards morphed from an incidental benefit to arguably their primary purpose.

Although a customer's purchases in a **public** shop were never private in the sense of being separate and unobserved by strangers, they were not previously public to an extent potentially spanning the global digital economy. Nissenbaum highlights this incongruity as an example of the inadequacy of the public–private sphere model to account for contemporary expectations of privacy. Zuboff, on the other hand, focuses more on the dehumanisation and dignitary **harm** of turning customers into products through the secondary marketing of their **data**.

Further reading:

Lauer, J., 2020. Plastic surveillance: payment cards and the history of transactional data, 1888 to present. *Big Data and Society*, 1–14, https://doi.org/10.1177/2053951720907632.

Nissenbaum, H., 2004. Privacy as contextual integrity. *Washington Law Review*, 79(1), 119–58, https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/.

Zuboff, S., 2019. *The age of surveillance capitalism: the fight for the future at the new frontier of power*. London: Profile Books.

See also: CONTEXTUAL INTEGRITY, CUSTOMER RELATIONSHIP MANAGEMENT, CUSTOMER TRACKING, DIGITAL ECONOMY, E-COMMERCE, PRIVATE SPHERE, PUBLIC SPHERE

LSO

See: LOCAL SHARED OBJECT

M

MAC

See also: MANDATORY ACCESS CONTROL, MEDIA ACCESS CONTROL

Machine Learning (ML)

Machine learning (ML) is one of the most central technologies in the modern economy. It refers to the rapidly growing field of using computer **algorithm**s that learn to discern signals, patterns or other significant structures from **data**. As the amount of data available has grown alongside computing power, so ML has become more adept at extracting weak signals from noisy data. Typically, an ML system finds a complex of rules, or a model, that connects inputs in a real-world problem with outputs that would be extremely hard, if not impossible, for human analysts to detect. ML performed on well-structured **dataset**s is usually referred to as **data mining**, but it is most valuable when used on unstructured data or data from heterogeneous sources. When a lot of data has been brought together in this way, it is sometimes referred to as **big data**.

The regularities ML discovers fall under a few classes, but the combination of these techniques is very powerful. ML can be used to cluster items into significant groups; for association-mining to predict the behaviour of an item given the behaviour of similar items (e.g., in a recommender system); to fill gaps in the model of an item (e.g., to suggest how a person with various known characteristics would vote); to spot outliers or unusual elements (e.g., an unlikely pattern of spending money); or for pattern recognition.

ML raises several **privacy** and other ethical issues. First, it may be trained on data which includes **personal data**; in that case, as all machine learning processes are **data processing**, there needs to be a legal basis for doing so. Second, it underpins important privacy-impacting disciplines and areas, such as the **digital economy**, **AI**, **face recognition**, voice recognition, natural language processing, organisational decision-making, and so on. Third, as systems improve, there is increasing temptation to take humans out of the loop. Quite often this can improve performance, but of course at the cost of reducing oversight. Fourth, because of this, the algorithm's output may also be surprising and unpredictable. New connections may be made that could compromise **privacy**. Fifth, a system run on big data may be facilitated by the amalgamation of separate datasets in a central store, increasing the **security** risks. ML methods that analyse such data without amalgamation are called **privacy-preserving machine learning**. Finally, an ML algorithm can only find regularities in data, but if the data itself is biased or otherwise unsatisfactory, then so will the outcomes be.

ML generally works by a system being trained on *training data*, to establish the rules connecting input and output. In *supervised learning*, the inputs and outputs are already defined in the training data, and the role of the algorithm is to approach that standard. In *unsupervised learning*, there are no predefined outputs, so the algorithm takes the inputs and looks for regularities across the data. In *reinforcement learning*, the algorithm is trained only by being evaluated after its performance. Following its training, the algorithm is then used in real-world situations (and may be adapted constantly by feedback). Training can be improved by pitting an algorithm against itself (e.g., a chess-playing system might learn by playing itself, over many more games in a short space of time than humans could manage in centuries). In *meta learning*, models automatically learn how to learn new tasks or adapt over time. These algorithms are designed to acquire skills that enable them to generalise from a set of related tasks. Possible privacy risks include the difficulty in detecting the susceptibility of meta learning models to model **inversion attack**s and **membership inference attack**. Each of these developments increases the power and flexibility of the technique, at the cost of sacrifice of some human control and understanding of the process.

In recent developments, **Generative AI** uses ML techniques trained over giant quantities of data in an unsupervised fashion. As a result, it has produced very convincing and unpredictable output in chatbots and image generation, leading to serious concerns about **deepfake**s and fraud.

Further reading:

Papernot, N., McDaniel, P., Sinha, A. and Wellman, M.P., 2018. Sok: security and privacy in machine learning. *In*: *2018 IEEE European Symposium on Security and Privacy*, 399–414, https://doi.org/10.1109/EuroSP.2018.00035.

AlRubaie, M. and Chang, J.M., 2019. Privacy-preserving machine learning: threats and solutions. *IEEE Security & Privacy,* 17(2), 49–58. https://doi.org/10.1109/MSEC.2018.2888775.

See also: DEEP LEARNING, EXPLAINABLE AI, FACIAL RECOGNITION TECHNOLOGY, INPUT PRIVACY, OUTPUT PRIVACY

Magnitude Data

Data in the form of the sum (or average) of a continuous variable over a set of **data unit**s; often conditioning on some other attribute(s) to create tables of magnitude.

Generally, magnitude data is generated on organisations or administrative units and therefore is not a matter of individual **privacy**, although such data may be a concern for business **confidentiality**.

See also: CONTINUOUS DATA

Main Establishment

When a **data controller** operates in more than one country, the regulator they deal with will be determined according to the location of their main base of business activities (e.g., their headquarters, and/or where they pay tax). This is known as their main establishment.

See also: JURISDICTION

Male Gaze, The

The male gaze expresses the idea that many media are structured so that women become passive objects of **voyeurism**, inclining the audience to a male perspective. Such objectification extends into celebrity culture, and some of its online manifestations range from creepshots (photos of women taken without **consent**) to upskirt photos to **revenge porn**. Socially, the male gaze manifests itself as a generalised and endemic heterosexual-male-oriented **surveillance** of women conceived as objects of desire relative to a constructed standard of beauty.

Further reading:

Gervais, S.J., Vescio, T.K., Förster, J., Maass, A. and Suitner, C., 2012. Seeing women as objects: the sexual body part recognition bias. *European Journal of Social Psychology*, 42(6), 743–53, https://doi.org/10.1002/ejsp.1890.

Mulvey, L., 2009. Visual pleasure and narrative cinema. *In*: Mulvey, L., ed.,*Visual and Other Pleasures*, 2nd edition. Basingstoke: Palgrave Macmillan, 14–29, https://doi.org/10.1007/978-1-349-19798-9.

Malicious Proxy Server

A proxy server that has been infiltrated by an **adversary** and is being used to intercept and change communications between a client and a server. Because they function as intermediaries between the client and server, **proxy** servers are often employed to improve **security** and **privacy**. However, a malicious proxy server can be used to carry out a variety of **attack**s. So, it remains crucial to secure **data in transit** by using **encryption** technologies such as SSL/TLS, as well as monitoring **network** traffic for suspicious activity.

Further reading:
Callegati, F., Cerroni, W. and Ramilli, M., 2009. Man-in-the-middle attack to the HTTPS protocol. *IEEE Security & Privacy*, 7(1), 78–81. https://doi.org/10.1109/MSP.2009.12.

See also: MAN-IN-THE-MIDDLE ATTACK, TRANSPORT LAYER SECURITY, REMOTE ACCESS, NETWORK SECURITY, TRAFFIC DATA

Malware

Any software program or code that is intended to damage or interfere with computer systems or **network**s. Malware may present in a wide variety of forms, including **spyware**, **virus**es, **trojan horse**s, **worm**s and **ransomware**. Malware can facilitate a variety of malicious activity, such as stealing confidential **data**, interfering with regular system processes and obtaining unauthorised access to networks or systems. Multiple channels, including fraudulent websites, **software** downloads and email attachments, can be used to spread it.

Further reading:
Rieck, K., Holz, T., Willems, C., Dussel, P. and Laskov, P., 2008. Learning and Classification of Malware Behavior. *In*: Zamboni, D. ed. *Detection of Intrusions and Malware, and Vulnerability Assessment*, Berlin: Springer Berlin Heidelberg, 108–25. https://doi.org/10.1007/978-3-540-70542-0_6.

Management Information System

A system for coordinating the operations of an organisation. Such systems are used widely and for many purposes in both the public and **private sector**s. Because they routinely deal with **personal data** about employees, contractors and customers – for example in human resources, pay and billing systems – they create a lot of requirements for **privacy** and **security**, which are complicated further by the increasing number of such systems being entrusted to cloud service providers. Failures in security can lead to serious **reputation**al impacts for the organisation.

Further reading:
Bélanger, F. and Crossler, R.E., 2011. Privacy in the digital age: a review of information privacy research in information systems. *MIS Quarterly*, 35(4), 1017–41, https://doi.org/10.2307/41409971.

See also: FILING SYSTEM, CUSTOMER RELATIONSHIP MANAGEMENT, RECORDS MANAGEMENT, EMPLOYEE INFORMATION

Mandatory Access Control (MAC)

A **security** tool called mandatory **access control** (MAC) imposes limitations on access to resources in accordance with a set of predetermined rules and regulations. Based on a user's security clearance, job function or other factors, the operating system or security administrator determines the access restrictions for each user or process in a MAC system. These restrictions are configured using a collection of labels or categories connected to each **user**, process or resource. A resource may be designated as 'sensitive' or 'unclassified', for instance, whereas a user might be designated as 'confidential' or 'top secret'.

By implementing stringent access rules that cannot be overturned by individual users, MAC systems offer a better level of security than **discretionary access control** (DAC), which gives individuals greater control over access to their own resources.

Mandatory Decryption

Mandatory or compelled decryption is the obligation by law for individuals to surrender their private keys to allow officials to decrypt encrypted **data** (also known as **key disclosure**), or alternatively to decrypt the data themselves and hand it over. It usually applies in limited (criminal) law enforcement contexts, akin to obtaining a **search** warrant for a suspect's physical home. Key disclosure is the stronger of the two measures, because the **private key** itself is compromised as a result.

Many jurisdictions have mandatory decryption laws, although in the United States constitutional protection against self-incrimination has prevented these being enacted, leading to complex debates between law enforcement agencies and technology companies (the **crypto wars**). Murphy has argued that the UK may have a legal basis for forcible **decryption** via the Investigatory Powers Act 2016, but that no equivalent power has yet made its way into EU legislation.

Further reading:

Murphy, C.C., 2019. EU counter-terrorism law: What kind of exemplar of transnational law? *The Cambridge Yearbook of European Legal Studies*, 21, 217–42, https://doi.org/10.1017/cel.2019.7.

Palfreyman, B.M., 2009. Lessons from the British and American approaches to compelled decryption. *Brooklyn Law Review*, 75(1), 345–78, https://brooklynworks.brooklaw.edu/blr/vol75/iss1/7.

See also: CRYPTOGRAPHY, ENCRYPTION

Mandatory Key Disclosure

See also: MANDATORY DECRYPTION

Man-in-the-Middle Attack

A man-in-the-middle **attack** involves an **adversary** intercepting and modifying **communication**s between two parties without either party noticing. The adversary acts as an intermediary between the two parties, sending and receiving messages in such a way that it appears to be one of the parties involved in the communication. In this way, the adversary can intercept and manipulate the **information** exchanged between the two parties or can prevent the **information** from reaching its destination. Man-in-the-middle attacks can be carried out using a variety of techniques, such as **eavesdropping** on wireless signals or redirecting **data** packets across a **network**.

Further reading:

Callegati, F., Cerroni, W. and Ramilli, M., 2009. Man-in-the-middle attack to the HTTPS protocol. *IEEE Security & Privacy*, 7(1), 78–81, https://doi.org/10.1109/MSP.2009.12.

See also: COMMUNICATION PRIVACY, EAVESDROPPING ATTACK, MALICIOUS PROXY SERVER, SECURE COMMUNICATION

Manual Key Transport

Manual key transport involves copying **cryptographic key**s onto a physical medium, such as paper or a USB drive, and then physically transferring them between two parties in a safe manner, such as in-person delivery or a secure courier service.

Further reading:
Lenstra, A.K. and Verheul, E.R., 2001. Selecting cryptographic key sizes. *Journal of cryptology*, 14, 255–93, https://doi.org/10.1007/978-3-540-46588-1_30.

Mash Attack

An **attack** on an anonymised **dataset** which involves combining multiple **data** sources into a single attack vector (**identification file**). For a mash attack to be effective the linkages between the data sources need to be a reasonably high **confidence**, which usually implies that they contain the same **direct identifier**s.

Mask

A mask is a covering that obscures the face from vision, but typically has holes or transparencies to allow the masked person to see, breathe and talk. Masks therefore disguise the face and make identification difficult, although this need not be their prime purpose. Some masks are used for protection, as with surgical or medical masks, welding masks, or the masks used in the sport of fencing. Others are used in performance and ritual in many cultures; actors may efface their own personality with a mask.

From the **privacy** perspective, masks may be worn deliberately to **disguise**, by protestors, criminals, law enforcement agents and participants in sexual imagery. The vocabulary of **masking** and unmasking is often used metaphorically to refer to disguises and identification in other areas, such as the anonymisation of **data**.

Further reading:
Hollis, M., 1985. Of masks and men. *In*: Carrithers, M., Collins, S. and Lukes, S., eds, *The category of the person*: *anthropology, philosophy, history*. Cambridge: Cambridge University Press, 217–33.
Mauss, M., 1985. A category of the human mind: the notion of person; the notion of self. *In*: Carrithers, M., Collins, S. and Lukes, S., eds, *The category of the person*: *anthropology, philosophy, history*. Cambridge: Cambridge University Press, 1–25.

Masking

The practice of hiding one’s **identity** in social interactions. This is a recognised phenomenon for neuro-atypical individuals, who consciously or unconsciously disguise their neurotype to ‘fit in’. In this sense masking can be seen as the opposite of **self-disclosure**.

In **statistical disclosure control**, the application of **perturbation** of any sort to increase uncertainty about whether any piece of **data** is accurate, thus reducing **disclosure risk**.

Further reading:
Bennie, M., 2022. *What is autistic masking?* Autism Awareness Centre https://autismawarenessawarenesscentre.com/what-is-autistic-masking/.

See also: ANONYMISATION, DISGUISE

Masquerade

In **network security**, this technique is used for gaining unauthorised access to specific resources by assuming the **identity** of a trusted entity. It is usually done by utilising **IP address**es or a website spoofing and it is useful for carrying out **phishing** and **man-in-the-middle attack**s.

Further reading:
Marin, G.A., 2005. Network Security Basics. *IEEE Security and Privacy*, 3(6), 68–72. https://doi.org/10.1109/MSP.2005.153.

Matching

Often used as a synonym for **record linkage**. However, within standard record linkage, a ‘match’ is also used to mean a true link.

Further reading:
Christen, P., 2012. *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*. New York: Springer, https://doi.org/10.1007/978-3-642-31164-2.

Material Scope

The EU’s GDPR has two dimensions to its scope: material and territorial. **Territorial scope** determines where the regulation applies, in terms of the geographical location of the processing (or the **data subject**s). Material scope, on the other hand, refers to the types of data processing governed by the **GDPR**.

Key to the material scope of the GDPR is the definition of **personal data**. Any activity involving personal **data** will constitute **processing**, including the mere act of possession and retention. The most common challenge to the material scope of the GDPR is therefore whether the data in question constitute personal data. This, in turn, translates into a question of whether natural living people can be identified by the data using any means reasonably likely to be used.

The material scope of the GDPR, and the inverse scope of ‘anonymous’ data, can be a highly contested issue, with some arguing that **proportionality** of scope is essential and others suggesting that claimed **anonymisation** can become a means of escaping regulation.

Further reading:

Mourby, M., 2020. Anonymity in EU healthcare law: not an alternative to information governance. *Medical Law Review*, 28(3), 478–501, https://doi.org/10.1093/medlaw/fwaa010.

Sharma, S., 2019. GDPR’s scope of application. *In*: Sharma, S., *Data privacy and GDPR handbook.* Hoboken: John Wiley & Sons Inc, 45–60, https://doi.org/10.1002/9781119594307.ch3.

Material Transfer Agreement

Many legal jurisdictions require any donated human tissue used in medicine and research to be trackable. A Material Transfer Agreement is an example of a legally binding contract which ensures all human tissue changes hands in a traceable way, so findings from any analysis can be linked back to the original donor.

Under European Union law, the need to maintain the traceability of human tissue must also be balanced with the principle of ‘donor **anonymity**’. This means that the **identity** of the original donor must be kept confidential, and only known to those who might need to re-contact them (e.g., in light of incidental findings). Material Transfer Agreements also help to ensure that human tissue is protected in a way which respects the **dignity** of the donors, even if they are unaware of its subsequent uses.

Further reading:
Mourby, M., 2020. Anonymity in EU Healthcare Law: Not an Alternative to Information Governance. *Medical Law Review*, 28(3), 478–501, https://doi.org/10.1093/medlaw/fwaa010.

See also: BODILY PRIVACY, PSEUDONYMISATION, UNIQUE IDENTIFIER

Maximum Knowledge Intruder

A term, coined by Domingo-Ferrer et al., to describe a hypothetical **adversary** who is practically omniscient. The term has mostly been applied in the **disclosure risk** context, where it denotes an adversary who has already has all the **data** in the **target dataset** apart from one piece (which they wish to learn).

Such an adversary is presupposed in the **privacy guarantee** made in **differential privacy** and other formal **privacy model**s. Although no such claims are explicitly made by proponents of differential **privacy**, defence against such an adversary is necessary for the guarantee to be sound. Opponents argue that because the maximum knowledge intruder is an unrealistic adversary, differential privacy is biased (and over-cautious) in its implicit attitude to risk.

Further reading:
Domingo-Ferrer, J., Ricci, S. and Soria-Comas, J., 2015. Disclosure risk assessment via record linkage by a maximum-knowledge attacker. *In*: *2015 13th Annual Conference on Privacy, Security and Trust*, 28–35. IEEE, https://doi.org/10.1109/PST.2015.7232951.

Media Access Control (MAC) Address

An individual identification code for a particular hardware device given to a **network** interface controller (NIC) for use in network **communication**. It and its manufacturer are identified by the 12 hexadecimal characters that make up the MAC address. They can be used to monitor device activities on a network privately. The ability to trace device activity and potentially connect it to a specific user's behaviour raises **privacy** concerns while also being valuable for network administration and troubleshooting.

There has been an increase in **privacy concern**s over the use of MAC addresses for monitoring in recent years, particularly in the context of mobile devices and wireless networks. To avoid being tracked, some mobile devices generate random MAC addresses, although this might affect network speed and make some network **protocol**s incompatible.

Further reading:
Ye, W. and Heidemann, J., 2004. Medium Access Control in wireless sensor networks. *In*: Raghavendra, C.S., Sivalingam, K.M. and Znati, T. eds, *Wireless Sensor Networks*, Boston, MA: Springer US, 73–91, https://doi.org/10.1007/978-1-4020-7884-2_4.

See also: ACCESS CONTROL, COMMUNICATION PRIVACY

Medical Record

A dossier of an individual's medical history.

See: ELECTRONIC HEALTH RECORD

Membership Inference Attack

The **inference** by an **adversary** that an individual is present in a **dataset** or has contributed to a statistic or model. If membership of the dataset is informative (e.g., if all members have a particular illness) then membership inference is a form of **attribute disclosure**.

One of the central tenets of **differential privacy** is to make membership inference very difficult for an adversary to achieve.

Further reading:
Shokri, R., Stronati, M., Song, C. and Shmatikov, V., 2017, Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy*, IEEE, 3–18, https://doi.org/10.1109/SP.2017.41.
Pyrgelis, A., Troncoso, C. and De Cristofaro, E., 2017. Knock knock, who's there? Membership inference on aggregate location data. *arXiv preprint,* https://doi.org/10.48550/arXiv.1708.06145.

See also: INFERENCE ATTACK

Mental Capacity

Mental capacity is the legal status in which one is taken to have the ability to make a specific decision.

Privacy is often understood as encompassing individual **autonomy**, particularly as regards decision-making. This assumes that the individual in question has the mental capacity to make the decision, yet not all individuals are physically and mentally autonomous, perfectly capable of functioning independently if they are 'let alone'. Nussbaum argues that this has led to existing theories of social justice neglecting the rights and **dignity** of people with physical and mental impairments.

Standard **information security** technology is not easily accessible for people whose cognitive functioning makes it difficult – for example – to remember complex passwords, or the answers to **security** questions. Lazar and colleagues argue that without accessible security mechanisms, people with disabilities will either be shut out of their **information** or forced to entrust **password**s and keys with others, against advice (or even contractual terms) from banks and other service providers.

Herring has therefore argued for a less individualistic and more interconnected understanding of human rights such as **privacy** based on an ethic of care. Where both 'caring' and 'cared for' people are affected by a decision, he suggests the law should promote the rights and interests of both and treat the two as mutually interdependent.

Further reading:
Herring, J., 2013. *Caring and the law*. Oxford: Hart Publishing.
Lazar, J., Wentz, B. and Winckler, M., 2017. Information privacy and security as a human right for people with disabilities. *In*: Lazar, J. and Stein, M.A. eds, *Disability, human rights, and information technology.* Philadelphia: University of Pennsylvania Press, 199–211, http://dx.doi.org/10.9783/9780812294095-014.
Nussbaum, M.C., 2007. Frontiers of justice: disability, nationality, species membership. *Scandinavian Journal of Disability Research*, 9(2), 133–6, https://dois.org/10.1080/15017410601003171.

See also: DECISIONAL PRIVACY, EUROPEAN CONVENTION ON HUMAN RIGHTS, FEMINIST CRITIQUE OF PRIVACY, MENTAL PRIVACY, PHYSICAL PRIVACY

Mental Privacy

See also: INTELLECTUAL PRIVACY, PSYCHOLOGICAL PRIVACY

Mesh Network

Each node in a mesh **network** serves as a relay for other nodes, creating a decentralised network structure. All nodes in a mesh network are interconnected, enabling **data** to be sent via various channels between nodes. Compared to centralised networks, mesh networks can provide a better level of **privacy**, making it more challenging for an **adversary** to collect or eavesdrop on **communication**s between nodes since data is carried across several channels. Moreover, there is no single point of failure or **vulnerability** that may be exploited.

There may still be issues with privacy that need to be resolved in mesh networks. For instance, hacked network nodes may be able to intercept or alter data being transported via the network.

Further reading:
Akyildiz, I.F., Wang, X. and Wang, W., 2005. Wireless mesh networks: a survey. *Computer Networks*, 47(4), 445–87, https://doi.org/10.1016/j.comnet.2004.12.001.

See also: AD HOC NETWORK

Message Digest

A message digest is a fixed-length string of characters created by applying a mathematical function to a message or set of **data**. It is sometimes referred to as a hash value or **checksum**. As the message digest is made to be specific to the original message, even minor changes will produce a different digest. Applications for data **integrity** and **cryptography** frequently employ **message digest**s. The receiver may ensure that a message hasn’t been tampered with or damaged in transit by comparing the message’s received digest to its original digest.

The digest of a message is signed using a **private key** in the process of creating a **digital signature**, which can then be validated using the matching **public key**. **Password** storage and verification employ message digests. The system saves the message digest of a **user**’s password rather than the **plaintext** version. When a user enters a password, the system creates a password digest and compares it to a previously saved digest. The system knows the password entered is correct if the digests match. To create message digests, a variety of **algorithm**s, including MD5, SHA-1 and SHA-256, can be utilised, to produce a digest of the desired complexity and probability of **uniqueness**.

Further reading:

Rivest, R., 1992. *RFC1321*: *The MD5 message-digest algorithm*. Boston: MIT Laboratory for Computer Science, https://dl.acm.org/doi/pdf/10.17487/RFC1321.

See also: CRYPTOGRAPHIC KEY, HASHING, DATA IN TRANSIT

Metadata

Metadata is **information** that describes other **information**, data or digital resources, and is typically associated with or appended to the resource. For example, a document's **metadata** may include its title, author, creation date, intended readership and keywords. A dataset's metadata may include the representation format and a link to its ontology. Metadata can be used to facilitate both the **search** for, and organisation of, data and to provide context and meaning to the data itself.

Metadata can reveal personal **information** about individuals, such as their location, behaviour and online activities (as for example with the metadata associated with a call or email, which links two parties in **communication**s at a time, even if the content of the communication is not included). Moreover, it can be stored and used for extended periods of time, increasing the risk of privacy violations, and can be used to profile and monitor the online activity of the **user**.

Further reading:

Pomerantz, J., 2015. *Metadata*. Boston: MIT Press, https://mitpress.mit.edu/9780262331203/metadata/.

Riley, J., 2017. *Understanding metadata: what is metadata, and what is it for? A primer.* Washington DC: National Information Standards Organisation, www.niso.org/publications/understanding-metadata-2017.

Metadata-Level Controls

Statistical disclosure control methods based on **data** restriction (rather than distortion). Examples include **sampling**, variable deletion and aggregation/recoding.

Further reading:

Elliot, M., Mackey, E. and O'Hara, K., 2020. *The anonymisation decision-making framework, 2nd edition*: *European practitioners' guide*. Manchester: UKAN Publications, https://ukanon.net/framework.

See also: ANONYMISATION, DELETION, DISCLOSURE CONTROL METHODS

Metaverse

A loosely defined term, the core meaning of which revolves around the totality of all virtual environments. In common usage, it tends to be tied specifically to immersive environments mediated by virtual reality hardware.

In some visioning, the metaverse is regarded as the natural development of the **Internet**. The possibility raised by the concept is an always-on, virtual layer of reality. It is also linked to potential societal transformations such as transhumanism and the s**ingularity** hypothesis, accelerating humanity's deepening relationship with advanced technology and **artificial intelligence** in particular.

The concept and its development raise concerns across all four of the **TIPS** areas: the metaverse brings into scope psychological, **biometric** and physiological **data** as **privacy risk** vectors; the impact on human identities of a parallel virtual existence is difficult to estimate; **security** as a practical concept will be something very different and is as yet ill-defined.

Further reading:

Ball, M., 2021. *Framework for the metaverse*: *the metaverse primer*, www.matthewball.vc/all/forwardtothemetaverseprimer.

Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T.H. and Shen, X., 2023. A survey on metaverse: fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials*, 25(1), 319–352, https://doi.org/10.1109/COMST.2022.3202047.

Huang, Y., Li, Y.J. and Cai, Z., 2023. Security and privacy in metaverse: a comprehensive survey. *Big Data Mining and Analytics*, 6(2), 234–47, https://doi.org/10.26599/BDMA.2022.9020047.

MFA

See also: MULTI-FACTOR AUTHENTICATION

Microaggregation

Microaggregation is an approach to **statistical disclosure control** primarily for continuous **microdata** in which records are aggregated into groups.

Instead of releasing the actual values for the continuous variable, the mean (typically) of the group is overimputed. **Confidentiality** is protected by each group having at least a minimum number of observations. Microaggregation can therefore be a method for implementing **k-anonymity**.

Further reading:

Domingo-Ferrer, J., 2009. Microaggregation. *In*: Liu, L. and Özsu, M.T., eds, *Encyclopedia of Database Systems*. Boston, MA: Springer, https://doi.org/10.1007/978-0-387-39940-9_1496.

See also: CONTINUOUS DATA, OVERIMPUTATION

Microdata

A term commonly used by the National Statistical Institutes (NSIs) to denote **data** in the form of sets of records of individual **population unit**s. This distinguishes them from aggregate statistics, which are the staple data format produced by NSIs. Since they first appeared in the 1980s, microdata from national **census**es have been the subject of considerable disclosure **risk assessment** exercises during each census round and have consequently become something of a gold **standard** when measuring **risk** on other **dataset**s.

Minimal Unique

For a **record**, within a **microdata** file, a minimal unique is a set of variables on which the record is unique but for all subsets of those variables the record is not unique. This is used as the basic building block for an operationalisation of the concept of a **special unique**.

Further reading:

Elliot, M.J., Manning, A.M. and Ford, R.W., 2002. A computational algorithm for handling the special uniques problem. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05), 493–509, https://doi.org/10.1142/S0218488502001600.

Haglin, D.J., Mayes, K.R., Manning, A.M., Feo, J., Gurd, J.R., Elliot, M. and Keane, J.A., 2009. Factors affecting the performance of parallel mining of minimal unique itemsets on diverse architectures. *Concurrency and Computation: Practice and Experience*, 21(9), 1131–58, https://doi.org/10.1002/cpe.1379.

See also: UNIQUENESS

Missing Data

Within **microdata** or some other form of **database**, missing **data** denotes a piece of **information** or possibly even a whole **record** which is missing and should be present given the rationale for the database.

Missing **data** is the bane of the analyst's work. However, in general it has a side benefit of reducing risks of **reidentification** and other forms of **statistical disclosure**.

Further reading:
Allison, P.D., 2001. *Missing data*. Thousand Oaks: Sage Publications, https://doi.org/10.4135/9781412985079.

See also: IMPUTATION

Mission Creep

Mission creep is the tendency of organisations or projects to extend their remit beyond the announced intention of their creators. As bureaucracies grow, they have a vested interest in increasing the scope of their operations, especially as their competences and **information** sources will most likely overlap with adjacent areas. As such, they can redefine problems as issues they can deal with, for instance taking on responsibilities for which they were not originally conceived, and for which they may not be well equipped. In general, to the extent that mission creep involves a change in the type or intensity of interactions with people, it also inevitably involves further encroachment on the **privacy** of those people. This can be a particular problem with government agencies, which tend to extend the scope of the **public sphere** by intruding further into the **private sphere**, thereby diminishing it.

Further reading:
Koops, B.-J., 2021. The concept of function creep. *Law, Innovation and Technology*, 13(1), 29–56, https://doi.org/10.1080/17579961.2021.1898299.

See also: FUNCTION CREEP

Misuse of Private Information

Misuse of private **information** is a tort that has been recognised in British courts following the 1998 Human Rights Act, that applies to information that is private in nature, such as information about someone's health or sexual activities. The tort was first referenced by Lord Nichols in the 2004 judgment in *Campbell v MGN Ltd,* and confirmed more recently in the 2015 case of *Vidal-Hall v Google Inc*. It covers exposure of such information by the media, as well as the sharing of information online.

The test for misuse of private information is that, following an unwanted **intrusion**, access has been gained to information about which the subject would have a **reasonable expectation of privacy**. However, for the tort to be established, it also must be balanced against the intruder's rights to **freedom of expression**, and against the **public interest** in **exposure** of the information.

Further reading:
Mo, J.Y.C., 2017. Misuse of private information as a tort: the implications of Google v Judith Vidal-Hall. *Computer Law and Security Review*, 33(1), 87–97, https://doi.org/10.1016/j.clsr.2016.11.004.

See also: BREACH OF CONFIDENCE, FREEDOM OF INFORMATION, PRIVACY TORT

ML

See also: MACHINE LEARNING

Mobility Traces

Mobility traces are computerised logs of a user's historical mobility patterns. These tracks are often made using location-based technologies that may follow a person's activities and whereabouts throughout the day, including GPS **tracking** or cellular network **data**, and reveal human mobility patterns, such as the frequency and length of trips to specific sites, the distance and speed of travel and the general pattern of movement. Their numerous uses include urban planning, the improvement of transportation **network**s and the study of public health, but they naturally also pose **privacy** issues.

Further reading:
De, M.Y.A., Hidalgo, C.A., Verleysen, M. and Blondel, V.D., 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3(1), 1–5. https://doi.org/10.1038/srep01376.

See also: LOCATION DATA

Model Inversion Attack

An **attack** which attempts to reconstruct the training **data** used to train a **machine learning** model. It does this exploiting the fact that the parameters in the model must necessarily contain some **information** about the training data. For example, the parameters of a trained regression model will be better attuned to the training data than any other data drawn from the same **population**. Model inversion will be more problematic if the model is overfitted. Similarly, small **dataset**s are more likely to create an issue than large datasets.

Further reading:

Fredrikson, M., Jha, S. and Ristenpart, T., 2015, October. Model inversion attacks that exploit confidence information and basic countermeasures. *In*: *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 1322–33, https://doi.org/10.4236/jcc.2021.95007.

Veale, M., Binns, R. and Edwards, L., 2018. Algorithms that remember: model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180083, http://dx.doi.org/10.1098/rsta.2018.0083.

See also: RECONSTRUCTION ATTACK

MOLKA

Molkas are a style of miniaturised concealed camera which originated in South Korea, which are used for a type of voyeuristic pornography filming – usually of women – in bathrooms, bedrooms and changing rooms. The videos have become known as a result as molka videos, and, despite being illegal, are widespread online.

Further reading:

Armesto-Larson, B., 2020. Nonconsensual pornography: criminal law solutions to a worldwide problem. *Oregon Review of International Law*, 21, 177–213, https://scholarsbank.uoregon.edu/xmlui/handle/1794/25394.

See also: VOYEURISM

Monetary Equivalent Burden (of Privacy)

The economic costs incurred as a result of maintaining or protecting **privacy**. For organisations these may include the costs of **cybersecurity** measures or hiring privacy professionals, and the productivity costs arising from **data protection** measures. These may be passed onto individuals (customers) resulting in a privacy premium, raising concerns that privacy may become too expensive for some.

Further reading:

Matzner, T., Masur, P.K., Ochs, C. and von Pape, T., 2016. Do-It-yourself data protection – Empowerment or burden? *In*: *Data protection on the move*: *Current developments in ICT and privacy/data protection*, 277–305, https://doi.org/10.1007/978-94-017-7376-8_11.

See also: ECONOMICS OF PRIVACY, PRIVACY PREMIUM

Mosaic Identification

See: JIGSAW IDENTIFICATION

Motivated Intruder

In **scenario analysis**, a motivated **intruder** is an **adversary** who is presumed to have reasons, goals or desires to **breach** the **security** or **confidentiality** of a system or **dataset**. The reasons are presumed to mitigate the costs that intruder might face.

Motivated Intruder Test

See also: INTRUDER TESTING

M-Probability

One of the two key parameters of **probabilistic record linkage**, the *m*-probability is the probability of two records agreeing on the values of a given variable for a pair of **record**s that do belong to the same **population unit**.

Due to **data divergence**, the m probability is invariably less than 1 and unlike the ***u*-probability** is not estimable through a simple calculation. Instead, the value of the *m*-probability is iteratively estimated based on prior **information** and the proportion of agreements among the comparison pairs that are accepted as links.

Further reading:
Fellegi, I.P., and Sunter, A.B., 1969. A theory for record linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210, https://doi.org/10.1080/01621459.1969.10501049.

Multi-Factor Authentication (MFA)

A form of access **authentication** whereby a **user** is required to present to the authentication system multiple pieces of evidence that they are a *bona fide* user, and specifically that they are who they claim to be. Each piece of evidence will typically be in a different form: something they know (e.g., a **password**), some artefact that they possess (e.g., an **ID card**), or something intrinsically part of them (e.g., a fingerprint). The minimal pattern of MFA, increasingly used in commercial contexts, is two-factor authentication (TFA), where the user is required to present two different **credentials** (for instance, a password and a **security** number which will be texted to their mobile phone on presentation of the password).

See also: INHERENCE, BIOMETRICS

Multimodal De-Identification

A form of disclosure control specifically aimed at multimedia **data**.

Data is increasingly multimodal, comprising, for example, text and images as well as structured data. Consequently, the need to consider the complex interactions between those data forms in terms of **identifiability** and **disclosure control** becomes more important.

Further reading:
Ribaric, S., Ariyaeeinia, A. and Pavesic, N., 2016. De-identification for privacy protection in multimedia content: a survey. *Signal Processing*: *Image Communication*, 47, 131–51, https://doi.org/10.1016/j.image.2016.05.020.

Multiple Imputation

See: IMPUTATION

Multi-Vector Attacks

A multi-vector **attack** makes use of multiple techniques to target system weaknesses. The attack may use **social engineering, malware** and **network** vulnerabilities to get past a system's defences as opposed to depending on a single attack vector, such as **phishing** emails or malware.

A phishing email, for instance, may be the first part of a multi-vector assault since it tempts the **user** to click on a dangerous link or download a dangerous file. Once a system has been breached, the **adversary** may employ malware to spread throughout the network or **security** flaws in the system's protocols to obtain more access.

Multi-vector attacks will be necessary for the adversary to succeed if the target organisation is using a **layered security model**.

See also: NETWORK SECURITY

Mutual Assistance

Under the EU's GDPR, national **regulator**s (known as 'Supervisory Authorities') in each member state are required to cooperate and provide each other with mutual assistance. This includes sharing **information** to assist investigations, particularly in the context of **cross-border data processing**.

Further reading:
Gentile, G. and Lynskey, O., 2022. Deficient by design? The transnational enforcement of the GDPR. *The International and Comparative Law Quarterly*, 71(4), 799–830, https://doi.org/10.1017/S0020589322000355.

See also: ONE-STOP-SHOP, SUPERVISORY AUTHORITY

Mutual Authentication

Before any **data** is shared, two parties may mutually authenticate each other as part of a **security** procedure. Unlike conventional authentication systems, which only need one side to validate the other, mutual **authentication** requires each party to confirm the **identity** of the other.

The client and server exchange digital certificates during a mutual authentication procedure to confirm their identities. These certificates provide data to verify the identification of the entity, such as the **public key** and **digital signature**. The client and server can safely exchange data after exchanging and verifying the **digital certificate**s.

Online banking, e-commerce and other applications with a need for secure two-way **communication** frequently employ mutual authentication. Verifying that both parties are who they say they are lowers the risk of **man-in-the-middle attack**s or other means of intercepting communications.

Further reading:

Otway, D. and Rees, O., 1987. Efficient and timely mutual authentication. *ACM SIGOPS Operating Systems Review*, 21(1), 8–10, https://doi.org/10.1145/24592.24594.

See also: SECURE COMMUNICATION

N

National Security

National security is the **security** of a sovereign state. Originally developed as a military concept, it has expanded to cover the protection of items of national interest, including the national **information** space and digital infrastructure. The types of protection have also evolved beyond military ideas; while the original idea of the **Internet** was of a distributed information system that could withstand degradation following a nuclear attack, protection in the national interest now generally covers **hacking**, to prevent the **exposure** of **classified information**, cyberespionage and even **cybercrime**. While cybercrime is largely considered a private interest, it is assumed that the expertise involved in combating it, and the effects that poor **cybersecurity** might have on the knowledge economy, together mean that governments should take the lead.

Further reading:
Goldman, J. and Maret, S.L., 2016. *Intelligence and information policy for national security*: *key terms and concepts*. Lanham, MD: Rowman & Littlefield.

See also: CYBERWARFARE, CYBERTERRORISM

Natural Person

Data protection laws only protect ‘natural’ people, as opposed to corporations. While commercial organisations enjoy many rights in law, particularly relating to **intellectual property**, human rights are generally reserved for humans, including the right to **privacy**.

Companies are thus obliged to comply with **data** protection laws, but they cannot themselves be considered **data subject**s or claim any right to privacy or **right to data protection**.

Further reading:
Adriano, E.A.Q., 2015. The natural person, legal entity or juridical person and juridical personality. *Penn State Journal of Law & International Affairs*, 4, 363, https://heinonline.org/HOL/LandingPage?handle=hein.journals/pensalfaw4&div=20&id=&page=.

See also: IDENTIFIABLE NATURAL PERSON, PERSONHOOD

Necessity

It is a general tenet of **privacy** laws in most jurisdictions that an **interference** with privacy, **confidentiality** or **data protection** rights can be justified if it is necessary. **Common law** duties of confidentiality, for example, can be overridden in cases of **public interest**, for example when a doctor believes that their patient is at **risk** of **harm**.

In Europe, one of the key articulations of the **right to privacy** comes from Article 8 of the **European Convention on Human Rights**. This right has been interpreted by the European Court of Human Rights ('ECtHR') as requiring an interference to be in accordance with the law, in pursuit of a legitimate aim and both necessary and proportionate. The final requirement points to whether it would be possible to achieve the aim with a less intrusive alternative.

The EU's **GDPR** also has a similar necessity requirement. Except where the **data subject**'s **consent** has been obtained, **personal data** cannot be processed unless it is 'necessary' to do so for one of the reasons listed under Article 6 GDPR. (i.e., unless it is not possible to fulfil those functions if the **processing** does not take place). The idea of using the minimum **identifiable data** needed for a particular purpose, as specified by the ECtHR, is also present in the GDPR's **data minimisation principle**.

Further reading:

Barczentewicz, M. and Stout, K., 2023. GDPR decision against meta highlights that privacy regulators don't understand 'necessity'. *Truth on the Market*, Newstex, Washington, https://truthonthemarket.com/2023/01/11/gdpr-decision-against-meta-highlights-that-privacy-regulators-dont-understand-necessity/.

Romero-Moreno, F., 2016. The Digital Economy Act 2010: subscriber monitoring and the right to privacy under Article 8 of the ECHR. *International Review of Law, Computers & Technology*, 30(3), 229–47, https://doi.org/10.1080/13600869.2016.1176320.

See also: LAWFUL BASIS

Need to Know

An informal principle of providing **information** to people only if that information is required for them to do their job effectively. There is an alignment between need to know and **GDPR**'s principles of fair and lawful processing. The main issue with the principle is its minimalism; it carries no information about the process by which need to know is determined which therefore can be prone to subjectivity and accompanying biases, power imbalances and inefficiencies. This in turn may cause tension with GDPR's principles of **accountability** and **transparency**.

See also: LEAST PRIVILEGE

Negative Externalities of Disclosed Data

An externality, in economics, is an unintended by-product of a transaction, typically one that affects those other than the parties to it. These effects can be *positive* or *negative* for those affected.

In a typical **data protection** model, a **data controller** can get access to a **data subject**'s **personal data** via **notice and consent**; they give notice as to how they intend to gather and use the **data**, to which the subject **consent**s. This description affects only the controller and subject. However, such transactions also have negative externalities for uninvolved third parties. For example, uploading a photo to a photo-sharing site will expose others in the photo to view (and their images may be tagged by **face-recognition** systems on the site). Disclosing **genetic data** also provides genetic **information** about one's relatives. Allowing access to contact lists gives the data controller access to the contacts themselves. More subtly, externalities fall upon data subjects in a notice-and-consent scheme, as they become responsible for their own privacy management.

Further reading:
De Brouwer, S., 2020. Privacy self-management and the issue of privacy externalities: of thwarted expectations, and harmful exploitation. *Internet Policy Review*, 9(4), https://doi.org/10.14763/2020.4.1537.

See also: GROUP PRIVACY, GENETIC PRIVACY, INFORMED CONSENT, NEGATIVE EXTERNALITIES OF PRIVACY, PRIVACY NOTICE

Negative Externalities of Privacy

An externality, in economics, is an unintended by-product of a transaction; typically one that affects those other than the parties to it. These effects can be *positive* or *negative* for those affected.

The argument that **privacy** was a source of unjustified economic inefficiency, and therefore imposed a cost on society as a negative externality, was formulated in a series of papers by Posner. He characterised privacy as **secrecy**, the avoidance of public **disclosure** of concealed **information**. Subjects want to control the flow of information about them, to manipulate the world with selective disclosure of facts. This, Posner thought, was morally dubious enough to warrant the law's support of the rights of others to gain relevant information. He noted that the right to misrepresent one's character (a corollary, according to Posner, of privacy) creates legitimate interests in others to receive a less misleading picture. Law should support freedom of speech and information, and only restrict misleading communications, via **libel**, **slander** and forgery.

The result of secrecy is **information asymmetry**, reducing the ability of participants in the marketplace to judge the worth of agents, correspondingly decreasing the value of the market itself as a signalling system. Privacy/secrecy brings costs to others that the subject does not have to pay – a negative externality. Posner therefore argued that privacy should be purchased by subjects for the amount that secrecy is worth to them, thereby rationing it and mitigating its social cost by the exchange.

On the other hand, Posner accepted that a totally transparent environment would result in guarded speech, and less frank and effective **communication**.

Further reading:

Posner, R.A., 1983. *The economics of justice*. Cambridge, MA: Harvard University Press.

See also: ECONOMICS OF PRIVACY, FREEDOM OF INFORMATION, NEGATIVE EXTERNALITIES OF DISCLOSED DATA, VALUE OF PRIVACY

Negligence

Negligence within tort law has limited application to **privacy** rights. It is true that, in some jurisdictions, the right to privacy is recognised within tort law (i.e., the **common law** of obligations between private citizens). However, privacy and negligence are still separate aspects of tort law in these countries.

Negligence is also a cause of action under tort law, but not one commonly used to enforce privacy rights. The first requirement of negligence is a duty of *care*, whereas a defendant does not need a particular relationship with a claimant to have a duty to use their private **information** appropriately. For example, a tabloid newspaper does not owe celebrities a duty of care, but it can still infringe their privacy. Negligence therefore follows different principles of proximity of relationship and standards of foreseeability, that do not apply to torts such as misuse of private information, which are commonly referred to as **privacy tort**s.

However, where a defendant *does* owe a claimant a duty of care, it is possible that careless **disclosure** of confidential information could form the basis of liability in negligence. For example, in *Swinney v Chief Constable of the Northumbria Police* [1997] QB 464 (CA), the English police were held to have a duty of care towards an informant in a murder investigation, including a duty to keep the information they disclosed confidential.

Further reading:

Nolan, D., 2013. Negligence and human rights law: the case for separate development. *The Modern Law Review*, 76(2), 286–318, https://doi.org/10.1111/1468-2230.12013.

See also: MISUSE OF PRIVATE INFORMATION

Network

A set of interconnected nodes. The connections are usually **communication** channels, and the nodes might be people, devices, systems or organisations. How big the set of nodes must be to qualify as a **network** is unclear, but the term usually implies some sort of scale – that is, more than a few nodes.

See also: INTERNET, INTRANET, AS, SOCIAL NETWORK, SOCIAL NETWORK ANALYSIS

Network Encryption

Network **encryption** protects **data-in-transit** across **network**s. **Transport layer security** is the **standard** for network **data protection** for **Internet** communications; some organisations also encrypt their internal networks.

See also: NETWORK SECURITY

Network Layer Attack

Attacks against the **network** layer of the OSI (Open Systems Interconnection) model. This layer offers addressing, routing and traffic management services in addition to moving **data** packets between networked devices. Attacks on the network layer aim to obstruct or impair its functioning, resulting in data loss, network downtime or unauthorised access. **Distributed Denial of Service**, spoofing and **Man-in-the-Middle attack**s are a few examples of network layer **attack**s.

The detection and prevention of network layer attacks can be challenging, but there are a number of **security** measures that can be put in place, including **network segmentation**, **access control** lists (ACLs) and **intrusion detection system**s (IDSs).

Further reading:
Nadeem, A. and Howarth, M.P., 2013. A survey of MANET intrusion detection & prevention approaches for network layer attacks. *IEEE Communications Surveys & Tutorials*, 15(4), 2027–45, https://doi.org/10.1109/SURV.2013.030713.00201.

See also: NETWORK SECURITY

Network Security

Network security is the process of defending computer networks against **intrusion**s, **attack**s and other harmful actions, with **security** procedures and safeguards to guarantee the **availability**, **confidentiality** and **integrity** of **network** resources.

Network Segmentation

The process of segmenting a computer **network**. Network segmentation aims to improve network management, **security** and performance by limiting **communication** between segments, which can lessen the impact of network **attack**s and stop **malware** from spreading across the whole network.

A network can be segmented such that each segment has its own security rules and access restrictions, preventing unauthorised access and minimising the effects of security **breach**es.

Further reading:
Mhaskar, N., Alabbad, M. and Khedri, R., 2021. A formal approach to network segmentation. *Computers & Security*, 103, 102162, https://doi.org/10.1016/j.cose.2020.102162.

See also: NETWORK LAYER ATTACK, NETWORK SECURITY

Neural Prosthesis

A device which replaces a motor, sensory or cognitive function damaged through injury or disease. Through such replacement (or augmentation), neural prostheses aim to improve the quality of life for those with disabilities. The earliest forms of such devices were cochlear implants, which date from the 1950s, but the field really began to expand in the early 2000s, with now virtually every area of human functioning being investigated.

Ethical issues arise from questions of **security**; wireless neural implants can have the same **cybersecurity** vulnerabilities as any other IT system. Where cognitive function is being corrected, the issue of how to obtain informed consent is also significant. Finally, it is a short step from correcting a disability to enhancing normal functioning, and so the nature of the technology is relevant to debates about transhumanism. It can be further argued that **neuroprosthetics** is a gateway to the technological **singularity**.

Further reading:

Kansaku, K., 2021.Neuroprosthetics in systems neuroscience and medicine. *Science Reports*, 11, 5404, https://doi.org/10.1038/s41598-021-85134-4.

Krucoff, M.O., Rahimpour, S., Slutzky, M.W., Edgerton, V.R. and Turner, D.A., 2016. Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation. *Frontiers in Neuroscience*, 10, 584, https://doi.org/10.3389/fnins.2016.00584.

See also: NEUROETHICS, BRAIN-COMPUTER INTERFACE

Neurocapitalism

An – at present – imagined political system where **neurodata** is a secondary **currency**. This vision is constructed through an extrapolation of current trends in **data** use and technological development – most particularly **brain–computer interface**s.

Further reading:

Lesaja, S. and Palmer, X.L., 2020. Brain-computer interfaces and the dangers of neurocapitalism. *arXiv preprint* 2009.07951v1, https://doi.org/10.48550/arXiv.2009.07951.

See also: SURVEILLANCE CAPITALISM, SECONDARY DATA, SECONDARY USE

Neurodata

Data extracted from neural systems. This could be measured as part of a medical process (e.g., a brain scan) or directly through neurotechnology such as **brain–computer interface**s.

Neurodata arguably threatens **privacy** in a manner quite unlike any other data form. The possibility of one's thoughts being read has led some, such as Unger, to argue that there is a tension between the existence of neurodata and the First Amendment to the US Constitution, which concerns freedom of thought and expression rather than privacy.

Further reading:

Hallinan, D., Schütz, P., Friedewald, M. and De Hert, P., 2014. Neurodata and neuroprivacy: data protection outdated? *Surveillance & Society*, 12(1), 55–72, https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata.

Unger, J.D., 2022. Stay out of my head: neurodata, privacy, and the First Amendment. *Washington and Lee Law Review*, 1439, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4293432.

See also: NEUROTECHNOLOGY, FREEDOM OF EXPRESSION, NEUROPRIVACY

Neuroethics

A branch of applied **ethics** which considers the ethical implications of advances in neuroscience and **neurotechnology**, especially covering the actual and potential impact of neuroscientific knowledge and techniques upon society, individuals, and our understanding of morality, **identity**, free will and **privacy**.

Further reading:

Amadio, J., Bi, G.Q., Boshears, P.F., Carter, A., Devor, A., Doya, K., Garden, H., Illes, J., Johnson, L.S.M., Jorgenson, L. and Jun, B.O., 2018. Neuroethics questions to guide ethical research in the international brain initiatives. *Neuron*, 100(1), 19–36, https://doi.org/10.1016/j.neuron.2018.09.021.

May, J., 2023. *Neuroethics*: *agency in the age of brain science.* Oxford: Oxford University Press.

Neuroprivacy

At a superficial level, advances in **neurotechnology** such as **brain–computer interface**s lead to increased collection, storage and analysis of **neurodata** (data related to brain activity, cognitive processes and emotions). However, neurotechnology’s potential impact on **privacy** is much deeper than simply generating a new form of data. The possibility of technological controls on brain processes (as well as vice versa) implies a level of **intrusion** risk which far exceeds that which derives from data **confidentiality breach**es.

Further reading:

Unger, J.D., 2022. Stay out of my head: neurodata, privacy, and the First Amendment. *Washington and Lee Law Review*, 1439, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4293432.

Neuroprosthetics

An interdisciplinary area at the intersection between neuroscience and biomedical engineering which explores the possibility and engineering requirements of neural prostheses.

See also: NEUROETHICS, BRAIN–COMPUTER INTERFACE, NEURAL PROSTHESIS

Neurotechnology

The interface of technology and the nervous system, where the technological aim is to monitor and manipulate the latter’s functioning. Neurotechnologies include brain–computer interfaces (BCIs), neuroimaging techniques (such as functional magnetic resonance imaging or fMRI), neurostimulation methods, neurofeedback and **neuroprosthetics**. Neurotechnology has been driven by the development of techniques and tools for studying, diagnosing and treating various neurological conditions and pathologies. However, its use for enhancement of human brain function and sensory and cognitive extension and augmentation is now a real possibility.

The ethical, legal and social implications of neurotechnology, and most particularly its potential impact on **privacy** and **security**, are significant, and existing legislative frameworks may need to be adapted or updated for the new socio-technical environment.

Further reading:

Hallinan, D., Schütz, P., Friedewald, M. and De Hert, P., 2014. Neurodata and neuroprivacy: data protection outdated? *Surveillance & Society*, 12(1), 55–72, https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata.

Unger, J.D., 2022. Stay out of my head: neurodata, privacy, and the First Amendment. *Washington and Lee Law Review*, 1439, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4293432.

See also: NEURODATA, BRAIN–COMPUTER INTERFACE, NEUROPRIVACY

(N,K) Rule

See: DOMINANCE RULE

Noise Addition

A form of **disclosure control**: the distortion of **data** through some random process. Noise addition underpins the implementation of **differential privacy**.

Nom de Guerre

A nom de guerre, literally 'name of war' or 'war name', is a **pseudonym** used by combatants in violent conflict. This may be to conceal a peacetime **identity**, to protect family from reprisals, to confuse enemy intelligence, or to establish a **reputation** with an intimidating name (as with Stalin, meaning 'man of steel').

Further reading:

Pfukwa, C., 2003. Onomastic innovation in Zimbabwean noms de guerre. *Language Matters*, 34(1), 13–23, https://doi.org/10.1080/10228190308566189.

Nom de Plume

A nom de plume or *pen name* is a **pseudonym** adopted by an author of a work to conceal their **identity** as author. Noms de plume can be adopted for numerous reasons: authors may wish to distance themselves from their work, to avoid retribution or embarrassment, to conceal aspects of themselves (such as gender or nationality), to name a group of writers working together, to change complex or hard-to-spell names, or to create different writing identities for works in different genres.

Further reading:
Room, A., 2010. *Dictionary of pseudonyms*: *13,000 assumed names and their origins*, 5th edition. Jefferson, NC: McFarland.

See also: NOM DE GUERRE, PSEUDONYM

Nominal Data

See: CATEGORICAL DATA

Non-Disclosure Agreements

Non-**Disclosure** Agreements (‘NDAs’) are a common arena in which rights of **privacy** and confidentiality can come into conflict with the right to **freedom of expression**. Following the Harvey Weinstein scandal, the use of NDAs as a tool to silence victims of misconduct and abuses of power has come under **scrutiny**.

However, it should also be stressed that the courts have emphasised the **public interest** in upholding the **confidentiality** set out in NDAs, and thus that privacy and the public interest are not always adversarial interests.

Further reading:
Mendonca-Richards, A., 2019. Privacy and NDAs: the use of Non-Disclosure Agreements in the wake of the #MeToo movement, *Entertainment Law Review*, 30(4), 109, www.farrer.co.uk/news-and-insights/privacy-and-ndas-the-use-of-non-disclosure-agreements-in-the-wake-of-the-metoo-movement/.
Smith, P., 2017. Court of Appeal upholds non-disclosure order and stresses depth of the public interest. *Entertainment Law Review*, 28(8), 270.

See also: CONFLICT OF RIGHTS, FEMINIST CRITIQUE OF PRIVACY

Non-Discrimination Law

Also referred to as ‘anti-discrimination law’, non-discrimination laws protect people from unfair treatment based on **protected characteristics**.

The rise of **big data** has heightened concerns about discrimination being enabled by the **processing** of **personal information**, as individuals can be profiled in a way that is automated, performed at scale and determined without human reflection.

The fear of discrimination on the basis of **genetic data** led to one of the few pieces of federal **privacy** legislation in the United States: the Genetic **Information** Nondscrimination Act of 2008. Nevertheless, discrimination based on algorithmic groupings is (arguably) insufficiently addressed in most legal **jurisdiction**s, as belonging to a group assigned by an **algorithm** is not necessarily a protected characteristic.

Further reading:
Lyon, D., 2003. *Surveillance as social sorting*: *privacy, risk, and digital discrimination*. New York: Routledge.
Wachter, S., 2022. The theory of artificial immutability: protecting algorithmic groups under anti-discrimination law. *Tulane Law Review*, 97(2), 149, https://heinonline.org/HOL/LandingPage?handle=hein.journals/tulr97&div=11&id=&page=.

See also: GENETIC PRIVACY, PROFILING, SURVEILLANCE CAPITALISM, GROUP PRIVACY

Non-Invasive BCI

A type of **brain–computer interface** that does not involve surgical implanting of electrodes into or on the cortex. The majority involve electroencephalographic (EEG) detection via skull caps containing multiple electrodes.

Further reading:
Vidal J.J., 1973. Toward direct brain–computer communication. *Annual Review of Biophysics and Bioengineering*, 2(1), 157–80, https://dpi.org/10.1146/annurev.bb.02.060173.001105.

Notice and Consent

Also known as 'notice and choice'. this is a **consent** model in which a **data subject** is usually presented with a **privacy policy**, or other form of public-facing text, which they must accept before proceeding to a website.

Further reading:
Craig, T. and Ludloff, M., 2011. *Privacy and Big Data*. Sebastopol: O'Reilly Media.

Okoyomon, E., Samarin, N., Wijesekera, P., Elazari Bar On, A., Vallina-Rodriguez, N., Reyes, I., Feal, Á. and Egelman, S., 2019. On the ridiculousness of notice and consent: contradictions in app privacy policies. *In*: *Workshop on Technology and Consumer Protection, in conjunction with the 39th IEEE Symposium on Security and Privacy*, http://hdl.handle.net/20.500.12761/690.

See also: COOKIE, PRIVACY NOTICE

Nudge Theory

Nudge theory or *nudging* is an application of behavioural science where the environment for an individual's decision-making is subtly altered, without their knowledge, to manipulate their incentives for action, and the choice heuristics that they are likely to apply. The result is that their decision-making will be influenced in ways determined by the designer of the environment. In nudge theory, the environment is referred to suggestively as the **choice architecture**.

Nudging is a type of paternalism but is deemed by many to be compatible with liberalism because the ultimate choice is still down to the individual. However, it is a deliberate **breach** of the individual's **decisional privacy**.

Further reading:

Thaler, R.H. and Sunstein, C.B., 2008. *Nudge*: *improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

See also: INTENTION–BEHAVIOUR GAP

Nudging

See: NUDGE THEORY

O

Obfuscation

The practice of hiding one’s **identity** by hiding, masking or fabricating identity **information** demanded by a service (most commonly online). This might be done to preserve **privacy**, especially where the **information** is not necessary for the delivery of that service. Obfuscation techniques include using fake names or addresses and creating multiple email accounts. The same techniques may also be used by malevolent actors (most criminal activity requires some degree of obfuscation) and for undercover work by security and law enforcement services.

Further reading:
Brunton, F. and Nissenbaum, H., 2015. *Obfuscation*: *a user’s guide for privacy and protest.* Cambridge, MA: MIT Press.

See also: MASKING

Objective Harm

While much of the damage that stems from a violation of an individual’s **privacy** lies in their subjective experience, courts of law can also compensate people for more tangible or obviously quantifiable forms of damage. This could include damage to **private property** from an act of **trespass**, or loss of income from **reputation**al damage.

Further reading:
van der Sloot, B., 2017. Where is the harm in a privacy violation? Calculating the damages afforded in privacy cases by the European Court of Human Rights. *JIPITEC*, 8(322), www.jipitec.eu/issues/jipitec-8-4-2017/4641.

See also: HARM, SUBJECTIVE HARM

Oblivious Transfer (OT)

Oblivious transfer is a cryptographic technique that enables **information** transmission between two parties, without revealing any information about the other party to either side. In an oblivious transfer protocol, the sender communicates one piece of a set of information to the receiver but does not know which information is sent. A specific piece of information from the sender is what the receiver wishes to receive at the same moment, but they do not want the sender to be aware of what information they have received. The recipient should also remain unaware of all the elements in the set of information that have not been sent.

Oblivious transfer is frequently used in secure multiparty computing, where parties wish to transmit useful information without disclosing their inputs to the other parties. Applications such as electronic voting, secure auctions and **secure communication** are other use cases.

Further reading:
Rabin, M.O., 2005. *How to exchange secrets with oblivious transfer.* Cryptology ePrint Archive, https://eprint.iacr.org/2005/187.pdf.

See also: CRYPTOGRAPHY, CRYPTOGRAPHIC PROTOCOL

Obscurity

Obscurity is a state of being inconspicuous or unremarked upon, potentially visible but, as a matter of (contingent) fact, unknown. An obscure object is difficult to apprehend or understand, not for any intrinsic reason, but because observers are either unaware of its existence or nature, or it is protected by some accidental cover. Something is *obscured* when something else conceals it from view.

The United States Supreme Court recognised an interest in what it termed *practical obscurity* in the case of *Department of Justice v Reporters' Committee for a Free Press* (1979), which held that FBI criminal records ('rap sheets'), while technically public **information**, were not easily available and were very difficult to get hold of, thus contingently protecting the **privacy** of those named on them. Practical obscurity was a property of paper-based filing systems, where assembling a dossier on someone might be difficult because it would require a large amount of physical **search** and trips to many different buildings, and could be stymied by misfiling, documents lost in floods or fires, closed buildings and so on.

The importance of practical obscurity therefore diminished in the digital age, where search technology, electronic **database**s and the **Internet** meant that large amounts of information could be speedily and effectively assembled. However, it has been argued by Hartzog and Stutzman that obscurity still has a part to play in protecting privacy – and indeed perhaps even a more important part than legal regulation. They identified four factors affecting the obscurity of online information: its visibility to search engines; the existence or otherwise of **access controls**; the ability to associate the information with individuals; and the clarity of the message (including whether it is machine-readable or free text).

Further reading:

File, P.C., 2017. A history of practical obscurity: clarifying and contemplating the twentieth century roots of a digital age concept of privacy. *University of Baltimore Journal of Media Law and Ethics*, 6(1/2), 4–21, https://heinonline.org/HOL/LandingPage?handle=hein.journals/ubjmleth6&div=5&id=&page=.

Hartzog, W. and Stutzman, F., 2013. The case for online obscurity. *California Law Review*, 101(1), 1–49, https://heinonline.org/HOL/LandingPage?handle=hein.journals/calr101&div=4&id=&page=.

See also: ANONYMITY, PUBLIC RECORDS

Obtrusion

Obtrusion is the invasion of one’s experience by unwelcome distractions, such as a mobile phone conversation conducted by someone close by, or loud music being played elsewhere. Obtrusion is often seen as a type of **privacy** invasion, and O’Hara refers to it as **extrinsic privacy**. Its **breach** means that one cannot retreat into oneself to achieve the kind of **solitude** and freedom from influence that was an ideal of stoicism and medieval monasticism. **Intimacy** is also rendered difficult with external **intrusion**s, as Robert Gerstein argued. However, others, such as Ruth Gavison, have argued that it is not a type of privacy, and that the language of privacy is only used by false analogy with other types of intrusion.

Further reading:

Gavison, R., 1980. Privacy and the limits of law. *Yale Law Journal*, 89(3), 421–71. https://search.informit.org/doi/10.3316/agispt.19810026.

Webb, D., 2007. *Privacy and solitude*. London: Hambledon Continuum.

See also: INTRUSION UPON SECLUSION, SECLUSION

OECD Guidelines (On Privacy)

The Organisation for Economic Co-operation and Development (‘OECD’) was formed in 1960, as a reconstitution (and expansion) of the Organisation for European Economic Co-operation, which originally formed in 1948 as part of the reconstruction of Europe.

In 1980, the OECD adopted the Guidelines governing the Protection of **Privacy** and Transborder Flows of **Personal Data**. These guidelines set out principles which influenced the EU **Data Protection Directive** in 1995, as well as the Council of Europe's **Convention 108** on the Automatic Processing of Personal Data in 1981 (with which there was some mutual influence, via cross-over between the writing groups). The Guidelines stemmed from a general recognition that **privacy** laws based on manual **data** were inadequate to regulate the development of digital technology.

The principles set out in the 1980 OECD Guidelines continue to influence data protection legislation and practice across the world. Their basic requirements for the lawful **processing** of personal data have been carried forward in the EU's **GDPR** and can provide some 'back-up' for OECD member states with no comprehensive privacy legislation (such as the United States). However, as the **Schrems** litigation has demonstrated, the OECD guidelines alone are not enough to ensure that a signatory state is compliant with EU **data protection** requirements, particularly as they do not prevent broad exemptions for **national security** purposes.

Further reading:

Kirby, M., 2011. The history, achievement and future of the 1980 OECD guidelines on privacy. *International Data Privacy Law*, 1(1), 6–14, https://doi.org/10.1093/idpl/ipq002.

Organisation for Economic Co-operation and Development, 2002. *OECD guidelines on the protection of privacy and transborder flows of personal data*. Paris: OECD Publishing, https://doi.org/10.1787/9789264196391-en.

See also: DATA PROTECTION PRINCIPLES, HISTORY OF PRIVACY

Offline Dictionary Attack

Instead of repeatedly attempting various **password** combinations in real time against a live system, an **adversary** utilising an offline dictionary attack tries to guess a password using a precomputed collection of password hashes.

The adversary in an offline dictionary attack requires a copy of the hashed passwords from a system or **database**, typically by taking advantage of a flaw and copying or stealing the password file. The adversary then creates a dictionary or wordlist of probable passwords and calculates the hash value for each password on the list. The adversary next checks to see whether there is a match between the computed hash values and the hashed passwords acquired from the system or database.

Further reading:

Nam, J., Paik, J., Kang, H., Kim, U.M. and Won, D., 2009. An off-line dictionary attack on a simple three-party key exchange protocol. *IEEE Communications Letters*, 13(3), 205–7, https://doi.org/10.1109/LCOMM.2009.081609.

One-Stop-Shop

The EU's **GDPR** introduced a 'one-stop-shop' mechanism, which is the name given in Recitals 127–128 for the process whereby one regulator takes the lead for investigation of **cross-border processing**.

One of the key challenges for **data protection** enforcement is the global reach of many online services. The GDPR therefore sets out rules to determine which national data protection regulator should take the lead where **data processing** affects citizens in multiple member states. The **regulators** (or 'supervisory authorities') are expected to cooperate and provide each other with **mutual assistance**. In the case of conflict, however, the **European Data Protection Board** can arbitrate under the **consistency mechanism**.

Further reading:

Woods, L., 2021. 'Facebook Ireland' and the one stop shop under the GDPR. *European Law Review*, 46(5), 685–91, https://dialnet.unirioja.es/servlet/articulo?codigo=8121910.

See also: SUPERVISORY AUTHORITY, DATA PROTECTION AUTHORITY

One-Way Hash Function

A mathematical operation known as a one-way hash function produces a fixed-size output known as the hash value or digest from a **plaintext**. Such a function is deterministic, so that the same input produces the same output. In a one-way hash, it should be computationally impossible to reverse the process and decode the original message from the hash value, making it secure, and allowing the hash to stand for the original message in many applications.

In **cryptography** and computer **security**, one-way hash functions are frequently used for a range of tasks, such as **password** storage, **digital signature**s and message **authentication**. Systems can check passwords without retaining sensitive **information** that can be revealed in a **data breach**, for instance by storing just the hash values of passwords rather than the passwords themselves.

Because one-way hash **algorithm**s are intended to be collision-resistant, it should be near impossible to generate two messages with the same hash value.

Further reading:

Naor, M. and Yung, M., 1989. Universal one-way hash functions and their cryptographic applications. *In*: *Proceedings of the twenty-first annual ACM symposium on theory of computing*, 33–43, https://doi.org/10.1145/73007.73011.

Onion Routing

See: TOR

Online Vetting

Online vetting is the practice of searching the **Internet** for evidence of the character of a person. Obvious sources for such evidence are **social media** sites; vehicles for the expression of opinion, such as blogs; and **search** engines. Based on such evidence, the searcher can then decide whether and how to engage with the person. Online vetting is a prominent practice for employers thinking of employing someone, but it is used in many other circumstances. A potential employee may conversely wish to research their prospective employer; customers or investors may vet the companies they are thinking of dealing with; prospective partners (or their parents) may vet someone at the outset of a romantic relationship.

Further reading:

Berkelaar, B.L. and Buzzanell, P.M., 2015. Online employment screening and digital career capital: exploring employers’ use of online information for personnel selection. *Management Communication Quarterly*, 29(1), 84–113, https://psycnet.apa.org/doi/10.1177/0893318914554657.

See also: DIGITAL FOOTPRINT

Ontological Security

A state of equilibrium where a (human) entity has an enduring sense of self over time and across contexts. Ontological **security** is enhanced by coherence of experience and diminished by disjunctive changes.

Ontological security underpins theories of mental health, such as Laing's, but has also been applied to fields as diverse as housing, education, international politics and climate change.

Privacy and ontological security may be considered as interdependent. Ontological security may be required to underpin meaningful **intimacy** in relationships as it both supports and is supported by **self**-expression. Neurological privacy as a space for enabling self-reflection may be necessary for processing and integrating experience to ensure the necessary continuity of self.

Acute privacy **breach**es in many cases represent a discontinuity of experience of the type that damages ontological security. Similarly, overt **surveillance** can be considered to have an adverse effect on ontological security. Specifically, the effect of self-censoring that invariably accompanies **awareness** of surveillance means that a **mask** is then presented which is discordant with the underlying self.

Further reading:

Ejdus, F., 2018. Critical situations, fundamental questions and ontological insecurity in world politics. *Journal of International Relations and Development*, 21(4), 883–908, https://doi.org/10.1057/s41268-017-0083-3.

Hiscock, R., 2013. Ontological security and psychosocial benefits from the home: qualitative evidence on issues of tenure. *Housing, Theory and Society*, 18(1–2). http://dx.doi.org/10.1080/14036090120617.

Laing, R.D., 1994. The divided self. *The British Journal of Psychiatry*, 165(3), 420–3. https://doi.org/10.1017/S0007125000072986.

See also: CHILLING EFFECT, PERSONHOOD, MENTAL PRIVACY

Onward Transfer

A transfer of **personal data** received from another party to a **third party** (and beyond). If the **data** are personal data and such an onward transfer occurs, then the original **data controller** who shared the **information** may remain accountable for any downstream **data processing** that occurs. In such cases, prospective **data provenance** information might need to be generated to facilitate decision making. There are, however, situations in which a data controller shares personal data with another organisation and is not legally responsible for the onward processing of that separate controller – for example, a public authority may need to collect data from other public bodies to generate national statistics, but in that instance they are responsible for their processing, and the contributing bodies are only responsible for the **accuracy** of their own data.

Further reading:

European Data Protection Board, 2021. *Guidelines 07/2020 on the concepts of controller and processor in the GDPR, version 2.1*. Brussels: EDPB, https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf.

Regan, P.M., 2003. Safe harbors or free frontiers? Privacy and transborder data flows, *Journal of Social Issues*, https://doi.org/10.1111/1540-4560.00064.

See also: DATA FLOW, DATA TRANSFER, JOINT DATA CONTROLLER

OPE

See: ORDER-PRESERVING ENCRYPTION

Open Access

Open access is a model of publishing that imposes no costs on readers to access, copy or disseminate, requiring an alternative income stream for covering costs (such as managing peer review). Open access has been thoroughly explored in the context of research outputs and may include access to **data** as well as written research papers. The legal position of open access is unclear in several respects. Clearly, if the **dataset** contains **personal data**, then measures need to be taken to ensure that access to it is restricted. Inadequate **data** protection could undermine research **ethics**, but strong **data protection** could require complex pseudonymisation measures, or even prevent full open access. It would be problematic if **data subject**s' consent was required for access. Hence, a typical funding model for open access publishing would mandate open access for research **publication**s, but provide opt-outs for research data, to protect **privacy** and other rights.

Further reading:

Wessels, B., Finn, R.L., Linde, P., Mazzetti, P., Nativi, S., Riley, S., Smallwood, R., Taylor, M.J., Tsoukala, V., Wadhwa, K. and Wyatt, K., 2014. Issues in the development of open access to research data. *Prometheus*, 32(1), 49–66, http://dx.doi.org/10.1080/08109028.2014.956505.

See also: OPEN DATA, OPEN SOURCE

Open Data

The creators of **data**, and sometimes other stakeholders, usually have rights to restrict the use of data. Open data is data that is made freely available for anyone to use, edit, share or exploit commercially, without serious restriction or requirement to pay. The terms of use are usually specified by an *open licence*, which may, for example, require users to cite the source of the data in any **publication**s they produce. The difference between open data and **freedom of information** (FOI) is that, with FOI, the **user** has a right to request certain **information**, whereas when the data is open, it is there to be downloaded by any users without notice.

The chief restriction on open data is its commercial value, which may deter private companies from publishing. Opening up data may also erode competitive advantage. Those using data for free, that has been laboriously acquired by someone, are in effect free-riding. Governments may be concerned that releasing data could impact on their standing (for example, revealing low standards of health or education). There is also a cost to a provider in getting data into a state where it can easily be reused by others (e.g., checking **data quality**, providing **metadata**, using well-known ontologies). For the user, **risks** are the **data provenance** and quality of the data.

As the data is open, it is impossible for **privacy** to be preserved by controlling access. Elliot et al carried out an extensive red team attack on two of the datasets held by the UK's Office for National Statistics which are available under licence. The aim was to assess additional risk of making the licensed datasets open. They found a significant increase in the risk with a high rate of successful identifications. Hence, if **personal data** is involved, either it will have to have further disclosure controls applied to it before publication, sufficiently strong to reduce the risk of identification by an arbitrary **intruder** to a negligible level for ever (in many cases an impossibly high bar), or **consent** of the **data subject**s must be sought, or there must be a regulatory requirement to publish. **Compliance** with **GDPR** would be a high bar for anyone wishing to make personal data open.

Further reading:

Elliot, M., Mackey, E., O'Shea, S., Tudor, C. and Spicer, K., 2016. End user licence to open government data? A simulated penetration attack on two social survey datasets. *Journal of Official Statistics*, 32(2), 329–48, https://doi.org/10.1515/jos-2016-0019.

Kitchin, R., 2014. *The data revolution*: *big data, open data, data infrastructures and their consequences*. London: Sage.

O'Hara, K., 2011. *Transparent government, not transparent citizens*: *a report on privacy and transparency for the Cabinet Office*. London: Cabinet Office, https://eprints.soton.ac.uk/272769/.

See also: DATA RELEASE, OPEN ACCESS, OPEN SOURCE, VALUE OF DATA

Open Source

Open source is a decentralised model for product development, most notably **software**. The *source code* is published and made publicly available for modification, leading to open collaboration and software development by a distributed peer group. The open source model contrasts with a proprietary model in which software is developed within companies, the code protected, and released under strict licence conditions (as a *black box*). Examples of influential open source systems include the Linux operating system kernel, and the Android mobile operating system, acquired by Google in 2005.

Because it is decentralised, open source is usually thought to be more secure than proprietary software in the long run, because vulnerabilities are more easily spotted and corrected by a wider and more diverse group of designers. While it might be thought that hackers might also benefit from access to the source code, it appears that vulnerabilities are usually found in both proprietary and open source software by creative probing from outside. Since open source code is transparent, any **privacy**-threatening aspects might also, in principle, be detected by an expert peer community. However, the effectiveness of these protections of **security** and privacy is proportional to the size of that community and the diligence with which it scrutinises the code.

Open source intelligence (OSINT) is the collection of intelligence from open sources of **information**, such as **social media**, local or citizen journalism and commercially available satellite imagery (Russia's invasion of Ukraine in 2022 was partly monitored by OSINT, as Russian soldiers posted their activities on social media). OSINT poses a privacy threat (as well as other issues, such as **breach**es of copyright), but in compensation is cheap and **risk**-free to gather. It has become clear that social media accounts of tourists and foreign students are often monitored by governments to assess the desirability of letting visitors in, while the same happens in the workplace to employees and prospective employees. Currently, there seems relatively little **accountability** for the use of OSINT.

Further reading:

Eijkman, Q. and Weggemans, D., 2012. Open source intelligence and privacy dilemmas: is it time to reassess state accountability? *Security and Human Rights*, 23(4), 285–96, https://heinonline.org/HOL/LandingPage?handle=hein.journals/helsnk23&div=46&id=&page=.

Hansen, M., Köhntopp, K. and Pfitzmann, A., 2002. The Open Source approach – opportunities and limitations with respect to security and privacy. *Computers and Security*, 21(5), 461–71, https://doi.org/10.1016/S0167-4048(02)00516-3.

See also: OPEN ACCESS, SOFTWARE DEVELOPMENT LIFECYCLE, COMMUNITY PRIVACY, PROPRIETARY PRIVACY

Opt-In

Opt-in is a model of acquiring assent favoured by those who see consent as a key safeguard of individual **autonomy**. It requires – in the words of the **GDPR** – that **consent** be signalled by a ‘clear affirmative act’. In other words, acceptance cannot be inferred by the **data subject**’s passivity, as is the case under an **opt-out** model.

The GDPR’s opt-in consent requirements have proven influential in other **jurisdiction**s. For example, the proposed American Data **Privacy** and Protection Act also requires consent to be ‘affirmative’ and ‘express’.

See also: US PRIVACY LAWS

Opt-Out

In discussions of **information governance**, organ donation and research recruitment, **consent** can be characterised as requiring a positive act (**opt-in**) or as inferred in the absence of an active objection or withdrawal (opt-out). The space between the two approaches reflects the diversity (at times, controversy) in views as to the scope of evidence required to say that an individual has legally and/or ethically accepted an intervention. This evidential scope encompasses the signalling standards, which vary for particular ‘types’ of consent.

In some cases, the law is clear as to when opt-out consent is an acceptable basis for interfering with someone’s private **information** or bodily **privacy**. **GDPR**, for example, states that the processing of sensitive **data** requires **explicit consent**, and as such an opt-out model would not secure this condition for processing. Other areas of law are more disputed in their requirements: for example, whether the failure to object is sufficient to authorise the **disclosure** of confidential **information**.

The requirements of a valid *ethical* consent are also a matter of scholarly debate. Mackay breaks the term down into two sub-types: **implicit consent** – in which silence can (in some circumstances) be construed as genuine acceptance – and *presumed* consent, which does not require any evidence that the subject has understood that their inaction will be taken as consent to a particular intervention. He argues convincingly that implicit consent can be a sufficiently autonomous **authorisation** to constitute a consent for bioethical purposes, but presumed consent cannot.

Some **jurisdiction**s, such as Australia, have gone beyond academic and policy debate, and codified the conditions for acceptable opt-out consent in research.

Further reading:

Dove, E.S. and Taylor, M.J., 2021. Signalling standards for progress: bridging the divide between a valid consent to use patient data under data protection law and the common law duty of confidentiality. *Medical Law Journal*, 29(3), 411–45, https://doi.org/10.1093/medlaw/fwab014.

MacKay, D., 2015. Opt-out and consent. *Journal of Medical Ethics*, 41, 832–5, http://dx.doi.org/10.1136/medethics-2015-102775.

See also: AUTONOMY, INFORMED CONSENT, LEGAL BASIS FOR PROCESSING

Order-Preserving Encryption

A type of **encryption** that keeps the **plaintext data**'s order in the encrypted version. In other words, if two values in plaintext have a connection in terms of their order (for example, one is greater than the other), then the same relationship will exist between their encrypted values.

OPE is frequently used when it is necessary to safeguard the **data** from unauthorised access while still maintaining the order of the data for analysis or searching. OPE, for instance, may be used to encrypt a **database** of credit card transactions such that the sensitive **information** is still shielded from unauthorised access, yet the sequence of the transactions is retained for analysis.

OPE, however, has several restrictions and significant **security** hazards. One of the key dangers is that an **adversary** may utilise the **encryption**'s order-preserving property to deduce important **information** from the encrypted data. By examining the sequence of the encrypted data, an adversary could exploit OPE, for instance, to identify which credit card transactions were for bigger sums or which users had the biggest balances.

Another drawback is that OPE is susceptible to frequency attacks, in which an adversary may guess the encrypted values for some plaintext values by counting how often such values appear. If the range of the plaintext values, such as the numbers in the range of 1 to 100, is limited, this might be particularly difficult.

Further reading:

Agrawal, R., Kiernan, J., Srikant, R. and Xu, Y., 2004. Order preserving encryption for numeric data. *In*: *Proceedings of ACM SIGMOD international conference on management of data*, 563–74, https://doi.org/10.1145/1007568.1007632.

Ordinal Data

See: DISCRETE DATA

Orwell Attack

A form of **paparazzi attack** where the number of **Bluetooth sensor**s is high, and a track and trace server colludes in the **attack**. The combination of ubiquitous **surveillance** and a centralised collaborator is the reason for the name. The additional payload is that all **data** available to the server can be used in the linkage, facilitating both the **sensitivity** of the linkage process and the scale of surveillance possible for the **adversary**.

Further reading:

Avitabile, G., Botta, V., Iovino, V. and Visconti, L., 2020. Towards defeating mass surveillance and Sars-Cov-2: The pronto-c2 fully decentralized automatic contact tracing system. *Cryptology ePrint Archive Report 2020/493*, https://eprint.iacr.org.

Buccafurri, F., De Angelis, V. and Labrini, C., 2020. A privacy-preserving solution for proximity tracing avoiding identifier exchanging. *In: International Conference on Cyberworlds*, 235–42, https://doid.org/10.1109/CW49994.2020.00045.

See also: DATA LINKAGE, TRACKING, LINKAGE ATTACK

OT

See: OBLIVIOUS TRANSFER

Other

In psychology, a companion concept to the self. Except in the most extreme forms of solipsism, 'self' only meaningfully exists in relation to an 'other'.

From a **privacy** perspective, the **self** might be regarded as the fundamental unit to be protected and 'other' would be that from whom the self needs protection. **Self-disclosure**, as the term is usually employed, implies some form of selection of 'other' with concomitant confidentiality boundaries being set up through a social contract between the conversational participants.

Further reading:

Greene, K., Derlega, V.J. and Mathews, A., 2006. Self-disclosure in personal relationships. *In: The Cambridge handbook of personal relationships*, Cambridge: CUP, 409, 427.

Laing, R.D., 1961. *Self and others*. London: Tavistock Publications.

See also: CONFIDENTIALITY, BOUNDARY, DISCLOSURE

Outing

Outing is the practice of disseminating sensitive **information** about a **person** that they themselves wish to keep concealed. In particular, it is often used to refer to the public revelation that a person is gay, thereby forcing them '*out* of the closet'. Gay people have been outed maliciously, but also by activists for political reasons – for example, when the outed person is perceived to have supported anti-gay measures. In Western cultures, the rapid decline in prejudice against gay people from the 1980s onwards has led to the practice reducing in its impact.

Further reading:

Chekola, M., 1994. Outing, truth-telling, and the shame of the closet. *Journal of Homosexuality*, 27(3–4), 67–90, https://doi.org/10.1300/J082v27n03_05.

See also: INFORMATION PRIVACY, HARASSMENT

Outlier

An unusual (possibly extreme) **data unit**. Outliers are problematic for **confidentiality** reasons because they are easy to **single out**. Statistically they may be **special unique** and may be vulnerable to **spontaneous recognition**.

Output Checking

The process of applying **disclosure risk assessment** to analytical outputs, usually in the context of trusted research environments or other forms of **safe setting**. The process can either be managed (which usually implies a set of hard and fast rules – such as the **threshold rule** – by an output checker) or collaborative (where the researcher and output checker work together to produce 'safe' output that meet a set of defined disclosure control principles).

Further reading:

Arbuckle, L., and Ritchie, F., 2019. The five safes of risk-based anonymization. *IEEE Security & Privacy*, 17(5), 84–9, https://doi.org/10.1109/MSEC.2019.2929282.

Griffiths, E., Greci, C., Kotrotsios, Y., Parker, S., Scott, J., Welpton, R., Wolters, A. and Woods, C., 2019. *Handbook on statistical disclosure control for outputs*. Safe Data Access Professionals Working Group, https://securedatagroup.files.wordpress.com/2019/10/sdc-handbook-v1.0.pdf.

See also: SAFE OUTPUT, FIVE SAFES, OUTPUT STATISTICAL DISCLOSURE CONTROL

Output Privacy

Output **privacy** means ensuring that the output of any **data** analysis does not compromise the privacy of persons represented in the input data. Where there is output privacy, the disclosed output is designed to prevent **inference** backwards to the input, so that an **adversary** cannot infer anything about the input that is privacy-**breach**ing. For instance, given a statistical analysis of confidential **census** data, could an adversary infer something about individuals? A secondary issue is that even if it is not possible to reverse engineer the output, it may still be possible to perform attribute disclosure on the output itself.

Output privacy also applies to **secure multi-party computation**, where different parties contribute their own data to a **machine learning** effort but wish to keep their data confidential from their partners. Output privacy here means that the output of the learning, distributed to all the partners, does not allow backwards inference to other partners' data.

Sanitising the input data may not be sufficient to prevent reverse engineering sensitive inputs if the adversary can bring auxiliary **information** to the output and the outputs are sufficiently detailed. Furthermore, where output is streamed rather than a one-off analysis (so that the mining output is published on a continuous or regular basis), multiple releases may be viewed by the adversary in combination, revealing temporal patterns.

To secure output privacy, the results of the analysis may also need to be perturbed or manipulated to prevent backwards inference (e.g., **output statistical disclosure control**, or **differential privacy**), and/or access/query control must be applied.

Further reading:

Ricciato, F., Bujnowska, A., Wirthmann, A., Hahn, M. and Barredo-Capelot, E., 2019. A reflection on privacy and data confidentiality in Official Statistics. *Presented at: 62nd ISI World Statistics Conference.* https://ec.europa.eu/eurostat/cros/content/reflection-privacy-and-data-confidentiality-official-statistics-0_en.

Smith, D. and Elliot, M., 2008. A measure of disclosure risk for tables of counts. *Transactions on Data Privacy*, 1(1), 34–52, www.tdp.cat/issues/tdp.a003a08.pdf.

See also: PRIVACY-PRESERVING MACHINE LEARNING, INPUT PRIVACY, OUTPUT CHECKING

Output Statistical Disclosure Control

A process by which analytical outputs are manipulated to minimise the **disclosure risk**. This is most relevant to **safe settings** where access is controlled but the **data** are highly detailed and would be personal if released as open data.

Output disclosure control is usually managed by a process of output checking and comes in two forms: rules-based, where a set of hard and fast rules such as the threshold rule are applied to the outputs; and principles-based, where the rules are seen as guidelines and the decision is made through a dialogue between the output checker and the researcher who has produced the output. One approach to this is to allow anything that meets the rules, while for everything else the research must demonstrate that it is 'safe'.

Further reading:

Arbuckle, L., and Ritchie, F., 2019. The five safes of risk-based anonymization. *IEEE Security & Privacy*, 17(5), 84–9, https://doi.org/10.1109/MSEC.2019.2929282.

Griffiths, E., Greci, C., Kotrotsios, Y., Parker, S., Scott, J., Welpton, R., Wolters, A. and Woods, C., 2019. Handbook on statistical disclosure control for outputs. *Safe Data Access Professionals Working Group*, https://securedatagroup.files.wordpress.com/2019/10/sdc-handbook-v1.0.pdf.

See also: FIVE SAFES, OUTPUT PRIVACY, P/Q RULE, P% RULE, THRESHOLD RULE, TRUSTED RESEARCH ENVIRONMENT

Overimputation

Imputation is a suite of techniques that are commonly used to replace missing values within **microdata** in order to improve analytical completeness. Overimputation is a method of **statistical disclosure control** that uses imputation to overwrite non-missing values for record-variable pairs. The idea is that some indirect identifiers are selectively overwritten to increase the uncertainty of any **reidentification attack**.

P

P3P

See: PLATFORM FOR PRIVACY PREFERENCES

Packet Filter

A packet filter is a **network security** mechanism which keeps track of incoming and outgoing **network** traffic and permits or denies **data** packets in accordance with a predetermined set of criteria.

At the network layer of the Open Systems Interconnection model of **network communication**, packet filters operate by inspecting each data packet as it travels across the network. The filters can be implemented in hardware or software, and they can be set up to permit or deny packets depending on several factors, including source and destination **IP addresses**es, ports, **protocol**s and packet content.

Network security rules can be implemented by using packet filters, such as banning traffic from known malicious IP addresses or access control of ports or services. By prohibiting unneeded or unauthorised traffic from using up network resources, they may also be used to control network traffic and enhance network performance.

Further reading:
Chapman, D.B., 1992. Network (in)security through IP packet filtering. *In*: *USENIX UNIX Security Symposium III*, www.usenix.org/legacy/publications/library/proceedings/sec92/full_papers/chapman.pdf.

See also: TRAFFIC DATA

Packet Sniffing

The technique of intercepting and analysing **network** traffic in real time to collect and investigate the contents of individual **data** packets as they transit over a computer network is known as packet sniffing, sometimes known as packet capture or packet analysis. It is a method for network monitoring, troubleshooting and **security** analysis, used for recognising and diagnosing network problems, tracking network performance and studying network behaviour.

Packet sniffing attacks use the technique maliciously to obtain **password**s or sensitive data. Precautions against them include the use of **encryption** to safeguard **sensitive data**, the implementation of access restrictions to limit network access, the use of **network segmentation** to isolate various network components and the use of secure network **protocol**s like **HTTPS**.

Further reading:
Chapman, D.B., 1992. Network (In) Security Through IP Packet Filtering. *In: USENIX UNIX Security Symposium III*, www.usenix.org/legacy/publications/library/proceedings/sec92/full_papers/chapman.pdf.

See also: DATA IN TRANSIT, NETWORK SECURITY, PACKET FILTER

Panel Data

See: LONGITUDINAL DATA

Panopticon

The original Panopticon was a plan suggested by philosopher Jeremy Bentham in 1787 for a House of Correction (prison) whose design ensured that every prisoner could potentially be under covert surveillance. This would ensure that prisoners would self-correct or self-censor their behaviour to avoid punishment, even when they were in fact unobserved. Bentham believed that this would both inflict a type of punishment without the evil of actually inflicting pain, while helping the prisoners become more virtuous people. Though the Panopticon was never actually built, its principles were borrowed by George Orwell in his novel *Nineteen Eighty-Four* and by philosopher Michel Foucault, who saw in it a mechanism of power characteristic of the modern age, in which society was suffused with disciplinary processes.

Further reading:
Bentham, J., 1995. *The Panopticon writings*. London: Verso.
Foucault, M., 1977. *Discipline and punish*. New York: Pantheon.
Orwell, G., 1949. *Nineteen eighty-four*. London: Secker & Warburg.

See also: SURVEILLANCE, BIG BROTHER, CHILLING EFFECT

Paparazzi

Paparazzi refers collectively to photographers or journalists who specialise in taking pictures of celebrities and public figures, often in their private moments or without their **consent**. The term is the Italian pluralisation of ‘Paparazzo’, the name of a character in the film *La Dolce Vita* (1960), a pushy and intrusive freelance photographer.

Paparazzi are known for their aggressive and persistent tactics, often using long-range lenses, hiding in bushes or other covert locations and following celebrities. Their photographs are often sold to tabloid newspapers and celebrity gossip magazines and can fetch high prices depending on the fame and popularity of the person in the photograph.

Further reading:

McNamara, K., 2011. The paparazzi industry and new media: the evolving production and consumption of celebrity news and gossip websites. *International Journal of Cultural Studies*, 14(5), 515–30, https://doi.org/10.1177/1367877910394567.

See also: ATTENTIONAL PRIVACY, CELEBRITY PRIVACY, INTRUSION

Paparazzi Attack

A type of **linkage attack** specifically aimed at **location data** which exploits the fact that **Bluetooth** systems exchange **identity information**. This type of attack is specifically relevant to downstream **dataset**s generated by **tracking** and tracing **application**s and became salient during the COVID-19 pandemic. The **attack** involves installation of covert Bluetooth sensors across an area of interest to collect the ephemeral **identifier**s of users as they pass by. This is then combined with the seed **information** generated when a **user** tests positive.

Further reading:

Avitabile, G., Botta, V., Lovino, V. and Visconti, I., 2020. Towards defeating mass surveillance and Sars-Cov-2: The pronto-c2 fully decentralized automatic contact tracing system. *Cryptology ePrint Archive Report* 2020/493, https://eprint.iacr.org.

Buccafurri, F., De Angelis, V. and Labrini, C., 2020. A privacy-preserving solution for proximity tracing avoiding identifier exchanging. *International Conference on Cyberworlds*, 235–42, https://doi.org/10.1109/CW49994.2020.00045.

See also: DATA LINKAGE

Parental Controls

Software which enables third parties (usually parents) to control the content to which children may have access on digital devices (including computers, mobile devices, television and games), their usage of that content or the time of their **exposure**.

Further reading:
Livingston, S. and Helsper, E.J., 2008. Parental mediation of children’s internet use. *Journal of Broadcasting and Electronic Media*, 52(4), 581–99, https://doi.org/10.1080/08838150802437396.

See also: ACCESS CONTROL, INTERNET

Partially Homomorphic Encryption

A type of **encryption** that permits some calculations to be made on the encrypted material without first having to decode it. Since PHE only allows one kind of mathematical operation – addition or multiplication – but not both, it is only partially homomorphic.

PHE transforms the encrypted **data** using mathematical operations that enable the **ciphertext** to be used for operations, such as addition and multiplication of the encrypted numbers. The outcome is a result that has been encrypted and may be decoded to show the outcome of the calculation on the **plaintext** data.

Further reading:
Acar, A., Aksu, H., Uluagac, A.S. and Conti, M., 2018. A survey on homomorphic encryption schemes: theory and implementation. *ACM Computing Surveys*, 51(4), 1–35, https://doi.org/10.1145/3214303.

See also: HOMOMORPHIC ENCRYPTION

Participant Information Sheets

In research using human **data**, Participant Information Sheets are a common way of communicating **privacy**-relevant **information** at the time of collecting data and obtaining **consent**. Where **data protection** laws require **data subject**s to be provided with a defined list of information at the point of collection, a participant information sheet can be an appropriate way of providing this information.

See also: PRIVACY NOTICE, TRANSPARENCY

Participatory Surveillance

Participatory **surveillance** is the monitoring or surveying of communities with their express cooperation, asking them to provide the bulk of the **data**. Although the approach dates back to the 19th century, usually some kind of digital technology is used to monitor behaviour. Community members can either upload **information** of interest using smartphones or other digital devices, or alternatively allow data from their phones to be monitored – for example, so that **population** movements can be tracked using **location data**. With sufficient participation, large high-quality **dataset**s may be assembled, visualised and used in near-real-time policy formation. Participatory surveillance has advantages of lower cost and higher scale than other types of survey and is perceived as particularly valuable in monitoring epidemics of transmissible diseases.

Further reading:
Wójcik, O.P., Brownstein, J.S., Chunara, R. and Johansson, M.A., 2014. Public health for the people: participatory infectious disease surveillance in the digital age. *Emerging Themes in Epidemiology*, 11, article no.7, https://doi.org/10.1186/1742-7622-11-7.

See also: DATAVEILLANCE, GEO-SOCIAL DATA

Passive Collection

Passive collection is a kind of **data collection** performed without using any active methodology. Its key characteristic is the non-intrusive nature of the data collection process. It relies on technology such as **smart device**s (**sensor**s, cameras, etc.) or **automated system**s for collecting data without any human interaction.

Passive collection may raise **privacy** concerns, particularly when **personal information** is involved, as the passivity means any conscious decision to provide the **data** will have been taken in a different context.

Password

A password is a sequence of characters used to authenticate the **identity** of a user when accessing an application or system. A password is usually associated with a **username**. The level of **security** provided by a password depends on (a) whether users can keep it **secret**, and (b) how easily it can be guessed or cracked by an **adversary**. The latter depends in turn on the password’s length and complexity. However, the longer and more complex it is, the harder it is to remember, a problem for which **password manager**s have been designed.

Further reading:

DellAmico, M., Michiardi, P. and Roudier, Y., 2010. Password strength: an empirical analysis. *In*: *2010 Proceedings of IEEE INFOCOM*, 1–9, https://doi.org/10.1109/INFCOM.2010.5461951.

See also: AUTHENTICATION, MULTI-FACTOR AUTHENTICATION

Password Manager

Software that securely manages and stores passwords and other sensitive login **information** is known as a password manager.

In order to access their encrypted **password** vault, which holds all of their login **information** for multiple websites and applications, users of password managers only need to remember one master password. It is usually simple to generate complicated, unique passwords for each account without having to remember them all thanks to browser extensions or mobile applications that can autofill login forms with the saved **credentials**.

Password managers utilise a variety of **security** techniques, including AES **encryption**, **multi-factor authentication** and biometric **authentication**, to safeguard the sensitive **data** they contain.

See also: BIOMETRIC DATA

Patch

A piece of code that is added to existing **software** or cybersystem to fix a **vulnerability** or **bug**.

Patch Management

The process of locating, obtaining, testing and delivering patches or updates to **software**, operating systems, **firmware** and other components of a computer **network** is known as patch management. To decrease vulnerabilities, reduce downtime and avoid cyberattacks, patch management keeps software and systems up to date with the most recent **security** updates, **bug** fixes and performance improvements. Patch management includes compiling a comprehensive inventory of software assets, keeping an eye out for patch releases on vendor websites and security warnings, testing patches in a supervised setting and installing fixes according to a standardised procedure to guarantee minimal network downtime.

Further reading:
Hasan, C., Huseyin, C. and Jun, Z., 2008. Security patch management: share the burden or share the damage? *Management Science*, 54(4), 657–70, https://doi.org/10.1287/mnsc.1070.0794.

PDS

See: PERSONAL DATA STORE

Peeping Tom

The archetypal name for a voyeur, especially of sexual activities, named after a character in a folk tale who spied on Lady Godiva as she rode naked through the streets of Coventry.

Further reading:
Hartland, E.S., 1890. Peeping Tom and Lady Godiva. *Folklore*, 1(2), 207–26, https://doi.org/10.1080/0015587X.1890.9720007.

See also: VOYEURISM

Penetralia

The innermost, often **secret**, element of an entity. Used about buildings but also about people (referring perhaps to subconscious thoughts and beliefs).

See also: SECRECY

Penetration Test

Penetration testing (often shorted to ‘pen testing’) is a **cybersecurity** activity in which a **cybersecurity** expert conducts one or more simulated **attack**s on an organisation cyber system. The purpose of the simulation is to identify vulnerabilities so that they can be addressed by the organisation.

One variant of a pen test is the **motivated intruder test**, where, rather than cybersystems, **dataset**s are subjected to simulated attacks.

See also: INTRUDER TESTING, RED TEAM, WHITE HAT ATTACK

Persistent Cookie

See: SUPER COOKIE

Persistent Pseudonym

With **pseudonymisation**, a persistent **pseudonym** is one which refers to the same person or entity through time on the same data system. The persistence of the pseudonym may be limited to a certain period of time (such as with a **sessional cookie**) or may extend indefinitely. Because of this, the pseudonym allows instances of data about that person or entity to be linked, and therefore for their behaviour to be traced through time. This makes the pseudonymised data more useful for researchers (e.g., it would allow a medical patient’s progress to be monitored without directly identifying them), but at the same time gives richer **information** to an **adversary**, increasing the **risk** of a **reidentification** of a pseudonymous entity as a result.

Further reading:
Elliot, M., Mackey, E. and O’Hara, K., 2020. *The Anonymisation Decision-Making Framework*: *European practitioners’ guide*, 2nd edition. Manchester: UKAN Publications, https://ukanon.net/framework/.

See also: DATA LINKAGE, LINKABILITY, PERSON, RECORD LINKAGE

Person

See: NATURAL PERSON, PERSONHOOD

Personal Data

This term **personal data** is used within EU law to denote **information** relating to an identified or identifiable individual – the **data subject**. It is thus the subject matter of **data protection** law in European jurisdictions. The parameters of the concept are a key touchstone for what the **GDPR** calls its **material scope** – that is, scope as defined by the *nature* of the **data processing**, as opposed to scope based on the location of the processing, which is instead the **territorial scope**.

The importance of the concept thus lies in its impact on when **data** protection laws do (not) apply to information processing. When it is not clear whether information constitutes ‘personal data’, it is not clear whether data protection rights and obligations apply to its use. Any ambiguity in the term thus has a significance beyond the purely academic, with implications for the clarity of **privacy** rights in a digital context.

The main ambiguity lurking within the GDPR’s definition of personal data is when natural, ‘living’ people can be said to be ‘identified’ or ‘identifiable’ from information. These words are not, themselves, defined within the GDPR. The reader is left to question whether the ‘**identity**’ revealed by an ‘identification’ needs to provide direct access to the ‘real-world’ individual, or if it is sufficient that they can be singled out and scrutinised within a **dataset**? The degree of **risk** required to make someone ‘identifiable’ is also contentious within academic and regulatory circles.

Given the stated aim of the GDPR to preserve privacy and data protection rights, a purposive understanding of the ‘identity’ revealed (or potentially revealed) by personal data is usually broad. This includes information sufficiently unique to allow a person’s characteristics to be evaluated, as this alone can be enough to engage privacy rights. A false profile that leads to personalised adverts still has privacy implications for the ‘misidentified’ individual – for example, parents who have experienced a recent miscarriage receiving adverts for baby products.

The EU’s definition of personal data has proven influential. The term is used in other **jurisdiction**s, and data protection frameworks currently proposed for the United States and the African Union use an almost identical definitions for their ‘covered data’.

Further reading:

Mourby, M. and Mackey, E., 2023. Pseudonyms, profiles and identity in the digital environment. *In*: van Der Sloot, B. and van Schendel, S., eds, *The boundaries of data*: *technical, practical and regulatory perspectives.* Amsterdam: Amsterdam University Press.

Prince, A.E.R., 2022. *I tried to keep my pregnancy secret.* The Atlantic: www.theatlantic.com/ideas/archive/2022/10/can-you-hide-your-pregnancy-era-big-data/671692/.

See also: IDENTIFIABLE INDIVIDUAL, UNIQUE IDENTIFIER, IDENTIFIABLE DATA, IDENTIFIED DATA, IDENTIFIER

Personal Data Cloud

See: PERSONAL DATA STORE

Personal Data Store (PDS)

A Personal Data Store (PDS) is a type of **Personal Information Management System** (PIMS) designed to allow individuals to store, retrieve and share their **personal data**, as well as to decide who receives access to data and on what terms. A PDS is a centralised system from the user's perspective, although actual data storage may be distributed across multiple servers, and possibly in the cloud. Cloud-based PDSs are sometimes called **Personal Data Cloud**s.

The idea of using PDSs to support individuals' maintenance of their **privacy** in the **digital economy** has to meet a number of challenges, as described by Van Kleek and O'Hara. They need to be able to store data, and not become obsolete, for arbitrarily long periods; they must be usable for non-specialists; they need to navigate complex regulatory environments, particularly for third-party access to **information**, and possibly across jurisdictions; they need to meet international data-handling and **security** standards; they need to integrate **Privacy-Enhancing Technology** (PET); and they will need to be future-proofed against technological and social change.

Further reading:

Van Kleek, M. and O'Hara, K., 2014. The future of social is personal: the potential of the personal data store. *In*: Miorandi, D., Maltese, V., Rovatsos, M., Nijholt, A. and Stewart, J., eds, *Social collective intelligence*: *combining the powers of humans and machines to build a smarter society*. Cham: Springer, 125–58, https://doi.org/10.1007/978-3-319-08681-1_7.

See also: CLOUD STORAGE, DATA STORAGE, INFORMATION SECURITY

Personal Identification Number (PIN)

A (usually) user-chosen number used for **authentication**, effectively a numerical **password**. Most online systems that use PINs now only do so as part of a **multi factor authentication** system.

Further reading:
Martin, K., 2012. *Everyday cryptography*: *fundamental principles and applications*. Oxford: Oxford University Press.

Personal Information

While data protection law in most European countries governs '**personal data**', meaning **information** which identifies natural living people, other jurisdictions (such as Australia and China) use the phrase 'personal information' instead.

The term also occurs outside data protection law, such as in the UK Statistics and Registration Service Act (2007), which defines 'personal information' as **data** released by the UK's Office for National Statistics (ONS) that either directly identifies an individual (living or dead) or does so in conjunction with other information that is already in the **public domain**.

The term can also be used more informally to mean information that is significant to the person whom that data is about. I might regard my health data as important and therefore care about who has access to it but be quite relaxed about who knows that I am a member of a trade union, even though both are regarded as **special category data** under the GDPR. Personal information in this sense incorporates the notions of **privacy preferences** and sensitivity beyond that which is expressed by the (more precise) technical definitions.

Further reading:
UK Data Service, 2012. *Statistics and Registration Services Act*, https://ukdataservice.ac.uk/learning-hub/research-data-management/data-protection/data-protection-legislation/statistics-and-registration-services-act/.

Personal Information Management System (PIMS)

Personal Information Management is the set of tasks and activities that carried out by individuals to control their **information**. Such information may be recorded on paper or digitally, and typically a personal collection will include photographs (both digital and prints), emails, address books, music (on hard media or digital), video (hard media or digital), letters, documents (both paper and digital), and so on. Management of this information includes safe storage, organising and filing, archiving, retrieving when needed, using, reusing and deletion.

Typically, the information under management will contribute to one or more purposes: perhaps pleasure, *aides-memoires*, records of family, friends and events, organising one's personal life, dealings with government, work-related tasks, treatment of medical conditions and so on. Such information may be of very short-term interest only or may potentially be needed years or decades hence. Some, for example family **record**s, may be bequeathed to future generations. Some may be of historic significance.

Where information is digital, Personal Information Management Systems (PIMS) may be used to provide a central point for storing and indexing an individual's information. They can be straightforward cloud-based file hosting systems such as Dropbox and Microsoft OneDrive. Where PIMS have been a research topic, they have focused on the preservation and storage of personal data, affording the individual control over their **data**, either deciding who can access and use the data (**consent** management), or at a minimum keeping track of who accesses it (traceability).

With a PIMS, the individual's gains in control and **privacy** are offset by increasing responsibility for **security** and management; it must therefore be well-designed and usable. In a PIMS there will be a balance between the subjective importance of information for the individual, which will influence its visual salience and accessibility, and the objective importance of information, which will determine how often it needs to be found and retrieved.

Further reading:

Bergman, O., Beyth-Marom, R. and Nachmias, R., 2003. The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology*, 54(9), 872–8, https://doi.org/10.1002/asi.10283.

Jones, W., 2008. *Keeping found things found*: *the study and practice of Personal Information Management*. Burlington: Morgan Kaufmann.

See also: PERSONAL INFORMATION, PERSONAL DATA STORE, DIGITAL FOOTPRINT, DELETION, RECORDS MANAGEMENT, DATA STORAGE, INFORMATION SECURITY

Personalisation

Personalisation (also *customisation*) is the creation of a product or service for an individual **user**, as opposed to the mass production of commoditised items to be provided for a range of users. In **e-commerce**, personalisation may take the form of personalised items to be traded, or alternatively for **recommendation system**s to personalise the choice function, placing the choices most like to appeal to the buyer at the head of the list. Online **communication**s can also be personalised, such as news feeds, webpages and maps, as well as presentational items such as layouts, backgrounds and ringtones. Personalised items, in theory, increase user satisfaction, but potentially at the cost to the provider of economies of scale.

Personalisation is generally carried out based on **information** gathered about, and thereby associated with, the individual user, plus information gathered about other users whose profiles are close to the user. Information that might be found valuable includes that about the behaviour of the user and their expressed preferences, and historical **data** about their interactions, the context of a transaction, and others (*collaborative filtering*). It follows that, for personalisation to be effective, the provider must know as much relevant information about the user as possible; **privacy** is therefore an impediment to personalisation. Personalisation may be provider-led, consumer-led or co-created, and hence the information about the user may be volunteered or gathered surreptitiously. Bias may mean that certain groups are better understood, and hence receive better personalised services, than others.

Furthermore, personalised services remove common elements to the experiences of users. Filtering of news and other information services mean that two individuals may not have many shared reference points to aid their discussions and debates. Political campaigns may make different, and even contradictory, policy pitches to different voters. The **public** space may therefore be impoverished, creating important issues for democratic politics.

Further reading:

Kuksa, I., Fisher, T. and Kent, T., eds., 2023. *Understanding personalisation*: *new aspects of design and consumption*. Cambridge: Chandos Publishing.

O’Hara, K., 2021. Personalisation and digital modernity: deconstructing the myths of the subjunctive world. *In*: Kohl, U. and Eisler, J., eds, *Data-driven personalisation in markets, politics and law*. Cambridge: Cambridge University Press, 37–54, https://doi.org/10.1017/9781108891325.004.

See also: PERSONALISATION REACTANCE, PERSONALISED SERVICES, PROFILING, TARGETTED ADVERTISING

Personalisation Reactance

Emails or other **communication**s from companies are often personalised to customers. **Personalisation** reactance is a customer's resistance to the message and reluctance to click through to the offer being made when the fit between the company's insight into their personal characteristics and the offer is not perceived to be justified.

Further reading:

White, T.B., Zahay, D.L., Thorbjørnsen, H. and Shavitt, S., 2008. Getting too personal: reactance to highly personalized email solicitations. *Marketing Letters*, 19(1), 39–50, https://doi.org/10.1007/s11002-007-9027-9.

See also: CREEPINESS

Personalised Medicine

Personal data can be used to inform automated diagnosis, prognosis or treatment advice in healthcare. Where **data** is used to tailor medical services to an individual patient, this is known as personalised medicine.

Personalised medicine poses similar **privacy** challenges to other forms of **personalised services**; the **risk** of discrimination through inaccurate profiling is a particularly acute issue, as global health research data skews in favour of people of White European ancestry, limiting clinical insight into other demographics. The secondary use of health data for purposes beyond individual care – a prerequisite for personalised medicine – also challenges the exclusivity of the conventional confidential relationship between clinician and patient.

Further reading:

Bodiroga-Vokobrat, N. et al., 2019. *Personalized medicine in healthcare systems: legal, medical and economic implications*. Cham: Springer, https://doi.org/10.1007/978-3-030-16465-2.

Hartlev, M., Gefenas, E., Mourby, M., O'Cathaoir, K. and Lukaseviciene, V. 2020. *EU-STANDS4PM report: legal and ethical review of in silico modelling*, www.eu-stands4pm.eu/publications.

See also: BIG DATA, CONFIDENTIALITY, PROFILING

Personalised Services

With the rise of **Big Data**, services can be increasingly tailored to an individual based on their automated profile. The neutral term for this is ‘personalised services’, although critics have warned of associated phenomena such as **price discrimination** and **surveillance capitalism**.

Cohen has argued that key to the ongoing **value of privacy** is its sheltering of the nebulous, fluid private self from the interest of market forces in rendering them fixed, transparent and predictable. Servers of **targeted advertising**, for example, will predict future purchases based on previous online behaviour, and thus nudge the **data subject** into remaining in the category of, for example, ‘yo-yo dieter’ or ‘weekend alcohol buyer’. For these insights to remain commercially valuable, the individual should ideally operate like Cohen’s fixed, transparent entity, and not behave too impulsively or opaquely according to their own mercurial will.

Further reading:
Cohen, J.E., 2013. What privacy is for. *Harvard Law Review*, 126(7), 1904–33.

Personally Identifiable Information (PII)

Personally identifiable information is the standard way in the US of referring to **data** from which someone is identifiable. Conceptually, this is equivalent to the **data protection** concept of **personal data**, but whereas the latter has been given a universal definition in EU law, the meaning of PII varies across contexts. Sometimes it refers to **information** that can be associated with an individual, and sometimes to information that uniquely designates an individual. Furthermore, its scope is almost always narrower than that of personal data.

PII is usually defined in the context of specific legislation in the US, resulting in certain types of identifying information (e.g., financial information, health information, information about children) receiving greater protection than others. As a result, many European commentators criticise the concept of PII as partial, unprincipled and contingent.

Further reading:
Erika McCallister, E., Grance, T. and Scarfone, K., 2010. *Guide to protecting the confidentiality of personally identifiable information (PII)*. Gathersburg, MD: National Institute of Standards and Technology, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf.

See also: IDENTIFIABLE DATA, IDENTIFIABLE INDIVIDUAL

Personal Space

Personal space is the physical space around an individual which they are not comfortable with another entering, because they would feel encroached upon. This varies across cultures and according to whom the **intruder** is: an intimate, an acquaintance, a stranger, an authority figure, a child, someone of a different gender, and so on. Theories vary according to whether the space itself changes (phenomenologically) depending on the intruder or whether the space is fixed but who is admitted to that space varies.

Even animals apparently have a sense of personal space, and indeed the concept emerged from zoology, only later being applied to humans.

Hall, who coined the term **proxemics** to denote the study of personal space, identified four regions or levels of personal space: Intimate, Personal, Social and **Public**, with increasing distance and social intensity implied as you move from the intimate space through to the public.

Some authors have observed the emergence of a digital equivalent to physical personal space through notions of digital overcrowding.

Further reading:

Altman, I., 1975. *The environment and social behavior: privacy, personal space, territory, crowding*. Monterey: Brooks/Cole.

Hall, E.T., 1966. *The hidden dimension*. New York: Doubleday.

Joinson, A.N., Houghton, D.J., Vasalou, A. and Marder, B.L., 2011. Digital crowding: privacy, self-disclosure, and technology. *In*: Trepte, S. and Reinecke, L. (eds) *Privacy online.* Berlin: Springer, https://doi.org/10.1007/978-3-642-21521-6_4.

See also: BODILY PRIVACY, SPATIAL PRIVACY

Personhood

The property of being a **person**.

The meaning of this term has been long debated. It is often discussed in terms of entities that deserve moral consideration; however, while this has intuitive force, it raises as many questions as it answers. Some, such as Noonan, want to tie personhood to human DNA. However, others want to exclude certain cases (e.g., dead bodies, biological samples, those in persistent vegetative states or foetuses) or include other species, and **Artificial Intelligence**, as candidates for personhood. Warren argues for six cognitive criteria for attributing personhood (consciousness, reasoning, self-motivated activity, capacity to communicate, self-awareness and moral agency).

Even discussion of legal personhood, which it might be argued is a more bounded term, reveals complexity. Legal personality accrues to entities that have acquired through some process the right to carry out legal processes that (some) humans can do – take ownership of property, sign contracts, and so on. **Legal person**s include some non-human entities such as companies and sovereign states. These non-human legal persons are also referred to as juridical persons. Legal personality is **jurisdiction** specific both in terms of how it is obtained and what it is the legal personhood enables the legal person to do.

Personhood is intrinsically tied to the concept of **privacy**. As Reiman observes, on the one hand, non-persons may not be accorded the right to privacy. On the other hand, (some) privacy may be necessary for personhood to (fully) develop.

Further reading:

Kurki, V.A.J., 2009. A theory of legal personhood. *Oxford Academic,* https://doi.org/10.1093/oso/9780198844037.001.0001.

Noonan, J.T., 2002. *Persons and masks of the law*: *Cardozo, Holmes, Jefferson, and Wythe as makers of the masks*. Berkeley: University of California Press.

Warren, M.A., 1997. *Moral status*: *obligations to persons and other living things*. Oxford: Clarendon Press.

Reiman, J.H., 2017. *Privacy*. Abingdon: Routledge.

See also: NATURAL PERSON, SELF, OTHER

Perturbation

Any **statistical disclosure control** method which alters values within a **dataset** rather than the overall **data** structure. Examples include **rounding**, **micro-aggregation**, **noise addition** and **record swapping**.

Pervasive Computing

See: UBIQUITOUS COMPUTING

PET

See: PRIVACY-ENHANCING TECHNOLOGY

Pharming

A cyberattack that redirects an internet user from a genuine website to a fake site by installing **malware** on the victims' computer. Pharming can be conducted either by changing the website host's file on a victim's computer or by exploiting vulnerabilities in the **DNS server**s responsible for resolving **Internet** names into their underlying **IP address**es. Compromised servers are sometimes referred to as 'poisoned'.

The term is a portmanteau of 'farming' and the related concept of **phishing**; both pharming and phishing can be used to gain **information** for online **identity theft** or to download (further) **malware** to the victim's computer. Pharming is a potential concern for **e-commerce** and online banking, although to date the number of publicised pharming attacks is small, perhaps reflecting the relative difficulty of executing such an attack compared to phishing.

Further reading:
Brody, R., Mulig, E. and Kimball, V., 2007. Phishing, pharming and identity theft. *Academy of Accounting and Financial Studies Journal*, 11, 43–56, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c77207b960d972c300f0ba5fc50a90092d4444f9.

PHE

See: PARTIALLY HOMOMORPHIC ENCRYPTION

Philosophy of Information

The philosophy of **information** is concerned with conceptual issues thrown up by information processing, whether by digital technology, the mind, the human brain or the natural world (e.g., in the genetic instructions carried by DNA). In particular, it must look at questions of how information is represented, and what characteristics it (or its representations) must have. What is it about a particular physical setup, whether marks on paper, voltages in electronic circuits, states of neurons or arrangements of physical objects, that makes it a representation of a particular piece of information? The important aspects of the physical setup are the *signal*, while all else is *noise*. A signal typically requires an interpretational system, in whose absence information cannot be deciphered. For instance, a lost culture's writings may be meaningless to future civilisations.

The central defining idea of ‘information’ is still that of Claude Shannon, that information conveys something through a state; by occupying that state rather than another (e.g., a 1 rather than a 0) it conveys some meaning as opposed to an alternative. The more unlikely such a state is, the more information is conveyed.

Some, such as Luciano Floridi and Fred Dretske, have argued that information must have some factual content, while others, such as Ludwig Wittgenstein in his *Tractatus Logico-Philosophicus*, have argued the related point that a tautologous sentence that is always true (e.g., 2+2=4) carries no information. On such a theory, a state of or event in the world causes some trace, which carries the information that the state/event has happened. However, this view of information as a statement of something true about the world is hard to square with a view of information as a signal. For instance, a computer program is often called information, but is only a series of instructions and says nothing true or false. A timetable is an ideal, but expresses information about the future behaviour of an object that may or may not turn out to be true. A sensor sends information to the outside world when it works properly, but if its outputs become unreliable, do its outputs cease to be information because they are misleading? It may be that the term ‘information’ is multiply ambiguous, making such paradoxes difficult to resolve outside of specific theories.

Floridi’s *ontological theory of* ***privacy*** is related to his philosophy of information; privacy is a function of the forces that hinder or promote the flow of information. As Floridi also argues that people are at least partly constituted by their information, so privacy has a direct effect on personal **identity**.

Further reading:

Dretske, F.I., 1981. *Knowledge and the flow of information*. Oxford: Blackwelll.

Floridi, L., 2005. The ontological interpretation of informational privacy. *Ethics and Information Technology*, 4(4), 287–304, https://doi.org/10.1007/s10676-006-0001-7.

Floridi, L., 2013. *The ethics of information*. Oxford: Oxford University Press.

Shannon, C.E. and Weaver, W., 1949. *The mathematical theory of information*. Urbana: University of Illinois Press.

See also: INFORMATIONAL PRIVACY, INFORMATION ETHICS, INFOSPHERE, ETHICS

Phishing

Phishing is a sort of cyberattack wherein users are tricked into disclosing sensitive **information**, such as **username**s, **password**s, credit card numbers or other **personal information**, via the use of false or deceptive approaches. Adversaries frequently fabricate emails, websites and other forms of **communication** that seem to be coming from a reliable source, such as a bank, social media site or well-known organisation. Such communications will typically contain a link or attachment which, if clicked on or opened, will facilitate the **adversary**'s access to sensitive information.

See also: SOCIAL ENGINEERING, PHARMING

Phone Hacking

Phone **hacking** is the practice of stealing **information** from mobile phones and **communication** devices. The hacker gains access to **application**s running on the phone, via vulnerabilities, **social engineering** or installing **spyware**, and can extract information or to modify settings. In a notorious series of cases that prompted the Leveson Inquiry into journalistic practice of 2011–12, the voicemails of **public figure**s were hacked by British journalists, who were able to gain access from landlines having established the **PIN**s.

Further reading:
Mills, E., 2011. Kevin Mitnick shows how easy it is to hack a phone. *CNET*, 7, www.cnet.com/news/privacy/kevin-mitnick-shows-how-easy-it-is-to-hack-a-phone/.

See also: COMMUNICATION PRIVACY, CELEBRITY PRIVACY, CHILDREN'S PRIVACY, HARASSMENT, PAPARAZZI, TELEPHONE TAPPING

Physical Privacy

Physical **privacy** is freedom from being looked at, listened to or recorded against one's wishes. As such, it is a type of what has also been called **attentional privacy**.

Further reading:
Moreham, N.A., 2014. Beyond information: physical privacy in English law. *Cambridge Law Journal*, 73(2), 350–77, https://doi.org/10.1017/S0008197314000427.

See also: SURVEILLANCE

PIA

See: PRIVACY IMPACT ASSESSMENT

PII

See: PERSONALLY IDENTIFIABLE INFORMATION

PIMS

See: PERSONAL INFORMATION MANAGEMENT SYSTEM

PIN

See: PERSONAL IDENTIFICATION NUMBER

Pixelisation

The blurring of all or part of an image by reducing the granularity of the pixels.

Pixelisation is used for multiple purposes, one of which is the protection of individual identities. It is commonly used to blur out faces and car licence plates.

Further reading:
Fan, L., 2018. Image pixelization with differential privacy. *In: Proceedings of data and applications security and privacy XXXII*: *32nd annual IFIP WG 11.3 conference*, Bergamo, Italy, 16–18 July 2018, 32, Springer, 148–62.

Plaintext

Plaintext is any **data** or **information** that has not been encrypted. In **communication** and storage systems that do not require high levels of **security** or where security is not the main concern, plaintext is frequently used.

However, if it can be intercepted or accessed by unauthorised parties, plaintext can also present a security **risk** if it contains sensitive or confidential information. Even a shredded plaintext document may be readable if enough of it is recovered and reassembled.

See also: CIPHERTEXT, CRYPTOGRAPHY, ENCRYPTION

Platform for Privacy Preferences (P3P)

Platform for Privacy Preferences (P3P) – created by the World Wide Web Consortium (W3C) – was a **protocol** that standardised how websites provide their **privacy** policies to browsers. P3P gave websites a machine-readable means to communicate their privacy policies, empowering users to decide whether to share **personal information**.

P3P described a website's privacy rules, such as the types of **data** gathered, how they are used and if they are shared with third parties, using an XML-based language. A user's Web browser could then automatically scan the **privacy policy** of a website that supports P3P and compare it to the user's **privacy preference**s.

However, its focus on negotiation and **transparency** rather than enforcement undermined its approach. P3P was perceived as neither supporting user privacy (because a website's preferences might not be respectful of privacy) nor even ensuring that websites adhered to the P3P policies they posted. Its perceived complexity also hindered its use by non-experts. It eventually fell out of use and became obsolete.

Further reading:

Reagle, J. and Cranor, L.F., 1999. The Platform for Privacy Preferences. *Communications of the ACM*, 42(2), 48–55, https://doi.org/10.1145/293411.293455.

Electronic Privacy Information Center, 2000. *Pretty poor privacy*: *an assessment of P3P and Internet privacy*, https://archive.epic.org/reports/prettypoorprivacy.html.

Poisoning Attack

Poisoning involves an **adversary** purposefully introducing harmful **data** or manipulating valid data to undermine the **accuracy** and **integrity** of a **machine learning** model. In most poisoning attacks, the machine learning model's training data are altered by the adversary, to manipulate the model's output in the adversary's favour. The adversary can get the model to anticipate or categorise things incorrectly by adding the malicious input or changing already existing data.

In applications such as image recognition, **spam** filtering or fraud detection, where inaccurate predictions might have serious repercussions, poisoning assaults can be particularly damaging. Measures including **data sanitisation**, **anomaly detection** and robust model training procedures are used to prevent poisoning attacks.

Further reading:
Tian, Z., Cui, L., Liang, J. and Yu, S., 2022. A comprehensive survey on poisoning attacks and countermeasures in machine learning. *ACM Computing Surveys*, 55(8), 1–35, https://doi.org/10.1145/3551636.

Population

A set of all entities (**population unit**s) which share a common characteristic. Classically the characteristic would be geographical, and the entities would be people (e.g., the population of London, UK), but in principle any set of characteristics and any type of entity could form a population.

In the context of **data**, a population can be viewed as the set of population units that could appear in a **dataset**. The dataset could be a sample and so not all units within the population will necessarily be in the dataset.

Population Unique

A term coined by Bethlehem et al which denotes a population unit which is unique within its **population** on a given set of **attribute**s or a **record** within a **dataset** that is unique within the population on a given set of **key variable**s.

Further reading:

Bethlehem, J.G., Keller, W.J. and Pannekoek, J., 1990. Disclosure control of micro**data**. *Journal of the American Statistical Association*, 85(409), 38–45. https://doi.org/10.2307/2289523.

See also: UNIQUENESS

Population Unit

A real-world entity; the fundamental unit of a **population**. The term is distinguished from **data unit**, which is a representation of a population unit in a **dataset**.

Port Scan

A port scan is a method for finding open ports on a **network** device or system. Computers utilise ports as **communication** endpoints when interacting across a network. Sending network packets to a system or device and waiting for it to respond with details about the state of its ports constitutes a port scan. There are three possible states for a port: open, closed or filtered. An open port denotes the presence of a service or **application** that is prepared to receive connections. A filtered port indicates that a firewall or other **security** measure is preventing access to that port, whereas a closed port indicates that no service is currently operating on that port.

In addition to being used for harmful reasons by hackers to find weak systems that may be attacked, port scanning can be used for lawful purposes such as network diagnostics and security testing. Network administrators can put security rules in place that restrict access to important network resources, configure **intrusion detection system**s to look for unusual network activity and employ **firewall**s to deny access to unused ports as a defence against port scans.

Further reading:

Gadge, J. and Patil, A.A., 2008. Port scan detection. *In*: *2008 16th IEEE international conference on networks*, IEEE, 1–6. https://doi.org/10.1109/ICON.2008.4772622.

See also: NETWORK SECURITY

Positive Externalities of Disclosed Data

In economics, an externality is an effect on someone that is not included in the price. They can be positive (for example, a person may be able to see a sporting event from the balcony of their flat, despite not having a ticket) or negative (such as having to breathe the pollution from a car). Because they are not priced in, the person does not have to pay for positive externalities and does not receive compensation for the negative ones.

Posner argued that the disclosure of private **data** is, perhaps paradoxically, associated with many positive externalities, because the purpose of keeping data private is to bolster **information** asymmetries between the **data subject** and others, and such asymmetries prevent markets achieving efficient and optimal distributions of resources. More information in the **public domain** benefits the majority because decisions will be more informed. **Privacy**, on the other hand, tends to bring negative externalities, on this view, benefiting only the person whose privacy is protected and bringing costs to the rest.

Further reading:

Posner, R.A., 1981. The economics of privacy. *American Economic Review*, 71(2), 405–9. https://www.jstor.org/stable/1815754.

See also: ECONOMICS OF PRIVACY, PRIVACY AS REDISTRIBUTION OF COSTS, BENEFITS OF PRIVACY, DISCLOSURE, NEGATIVE EXTERNALITIES OF PRIVACY

Post Quantum Cryptography (PQC)

A type of **encryption** capable of withstanding attacks from **quantum computing**. Several cryptographic methods now in use might be vulnerable, including **RSA Encryption** and Elliptic Curve **Cryptography**.

PQC is becoming more and more significant as quantum computers develop in power. Although it is still unclear when quantum computers will become strong enough to seriously threaten the present cryptographic methods, some experts think that the rapid development of PQC algorithms and protocols is vital to prevent widescale disclosure of sensitive and confidential **data** and significant damage to the digital economy.

PQC algorithms use multivariate cryptography, lattice-based encryption and other techniques that are thought to be difficult even for quantum computers to solve mathematically.

Further reading:

Kumar, M. and Pattnaik, P., 2020. Post quantum cryptography (PQC) – an overview. *In*: *2020 IEEE High Performance Extreme Computing Conference*, 1–9. https://doi.org/10.1109/HPEC43674.2020.9286147.

Post Randomisation (PRAM)

A **statistical disclosure control** method which replaces categorical values in a **microdata** file with new values drawn from a transition probability matrix.

Further reading:

Gouweleeuw, J.M., Kooiman, P. and De Wolf, P.P., 1998. Post randomisation for statistical disclosure control: theory and implementation. *Journal of Official Statistics*, 14(4), 463, www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/post-randomisation-for-statistical-disclosure-control-theory-and-implementation.pdf.

See also: K-ANONYMITY

PPDA

See: PRIVACY-PRESERVING DATA ANALYTICS

PPDM

See: PRIVACY-PRESERVING DATA MINING

PPML

See: PRIVACY-PRESERVING MACHINE LEARNING

P/Q Rule

A rule used in **statistical disclosure control** for summary statistics (usually volumes or averages). The presupposition is that – using already available **information** – the contribution of any **population unit** to the statistics can be estimated to within p per cent and after the **publication** of the statistic the value can be estimated to within q per cent. The ratio p/q represents the information gain for an **adversary** through publication of the statistic. If that information gain is unacceptable the cell is deemed to be disclosive. The setting of the parameters p and q will be determined empirically; the threshold ratio is a matter of judgement, though, and will usually be set by the **data controller**.

There is a conceptual relationship between the p/q ratio and the **epsilon** parameter used in **differential privacy**. The parameter p also relates to the baseline level used in the **correct attribution probability** metric.

A variant of the p/q rule is known as the **prior posterior ambiguity rule**, where, rather than the ratio, the difference between p and q is calculated.

Further reading:
Willenborg, L. and De Waal, T., 2012. *Elements of statistical disclosure control*. Vol. 155. Springer. https://doi.org/10.1007/978-1-4613-0121-9.

See also: DOMINANCE RULE, OUTPUT PRIVACY

PRAM

See: POST-RANDOMISATION

Predictive Analytics

The practice of using **machine learning** to model future behaviour of some system, market, **population** or phenomenon. Use cases are varied, from predicting flu prevalence rates to stock market prices, but a critical feature of most predictive analytics is the ingestion of large amounts of **data**.

Predictive analytics throw up numerous **privacy** (and other ethical) concerns. The ingested **data** may well be personal, and some applications (e.g., predictive policing) may themselves lead to privacy invasions. They may also treat individuals as responsible for decisions they have not yet made or actions they have not yet performed, or submit populations to measures that are not justified by the current situation (e.g., a heightened level of **surveillance** in advance of a predicted surge in crime).

Furthermore, they are often able to discover weak signals in noisy data, thereby making explicit relationships in the data that were not discernible with less computationally intensive analysis. Such **information** may well be 'new' to a human audience or system, therefore bringing it out of **obscurity**.

Further reading:

Parikh, R.B., Obermeyer, Z. and Navathe, A.S., 2019. Regulation of predictive analytics in medicine. *Science*, 363(6429), 810–12, https://doi.org/10.1126/science.aaw0029.

Siegel, E., 2013. *Predictive analytics: the power to predict who will click, buy, lie, or die*. Hoboken, NJ: John Wiley & Sons.

See also: BIG DATA

Predictive Modelling

The use of a statistical model to predict an outcome based on observed **data**; the outcome is often in the future, but predictive models can be used to estimate current or even unknown past events. The general form of a predictive model is a set of explanatory variables used to predict a single **response variable**; however, more complex types have been developed, including auto regression, multilevel and multiple indicators multiple causes models, which vary from that general form.

As the idea of predictive modelling is to predict the values of unknown **data** points, **attribute disclosure risk** may arise if the model is good enough in its predictions. The model owner might collect the values for the **explanatory variable**s from a **data subject** (perhaps for legitimate reasons) and then use this to predict the value of the response variable (perhaps without the data subject's **consent** or even **awareness**).

See also: PREDICTIVE ANALYTICS

Presence Detection

See: MEMBERSHIP INFERENCE ATTACK

Price Discrimination

Price discrimination refers to the practice of adjusting the price of a product based on an assessment of the customer's perceived ability to pay. Also known as personalised pricing or differential pricing, it uses **personal data** to estimate the level of a customer's demand and to set the price as high as possible to maximise revenue whilst still making the sale. It is also an example of the potential opacity of **algorithm**ic discrimination, which the law may not be equipped to address.

Price discrimination was not always seen as unfair. Until standardised pricing became common in the 19th century, all markets were subject to price discrimination, via techniques such as haggling, although these were rarely **privacy**-threatening. Standardised pricing was initially a means of greater efficiency and was not thought of as fairer.

Further reading:

Odlyzko, A., 2003. Privacy, economics, and price discrimination on the Internet. *In*: *Proceedings of the 5th International Conference on Electronic Commerce*, ACM, 355–66, https://doi.org/10.1145/948005.948051.

Iordanou, C., Soriente, C., Sirivianos, M. and Laoutaris, N., 2017. Who is fiddling with prices? Building and deploying a watchdog service for e-commerce. *In*: *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, ACM, 376–89. https://doi.org/10.1145/3098822.3098850.

See also: E-COMMERCE, NON-DISCRIMINATION LAW, ECONOMICS OF PRIVACY

Primary Data

Data which have been gathered or created for their current use.

See also: SECONDARY DATA

Prior Posterior Ambiguity Rule

A variant of the **p/q rule** where, rather than the ratio, the metric is the difference p–q.

Further reading:

Willenborg, L. and de Waal, T., 1996. *Statistical disclosure control in practice*. Springer: New York. https://doi.org/10.1007/978-1-4612-4028-0.

Privacy

Privacy is a concept that covers an enormous range of connected and disparate phenomena, as this dictionary attests. Lexicographical dictionaries emphasise withdrawal of or lack of access to a private person or matter, freedom from attention and **seclusion**. The difficulty in making such a complex idea pragmatically usable was cited by Daniel Solove, who argued that privacy was really a family resemblance term, with different uses of the term having various things in common between them, but nothing common to all of them.

Kieron O'Hara argues that standard usage of the English term 'privacy' typically covers a range of ideas: **informational privacy**, **decisional privacy**, **private property**, **psychological privacy**, **ideological privacy**, **spatial privacy**, **attentional privacy** and extrinsic privacy (or **obtrusion**). Each of these exhibits aspects of the lexicographical definition, while their range testifies to the abstraction and fluidity of privacy. The idea of privacy is fluid across time and context and can be significantly shaped by social movements and technological development. For example, the idea of **bodily privacy** took on a particular significance in North America following second-wave feminism and *Roe v Wade.* More recently, the common understanding of the **private sphere** has arguably been altered since the **World Wide Web** permeated our domestic lives.

Further reading:

Nissenbaum, H., 2004. Privacy as contextual integrity. *Washington Law Review*, 79, 119, https://heinonline.org/HOL/LandingPage?handle=hein.journals/washlr79&div=16&id=&page=.

O'Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Solove, D.J., 2008. *Understanding privacy*. Cambridge, MA: MIT Press.

See also: ASSOCIATIONAL PRIVACY, AUTONOMY, CELEBRITY PRIVACY, CHILDREN'S PRIVACY, COMMUNICATION PRIVACY, COMMUNITY PRIVACY, CONTEXTUAL INTEGRITY, CO-PRIVACY, EXTRINSIC PRIVACY, FINANCIAL PRIVACY, FAMILY RESEMBLANCE THEORY OF MEANING, GENETIC PRIVACY, GEOPRIVACY, GROUP PRIVACY, INPUT PRIVACY, INTELLECTUAL PRIVACY, INTIMACY, LOCATIONAL PRIVACY, MENTAL PRIVACY, NEUROPRIVACY, OUTPUT PRIVACY, PERSONHOOD, PHYSICAL PRIVACY, PRIVACY AS CONTROL, PROPRIETARY PRIVACY, SPATIAL PRIVACY, TERRITORIAL PRIVACY, TIPS

Privacy as Control

As outlined in our definition of **privacy**, the term captures a broad range of values, with significant variation in its characterisation by different theorists. The idea of 'privacy as control' is shorthand for one such theory – often attributed to Alan Westin – that privacy is the claim of individuals, groups or institutions to determine how **information** about them is communicated to others.

While Westin drew on 1960s anthropology to consider 'control' in physical and psychological senses – through dress, emotional **reserve** and community practices – the term is now commonly used in a narrower sense, to mean an individual's influence over uses of their **personal data**. Austin has argued that this approach is too narrow, and risks limiting privacy to individual control at the expense of addressing more systemic issues within the **Big Data** ecosystem.

O'Hara has suggested that the definition is subject to problematic paradoxes. For instance, he argues if one used one's control over information to broadcast it indiscriminately, then one would have control, but very little privacy as commonly understood; conversely, if one were prevented from so doing, then one would have privacy restored, but no control.

The centrality of **informed consent** within potentially intrusive practices, such as medical treatment and the use of **personal information**, and concepts such as **revocation** and **consent** do suggest control has a place in our thinking about privacy – but perhaps more as a means for achieving one's **privacy preference**s (which might include a preference for less seclusion) rather than as a definition.

Further reading:
Austin, L.M., 2019. Re-reading Westin. *Theoretical Inquiries in Law*, 20(1), 53–81, https://doi.org/10.1515/til-2019-0003.
O'Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.
Westin, A.F., 1968. *Privacy and freedom.* New York: Atheneum.

See also: AUTONOMY, GROUP PRIVACY, RIGHT TO BE FORGOTTEN, COMMUNITY PRIVACY, PHYSICAL PRIVACY, PSYCHOLOGICAL PRIVACY

Privacy as Redistribution of Costs

To the extent that **privacy** is seen as a restriction of the flow of **information**, in economic terms there will be costs and benefits for the subjects of the information. Posner argued that privacy therefore leads to a redistribution of resources from those who benefit from the **availability** of full information to those who benefit from its concealment (Posner's examples were those with more arrests and convictions, and those with poorer credit records than the typical person). As markets become less efficient, costs are raised for all, but they are borne disproportionately by employers, creditors, and so on. Acquisti et al remarked that not only is privacy redistributive, but by the same token so is lack of privacy.

Further reading:

Acquisti, A., Taylor, C. and Wagman, L., 2016. The economics of privacy. *Journal of Economic Literature*, 54(2), 442–92, https://doi.org/10.1257/jel.54.2.442.
Posner, R.A., 1981. The economics of privacy. *American Economic Review*, 71(2), 405–9, www.jstor.org/stable/1815754.

See also: ECONOMICS OF PRIVACY, POSITIVE EXTERNALITIES OF DISCLOSED DATA, BENEFITS OF PRIVACY

Privacy Avatar

In online interaction, the replacement of the actual image of a **person** with a graphical image which may be different from the person’s physical form; a form of **privacy** through **obfuscation**.

An **AI** system that would control access to a **personal data store**, to reduce **data subject** consent burden, enabling **just in time consent** while protecting **informational privacy**. The personalised technology to create such an avatar is more theoretical than real at the time of writing and training it to understand the data subject’s preferences would bring burdens of its own.

Further reading:

Elliot, M.J., 2018. AI: privacy problem or opportunity, https://eprints.ncrm.ac.uk/id/eprint/4308/.
Brunton, F. and Nissenbaum, H., 2015. Obfuscation: a user’s guide for privacy and protest. Cambridge, MA: MIT Press.

See also: ACCESS CONTROL, PERSONAL INFORMATION MANAGEMENT SYSTEM

Privacy, Benefits Of

See: BENEFITS OF PRIVACY

Privacy Budget

In the context of **differential privacy**, the privacy budget is an attempt to quantify what amount of **privacy** loss to data subjects in a query system is acceptable.

The privacy budget is usually represented by the parameter **epsilon**. The larger the value of epsilon, the more privacy loss is deemed acceptable, and the greater the accuracy of the results of any query. However, larger values also increase the **risk** of an individual **data subject** as being recognised as contributing to the underlying **database**.

Further reading:
Luo, T., Pan, M., Tholoniat, P., Cidon, A., Geambasu, R. and Lécuyer, M., 2021. Privacy Budget Scheduling. *In*: *15th USENIX Symposium on Operating Systems Design and Implementation OSDI*, 55–74, www.usenix.org/system/files/osdi21_full_proceedings.pdf.

Privacy-by-Design

A term coined by Ann Cavoukian, at the time the **Information** and **Privacy** Commissioner of Ontario, privacy-by-design refers to the anticipation and prevention of privacy issues in the design phase of any system, process or entity.

The related term **data-protection-by-design** is sometimes used interchangeably when the **privacy concern** in focus is framed as **data protection**. The related principle of **data-protection-by-default** has since been introduced as an obligation under the **GDPR**.

Privacy-by-design is sometimes understood in the narrow sense of setting up the **default settings** of ICT systems in favour of **transparency**, **user** control and **data minimisation**; however, properly conceived, privacy-by-design goes beyond this, proactively considering privacy in all aspects of a system or entity's operations, policies and procedures. Similarly, the GDPR frames the obligation as encompassing **technical and organisational measures**, suggesting a social as well as a technological dimension to its requirements.

Further reading:
Cavoukian, A., 2009. *Privacy by design, take the challenge*, Canadian Electronic Library, https://policycommons.net/artifacts/1202287/privacy-by-design-take-the-challenge/1755397/CID:20.500.12592/9965z2.

Koops, B.-J. and Leenes, R., 2014. Privacy regulation cannot be hardcoded: a critical comment on the 'privacy by design' provision in data-protection law. *International Review of Law, Computers & Technology*, 28(2), 159–71, https://doi.org/10.1080/13600869.2013.801589.

See also: DATA PROTECTION POLICY, PRIVACY POLICY, PRIVACY ENGINEERING, PRIVACY AS CONTROL, PRIVACY SETTINGS

Privacy Calculus

A model of the decision-making process that individuals use to determine whether to share personal **information**. It involves weighing the potential benefits of sharing against the potential **risk**s. In the model, individuals consider factors such as the sensitivity of the information being shared, the **trustworthiness** of the organisation requesting the information and the perceived likelihood and severity of potential **privacy** violations.

The model has been criticised as unrealistic because it can be difficult for individuals to understand the potential risks and consequences of sharing personal information fully, and to make an 'apples and oranges' comparison of risks and benefits. It also makes strong and controversial assumptions about the rationality of consumer decision-making.

Further reading:

Kehr, F., Kowatsch, T., Wentzel, D. and Fleisch, E., 2015. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*, 25(6), 607–35, https://doi.org/10.1111/isj.12062.

Plangger, K. and Montecchi, M., 2020. Thinking beyond privacy calculus: investigating reactions to customer surveillance. *Journal of Interactive Marketing*, 50(1), 32–44, https://doi.org/10.1016/j.intmar.2019.10.004.

See also: RATIONAL CONSUMER, RISK-UTILITY TRADE-OFF, BENEFITS OF PRIVACY

Privacy Concern

The level of regard for **privacy** (either in general or of a specific individual). This will consider, and be shaped by, perceived threats to privacy from the actions of others. Hence, for example, the privacy concerns of someone with a secret to keep will have a focus on the ways in which the secret may be betrayed, the privacy concern of somebody sharing their **data** with another entity might be whether and how those data are misused, while the privacy concerns of someone changing into a swimming costume on a **public** beach will be the threat of another observing them.

See also: PRIVACY FUNDAMENTALISTS, PRIVACY PRAGMATISTS, PRIVACY UNCONCERNED, PRIVACY THREAT

Privacy, Cultural Variation of

See: CULTURAL VARIATION OF PRIVACY

Privacy Elasticity

The impact of a change in **privacy** on some other (economic) variable.

Further reading:
Dekel, I., Cummings, R., Heffetz, O. and Ligett, K., 2022. *The privacy elasticity of behavior: conceptualization and application* (No. w30215). National Bureau of Economic Research, www.nber.org/papers/w30215.

See also: ECONOMICS OF PRIVACY

Privacy Engineering

The process of creating systems, goods and services that proactively safeguard **privacy** is known as privacy engineering. It entails incorporating privacy concerns into the engineering process at every stage, from the initial design stage through system development, testing and deployment.

Computer science, **information** technology, law, policy and **ethics** all contribute to the multidisciplinary practice of privacy engineering, although its outputs tend to be technical solutions and therefore it is more sharply defined that the related **privacy-by-design**.

Implementing privacy engineering entails recognising possible **privacy concern**s and putting in place the necessary safeguards to reduce **risk**s, including **encryption**, **access control**s, **Privacy-Enhancing Technology** and **anonymisation**. Instead of depending on remedial actions after privacy violations and **data breach**es have occurred, the aim is to anticipate them and neutralise them in advance.

Further reading:
Gurses, S. and Del, A.J.M., 2016. Privacy engineering: shaping an emerging field of research and practice. *IEEE Security & Privacy*, 14(2), 40–6, https://doi.org/10.1109/MSP.2016.37.

See also: DATA-PROTECTION-BY-DEFAULT, ENGINEERING ETHICS, INFORMATION ETHICS

Privacy-Enhancing Technology (PET)

Privacy-enhancing technologies are tools or systems designed to protect the **privacy** of individuals while online or using digital technologies. These technologies may include software, **network** protocols, algorithms and other solutions designed to minimise the unauthorised collection and use of personal or sensitive **information**. Some are designed to increase the transparency of the **data** processing, to enable informed decisions to be made by individuals, while others are designed to increase the opacity of individuals themselves as they interact online.

The inclusion criteria for the concept have not been formally defined and consequently lists of **PETS** tend to vary hugely in terms of what is included. Some examples that appear on some lists are **encryption**, **anonymisation**, **multi-factor authentication**, **Virtual Private Network**s, **tracker blocker**s, **Differential Privacy** and edge-based solutions.

All PETs serve to restrict or map the flow of **information** in some way. They are therefore focused on **information privacy**. Most of them do not directly involve the **data subject**s in their operation however and might be more appropriately described as **confidentiality** enhancing.

Further reading:
Shen, Y. and Pearson, S., 2011. *Privacy enhancing technologies*: *a review*. Hewlett Packard Development Technical report, HPL-2011-113, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6bf9f0a288dd496de6bca96f360702b028fa0b58.

See also: EDGE COMPUTING, NETWORK, DATA FLOW

Privacy First

An approach to disclosure control where a **privacy** model such as **differential privacy** or **k-anonymity**, or some baseline **disclosure risk** standard is used to drive the **data** specification and the **data utility** of the data is only considered once the **dataset** has been treated.

See also: UTILITY FIRST

Privacy Fundamentalists

Alan Westin, who pioneered 20th-century research into and legislation on **privacy**, also produced a series of surveys of attitudes to **privacy** in the United States for more than 30 years, from 1978. In these surveys he generated robust findings, although changes in definitions and methodology mean these are not always directly comparable.

However, they did tend to support his view that people naturally coalesced into three groups: *fundamentalists*, the *unconcerned* and *pragmatists*. Fundamentalists mistrust organisations that ask for their **information**, support privacy rights and regulations and are concerned about the power of computer systems to uncover facts about them. Westin estimated that about 25 per cent of the **public** fall into this category.

Further reading:

Kumaraguru, P. and Cranor, L.F., 2005. *Privacy indexes*: *a survey of Westin's studies* [online]. Institute for Software Research International, report CMU-ISRI-5-138, http://repository.cmu.edu/cgi/viewcontent.cgi?article=1857&context=isr.

See also: PRIVACY PRAGMATISTS, PRIVACY UNCONCERNED

Privacy Guarantee

A mathematically provable property of some **privacy model**s such as **differential privacy**, that is usually expressed in terms of how much **information** about individual **data** units could leak from the system following the release of some data or analytical output.

See also: DATA RELEASE, OUTPUT PRIVACY

Privacy Impact Assessment

A method for identifying and assessing **privacy risk**s throughout the life of a project or system or organisation. It will often be a key component of a privacy-by-design approach.

The EU **GDPR** has refined the concept of **Data Protection Impact Assessments**, which now provides a comprehensive framework setting out the circumstances when risks to individuals' rights and freedoms should be conducted prior to processing **personal data**, and the considerations such an assessment should cover. These 'DPIAs' have thus formed a benchmark for **Privacy** Impact Assessments in the broader context of **data processing** activities.

Further reading:
Clarke, R., 2009. Privacy impact assessment: its origins and development. *Computer Law & Security Review*, 25(2), 123–35, https://doi.org/10.1016/j.clsr.2009.02.002.

See also: HUMAN RIGHTS IMPACT ASSESSMENT

Privacy Insurance

See: CYBER INSURANCE

Privacy-Invasive Technology

Any technology or tool that can potentially compromise an individual's **privacy**, usually by processing their **personal data**. Common examples include **facial recognition technology**, geolocation tracking and web **cookie**s. But any new technology that humans interact with has the potential to be privacy-invasive.

One approach to mitigate the potential harms caused posed by privacy-invasive technology, is implementing **privacy-by-design** and **data-protection-by-default** from the beginning of the technology's development. This can include the use of **anonymisation**, **privacy-enhancing technology** and/or **privacy impact assessment**s.

Privacy Metric

A quantity which (attempts to) measure the **privacy** afforded or lost by a system. The best-known examples of privacy metrics are embodied in the parameters of **differential privacy** (**epsilon, delta**) and **k-anonymity** (k), but other examples use entropy or mutual **information**.

There are two important points to note. First, privacy metrics almost universally concern themselves with a single type of privacy: **informational privacy** in **data** systems. Elliot et al argue that they do not directly relate to privacy at all, rarely having anything to say about the control that an individual might have or the relevant norms and context, and are not even in their own terms able to say anything about the loss or gain in privacy for a particular individual. Rather, as they observe, these measurements relate to **confidentiality**, as they concern flows of data. Nevertheless, many proponents of privacy metrics would argue that their metrics can be used to operationalise approaches such as **contextual integrity**.

Second, as Wagner and Eckhoff observe, many so-called privacy metrics are not metrics in the mathematical sense, as they do not meet all four of the conditions to qualify as a metric (nonnegativity, identity of indiscernibles, symmetry and triangle inequality). However, they are all attempts to quantify a distance between a desired state, which is deemed to be 'private', and the current situation, and attempt to account for an **adversary**'s goals, capabilities and resources. Thus, they map onto Elliot and Dale's **scenario analysis** principles.

Wagner and Eckhoff's comprehensive survey indicates the breadth of application of privacy metrics, across at least six different domains: **database**s, **communication** systems, **location based services**, **smart meter**ing, **social networks** and **genomics data**.

Further reading:

Elliot, M. and Dale, A., 1999. Scenarios of attack: the data intruder's perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring), 6–10, https://tinyurl.com/scen-attack.

Wagner, I. and Eckhoff, D., 2018. Technical privacy metrics: a systematic survey. *ACM Computing Surveys (CSUR)*, 51(3), 1–38, https://doi.org/10.1145/3168389.

Privacy Model

A usually mathematical or statistical framework for handling **informational privacy**, where the **data** in question are compared to some standard and the manipulated until they reach that standard. Examples are **k-anonymity** and **differential privacy**.

Further reading:
Domingo-Ferrer, J., Sánchez, D. and Soria-Comas, J., 2022. *Database anonymization: privacy models, data utility, and microaggregation-based inter-model connections*. Morgan & Claypool, http://dx.doi.org/10.2200/S00690ED1V01Y201512SPT015.

Privacy Notice

Often (and incorrectly) used interchangeably with '**privacy** policy', a privacy notice is **information** provided to a **data subject** about the **processing** of their information, usually provided at the point that the **data** are collected.

A difficult tension exists between the corporate responsibility to provide detailed, technical information (particularly under the EU **GDPR**) and the means by which individuals are genuinely likely to be assisted in making informed choices about the use of their information. Partly because of this, privacy notices can often be perceived as being unintelligible, unenlightening and unlikely to empower individuals to manage their informational **autonomy**.

Further reading:
Craig, T. and Ludloff, M., 2011. *Privacy and big data*. Sebastopol: O'Reilly Media.

See also: INFORMED CONSENT, LAYERED NOTICE, NOTICE AND CONSENT, TRANSPARENCY

Privacy Officer

While the EU **GDPR** requires some organisations to appoint a '**Data Protection Officer**', the closest equivalent in the United States is a '**Privacy** Officer'.

At the time of writing, the US lacks equivalent comprehensive privacy legislation. As such, the requirement to appoint a Privacy Officer is sector-specific, along with the corresponding legislation. One example is healthcare providers, who are required to appoint a Privacy Officer to manager their **compliance** with the **Healthcare Insurance Portability and Accountability Act** (HIPPA).

Further reading:
Herold, R. and Beaver, K., 2014. *The practical guide to HIPAA privacy and security compliance*. Boca Raton: Auerbach Publications.

Privacy Paradox

Research in behavioural psychology reveals systematic discrepancies within populations between people’s positively expressed attitudes towards their own privacy and their behaviour, which is often careless of it. Relatively few people take many active steps to protect their **personal data**, even those who insist their privacy is important. This has become known as the privacy paradox.

The non-salience of **risk** seems to be important in explaining the phenomenon. The risks from privacy-compromising behaviour are complex and hard to understand, and risk perception itself does not seem to be a strong motivator for behaviour moderation. It is also hard to translate an intention to limit disclosure into an actual limitation of disclosive behaviour, and for many people the level of **disclosure** is far higher than intended. Furthermore, the **benefits of privacy** are intangible whereas the return for giving up privacy is often more tangible (e.g., cheaper goods or free services) even if small. Finally, survey questions about **privacy concern** tend to be decontextualised and abstract, whereas disclosive actions take place in a concrete context such as an e-commerce purchase or registering for an **application**.

Some scholars, for example Waldman, argue that the gap between attitudes and behaviours is not a paradox at all but the direct result of the design tactics that platforms use to manipulate users into disclosing **information** via their cognitive biases.

Further reading:

Acquisti, A., Brandimarte, L. and Loewenstein, G., 2015. Privacy and human behavior in the age of information. *Science*, 347(6221), 509–14, https://doi.org.10.1126/science.aaa1465.

Barth, S. and de Jong, M.D.T., 2017. The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review. *Telematics and Informatics*, 34(7), 1038–58, https://doi.org/10.1016/j.tele.2017.04.013.

Waldman, A.E., 2020. Cognitive biases, dark patterns, and the ‘privacy paradox’. *Current Opinion in Psychology*, 31, 105–9, https://doi.org/10.1016/j.copsyc.2019.08.025.

See also: ATTITUDE–BEHAVIOUR GAP, BOUNDED RATIONALITY PRIVACY FUNDAMENTALISTS

Privacy Policy

An organisation which controls **personal data** – usually on a larger scale – will often consider it appropriate to implement a formal, documented policy to both ensure and help demonstrate **compliance** with applicable **privacy**, **data protection** and **confidentiality** laws. Although the phrase is not legally defined, it is likely that the organisation adopting such a policy will be what EU law terms a **data controller**, that is, an actor which determines the purposes and manner of **data processing**.

Some organisations ask **data subject**s to ‘accept’ the terms of their privacy policy to elicit **consent** to **data** processing (e.g., for cookies). This is a slight misuse of the term: a privacy policy is generally a detailed means of internal governance, and a **transparency notice** is a more appropriate form of **communication** with data subjects.

See also: ACCOUNTABILITY, NOTICE AND CONSENT, PRIVACY NOTICE

Privacy Pragmatists

Alan Westin, who pioneered 20th-century research into and legislation on **privacy**, also produced a series of surveys of attitudes to privacy in the United States for more than 30 years, from 1978. In these surveys he generated robust findings, although changes in definitions and methodology mean these are not always directly comparable.

However, they did tend to support his view that people naturally coalesced into three groups: *fundamentalists*, the *unconcerned* and *pragmatists*. Pragmatists weigh the benefits to them of consumer services, public safety measures and enforcement of **public** morality against the increased power of government and corporations and intrusiveness of **information** systems, and decide accordingly whether they support a particular privacy-increasing or reducing measure. Westin estimated that about 57 per cent of the public fall into this category.

Further reading:

Kumaraguru, P. and Cranor, L.F., 2005. *Privacy indexes: a survey of Westin's studies*. Institute for Software Research International, report CMU-ISRI-5-138, http://repository.cmu.edu/cgi/viewcontent.cgi?article=1857&context=isr.

See also: PRIVACY CONCERN, PRIVACY FUNDAMENTALISTS, PRIVACY UNCONCERNED

Privacy Preference

The **value of privacy** to an individual may vary with their context. They are likely to have, and to pursue, specific preferences about how private they want to be, or alternatively how open to scrutiny and visible to others. Such preferences will also vary relative to those others (e.g., one will keep different things private from one's spouse, one's doctor and one's colleagues). The conception of **privacy as control** implies that individuals should have access to mechanisms, such as **consent**, to enable them to manage their own **privacy** in accordance with their preferences.

Most platforms and web sites now allow users at least a limited capacity to operationalise their privacy preferences through **privacy settings**.

Further reading:

Altman, I., 1975. *The environment and social behavior*: *privacy, personal space, territory, crowding*. Monterey: Brooks/Cole.

O'Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Privacy Premium

The increase in price paid for a service which offers **privacy** features against one which does not.

Further reading:

Mai, B., Menon, N.M. and Sarkar, S., 2010. No free lunch: price premium for privacy seal-bearing vendors. *Journal of Management Information Systems*, 27(2), 189–212, https://doi.org/10.2753/MIS0742-1222270206.

Winegar, A.G. and Sunstein, C.R., 2019. How much is data privacy worth? A preliminary investigation. *Journal of Consumer Policy*, 42, 425–40, https://doi.org/10.1007/s10603-019-09419-y.

See also: ECONOMICS OF PRIVACY

Privacy-Preserving Data Analytics (PPDA)

Analysing sensitive **data** while protecting the **privacy** of the people whose data is being analysed can be done using a collection of methodologies and procedures called privacy-preserving data analytics (PPDA). **Anonymisation**, **masking**, **encryption** and **secure multi-party computation** are PPDA approaches. These methods enable data analysis without disclosing the identities of the contributors to the data.

Further reading:
Chakravorty, A., Wlodarczyk, T. and Rong, C., 2013. Privacy preserving data analytics for smart homes. *In*: *2013 IEEE Security and Privacy Workshops*, IEEE, 23–7, https://doi.org/10.1109/SPW.2013.22.

Privacy-Preserving Data Mining (PPDM)

A collection of methods and techniques is used to mine **big data**sets for relevant patterns and trends while protecting the **privacy** of the people whose **data** is being examined. PPDM is an element of **privacy engineering**.

Perturbation and **encryption** are two of the techniques used. These techniques aim to modify the **data** to reduce the risk of reidentifying individuals in the **dataset** and the outputs of the **data mining** process whilst preserving the **data utility**.

Further reading:
Aldeen, Y.A.A.S., Salleh, M. and Razzaque, M.A., 2015. A comprehensive review on privacy preserving data mining. *SpringerPlus*, 4(1), 1–36, https://doi.org/10.1186/s40064-015-1481-x.

See also: PRIVACY-ENHANCING TECHNOLOGY, PRIVACY-PRESERVING MACHINE LEARNING

Privacy-Preserving Data Publishing

The set of techniques that might be used to protect **dataset**s prior to publishing them. This includes data synthesis, **statistical disclosure control** and the application of **privacy models** such as **k-anonymity** or **differential privacy**.

Further reading:
Fung, B.C., Wang, K., Chen, R. and Yu, P.S., 2010. Privacy-preserving data publishing: a survey of recent developments. *ACM Computing Surveys*, 42(4), 1–53, https://doi.org/10.1145/1749603.1749605.

See also: ANONYMISATION, PUBLISHING

Privacy-Preserving Machine Learning (PPML)

A set of methods and techniques used to build **machine learning** models while protecting the **privacy** of the people whose **data** is being used to train the model. PPML is a crucial part of **privacy engineering**, which aims to safeguard identities of **data subject**s when **processing personal data**. Methods used include **federated learning**, **homomorphic encryption** and **differential privacy**.

Applications for PPML are numerous, ranging from marketing and finance to healthcare. For example, PPML may be used in epidemiology to analyse patient data to identify possible risks to population health or trends in disease outbreaks (where the focus is on populations rather than individuals). Similarly, without disclosing personal financial data, PPML can be used to flag potentially fraudulent transactions.

Further reading:
AlRubaie, M. and Chang, J.M., 2019. Privacy-preserving machine learning: threats and solutions. *IEEE Security & Privacy*, 17(2), 49–58, https://doi.org/10.1109/MSEC.2018.2888775.

See also: FINANCIAL PRIVACY

Privacy-Preserving Record Linkage (PPRL)

A set of techniques used to link **datasets** without revealing any identifying **information** about the individuals represented in the **data**sets. This is usually carried out using **cryptography** to **mask** the data being matched, so that the data can only be linked together by parties who have the cryptographic keys.

There are several methods for performing PPRL, including **Secure Multi-Party Computation**, Cryptographic **Hashing** and Bloom Filters.

Further reading:
Vatsalan, D., Christen, P. and Verykios, V.S., 2013. A taxonomy of privacy-preserving record linkage techniques. *Information Systems*, 38(6), 946–69, https://doi.org/10.1016/j.is.2012.11.005.

See also: RECORD LINKAGE

Privacy Risk

The possible **harm** or unfavourable consequences from a privacy breach considered in the context of the likelihood of the **breach** occurring. The term is mostly commonly applied in the context of **information privacy** where it denotes the possibility that someone's **privacy** rights are at risk of violation by their **personal information** being disclosed, disseminated or utilised in a way that could hurt their **reputation** or their finances, or lead to emotional distress or physical harm.

Unauthorised access, **data breach**es, insecure **data transfer**s or misuse of personal **information** are examples of the issues that should be considered when assessing privacy risk. In sensitive **data situation**s such as healthcare, banking or law enforcement, the potential impacts can be significant, and so the probability of **breach** must be kept low for **risk** to be at an acceptable level.

Further reading:

Hong, J.I., Ng, J.D., Lederer, S. and Landay, J.A., 2004. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. *In*: *Proceedings of the 5th conference on designing interactive systems*: *processes, practices, methods, and techniques*, 91–100, https://doi.org/10.1145/1013115.1013129.

Jakobi, T., Alizadeh, F., Marburger, M. and Stevens, G., 2021. A consumer perspective on privacy risk awareness of connected car data use. *In*: *Proceedings of Mensch Und Computer 2021*, ACM, 294–302, https://doi.org/10.1145/3473856.3473891.

See also: HARM, PERSONAL INFORMATION

Privacy Screen

A filter placed over a monitor or laptop screen which makes it difficult for an overlooker to see what is on the screen.

A room divider, often used to provide **privacy** while changing.

Privacy Seal

A **third party trust**mark displayed on a company’s website, or **privacy policy**. They are intended to assure website visitors that the entity displaying the seal meets a prescribed standard of **informational privacy**.

Further reading:

Mousavi, R., Chen, R., Kim, D.J. and Chen, K., 2020. Effectiveness of privacy assurance mechanisms in users’ privacy protection on social networking sites from the perspective of protection motivation theory. *Decision Support Systems*, 135, 113323, https://doi.org/10.1016/j.dss.2020.113323.

Privacy Settings

Controls that allow individuals to manage what **information** they share and their level of visibility on online platforms, **application**s, devices and services. Privacy settings vary in how much control they give to users. As **data protection** regulation has developed, **regulators** have become involved in creating expectations for **privacy** settings. For example, where services are likely to be accessed by children, regulators will typically expect stricter **default settings**.

Further reading:

Information Commissioner's Office, 2023. *Privacy and data use settings*, https://ico.org.uk/for-organisations/uk-GDPR-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/how-to-use-our-guidance-for-standard-one-best-interests-of-the-child/children-s-code-best-interests-framework/privacy-and-data-use-settings/.

Liu, Y., Gummadi, K.P., Krishnamurthy, B. and Mislove, A., 2011. Analyzing Facebook privacy settings: user expectations vs. reality. *In*: *Proceedings of the 2011 ACM conference on Internet measurement conference*, 61–70, https://doi.org/10.1145/2068816.2068823.

See also: PRIVACY AS CONTROL

Privacy Threat

Every action, event or entity that has the potential to undermine the **privacy** of a person or group is referred to as a privacy threat. **Hacking**, **malware**, **spyware** and other cyber-attacks that might compromise **personal information** are examples of technology-related privacy hazards. However, the mere existence of a privacy threat does not by itself equate to an actual privacy **breach**. Considering the likelihood of a threat becoming actualised will produce an assessment of privacy **risk**, which can be managed using both standard and privacy-specific risk management techniques.

Even if not actualised, threats to privacy can have a variety of negative effects on people and organisations, such as loss of **confidence** and **trust**. They may also affect society more broadly, diminishing public confidence in institutions or threatening the free exchange of ideas.

See also: RISK ASSESSMENT, HARM

Privacy Tort

Tort is the law of wrongs between private citizens, as dealt with under **common law** (as opposed to statute or the historic Courts of Equity). Prosser identified four different types of interests protected in US law that could be invaded as part of a **breach** of **privacy**, leading to the following four torts:

1. **Intrusion on seclusion** or **solitude**;
2. **Public disclosure** of embarrassing facts;
3. **Publicity** placing someone in a **false light**;
4. **Appropriation of someone’s name or likeness**.

While these categories have been influential in the US courts, privacy torts do not necessarily map onto these categories in other jurisdictions. In England and Wales, for example, the main tort which expresses the right to privacy is misuse of private **information** (which does not need to be false, embarrassing or appropriative). Privacy tort in the UK is a relatively recent development, ushered in by the Human Rights Act 1998 which made the right to **private life** under Article 8 of the **European Convention on Human Rights** directly applicable in UK law. Prior to the 21st century, privacy wrongs in English law were dealt with under equitable principles of **breach of confidence**.

Further reading:

Giliker, P., 2015. A common law tort of privacy? The challenges of developing a human rights tort. Singapore Academy of Law Journal, 27, 761–88. https://research-information.bris.ac.uk/ws/portalfiles/portal/55674339/A_Common_Law_Tort_of_Privacy_final_.pdf.

Prosser, W.L., 1960. Privacy. *California Law Review*, 48, 383–423, https://lawcat.berkeley.edu/record/1109651?ln=en.

Richards, N.M. and Solove, D.J., 2010. Prosser’s privacy law: a mixed legacy. *California Law Review*, 98(6), 1887–1924, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1567693.

See also: CONFIDENTIALITY, MISUSE OF PRIVATE INFORMATION, NEGLIGENCE, PUBLIC DISCLOSURE OF PRIVATE FACTS

Privacy Trade-Off

If **privacy** is a human right, then it is inalienable. However, if it is treated as a personal preference for data subjects, then it may be traded off against other goods by individuals.

There are two commonly discussed trade-offs against **privacy**. First, **personal data** may be traded off against *free services* under the common model of **surveillance capitalism**, or other benefits. **Consent** to use of the **data** – by **data subject**s – diminishes their privacy, but they gain services as a result, and the data may be used by service providers to improve those services, provide others and better personalise the **user** experience. Medical data from wearable devices may be shared to provide health benefits. Vulnerable people may share data with carers in order that they may be alerted in an emergency.

The second trade-off is between the benefits of privacy to the individual and the *costs to the community*. An individual's privacy may come at a cost to **security** in a less well-policed society, or less valid medical research, or less effective policy interventions (e.g., for controlling carbon emissions, or traffic congestion).

Further reading:
Etzioni, A., 1999. *The limits of privacy*. New York: Basic Books.

See also: BENEFITS OF PRIVACY, ECONOMICS OF PRIVACY, PRIVACY PREFERENCE

Privacy Unconcerned

Alan Westin, who pioneered 20th-century research into and legislation on **privacy**, also produced a series of surveys of attitudes to privacy in the United States for more than 30 years, from 1978. In these surveys he generated robust findings, although changes in definitions and methodology mean these are not always directly comparable.

However, they did tend to support his view that people naturally coalesced into three groups: *fundamentalists*, the *unconcerned* and *pragmatists*. The unconcerned are comfortable with increased flow of **information** about them if it leads to tangible benefits such as better consumer services and **security** of citizens. They do not favour more regulation of privacy. Westin estimated that about 18 per cent of the **public** fall into this category.

Further reading:

Kumaraguru, P. and Cranor, L.F., 2005. *Privacy indexes*: *a survey of Westin's studies*. Institute for Software Research International, report CMU-ISRI-5-138, http://repository.cmu.edu/cgi/viewcontent.cgi?article=1857&context=isr.

See also: BENEFITS OF PRIVACY, PRIVACY FUNDAMENTALISTS, PRIVACY PRAGMATISTS

Privacy, Value of

See: VALUE OF PRIVACY

Private Army

A private army is an armed, organised force which has been created by a private person or organisation, rather than a legitimate nation state.

There is therefore an important qualitative difference in legitimacy between a national army and a private one: the national army is responsible to the government and the people of a state, whereas a private army has a more utilitarian premise, perhaps to defend or advance the interests of the employer, or to make a profit. The state army has rights to intervene in conflict and detain citizens, while also being restrained from attacking them. A private army, when it intervenes, does so based on the strength it can wield. It has no rights to enter any conflict, except on privately held territory with the **consent** of the owner.

Further reading:

McFate, S., 2014. *The modern mercenary*: *private armies and what they mean for world order*. New York: Oxford University Press.

See also: CONFLICT OF RIGHTS, PRIVATE SPHERE

Private Biometrics

A system that preserves the **privacy** of an individual's **biometric data** while still allowing for the **authentication** and verification of their **identity**. Private biometric systems often use **encryption** techniques to store and transmit **data** securely. There is a trade-off with **accuracy** and system efficiency and one concern is that biometric systems that are too heavily focused on privacy may be less accurate in identifying individuals.

Further reading:

Ratha, N.K., Connell, J.H. and Bolle, R.M., 2001. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3), 614–34, https://doi.org/10.1147/sj.403.0614.

Kumar, M. and Kumar, N., 2020. Cancellable biometrics: a comprehensive survey. *Artificial Intelligence Review*, 53, 3403–46, https://doi.org/10.1007/s10462-019-09767-8.

Private Browsing Mode

See: INCOGNITO MODE

Private Enterprise

Private enterprise is entrepreneurial economic activity not directed by the state, where individual investors own companies and receive a share of the profits of their commercial activities as a return on the **risk** to their investment. It relies on robustly defended **private property** rights, which are, in a modern capitalist economy, usually protected by the state, through regulation, and via respect for contract. In a mixed economy, where it is carried out alongside state-controlled resource management, private enterprise comprises the **private sector**. Government activity takes place in the **public sector**, and private but non-profit activity is sometimes called the *third sector*.

Further reading:

Spulber, D.F., 2009. *The theory of the firm*: *microeconomics with endogenous entrepreneurs, firms, markets, and organizations*. New York: Cambridge University Press.

Private Key

See: ASYMMETRIC CRYPTOGRAPHY

Private Life

The aggregate of a person's activities and relationships which are not properly the object of **scrutiny** or disruption (at least in the absence of that person's **consen**t, or a commonly accepted **public interest** justification).

The extent of private life can be understood in terms of social norms, or with reference to more formalised legal and ethical rights, giving rise to two related meanings:

1. In terms of norms, it is those aspects of life that make up the **private sphere**.
2. Most countries have a broad, fluid and often contested ambit of human activity which the law will protect as a citizen's 'private life'. For example, the **European Convention on Human Rights** requires contracting states to protect private and family life, without providing a precise definition of its scope. The European Court of Human Rights has interpreted it on a case-by-case basis, tending to give it an inclusive reading. It has been applied to areas such as protection of **reputation**; protection of **integrity**; sexual identity and sexual life; limits to **search**; **self-determination**; and the recognition of an individual's legal civil status, among others.

Recent scholarship has attempted to trace the links between privacy as a variable set of social norms, and as a legal right, most notably within Nissenbaum's theory of **contextual integrity**.

Further reading:

Benn, S.I., 1971. Privacy, freedom, and respect for persons. *In*: Pennock, J.R. and Chapman, J.W., eds, *Privacy and personality*. London: Transaction, 1–26.

Roagna, I., 2012. *Protecting the right to respect for private and family life under the European Convention on Human Rights*. Strasbourg: Council of Europe, www.echr.coe.int/documents/d/echr/roagna2012_en.

Private Parts

Private parts are those parts of the body, especially the sex and excretory organs, which are traditionally or normally veiled or clothed in public places. **Public** revelation of private parts is often a source of shame or embarrassment, and in most countries is an offence. Forced revelation and touching private parts without **consent** are usually criminal sexual assaults.

Further reading:

Murphy, S.B., 1989. *State v Woodley*: defining 'intimate parts' under the Oregon criminal code. *Oregon Law Review*, 68(1), 255–9, https://heinonline.org/HOL/LandingPage?handle=hein.journals/orglr68&div=18&id=&page=.

See also: BODILY PRIVACY, INTIMACY

Private Property

Private property is a designation for objects, places, abstractions and other things which are *owned* by private entities, that is, individuals and non-governmental legal entities. Ownership conveys certain rights, such as the ability to exclude others from using the property, the ability to make it a gift to **other**s and (usually) the ability to exchange it for something valued more highly by the owner (including money). Such rights are defined by the legal systems of different jurisdictions, and therefore vary across countries.

While legally defined private property is a fundamental pillar of the capitalist system in particular, feelings and norms of possession are important in virtually all societies even when not legally defined and underlie many social arrangements. William James argued that feelings of possession are key in defining the (extended) self, while O'Hara has suggested that the first- and third-person possessive pronouns are informative linguistic markers of **privacy** interests.

Further reading:
James, W., 1890. *The principles of psychology*. Volume I. London: Macmillan.
Peck, J. and Shu, S.B., eds, 2018. *Psychological ownership and consumer behavior*. Cham: Springer, https://doi.org/10.1007/978-3-319-77158-8.

See also: FINANCIAL PRIVACY, INTELLECTUAL PROPERTY, PRIVATE SECTOR, PRIVATE SPHERE

Private School

In most countries, the state provides an education for all or most children. A private school lies outside this system and is either run as a charity or for profit. Its governors are not connected with the state and are not responsible to the state for the curriculum or methods of teaching. Private schools may thrive where they are perceived to confer educational or social advantages on their pupils, or where they provide education according to particular religious or moral values (including education for a single sex), for particular types of pupil or in specialist subjects such as music. Older schools also have strong traditions that may be found attractive by parents (especially former pupils).

Private schools are sometimes called *independent schools*. In England and Wales, confusingly, many elite fee-charging schools are called *public schools*, because, while entry is selective by entrance examinations (and ability to pay), they have no restrictions of denomination or residence, and so in that sense are open to the **public**.

Opponents of private schools argue that the education of children should not be a private matter for parents to decide, as it is fairer to provide a single standard of education, and that the existence of a **private sector** of education disadvantages those who receive education from the state (or no education at all). Supporters of private schools often make a pluralistic or libertarian case, and argue that, while the state should provide adequate education for all, it has no legitimacy to restrict diversity or impose its standards (which may not be high).

Further reading:

Dronkers, J. and Robert, P., 2008. Differences in scholastic achievement of public, private government-dependent, and private independent schools: a cross-national analysis. *Educational Policy*, 22(4), 541–77, https://doi.org/10.1177/0895904807307065.

Du Toit, J.L., 2004. *Independent schooling in post-apartheid South Africa: a quantitative overview*. Cape Town: HSRC Publishers, http://hdl.handle.net/20.500.11910/7676.

See also: PRIVATE SPHERE, PUBLIC SPHERE

Private Sector

See also: PRIVATE ENTERPRISE

Private Sphere

A conventional understanding of social life in Western culture rests on a threefold distinction between matters properly of interest to the state; matters properly of interest only to individuals or basic social units such as family, tribes, private associations or friends (the *private sphere*); and an intermediate area of general social interest (the **public sphere**). The private sphere, as described by philosopher Benn, consists of *private affairs* defined by social norms that can be invoked merely by pointing them out to an **intruder**. Even if someone's private affairs are publicised, they remain properly a private matter, as the private sphere is a normative concept.

Pre-modern accounts of the public sphere, such as Aristotle's *Politics*, tended to describe the private sphere as a mere *residuum* from public life, a domestic arena over which men presided, delegating its smooth running to women and slaves. Over time, the private sphere has been increasingly conceptualised as a place of refuge from a complex and challenging public space, as argued by Sennett. Whereas in public, one needs to adopt **mask**s to interact within well-understood but inauthentic social roles, in private one can ‘be oneself’.

The **boundary** between the private and public spheres constantly shifts. For example, Mill saw the household as a basic locus of the private sphere, which was not the proper subject of debate or of legislation, but contemporary feminists such as Rössler have argued that even within the household, the treatment of women and children can be a matter of legitimate public interest.

Further reading:

Arendt, H., 1998. *The human condition*, 2nd edition. Chicago: University of Chicago Press.
Benn, S.I., 1971. Privacy, freedom and respect for persons. *In*: Pennock, J.R. and Chapman, J.W., eds, *Privacy and personality*. Abingdon: Routledge, 1–26.
Mill, J.S., 1991. On liberty. *In*: Mill, J.S., *On liberty and other essays*. Oxford: Oxford University Press, 5–128.
Rössler, B., 2005. *The value of privacy*. Cambridge: Polity Press.
Sennett, R., 2002. *The fall of public man*. London: Penguin.

See also: FEMINIST CRITIQUE OF PRIVACY

Probabilistic Record Linkage

A form of **record linkage** which does not require perfect one-to-one matches to identify a link as a match but instead uses all the **information** in the records to assess the most likely matches. This allows the linkage system to deal with errors in the **data** and for many practical use cases gives a more accurate set of matches than pure deterministic (or exact) **matching**.

Further reading:

Fellegi, I.P., and Sunter, A.B., 1969. A theory for record linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210, https://doi.org/10.1080/01621459.1969.10501049.
Sayers, A., Ben-Shlomo, Y., Blom, A.W., and Steele, F., 2016. Probabilistic record linkage. *International Journal of Epidemiology*, 45(3), 954–64, https://doi.org/10.1016/j.jclinepi.2021.04.015.

See also: DATA LINKAGE, DETERMINISTIC RECORD LINKAGE

Processing

See: DATA PROCESSING

Profile

See: PROFILING

Profiling

Profiling is the use of **data** to create a dossier of critical **information** which could then be applied to an individual. The information may be discovered or inferred; the dossier is called a *profile*. When the information is collected by an organisation with a commercial relationship with the individual, it is a **user** *profile*. A profile may be built up by gathering and analysing the data about a service user’s interactions with the system to create a *user model*.

The profile associated with an individual is then used to determine interactions with that individual, including enabling them to identify themselves to the system. **Profiling** is a key technology underpinning **social media**, the **personalisation** of products and services for consumers and the effective targeting of adverts at those who will be most receptive. It is often used predictively, to make probabilistic assumptions about how individuals will react to phenomena, and thus how they will behave in future. The more data collected about, or related to, an individual, the richer the model will be, and the more confidence there will be in the **predictive analytics**.

The model may be augmented by data taken from elsewhere, and by making inferences about the profiled individual given their similarity to others (for instance, people of a particular age, gender, address and income may be taken to be more likely to exhibit similar consumption, voting or criminal behaviour). In these ways, a profile applied to an individual may not include much data that is directly about them (so-called *indirect profiling*). Profiles that capture the characteristics of a specified group are called *group profiles* and are applied to individuals that meet the specification to make predictions about their behaviour, regardless of whether any data about them personally was used to create the profile.

Profiling is used in sensitive areas, including policing, credit rating and advertising. Hence the quality of an individual’s experience with the profiling organisation may depend on their own past interactions, or even on the past interactions of similar people, impinging on their **autonomy**. This may involve extremely sensitive matters, such as discrimination on grounds of race, gender or sexual orientation. The function of a profile depends on the domain; in policing it is used as a mechanism for identifying the perpetrator of a crime; in advertising it might be used to target messages to particular groups; in credit rating it might be used to evaluate an individual’s suitability for credit.

Under the EU’s **GDPR**, ‘profiling’ refers to the automated **processing** of data to evaluate personal characteristics. This may include using information relating to an individual to predict their future behaviour, but the personal evaluation in question does not need to be predictive in nature.

Where profiling is used as a sole means of making significant decisions about people, it is subject to further **transparency** obligations under the GDPR.

Further reading:
Hildebrandt, M., 2008. Defining profiling: a new type of knowledge? *In*: Hildebrandt, M. and Gutworth, S., eds, *Profiling the European citizen*: *cross-disciplinary perspectives*. Dordrecht: Springer, 17–45, https://doi.org/10.1007/978-1-4020-6914-7.

See also: BEHAVIOURAL ADVERTISING, CUSTOMER TRACKING, DIGITAL FOOTPRINT, INFERRED DATA, PREDICTIVE MODELLING, SOCIAL PROFILING, USER MODELLING

Prolepticon

A form of **sousveillance** that specifically targets law enforcement agencies.

Further reading:
Singh, A., 2017. Prolepticon: anticipatory citizen surveillance of the police. *Surveillance & Society*, 15(5), 676–88, https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/6418.

Proportionate Security

See: PROPORTIONALITY

Proportionality

Proportionality is the broad principle, with applications in law, **ethics** and socio-technology, that the scale of an action, measure or restriction should be in balance with the cost or benefit. So, for example, in a human rights context, proportionality measures the **lawfulness** of an exercise of state power, by inquiring whether an **interference** with the fundamental right of an individual (such as the *right to **privacy***) exceeds the measures necessary to achieve a legitimate aim. The EU's **GDPR** refers to 'proportionality' in a different sense, as the calibration of a **data controller**'s obligations according to the **risk** posed to individual **data subject**s (e.g., from a **breach** in **data security**).

In a technical context, the **security** responses should be proportionate to the threat they are protecting against. For example, the fourth principle of functional **anonymisation** states: 'The measures you put in place to manage **disclosure risk** should be proportional to the likelihood and the likely impact of that risk.'

Further reading:

Elliot, M., O'hara, K., Raab, C., O'Keefe, C.M., Mackey, E., Dibben, C., Gowans, H., Purdam, K. and McCullagh, K., 2018. Functional anonymisation: personal data and the data environment. *Computer Law & Security Review*, 34(2), 204–21, https://doi.org/10.1016/j.clsr.2018.02.001.

See also: NECESSITY

Proprietary Privacy

Proprietary **privacy** was defined by Anita Allen as freedom from misrepresentation that damages **reputation**.

More broadly, there is a long philosophical tradition exploring the relationship between property and privacy. Locke for example asserted that a person's relationship with their body was one of self-ownership. Similarly, questions of **information privacy** can often be couched in terms of who owns the **data** in question.

Further reading:

Allen, A.L., 2011. *Unpopular privacy*: *what must we hide?* New York: OUP.
Goldie, M., ed., 2002. *Texts in the History of Political Thought*. Cambridge: CUP.

See also: APPROPRIATION OF NAME OR LIKENESS, DATA OWNERSHIP, PRIVATE PROPERTY

Protected Characteristics

Many countries have equality laws preventing discrimination based on protected characteristics. In Europe and North America, for example, these characteristics include sex, race, religion and genetic **information**.

There is an overlap between these protected characteristics and the human qualities documented in *sensitive* or **special category data**. The latter are **data protection** terms. **Data** protection and equality law may protect similar human characteristics, but they differ in scope and aims. Equality law regulates a greater range of activity, whereas data protection law serves a greater number of aims.

Sensitive **personal data** are only regulated in the context of **data processing**, requiring greater *justification*. Equality law, on the other hand, regulates a broader scope of activity than data processing, and will capture how someone is treated as an employee, customer, user of public services, and so on. At the same time, discrimination is one form of **harm** that **privacy** law seeks to minimise, but data protection law balances a broader range of objectives – including **accountability**, **transparency**, and **data minimisation**.

Further reading:
Malleson, K., 2018. Equality law and the protected characteristics. *Modern Law Review* 81(4), 598–621, https://doi.org/10.1111/1468-2230.12353.

See also: HARM, NON-DISCRIMINATION LAW, SENSITIVE VARIABLE

Protocol

An exchange of **information** between devices or systems is governed by a set of rules or standards known as a protocol in networking and **security**. The structure and flow of messages that devices use to create and maintain connections, exchange **data** and carry out other **network**-related tasks are defined by protocols specifying how data is transferred, routed and processed as well as how devices can connect with one another via a network. Examples include **TCP/IP**, DNS, **SSL/TLS** and **HTTPS**.

Security protocols are specifically intended to offer safe **communication** and shield confidential data from illegal access or **disclosure**. A variety of controls, including **encryption**, **authentication**, access limits and audit logging, may be included in security protocols to guarantee the **security**, **availability** and **integrity** of both **data in transit** and **data at rest**. SSL/TLS, IPSec and SSH are examples of common security protocols. Any procedure involving exchanges of data that need to be secure, such as authentication of people and devices for **access control**, encrypting data to avoid data theft, creating secure connections between devices and detecting and preventing network-based attacks rely on secure protocols to govern the safe **data transfer** and reception of data.

See also: INTERNET PROTOCOL, DATA FLOW, DATA IN USE, NETWORK SECURITY

Provable Security

A proof of **security** (of an **algorithm**, system or **data processing** mechanism) is a mathematical argument that, given certain conditions, the success of an **adversary** attempting to attack the system would be bounded. Examples of such **bounds** include the adversary having to solve some NP-hard problem to succeed, or that no more than *i* bits of **information** could leak, or the adversary requiring a minimum of *t* time to break into the system.

The proof will also be parametrised by the adversarial scenario, or **attack model**. To qualify as provable, the assumptions of such a model must be fully specified (not merely heuristic). Consequently, general proofs cannot be made for certain types of **attack**, such as side channel attacks, although it may be possible to prove that a given implementation is secure.

Provable security is most used in **cryptography**, which is amenable to mathematical analysis, but is also a factor in some **security by design** approaches, and **anonymisation** approaches such as **differential privacy**.

Proofs are subject to (human) error and there are known instances where proofs have been subsequently found to contain mathematical fallacies. Proofs are also sometimes claimed by software (e.g., anti-virus) vendors but usually with poor evidence.

Further reading:

Degabriele, J.P., Paterson, K. and Watson, G., 2010. Provable security in the real world. *IEEE Security & Privacy*, 9(3), https://doi.org/33-41.10.1109/MSP.2010.200.

He, D., Zeadally, S., Kumar, N. and Lee, J.H., 2016. Anonymous authentication for wireless body area networks with provable security. *IEEE Systems Journal*, 11(4), 2590–2601, https://doi.org/10.1109/JSYST.2016.2544805.

Koblitz, N. and Menezes, A., 2007. Another look at ‘provable security’. *Journal of Cryptology*, 20(1), 3–37, https://doi.org/10.1007/s00145-005-0432-z.

Proxemics

See: PERSONAL SPACE

Proxy

A proxy is a device that acts as an intermediary between a client and a server in the **Internet**, enabling additional **security** measures to be employed. It is typically used for **anonymity**, caching and security.

An anonymising proxy allows an Internet user to navigate anonymously. It is located between the user's client and the websites or online service being accessed. The anonymising proxy guarantees anonymity by **masking** the **IP address** of the user and replacing it with its own IP address. Examples of anonymising proxies include **HTTPS** proxies, which encrypt all traffic using the HTTPS protocol, distorting proxies (which pass false IP addresses in the **header information**), elite proxies (which appear to be the client), **VPN**s and **TOR**.

Further reading:

Jakobi, T., Alizadeh, F., Marburger, M. and Stevens, G., 2021. A consumer perspective on privacy risk awareness of connected car data use. *In*: *Proceedings of Mensch Und Computer 2021*, ACM, 294–302. https://doi.org/10.1145/3473856.3473891.

See also: INTERNET PROTOCOL

P% Rule

An **output statistical disclosure control** rule whereby a contributor to a **magnitude data** statistic should not be able to determine the value for any other contributor to within p%.

Pseudonym

A pseudonym is a temporary, artificial or fictitious identifier for an entity.

In a privacy and data protection context, a pseudonym is a piece of **information** which represents and signifies a **population unit** without directly identifying them for the **data user**. This definition can be arrived at from the EU's **GDPR**, which defines pseudonymisation as the processing of **personal data** in such a way that individuals cannot be identified without reference to further information (which is held separately).

To be an effective substitute for a **direct identifier**, a pseudonym should not be informative about an individual's personal characteristics. This would **risk** the representation falling within the **GDPR** definition of a profile, which would in turn make it the pseudonym a **key variable** (which could then be used linked to other personal data). Information organised according to pseudonyms should therefore enable analysis of patterns within the **data**, but not attributable details of individual **data subject**s.

Note that an effective pseudonym does need to be unique to the record and should itself be linkable to other data that have been pseudonymised under the same **pseudonymisation** scheme. This means that pseudonymised data fails two of the three tests for anonymous information (singling out and **linkage**) laid out by **Article 29 Working Party** and therefore is regarded as personal data under GDPR. Pseudonymous data may still achieve the standard of **functional anonymisation** for a given data user (in that it is not possible for that user to re-identify any **data unit**s), but that will only be determinable through a separate risk assessment and is not inherent to the pseudonymisation process.

Further reading:

Mourby, M. and Mackey, E., 2023. Pseudonyms, profiles and identity in the digital environment. *In*: van Der Sloot, B. and van Schendel, S., eds, *The boundaries of data*: *technical, practical and regulatory perspectives.* Amsterdam: Amsterdam University Press.

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S.E., Bell, J., Smith, H., Aidinlis, S. and Kaye, J., 2018. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, 34(2), 222–33, https://doi.org/10.1016/j.clsr.2018.01.002.

See also: PROFILING, UNIQUE IDENTIFIER

Pseudonymisation

As a category of **data** modification/minimisation techniques, pseudonymisation has been a feature of **information** management for some time. Historically the term has referred to the replacement of direct identifiers within a **dataset** with pseudonyms so that the data longer directly identifies individuals. It is noteworthy that while the adoption of a pseudonym by an individual is an intentional practice usually motivated by **privacy**, technically pseudonymisation is a **confidentiality** process.

Approaches to pseudonym generation include replacement of the direct identifiers with a serial number and various forms of cryptographic

hashing of the **direct identifier**s. There are two reasons why **pseudonym**s are used rather than **suppression**. First, it enables the option of reinstating the direct identifiers should they be needed and second, it enables **information** updates to take place through linking of identifiers. Both functions require some mechanism of recovery using either **cryptographic key**s to decrypt the hashes or a lookup table.

Pseudonymisation has found recent prominence (and a more complex definition) in the **GDPR**. The definition introduced in Article 4(5) of the Regulation is 'the **processing** of **personal data** in such a manner that it can no longer be attributed to a specific **data subject** without the use of additional information provided that such additional information is kept separately and is subject to **technical and organisational measures** to ensure that the personal data are not attributed to an identified or **identifiable natural person**'. It is thus both a **data minimisation** technique and one which involves some safeguards to ensure that **pseudonym reversal** is non-trivial. Critically, under the GDPR, pseudonymisation is not a process which takes information out of the category of 'personal data' (and thus the data remain within the scope of the GDPR).

The boundaries between **anonymisation** and pseudonymisation have become more contested since the GDPR entered into force. One argument holds that anonymisation only takes place if the direct identifiers within information (or from which the information originally derived) are permanently deleted, so that the data cannot be used by anyone to identify individuals. Mourby and colleagues have argued instead that the definition of pseudonymisation in the GDPR does not fundamentally alter the boundary between anonymous/personal data. They suggest that pseudonymised data can be shared with a third party, who has no access to the original identifiers, and with sufficient safeguards these data can be functionally anonymous for the **third party**.

Further reading:

De Moor, G.J.E., Claerhout, B. and De Meyer, F., 2003. Privacy enhancing techniques. *Methods of Information in Medicine*, *42*(02), 148–53, https://doi.org/10.1055/s-0038-1634326.

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S.E., Bell, J., Smith, H., Aidinlis, S. and Kaye, J., 2018. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, 34(2), 222–33, https://doi.org/10.1016/j.clsr.2018.01.002.

Mourby, M. and Mackey, E., 2023. Pseudonyms, profiles and identity in the digital environment. *In*: van der Sloot, B. and van Schendel, S. eds, *The boundaries of data*: *technical, practical and regulatory perspectives*, Amsterdam: Amsterdam University Press.

See also: FUNCTIONAL ANONYMISATION, MATERIAL SCOPE, PROFILING

Pseudonym Reversal

The transformation of a **pseudonym** into the original **direct identifier**s.

Further reading:
Veeningen, M., de Weger, B. and Zannone, N., 2012. Formal modelling of (de) pseudonymisation: a case study in health care privacy. *In*: *International Workshop on Security and Trust Management*, 145–60, Berlin: Springer, https://doi.org/10.1007/978-3-642-38004-4_10.

See also: PSEUDONYMISATION

Psychographic Advertising

A form of **targeted advertising** in which consumers are profiled based on their psychological and behavioural characteristics.

With the emergence of **social media**, this type of advertising has become more prevalent. Advertisers can use **data** from social media platforms, such as Facebook and Twitter, to analyse user behaviour and preferences.

Psychographic advertising has been criticised, with suggestions that it enables manipulation of consumer behaviour – the notorious Cambridge Analytica case, where digital election campaigning was allegedly deceptive and coercive, being one of the most egregious examples.

Further reading:
Bakir, V., 2020. Psychological operations in digital political campaigns: assessing Cambridge Analytica's psychographic profiling and targeting. *Frontiers in Communication*, 5, 67, https://doi.org/10.3389/fcomm.2020.00067.

See also: BEHAVIOURAL ADVERTISING, PROFILING, TARGETED ADVERTISING, PSYCHOLOGICAL PRIVACY

Psychological Privacy

Psychological **privacy** is described by O'Hara as the concealment from others of the beliefs, motives and attitudes that underlie visible action, and is similar to what Westin called **reserve**. Psychological privacy is often seen ambivalently – a necessary policing of interior space, but also the means of concealing unpleasant, unworthy or sordid thoughts. John Calvin coined the term *Nicodemite*, after the biblical Nicodemus, to describe those who misrepresented their true beliefs during the European religious wars to avoid persecution. In the post-Freudian world, there is also widespread concern about failures of psychological privacy through the involuntary **disclosure** of emotions and thoughts via unconscious slips of the tongue or other unintentional acts.

Further reading:

O'Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.
Westin, A., 1967. *Privacy and freedom*. New York: Ig Publishing.

See also: IDEOLOGICAL PRIVACY, INTEGRITY, INTELLECTUAL PRIVACY, MENTAL PRIVACY

Public

'Public' can be used adjectivally or as a noun. As an adjective, it implies the opposite of 'private'; a sense of being open, enabling access for outsiders. A **public figure** is a well-known person; *public opinion* is the collective opinion, measured via some methodology, of a population; *public law* regulates the interactions between citizens and government; a *public good* is an economic good whose use cannot be restricted to an owner; a *public company* is one whose shares are traded in an open market; a *public house* is an establishment which sells alcoholic drinks to an unlimited set of customers; a *public space* is open to everyone. Someone who is not in a restricted space is said to be *in public*.

As a noun, the (general) public is the wider **population** as a whole. Groups with specific common interests are often known as *publics*, if they are open and at least theoretically widespread.

Further reading:

Hannay, A., 2005. *On the public*. Abingdon: Routledge.

See also: PUBLIC SPHERE, PUBLICITY, PUBLISHING, OPEN ACCESS, PERSON, PUBLIC FIGURE

Publication

Publication is the act of communicating some content to the **public**, or, in other words, *publishing* it. The content, once published, is usually referred to as a publication. The *publisher* of a work, legally, should have the right to publish, and *copyright* is the term for an exclusive such right.

Publication involves the publisher, in some way, ceding control over the distribution of the content. It may involve the sale of the publication, or rental, or placing online. Merely passing content to friends, family or colleagues, or some naturally limited grouping, is not usually thought of as publishing, however large a circle this may be. Neither is public display of the content usually counted as publishing, as opposed to distributing the content for the purposes of public display. This has been re-emphasised in the **GDPR**, where the justification that **personal data** have been 'manifestly made public' is interpreted narrowly and does not include **information** which has only been disclosed to a limited audience.

Publishing private materials without permission is a serious **breach** of **privacy**; publishing artistic works without permission is a type of piracy.

Further reading:

Information Commissioner's Office, 2022. *What are the conditions for processing?* http://tinyurl.com/MadePublic.

World Intellectual Property Organization, 1979. *Berne Convention for the Protection of Literary and Artistic Works*. WIPO IP Portal, https://wipolex.wipo.int/en/text/283698.

See also: CONDITIONS FOR PROCESSING, INTELLECTUAL PROPERTY, PUBLIC SPHERE, PUBLICITY, THIRD PARTY DOCTRINE

Public Disclosure of Private Facts

Public **disclosure** was considered by Prosser to be separate from **breach of confidence**, which was discussed in depth by Warren and Brandeis. He focused on the embarrassment that the disclosure would cause to the plaintiff, and the offensive nature of the disclosure. A private fact is generally taken to be a factual detail of the plaintiff's life that is not generally known publicly, and would include medical, sexual or financial facts. The disclosure must be **public**, not merely telling secrets or **gossip**ing to a small group of people, and the standard of offensiveness is measured against the judgments of a reasonable **person** of ordinary sensibilities. Defences against the claim include the legitimate **public interest** in knowing the fact, its previous existence in the **public record**, and of course the **consent** of the plaintiff.

Further reading:

Prosser, W.L., 1960. Privacy. *California Law Review*, 48, 383–423, https://lawcat.berkeley.edu/record/1109651?ln=en.

Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220, https://doi.org/10.2307/1321160.

Public Domain

The public domain refers to **information** which is in some sense public. Most usually, this refers to information over which there are no exclusive rights, or alternatively that any rights such as copyrights, **database** rights or patents have expired or been waived, so that anyone can use the information without penalty or having to obtain a licence. One extreme position is that information should be in the public domain by default, as its value can be extracted more effectively by a wider range of **user**s, and more equitably distributed.

In a vaguer and more abstract sense, information in which there is **privacy** interest, but which is widely known, is sometimes said to be in the **public** domain (for example, a state **secret** or confidential information may be revealed in a newspaper, in which case it has been put into the public domain).

Finally, the term is sometimes used as a synonym for the **public sphere**.

Further reading:

Boyle, J., 2008. *The public domain*: *enclosing the commons of the mind*. New Haven: Yale University Press.

See also: PUBLIC INTEREST

Public Figure

A public figure is someone who has some kind of prominence, importance, fame, celebrity or notoriety within a society; their name is known to, and their affairs are discussed by, people with whom they have no direct connection. Many public figures have sought this status, but others are put in the position through luck (good or ill). It is an important factor in legal judgments about **privacy**, as norms and laws relating to public figures often vary from those concerning 'ordinary' members of society. Hence the **reasonable expectations of privacy** vary depending on the public status of individuals. As a result, much privacy case law has ultimately been generated following litigation featuring public figures.

There are several reasons for the different treatment of public figures. First, many activities of many public figures that would ordinarily be confidential are uncontroversially of **public interest**. For instance, the business dealings of an elected politician should be transparent, so that voters can judge their probity and incentives for their actions.

Second, public figures are often the targets of media interest because media consumers want to know details of their lives. Hence many issues emerge as the media try to reveal their activities, even those that are not obviously of pressing concern (such as their sex lives).

Third, many public figures ('celebrities') require media coverage of their activities to meet the demands of their fans. They will also send out public messages via **social media** and the like, courting a public image of which they wish to be in control.

Fourth, related to the previous point, many public figures have an interest in protecting their image or name, as they are important sources of income and branding, not unlike **intellectual property**.

Fifth, many public figures have relationships with non-public figures (most obviously, their children) whose privacy is also invaded by legitimate media interest in the public figure.

Further reading:

Shackelford, S.J., 2012. Fragile merchandise: a comparative analysis of the privacy rights for public figures. *American Business Law Journal*, 49(1), 125–208, https://heinonline.org/HOL/LandingPage?handle=hein.journals/ambuslj49&div=7&id=&page=.

See also: APPROPRIATION OF NAME OR LIKENESS, CELEBRITY PRIVACY, GOSSIP, MISUSE OF PRIVATE INFORMATION, PUBLIC SPHERE

Public Interest

From a sociological perspective, in which the plural publics is the norm, the legal concept of 'public interest' may appear simplistic. In essence, it constitutes the rights, freedoms and benefits of an unspecified 'everyone', often juxtaposed against the interests of a specific individual or entity. As such, it relies on the conventional binary between private vs public and the individual vs the collective that many have challenged. Dove and other commentators conceive **privacy** and **confidentiality** as being elements of the public good, acknowledging a shared interest in respect for our fundamental rights. The emergence of **group privacy** in recent literature has also introduced an intermediate tier between an individual and an unspecified 'everyone'.

Further reading:

Dove, E.S., 2023. Confidentiality, public interest, and the human right to science: when can confidential information be used for the benefit of the wider community? *Journal of Law and the Biosciences*, 10(1), 1–53, https://doi.org/10.1093/jlb/lsad013.

See also: CELEBRITY PRIVACY, GDPR, GROUP PRIVACY, LAWFUL BASIS, RIGHT TO PRIVACY, PUBLIC, BENEFITS OF PRIVACY, CONFIDENTIALITY, OTHER

Publicity

In its original usage, publicity refers to openness to scrutiny and comment – in a sense, the opposite of **privacy**. In a more common and specialised sense, publicity is attention deliberately sought, especially through advertising, marketing, **public** relations or pulling stunts to get noticed. In this sense, the term often refers to the output which 'spreads the story' – media articles, promotional literature or video, or advertisements. A *publicist* is a person employed to generate publicity, while *public relations* is the art of ensuring that independently produced publicity is favourable to a paying client.

While publicity about an identifiable person can be an invasion of that person's privacy, the relationship between privacy and publicity is complex. For instance, a celebrity may court attention; in that case, they may criticise publicity not because it invades their privacy, but because it is false or misleading, or that it is unfavourable to them, or because it subverts their rights (e.g., using a photograph of a film star in an advert).

Further reading:

Nimmer, M.B., 1954. The right of publicity. *Law and Contemporary Problems*, 19(2), 203–23, https://doi.org/10.2307/1190488.

See also: PUBLIC DOMAIN, PUBLIC DISCLOSURE OF PRIVATE FACTS, PUBLIC SPHERE, PUBLICATION, CELEBRITY PRIVACY

Public Key

See: ASYMMETRIC CRYPTOGRAPHY

Public-Key Cryptography

Public-key **cryptography** is a method that uses two different keys to encrypt and decrypt messages to be passed to a receiver: a **public key** for **encryption**, and a **private key** for decryption. The public key can therefore be widely known, and associated with a particular receiver, Alice. If she publishes the public key, then anyone can send her an encrypted message. Decryption is performed using Alice's private key, which, if she keeps it secure, means only she can decrypt the messages.

The use of two keys corrects for a weakness in so-called **symmetric-key cryptography**, where a single key is used both for encryption and **decryption**, because the sender must not only communicate the encrypted message securely, but also the key for decryption – a **vulnerability** if that key is intercepted, especially if the key is to be sent to a wide range of **user**s.

The technology relies on the use of *one-way functions*, mathematical functions that are easy to perform, but hard to reverse (i.e., to decide which input produced the output). Factoring very large primes is often used – it is simple to multiply two large primes together but requires unfeasibly large amounts of conventional computing power to factor the multiple back into its constituent primes. Public-key cryptography is based on two **algorithm**s: the *Diffie-Hellman key exchange algorithm* (1976), which showed that keys could be distributed securely, and **RSA encryption**, which was the first algorithm of several that defined a practical key generation system.

Although public-key cryptography is an advance over symmetric key cryptography, it still has some weaknesses. It is vulnerable to **brute force attack**s, but algorithms can be chosen which would render brute force highly unlikely to work in a reasonable time. However, **quantum computing** may upset this calculation, by facilitating the speedy reversal of one-way functions. Furthermore, if the aim of the **adversary** is merely to disrupt **communication**s, rather than to decrypt messages, then this may be achieved by changing or sabotaging the public key. This would mean Alice's messages could no longer be authenticated.

Further reading:
Katz, J. and Lindell, Y., 2008. *Introduction to modern cryptography*. Boca Raton: Chapman & Hall/CRC, https://doi.org/10.1201/b17668.

See also: PUBLIC-KEY INFRASTRUCTURE, DIGITAL SIGNATURE

Public-Key Infrastructure (PKI)

Public-key infrastructure (PKI) is a set of **protocol**s, hardware, **software** and **information governance** procedures to manage a **public key** system, for example by storing, authenticating and where necessary revoking **digital signature**s. Essentially, the PKI binds public keys with verified identities and will provide certificates of authenticity to that effect. **Certification**, registration and verification may be performed by separate entities. *Decentralised PKI* (DPKI) uses **distributed ledger** technology to avoid problems associated with centralisation of PKI functions.

The International Telecommunications Union provides a PKI standard X.509.

Further reading:

Buchmann, J.A., Karatsiolis, E. and Wiesmaier, A., 2013. *Introduction to public key infrastructures*. Berlin: Springer-Verlag, https://doi.org/10.1007/978-3-642-40657-7.

International Telecommunications Union, 2021. *X.509*: *Information technology – Open Systems Interconnection – The Directory*: *public-key and attribute certificate frameworks*, www.itu.int/rec/T-REC-X.509.

See also: CERTIFICATION AUTHORITY, ENCRYPTION, PUBLIC-KEY CRYPTOGRAPHY

Public Records

Public **record**s are documents used or created by **public** bodies during processes of government. If they are non-confidential, they are usually available to the public for **scrutiny**. They may be placed online, as **open data** or behind a **firewall**, or it may be that application has to be made to a **user** to view them. **Freedom of Information** laws regulate which documents must be made available to the public, and the conditions under which applications may be denied. As they may include criminal and court records, births, marriages and deaths, property registers and government contracts, they often raise issues of **privacy** and **confidentiality**.

Further reading:

Martin, K. and Nissenbaum, H., 2017. Privacy interests in public records: an empirical investigation. *Harvard Journal of Law and Technology*, 31(1), 111–43, https://jolt.law.harvard.edu/articles/pdf/v31/31HarvJLTech111.pdf.

See also: TRANSPARENCY, PUBLIC SPHERE, PUBLIC INTEREST

Public Sphere

A conventional understanding of social life in Western culture rests on a threefold distinction between matters properly of interest to the state, matters properly of interest only to individuals or basic social units such as family, tribes, private associations or friends (the **private sphere**), and an intermediate area of general social interest (the **public sphere**). The public sphere consists of social matters or issues of general interest, but which are not, cannot or should not be regulated. In the public sphere, a **public** (or *publics*) debates and contests matters of interest, helping form *public opinion*, and influencing wider political attitudes in the state.

Pre-modern accounts of the public sphere, such as Aristotle’s *Politics*, tended to see it as the area in which men (rarely women) excelled, and showed their virtue. The private sphere was a mere *residuum*, a domestic arena over which men presided, delegating its smooth running to women and slaves. Men were self-sovereign in the private sphere but needed to justify their interventions in the public sphere through argument.

The development of the public sphere in modern Europe was theorised by sociologist Jürgen Habermas, who traced its efflorescence to the Enlightenment period. Institutions are needed for the **communication**s necessary to the public sphere to take place, and Habermas traced these, for example, to the coffee houses of 18th-century England, or in France to aristocratic *salons*.

Matters anchored in the private sphere are not the proper subject for public debate, but over time the **boundary** between the two shifts. For example, in early modern Europe, religion was seen to be a matter of **public interest**, but it later came to be seen as a private matter. However, in more recent years, the profession of religious adherence in public (for example, wearing a religious symbol in the workplace) has once more become a topic of public debate, and indeed in many countries the state has seen fit to legislate.

Further reading:
Arendt, H., 1998. *The human condition*, 2nd edition. Chicago: University of Chicago Press.
Habermas, J., 1989. *The structural transformation of the public sphere*: *an inquiry into a category of bourgeois society*. Cambridge: Polity Press, https://courses.ischool.berkeley.edu/i218/s10/JH-STPS.pdf.
Sennett, R., 2002. *The fall of public man*. London: Penguin.

See also: FEMINIST CRITIQUE OF PRIVACY, HISTORY OF PRIVACY, PUBLIC INTEREST

Publishing

See: PUBLICATION

Purchase History

One of the most important **data** trails left by individuals and households is their purchase history, the **record** of their prior purchase behaviour. Even relatively short-term or non-comprehensive histories can prove valuable to companies, either in their ability to set different prices for different households (**price discrimination**) or in effectively targeting marketing. Purchase histories also facilitate **personalisation** and **recommendation systems**, a development which has been accelerated by the development of **e-commerce**. While purchase histories tend to empower sellers over consumers, leading to a strong material **privacy** interest for the latter, buyers can also publish aspects of their history via feedback or seller ratings on marketplaces, which are often informative for future buyers.

Further reading:

Acquisti, A. and Varian, H.R., 2005. Conditioning prices on purchase history. *Marketing Science*, 24(3), 367–81, https://doi.org/10.1287/mksc.1040.0103.

Rossi, P.E., McCulloch, R.E. and Allenby, G.M., 1996. The value of purchase history data in target marketing. *Marketing Science*, 15(4), 321–40, https://doi.org/10.1287/mksc.15.4.321.

Purple Team

In a cyber **security** context, a purple team is one which fuses the functions of blue and red teams. The purpose of a **red team** is to improve the functioning of its corresponding **blue team** by carrying out a simulated attack – but to do that it needs to be adversarial. The purple team construct suggests that this is optimised by treating the blue–red team dynamic as a single system. Opinions vary about whether the purple team should be an actual team which observes what both blue and red teams are doing or whether a well-functioning blue-red team dynamic should make this unnecessary. On the other hand, an **adversary** would not consider the organisation's learning, so removing this function from the red team's responsibilities makes the simulation more realistic.

Further reading:

Miessler, D., 2021. *The difference between red, blue, and purple teams*, https://danielmiessler.com/study/red-blue-purple-teams/.

See also: PENETRATION TEST, MOTIVATED INTRUDER TEST

Purpose Limitation

The EU **GDPR** sets six principles in Article 5 which should be observed to process **personal data** lawfully. The second of these is the purpose limitation principle. This essentially requires personal data to be used only for the purpose(s) for which they were originally collected, or for a compatible purpose. Recital 50 of the GDPR provides that the **reasonable expectations of privacy** of **data subject**s can be considered when determining whether a subsequent purpose is 'compatible' with the original. This places secondary uses of **information** under the GDPR on a similar footing to **common law privacy**, which also uses the reasonable expectations *of data subjects as a key benchmark.*

An area of potential controversy arises in the context of scientific research, which is designated a 'compatible purpose' under Article 5.1(b) GDPR. There is some debate in academic and policy circles as to whether this means that no further legal basis is required for research if the **data** were originally collected in a lawful manner, or whether the **secondary use** of data for research requires separate justification under the GDPR (e.g., satisfying the **public interest** basis and the scientific research condition). At the time of writing, further clarification is anticipated from the **European Data Protection Board** on this issue.

Further reading:

Hartlev, M., Gefenas, E. Mourby, M., O'Cathaoir, K., Lukaseviciene, V., 2020. *EU-STANDS4PM report*: *legal and ethical review of in silico modelling*, www.eu-stands4pm.eu/publications.

Jasserand, C., 2018. Subsequent use of GDPR data for a law enforcement purpose: the forgotten principle of purpose limitation? *European Data Protection Law Review*, 4(2), 152–67, https://doi.org/10.21552/edpl/2018/2/6.

See also: LAWFUL BASIS, TRANSPARENCY

Purpose Specification

Under the EU's **GDPR**, **personal data** should only be used for specified (and legally justifiable) purposes. Where a specific purpose for the use or retention of **personal data** cannot be identified, the **purpose limitation** principle will be **breach**ed, and the processing will be unlawful.

See also: DATA PROCESSING, DATA RETENTION

Q

Quantitative Privacy

A term with two distinct meanings.

1. Methods for capturing impacts on or protections of privacy using mathematical or statistical measurements. The term covers **privacy metric**s, privacy models (e.g., **differential privacy**) and statistical **risk assessment** methods.
2. A conceptualisation of privacy arising from US case law which focuses on the consequences the use of **big data** for mass **surveillance** by governments or their agents. Gray and Citron frame this as a right which invokes the Fourth Amendment to the US Constitution (which protects US citizens from **unreasonable search**es and seizures by the government).

Further reading:

Gray, D. and Citron, D., 2013. The right to quantitative privacy. *Minnesota Law Review*, 98, 62–144, www.minnesotalawreview.org/wp-content/uploads/2013/11/GrayCitron_MLR.pdf.

See also: PRIVACY RISK, RIGHT TO PRIVACY, US PRIVACY LAWS

Quantum Computing

A type of computing that manipulates **data** using quantum-mechanical phenomena.

While quantum computing is technically equivalent to conventional computing (i.e., quantum computers are Turing machine equivalent), it has the capacity to execute some types of computations exponentially faster than classical computing. This has ramifications for many areas, including **privacy** and **encryption**.

The practicality of quantum computing has not yet been established and at the least widespread deployment would require larger quantum computers than the prototypes currently available. However, if its use became routine, the foundations of **cybersecurity** would need to be re-imagined. Many **cryptographic** techniques now in use to safeguard **sensitive data** could be made redundant. It may be possible to employ quantum computers to reverse-engineer **anonymisation** methods, enabling an **adversary** to connect seemingly unconnected data to specific **identiti**es. This has led authors such as Bernstein to develop theories of **post-quantum cryptography**, but these will be untested until quantum computing has fully emerged.

Further Reading:

Bernstein, D.J. and Lange, T., 2017. Post-quantum cryptography. *Nature*, 549(7671), 188–94, https://doi.org/10.1038/nature23461.

Gyongyosi, L. and Imre, S., 2019. A survey on quantum computing technology. *Computer Science Review*, 31, 51–71, https://doi.org/10.1016/j.cosrev.2018.11.002.

Mosca, M., 2018. Cybersecurity in an era with quantum computers: will we be ready? *IEEE Security & Privacy*, 16(5), 38–41, https://doi.org/10.1109/MSP.2018.3761723.

Quasi-Identifier

See: INDIRECT IDENTIFIER

Query Logging

The process of keeping track of the queries that **user**s of a **database** system make on the database. The **information** captured can include user **identifier**s and **username**s, the timestamp of the query, the form of the query and any parameters. Query logs can be one tool in providing **security** for databases, such as **intrusion** detection capabilities, for example, in the detection of **differencing attack**s.

Query logs are themselves likely to be **personal data** and are also subject to privacy and **data protection** considerations.

See also: PRIVACY-BY-DESIGN, PRIVACY ENHANCING TECHNOLOGY

Query Overlap

In query systems, a **data user** might – for *bona fide* or malevolent reasons – create closely related queries with the consequence that a **differencing** attack may be possible. The **attack** would remove the common units in the results of either or both queries, thus creating a more attackable data fragment.

Quishing

See: PHISHING, SMISHING

R

Radical Transparency

An approach to **data sharing** which might be viewed as the opposite of secrecy, where organisations aim to practice maximal openness about process and **data** for ideological rather than pragmatic reasons. Radical transparency is a mode of **disclosure** that can circumvent the negative aspects of more selective or targeted disclosure types. Related to the legislative concept of **freedom of information**, radical transparency operationalises trustworthiness. There are questions about the risks to **privacy** from such an approach which are not dissimilar to the tensions between freedom of **information** and **data protection**.

Further reading:
Birchall, C., 2014. Radical transparency? *Cultural Studies Critical Methodologies*, 14(1), 77–88, https://doi.org/10.1177/1532708613517442.

See also: IDEOLOGICAL PRIVACY

Radio Frequency Identification (RFID)

Technology that wirelessly identifies and tracks items. RFID systems normally have two major parts: a tag that is affixed to the item being monitored and a reader that communicates with the tag through radio waves. **Access control**, supply chain tracking and inventory management are just a few of the uses for RFID technology.

While RFID technology has numerous advantages, it also has certain **security** and **privacy** risks. Because RFID tags can be read from a distance, it is possible for unauthorised people to follow the movements of items without their permission. Also, there is a chance that **sensitive data** contained on an RFID tag might be intercepted.

Further reading:
Garfinkel, S.L., Juels, A. and Pappu, R., 2005. RFID privacy: an overview of problems and proposed solutions. *IEEE Security & Privacy*, 3(3), 34–43, https://doi.org/10.1109/MSP.2005.78.

Randomised Response

An approach to **primary data** collection used to increase **respondent confidentiality** when the survey topic is sensitive; for example, drug consumption or sexual behaviour. A random mechanism (hidden from the researcher) decides whether the respondent will answer truthfully or with a pre-specified response (e.g., always 'yes') regardless of what their true response would have been. As the proportion of respondents that have answered with the pre-specified response is known, they can be discounted in calculating actual rates. It is not possible for anyone other than the **data subject** to know if they answered the way they did because it is their true response or because they were instructed to do so through the random mechanism.

Random Rounding

A **statistical disclosure control** method for **tabular data** that rounds each cell of a table up or down to a multiple of a given base b. It uses a probability function to decide the direction of the **rounding** for each value. A common form cell value i will go down with probability: $1-$ remainder(i/b)/b and up with probability: remainder(i/b)/b. The main advantages of random rounding are its simplicity and that the resulting table will be unbiased. The main disadvantage is that the resulting table may not be additive if the marginal cells are also modified with the same approach. If they are modified to respect **additivity** the marginal totals may vary considerably from their real values.

Random Unique

The companion concept to **special unique**. Random uniques are **data unit**s that are unique within a **dataset** on a given set of **key variable**s, but the **uniqueness** has occurred by chance because of the specifics of the **data** generating process (e.g., the sample that has been drawn) rather than because the corresponding **population unit** is intrinsically unusual.

Further reading:
Elliot, M.J., Skinner, C.J. and Dale, A., 1998. Special uniques, random uniques and sticky populations: some counterintuitive effects of geographical detail on disclosure risk. *Research in Official Statistics*, 1(2), 53–67, http://tinyurl.com/SPEC-UNIQ.

Elliot, M.J., Manning, A.M. and Ford, R.W., 2002. A computational algorithm for handling the special uniques problem. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05), 493–509, https://doi.org/10.1142/S0218488502001600.

Ransomware

Ransomware is a type of **malware** that threatens to block permanently the user/victim's access to their **data**, and/or publish the data, unless a ransom is paid to the ransomware's owner (often via an anonymous **cryptocurrency**). Simple ransomware may lock the system without affecting any files directly; more advanced variants encrypt the victim's files, making them inaccessible, and demand a ransom payment to decrypt them. Recovery from a successful ransomware attack without making the payment can itself cost substantial amounts of money, on top of the administrative disruption that often follows.

Although mostly aimed at organisations, ransomware **attack**s could also be targeted at individuals, especially as smart, **network**ed devices become more widely used.

Further reading:

Young, A. and Yung, M., 1996. *Cryptovirology*: *extortion-based security threats and countermeasures*. IEEE Symposium on Security and Privacy, 129–40, https://doi.org/10.1109/SECPRI.1996.502676.

Rational Consumer

A theoretical construct in economics of a consumer who makes economic decisions based on a systematic evaluation of the relevant **information**, with the objective of maximising their own utility (roughly, maximising their benefits and/or minimising their costs).

Its relevance to **privacy** is that it underpins **privacy calculus** models of decision making about **information** sharing. Economic approaches to privacy are also based on this construct and aim at pricing the social costs of privacy into privacy decisions.

Further reading:

McFadden, D., Machina, M.J. and Baron, J., 2000. Rationality for economists? *Elicitation of Preferences*, 73–110, https://doi.org/10.1007/978-94-017-1406-8_4.

See also: ECONOMICS OF PRIVACY, NEGATIVE EXTERNALITIES OF PRIVACY, VALUE OF PRIVACY

Reality Mining

A **data mining** technique that specifically focuses on **data** from mobile devices, to gain insights into human behaviour and social interactions. Data can include the owner’s location, their call and text logs, usage of apps, and output from **sensor**s in the mobile devices. Such data is only likely to increase in scope as mobile devices become smarter and the potential interface between mobile devices and **neurotechnology** throws up even greater possibilities. The implied possibility of mass surveillance raises significant **privacy** concerns.

Further reading:
Eagle, N. and Pentland, A., 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10, 255–68, https://doi.org/10.1007/s00779-005-0046-3.

See also: LOCATION DATA

Reasonable Expectation of Privacy

A reasonable expectation of **privacy** is a legal test for the application of privacy protections, usually defined as the expectations that a reasonable **person** would have in a particular social setting. If a reasonable person would expect privacy in that setting, then a **breach** would be against reasonable expectations; otherwise, it would not. The relativisation to a hypothetical and perhaps non-existent reasonable person is intended to produce a standard view which is more authoritative than the highly variant perceptions of individuals.

However, because such expectations differ over time (and perhaps within social milieux), the test is adaptable to new circumstances governed by new social norms (e.g., with respect to new digital technology) or unusual situations (e.g., with respect to the **privacy** of **public figure**s). Using this test, the reasonable expectations do not need to be defined in law but can be recalibrated by new judgments. It allows a nuanced approach to the circumstances of an individual case (e.g., balancing privacy against freedom of speech or the **public interest**).

The concept first emerged in legal doctrine as a result of the 1967 US Supreme Court judgment *Katz v United States*, which extended the Fourth Amendment’s prohibition on **unreasonable search**es by government agencies from a person’s **private property** and effects to anywhere where that person would have a reasonable expectation (assuming no search warrant had been obtained). The test in *Katz* was formulated in two stages. First, the person must have a subjective expectation of privacy; that is, the person must have taken steps to ensure the privacy of the item, ensuring that it was not open to public **scrutiny**. Second, that subjective expectation should be reasonable. In the *Katz* judgment itself, Katz, whose call from a public telephone box had been tapped by the FBI, did have reasonable expectations that his call would and should be private, even though the phone was a public one. Later American jurisprudence established that one does not have such reasonable expectations about papers thrown into an external dustbin, about activities taking place in plain view of the public or even in front of an open window, about **information** shared with third parties or about the telephone numbers one dials (because they are shared with the telephone company, a **third party**).

The test appeared in European law with judgments from the European Court of Human Rights, particularly *Halford v UK* (1997) and *Von Hannover (no.1) v Germany* (2004), connecting it explicitly with the **ECHR** Article 8 right of privacy. A House of Lords judgment in the case of *Campbell v Mirror Group Newspapers* (2004) brought the notion to prominence in Britain. In each case, passing the test brought a level of legal protection against the **publication** of information or **surveillance** (Halford, like Katz, had had a phone call tapped, while Von Hannover – Princess Caroline of Monaco – and Naomi Campbell were concerned about **paparazzi** photographs).

The test has been criticised for several reasons. In the American case, the test only applies to government searches which are regulated by the Fourth Amendment and does not extend to intrusions by private actors, leaving privacy protection partial at best. It is, of course, vague (how should we apply the notion to the privacy of children, for instance), and if reasonable expectations of privacy are declining – perhaps because of the prevalence of surveillance, **dataveillance** and media **intrusion** – then so will protections decline accordingly. The reasonable expectations test will not succeed in protecting a particular level or standard of privacy if social mores are perceived as shifting. To that end, the notion of who the 'reasonable person' is becomes crucial. Could such a role be defined in abstraction? Will it simply default to the beliefs of senior legal figures? Finally, it has been argued that the test is iniquitous, as it places the burden on plaintiffs to prove their own expectations are reasonable, as opposed to demanding that those breaching privacy show that their intrusion was reasonable.

Further reading:

Barendt, E., 2016. Problems with the 'reasonable expectation of privacy' test. *The Journal of Media Law*, 8(2), 129–37, https://doi.org/10.1080/17577632.2016.1209326.

Kerr, O.S., 2007. Four models of Fourth Amendment protection. *Stanford Law Review*, 60(2), 503–51, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=976296.
Wilkins, R.G., 1987. Defining the 'reasonable expectation of privacy': an emerging tripartite analysis. *Vanderbilt Law Review*, 5(5), 1077–1129, https://scholarship.law.vanderbilt.edu/vlr/vol40/iss5/2/.

See also: CELEBRITY PRIVACY, CONTEXTUAL INTEGRITY, MISUSE OF PRIVATE INFORMATION, PRIVACY, CULTURAL VARIATION OF

Reasonable Search

See: UNREASONABLE SEARCH

Recognition

See: EMOTION RECOGNITION, FACIAL RECOGNITION TECHNOLOGY, GAIT RECOGNITION, SPEECH RECOGNITION

Recommendation System

A recommendation, or recommender, system, is a system for **matching** consumers with products or services they will appreciate: perhaps entertainment, news stories, restaurants or online dating. The consumer is profiled, and the profile is used to generate recommendations. The technique of collaborative filtering assumes that consumers with similar profiles are more likely to have similar tastes than a random pair, and weights recommendations according to the **data** it has about purchases and satisfaction ratings. Alternatively, the system can use a profile of the product to match it against consumers (or use a combination of the two techniques).

Typically, recommendation systems work best with a large quantity of data about consumers, products and sales, using **machine learning** techniques. It follows that they tend to undermine consumers' **privacy** in terms of the **information** they hold about consumers to reason about them. They may also undermine **decisional privacy**, by steering consumers toward recommendations and discouraging their research into the market. Since recommendations are partially based on past choices, they may make it harder for consumers to experiment with new experiences.

An apt demonstration of how a complex series of business decisions can lead to breaches of privacy concerns the *Netflix prize*. The DVD-rental platform (as it then was) Netflix released a large amount of de-identified data about users' choices for a $1m annual prize (2007–9) for anyone who could create a collaborative filtering **algorithm** that outperformed Netflix's own. It was shown by privacy researchers that a **reidentification attack** was possible, and preferences of individuals could be reconstructed. Although this is often discussed in terms of the poorly conducted **anonymisation,** the driver for the decision making was the business value that improving the recommendation system represented.

Further reading:
Aggarwal, C.C., 2016. *Recommender systems*: *the textbook*. Cham: Springer, https://doi.org/10.1007/978-3-319-29659-3.

See also: PLATFORM FOR PRIVACY PREFERENCES, SURVEILLANCE CAPITALISM

Recommender System

See also: RECOMMENDATION SYSTEM

Reconstruction Attack

A method of attacking aggregated **data** to recover the underlying **database** in full or part. Such attacks will either use **auxiliary data** (possibly preparing for the reconstruction with a **subtraction attack**) or using repeated overlapping queries to an analysis server. Such attacks were the inspiration for the development for **differential privacy**.

Further reading:
Dinur, I. and Nissim, K., 2003. Revealing information while preserving privacy. *In: Proceedings of the twenty-second ACM symposium on principles of database systems*, 202–10, https://dl.acm.org/doi/10.1145/773153.773173.

Record

A form of **data** or document that captures an event, action or transaction or a sequence of the same. This includes artefacts as diverse as minutes of a meeting, medical history, films and audio recordings and transcripts of legal proceedings.

In the context of microdata, a specific technical meaning of record refers to all the **information** about a specific **data unit**, often represented by a single row in the **dataset**.

From a **privacy** perspective, **record**s create a problem by converting the transient experience of the event into a permanent form which can be stored, searched for and reproduced at an indefinite time in the future. They also enable the juxtaposition of remote events. This is exacerbated when they are sensitive, highly detailed and nearly impossible to effectively anonymise; for example, **electronic health record**s.

Record Linkage

A specific form of **data linkage** in which **record**s corresponding to the same **population unit**s in different **dataset**s are combined to produce a single **dataset**. There are multiple statistical and machine learning methods for carrying out record linkage, but most approaches are underpinned by a theory developed in the 1960s by Felligi and Sunter. The potential value to an analyst of linked **data** is that it increases the overall **data quality** and allows a larger range of more sophisticated models to be developed. For example, if I link a lifestyle **database** to a database of **electronic health record**s, I could then build a model to examine how lifestyle affects health.

Record linkage is an issue for **privacy** for several reasons. First, when data subjects have given over data about themselves to a **third party**, they are often unaware that that data might be linked to other data about them. Second, the temptation of linking is to continually add more and more data – a potential pathway to Ohm’s **database of ruin**. Third, if the data in question are supposedly anonymised, then then linkage may open a much larger set of **key variable**s and thus increase **disclosure risk**. Finally, record linkage is itself the mechanism through which an **adversary** can execute a **linkage attack** to enable **reidentification**.

Further reading:

Fellegi, I.P. and Sunter, A.B., 1969. A theory for record linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210, https://doi.org/10.1080/01621459.1969.10501049.

Christen, P., 2012. *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*. New York: Springer, https://doi.org/10.1007/978-3-642-31164-2_2.

Records Management

Records management is a process of **data lifecycle management** applied to records of personal **data**.

Record Suppression

A form of **statistical disclosure control** where whole **record**s are redacted from a **dataset**. This would normally be targeted at risky records such as **special unique**s.

Rectification

Under Article 16 of the EU **GDPR**, a **data subject** has a right to obtain from a **data controller** rectification of inaccurate **personal data**. This may also include completing incomplete personal **data**. Data controllers – for example, owners of a website – must therefore provide data subjects with a mechanism to enforce the GDPR's **accuracy** principle, at least in respect of **information** which identifies them. As Sharma and Menon point out, this means that the GDPR does not assist the family and friends of a deceased person who may wish to rectify the personal data of their relative, as the right cannot be exercised by a **third party**.

Further reading:
Sharma, S. and Menon, P., 2020. *Data privacy and GDPR handbook.* Hoboken: Wiley, 203–5.

See also: RIGHT TO RECTIFICATION

Redaction

A form of **suppression** for textual **information**, where sensitive, confidential or disclosive information is deleted from the text or masked. This process may happen in court cases where it might be deemed that irrelevant, private information in a document may prejudice proceedings. The other context where redaction is common is **national security**, where **disclosure** of a piece of **information** may be deemed to be against the national interest. In some cases, text may be redacted as a part of a disclosure control process. For example, pharmaceutical companies are required to make available to the European Medicines Agency (EMA) the reports from clinical studies, which the EMA then publishes online. However, the pharmaceutical companies are also responsible for ensuring that any such **publication** is **GDPR** compliant. For this reason, patient narratives are often redacted before the reports are shared.

See also: RESPONSIBLE DISCLOSURE

Red Team

A red team is a simulated **adversary** that attempts to **attack** an organisation's systems to test the **security** of those systems. Red team testing differs from a **penetration test** as the former tend to operate over longer time periods using an array of **attack vector**s and will be countered by a corresponding **blue team**. Red–blue team exercises therefore form part of the organisation's **security** infrastructure rather than simply tests of it.

Further reading:
Miessler, D., 2021. The difference between red, blue, and purple teams. *Unsupervised Learning*, https://danielmiessler.com/study/red-blue-purple-teams/.

See also: INTRUDER TESTING, MOTIVATED INTRUDER, PURPLE TEAM

Reference Monitor

A **security** device that mediates all access to the resources of a system to implement an **access control** policy. It is an abstract idea rather than a concrete implementation, and its goal is to guarantee that, independently from the system's implementation details, all access to resources is verified against a predetermined set of criteria. Typically, the reference monitor is implemented as a **software** module that sits in between the system's resources and any **application**s or processes that use them.

Regulation

See: REGULATORS

Regulators

A regulator is a body responsible for enforcing **compliance** with the law, usually civil law. Criminal law enforcement agencies are generally beyond the scope of what are collectively termed ‘regulators’. National police services may act against some **breach**es of **privacy** rights (e.g., **harassment**, physical violation, coercive control and some cases of **revenge porn**. Generally, however, regulators are public bodies (or private bodies with delegated powers) who investigate, enforce and provide guidance on civil law compliance.

The regulatory landscape of a given country will be determined by the composition of its legislative jigsaw. Countries which apply the **GDPR**, for example, will have a regulator responsible for **data protection** compliance across all sectors (public and private). In the US, by contrast, data protection and privacy laws are currently more fragmented, meaning that regulators such as the Office for Civil Rights will enforce privacy in healthcare, but not more broadly in other sectors.

In England and Wales, the Information Commissioner’s Office enforces data protection, **freedom of information** and privacy in electronic **communication** laws. This is a common arrangement, with the French National Commission on **Information** and Liberty (‘CNIL’) similarly regulating organisations in respect of multiple **information** laws.

A final role of national regulators is to cooperate internationally on information law enforcement and policy, for example through the **European Data Protection Board**.

Further reading:

Commission on Information and Liberty, 2015. *The CNIL’s missions*, www.cnil.fr/en/cnils-missions.

Information Commissioner’s Office, 2023. *What we do*, https://ico.org.uk/about-the-ico/what-we-do/.

Office for Civil Rights, 2023. *About us*, www.hhs.gov/ocr/about-us/index.html.

See also: HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT, SUPERVISORY AUTHORITY

Reidentification

Reidentification is often used as a generic term denoting the process of revealing the (previously concealed) **identity** of an individual. This could involve the reversal of anonymisation, pseudonymisation or de-identification. Where not defined within a particular **jurisdiction**, it could mean any of these. However, there are subtle variations in the usage of the term ‘reidentification’.

In 1981, when the Council of Europe passed Convention 108 on automated **data processing**, the law was not concerned with ‘*identification of persons by means of very sophisticated methods*’. However, as sophisticated identification techniques have become more accessible, a wider scope of **information** has become of legal concern as potentially (re)identifiable.

In the United Kingdom, the **Data Protection** Act 2018 defines reidentification as the process of reversing **de-identification**, with the latter falling slightly short of what would be conventionally understood as **anonymisation**. De-identification per s.171 Data Protection Act, is when an individual cannot be identified ‘without more’ (i.e., without further information). This definition is more akin to the GDPR’s category of data which have undergone **pseudonymisation**, rather than the elimination of all reasonable means of identification as required for anonymisation.

This nuance provides commentary on Paul Ohm’s much-cited characterisation of reidentification as ‘the surprising failure of anonymisation’. Mere removal of identifiers is not the same as managing all reasonable means of identification, and the sophistication of what Ohm terms ‘reidentification science’ is a reminder of the complex challenge that anonymisation represents. Where anonymisation has been implemented effectively, there should not be a reasonably likely risk of reidentification. This changes the question to ‘how do we carry out effective anonymisation?’ rather than ‘how effective is anonymisation?’

A more holistic understanding of identification and associated **risk**s would therefore cast reidentification not just as a computationally powered science, but also as a failure of **information governance**. To prevent unauthorised revelation of personal **identity**, it is necessary to anticipate not only the power of **algorithm**s to match people across **dataset**s, but also the human agency involved. Careful consideration of who has access to concealed identities, and for what purposes, is also required.

Further reading:

Council of Europe, 1981. *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*: *Explanatory Report*, https://apdcat.gencat.cat/web/.content/01-autoritat/normativa/documentos/712.pdf.

Ohm, P., 2009. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701, www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/.

Narayanan, A. and Shmatikov, V., 2010. Myths and fallacies of ‘Personally Identifiable Information’. *Communications of the ACM*, 53(6), 24–6, https://doi.org/10.1145/1743546.1743558.

See also: IDENTIFIABILITY, JIGSAW IDENTIFICATION, PERSONAL INFORMATION, PERSONAL DATA, REIDENTIFICATION ATTACK, RELEASE AND FORGET

Reidentification Attack

An attempt by an **adversary** to identify **population unit**s within **data** that have been through an **anonymisation** or **pseudonymisation** process.

The risk of such an **attack** being successful is the basis for most **disclosure risk** metrics.

See also: REIDENTIFICATION

Relational Autonomy

The concept of **privacy** encompasses a range of values at play as people co-exist across societies. **Autonomy** is one such value, which broadly expresses a person's ability to make decisions, shape their sense of self and govern their own lives.

The default characterisation of autonomy thus revolves (at least implicitly) around the individual. Relational autonomy is a qualification of the concept, designed to address the alleged individualism of autonomy as a political and ethical value. Feminist scholars have tried to understand personal agency in a way which is nonetheless grounded in our social embeddedness. A decision made by a 'relationally autonomous' actor is still uncoerced, but it will also consider the perspectives, preferences and interests of the people with whom the agent exists in relationship. The qualification makes explicit the nuance of autonomy as a lived experience, which is seldom exercised in any absolute sense of complete decisional **seclusion**. It may therefore facilitate the expression of conceptions of group, as well as individual, privacy.

Further reading:

Davy, L., 2019. Between an ethic of care and an ethic of autonomy. *Journal of Theoretical Humanities*, 24(3), 101–14, https://doi.org/10.1080/0969725X.2019.1620461.

Mackenzie, C. and Stoljar, N., 2000. *Relational autonomy*: *feminist perspectives on autonomy, agency, and the social self*. New York: Oxford University Press.

See also: DECISIONAL PRIVACY, FEMINIST CRITIQUE OF PRIVACY, GROUP PRIVACY

Release and Forget

‘Release and forget’ is a term coined by Paul Ohm to describe a type of bad **anonymisation** practice. If a **data controller** incorrectly believes that the risk of **reidentification** from anonymised **data** is zero, then they might believe it is safe to *release* into the **public domain**, and once that is done, it can be *forgotten* about. This is not the case, as useful anonymised **information** always carries a **risk** of reidentification. Hence, thorough anonymisation methodologies such as the **Anonymisation Decision-Making Framework** should include (a) steps to minimise reidentification risk, (b) a process of monitoring past releases for changes in risk and (c) plans to respond in the event of a post-release **data breach**.

Further reading:
Ohm, P., 2010. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701–77, www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/.

See also: PUBLICATION, DATA RELEASE

Reliability

See: TRUST

Reliance

See: TRUST

Reliance Authentication

The process used to identify the **identity** of **user**s to access a system. These methods usually involve the use of a password, biometric **authentication** or **two factor authentication**. Sometimes these methods require to store **personal data** of the user in a single **database**, making it a place for **attack**s to **breach** the system.

See also: RELIANCE

Remailing

The practise of sending emails to a recipient while maintaining the sender's **privacy**. Remailing services come in many forms, such as mixmaster remailers and nym servers. To make sure that the **identity** of the original sender cannot be connected to the message, mixmaster remailers employ a sophisticated **encryption** system. On the other hand, Nym servers let users create private email accounts that may be used to send and receive messages without disclosing their real identities. Remailing may be a helpful tool for maintaining **anonymity** and privacy, but it can also be used for malicious activities such as **spam**ming, **phishing** or spreading **malware**.

Remediation

See: REMEDIES

Remedies

The diverse nature of **privacy breach**es means that they have no single remedy, and remedies that are possible vary in their effectiveness. Alienation from **private property** can be remedied by restoring the property to the owner, and if someone suffers financial **harm** because of a breach, they can be compensated. Certain types of observation, such as **voyeurism** or **surveillance**, can only be ended; the past **intrusion** cannot be undone. **Record**s of the observation may be deleted, and a surveillant organisation may be ordered not to use them for decision-making purposes. However, if such records have been disseminated, especially on the **Internet**, it will be hard to suppress them entirely. In the case of **informational privacy**, the release of **information** into the **public domain** means that the privacy has effectively gone and cannot be regained (although an injunction may slow the spread, at least temporarily). The embarrassment and stress caused by **doxxing**, **outing** and the release of intimate **information** or images cannot be expunged. A breach of **group privacy** may affect the cohesion of the group irreparably.

Since many privacy breaches increase various **risk**s without causing immediate or tangible financial or physical harm, quantifying compensation can be extremely difficult. Courts in the United States have tended to focus on financial or physical harm, discounting effects such as anxiety or inconvenience, as argued by Citron and Solove. The EU has aimed to provide a more extensive set of rights to remedies, such as the **right to be forgotten**.

Further reading:

Citron, D.K., 2022. Privacy injunctions. *Emory Law Journal*, 71(5), 955–83, https://scholarlycommons.law.emory.edu/elj/vol71/iss5/3.

Citron, D.K. and Solove, D.J., 2022. Privacy harms. *Boston University Law Review*, 102(3), 793–863, https://www.bu.edu/bulawreview/files/2022/04/CITRON-SOLOVE.pdf.

Cofone, I.N., 2020. Online harms and the right to be forgotten. *In*: Cofone, I.N., ed., *The right to be forgotten*: *a Canadian and comparative perspective*. London: Routledge, 1–16, https://doi.org/10.4324/9781003017011.

See also: OBJECTIVE HARM, RECTIFICATION, SUBJECTIVE HARM

Remote Access

The ability to use a **network** connection to access a server from a distance. Users are now able to access resources and **data** that aren't really on their computer or on their local **network**. **Virtual Private Network**s (VPNs), Remote Desktop **Protocol** and cloud-based apps are just a few ways to gain remote access.

Businesses frequently utilise remote access to let workers work from home or other remote places. Although remote access has numerous advantages, it might provide unauthorised people access to critical data or systems if it is not adequately protected. To reduce the danger of unwanted access, firms that employ remote access should have robust **security** measures in place, such as **encryption**, **access control** and frequent **security audit**s.

Remote Access Server

See also: REMOTE ACCESS

Remote Analysis Server

In the context of **data** analysis, remote analysis servers provide an additional layer of **security** beyond remote access. The **user** in this situation does not access the data directly but rather submits code which can then be run by service staff. As the user does not see or directly manipulate the data, **attack vector**s which require this (such as **reidentification attack**s using statistical **matching**) are prevented. Remote analysis servers are still vulnerable to **differencing** attacks and to the submission of malicious code, so rigorous checking of both the inputs and outputs is still required.

See also: REMOTE ACCESS, DATA ENVIRONMENT

Remote Query

A request for **data** or **information** delivered across the **Internet**. A remote server receives a request from the requesting machine and responds by sending the data. Structured Query Language (SQL) and Web Services are just a few of the computer languages that may be used to carry out remote inquiries (RPCs). If remote inquiries are not adequately protected, they may provide unauthorised people access to systems or sensitive data.

Further reading:
Shar, L.K. and Tan, H.B.K., 2012. Defeating SQL injection. *Computer*, 46(3), 69–77, https://doi.org/10.1109/MC.2012.283.

See also: REMOTE ACCESS

Replay Attack

The purpose of a replay **attack** is to obtain unauthorised access to a system intercepting and replaying a **communication**. In a replay attack, the adverasary intercepts a genuine **data** transmission – such as an **authentication** request – and replays it later to pass as the original sender. Strong authentication systems that include additional **security** features, such as message **encryption** or **digital signature**s, can prevent replay attacks. Using time-based **protocol**s, which provide distinct tokens that cannot be reused or replayed, is another tactic for mitigating replay attacks.

Further reading:
Miao, F., Pajic, M. and Pappas, G.J., 2013. Stochastic game approach for replay attack detection. *In*: *52nd IEEE conference on decision and control*, 1854–9, IEEE, https://doi.org/10.1109/CDC.2013.6760152.

Repurposing

The use of **data** for one purpose that was originally collected for a different purpose. If the data in question are **personal data**, then that may throw up issues of whether the new purpose falls within the reasonable expectations of the **data subject**s, particularly where **consent** is relied upon as the **legal basis for processing**.

See also: GDPR, PURPOSE LIMITATION, SECONDARY USE OF DATA

Reputation

Reputation is an evaluative social judgement, measure or opinion about an agent (including individuals, social groups and organisations). It is usually derived from the history of the agent's interactions or activities but can also be affected by misinformation and selective consideration of past events and other forms of bias. It is in the agent's interests to restrict the input to the evaluation to actions that can be seen in a positive light. Actively influencing or curating one's reputation is called **reputation management**, a branch of public relations. An increase in the negativity of one's reputation following the revelation of **information** or misinformation is termed reputational **harm**.

One purpose of **privacy** is to enable individuals to curate their own reputations; to present themselves as they wish to be seen. One function of **transparency** is to prevent individuals crafting misleading reputations. In the past, reputation often fluctuated through time as certain events were forgotten, but one concern about digital technology is that such natural 'decay' of memory is prevented.

Further reading:

Solove, D.J., 2007. *The future of reputation: gossip, rumor, and privacy on the Internet*. New Haven: Yale University Press.

See also: DEFAMATION, GOSSIP, INFORMATIONAL SELF-DETERMINATION, RIGHT TO BE FORGOTTEN

Reputation Management

See: REPUTATION

Reserve

Reserve is one of the four states of **privacy** discussed in Alan Westin's *Privacy and Freedom*. It signifies resistance to unwarranted **intrusion**; the erection of a psychological barrier against, or withdrawal from, aspects of the surrounding context.

Further reading:
Westin, A.F., 1967. *Privacy and freedom*. New York: Ig Publishing.

See also: PSYCHOLOGICAL PRIVACY

Resilience

The capacity of an entity to withstand shocks and/or adverse changes in its environment and to be able to recover from such shocks.

In a **privacy** context, resilience is the capacity of an individual to mitigate the **harm**s of a **breach** and to recover from the breach quickly. The term **cyber resilience** is applied at the organisational level to denote the organisation's capacity to deal with cyberattacks.

See also: ONTOLOGICAL SECURITY

Respondent

A respondent is a **person** who has responded to a survey with **information**. If such a person has responded with some of their personal **data** in their answers, then they will become a **data subject**, and the researchers will need to adjust their methods and ethical practice to comply with **data protection** regulation. It is also important to note that a respondent may also provide **personal data** about **other** people, for example members of their household, which may impact their **privacy**.

Further reading:
Fowler, F.J., 2014. *Survey research methods*, 5th edition. Thousand Oaks, CA: Sage.

Response Knowledge

The knowledge that a given population unit is included in a **dataset**. This could be through private knowledge (e.g., somebody that I know has told me that s/he responded to a particular survey, or perhaps was in a particular clinical trial and the dataset contains **data** from that trial). Or it could be through background knowledge that a particular **population unit** is a member of a **population**, and the data is a full dataset for that population (e.g., a **census**).

In general, response knowledge of an **adversary** significantly increases the likelihood of a **reidentification** by that adversary. This is particularly true if the dataset is small.

Response Variable

In statistical models, a variable representing the outcome that is to be predicted. Also known as the ‘outcome variable’ and, in experimental contexts, the ‘dependent variable’.

If the statistical model is good enough and therefore the residual error term small enough, the value of the response variable might be disclosed – at a high level of probability – about specific **population unit**s.

See also: PREDICTIVE MODELLING

Responsible Disclosure

The process of disclosing a **security vulnerability** in a system to the owner of the system. It is usually carried out by security researchers or ethical hackers who discover a **bug** or flaw. The goal of the responsible **disclosure** process is to minimise the impact of the security issue on system owners and **user**s.

Further reading:
Ding, A.Y., De, J.G.L. and Janssen, M., 2019. Ethical hacking for boosting IoT vulnerability management: a first look into bug bounty programs and responsible disclosure. *In*: *Proceedings of the eighth international conference on telecommunications and remote sensing*, 49–55, https://doi.org/10.1145/3357767.3357774.

See also: ETHICAL HACKING, ENGINEERING ETHICS

Restricted Access

Cybersecurity measures that (seek to) limit who has access to **data** and the manner of that access.

See also: ACCESS CONTROL, SAFE SETTINGS, SAFE PEOPLE

Retention

See: DATA RETENTION

Revenge Porn

The term 'revenge porn' began to appear in mainstream lexicographical dictionaries in 2016. Common to most definitions is the non-consensual **publication** of sexual content (involving adults). A key area of divergence is whether the publisher must intend to cause their victim **harm**, or whether malice is simply a common feature of revenge porn.

This is not just a lexicographical distinction: intent is also a controversial question for lawmakers. Some countries (such as the UK, since 2015) require proof of malicious intent for non-consensual distribution of sexual imagery to constitute a crime, while others (such as Finland, since 1987) do not. At the heart of this distinction is whether the non-consensual publication is considered a sufficient harm to the victim's **privacy** that it should be treated as a crime in and of itself – that is, whether there should be a legal duty to obtain **consent** to share intimate content. The Finnish criminal law treats this as a **data protection** offence, whereas the UK offence is founded more on **harassment**, in which personal motivation has greater relevance.

Further reading:

Foley, K.G., 2021. 'But, I didn't mean to hurt you': why the first amendment does not require intent-to-harm provisions in criminal 'revenge porn' laws. *Boston College Law Review*, 62(4), 1365–1412, https://lira.bc.edu/work/ns/ccb906d7-400f-4ea3-a774-53f6beb5376e.

Kierkegaard, S., 2011. To block or not to block – European child porno law in question. *Computer Law & Security Review*, 27(6), 573–84, https://doi.org/10.1016/j.clsr.2011.09.005.

Pöysti, T., 2009. Judgment in the case of K.U v Finland. *Digital Evidence and Electronic Signature Law Review*, 6, 33–77, https://sas-space.sas.ac.uk/5452/1/1855-2575-1-SM.pdf.

Reverse Fishing Attack

An **attack** which starts from a *known to be unusual* **population unit** and attempts to find them in one or more de-identified **dataset**s.

See also: FISHING ATTACK

Reversibility

A key question raised in the context of **statistical disclosure control** is whether (or to what extent) the process is 'reversible'. If the effects of a disclosure control process or **privacy enhancing technology** can be reversed by an **adversary**, in most cases the **data** will not be deemed to be anonymous **information**. How feasible it is to reverse the concealment of identities is a question key to the distinction between **anonymisation** and **pseudonymisation**.

See also: PSEUDONYM REVERSAL

Revocation

The EU **GDPR** requires that valid consent must always be revocable. Article 7.3 states: 'The **data subject** shall have the right to withdraw his or her consent at any time. The withdrawal of **consent** shall not affect the **lawfulness** of **processing** based on consent before its withdrawal. Prior to giving consent, the **data** subject shall be informed thereof. It shall be as easy to withdraw as to give consent.'

The GDPR uses the word 'withdrawal' rather than 'revocation', but 'revocable' is often the preferred adjective outside the statutory language (the alternative being the more cumbersome 'capable of withdrawal'). Consent revocability is one of the more stringent innovations of the GDPR, making it an inappropriate basis for processing even in the contexts where consent is required for other purposes. For example, consent may be required for the release of health-related **information** under **common law confidentiality**, but if the **data** are required for – for example – medical services even if the patient withdraws consent, another basis for processing under the GDPR must be selected.

Further reading:

Dove, E.S. and Taylor, M.J., 2021. Signalling standards for progress: bridging the divide between a valid consent to use patient data under data protection law and the common law duty of confidentiality. *Medical Law Review*, 29(3), 411–45, https://doi.org/10.1093/medlaw/fwab014.

Politou, E., Alepis, E. and Patsakis, C., 2018. Forgetting personal data and revoking consent under the GDPR: challenges and proposed solutions. *Journal of Cybersecurity*, 4(1), 1–20, https://doi.org/10.1093/cybsec/tyy001.

See also: LAWFUL BASIS, INFORMED CONSENT, RIGHT TO BE INFORMED

Right of Access

The right of a **data subject** to request a copy of **personal data** about themselves, or to receive **information** about how it is used.

Article 15 of the GDPR, lists the categories of which data subjects are entitled to receive upon request to the **data controller**. These map very closely onto the information which **data** subjects should receive anyway at the point their information is collected (purposes of processing, the categories of data involved, recipients of categories of recipient with whom their information is shared, and so on).

The data subject also has a right under Article 15.3 to receive a copy of the information about them which are held by the data controller. This right should not adversely affect the rights and freedoms of others (e.g., redactions to protect the **privacy** of other data subjects may be appropriate).

Further reading:

Custers, B. and Heijne, A.S., 2022. The right of access in automated decision-making: the scope of article 15(1)(h) GDPR in theory and practice. *Computer Law & Security Review,* 46, 105727, https://doi.org/10.1016/j.clsr.2022.105727.

See also: DATA RECIPIENT, DATA SUBJECT ACCESS REQUEST, PRIVACY, RIGHT TO PRIVACY, TRANSPARENCY

Right to an Explanation

The right of an individual to receive an explanation of a decision made solely on the basis of automated processing of **personal data**, and that significantly affects them.

The EU **GDPR** does not provide a ‘right to explanation’ of personal **data processing** *per se*. Article 15 GDPR sets out **information** which the individual is entitled to request about their personal data processing (which is similar to the details they should receive under Article 13 GDPR anyway).

A particularly contested aspect of Article 15 GDPR is Article 15.1(h), as it intersects with Article 22 GDPR on **automated decision-making**. Wachter et al argued strongly against the idea that the GDPR offers any right to an explanation of decision-making. In response, Selbst and Powles suggested that while there is no single, neat ‘right to explanation’ in the GDPR, neither is such a right illusory. They point to the cumulative impact of Articles 15 and 22 GDPR, which provide rights to meaningful information about the ‘logic’ of automated processing, where such processing forms the basis of ‘significant’ decisions. Significant decisions include measures with legal impact, or which otherwise engage the rights and freedoms of the **data subject** in a more than trivial way.

The overall impact of the ‘right to explanation’ in the GDPR can be summarised as: (1) a general right to specified information about one’s personal data processing in Article 15 and (2) a right to receive meaningful information and human review when significant decisions are made about you through automated processing of your personal data.

Further reading:

Selbst, A. and Powles, J., 2018. ‘Meaningful information’ and the right to explanation. *International Data Privacy Law*, 7(4), 233–42, https://doi.org/10.1093/idpl/ipx022.

Wachter, S., Mittelstadt, B. and Floridi, L., 2017. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99, https://doi.org/10.1093/idpl/ipx005.

See also: RIGHT TO BE INFORMED, RIGHT TO PRIVACY, RIGHT TO DATA PROTECTION

Right to Be Forgotten

The ‘right to be forgotten’ in Europe, that is, the right to **erasure** of personal **information** of limited ongoing relevance, has a longer legacy than is commonly believed. As Sharma notes, in 1978, the ‘right to oblivion’ was introduced in French law. Versions of this right then travelled to other European jurisdictions, such as Spain (of which, see more below). The EU **Data Protection Directive** (‘the Directive’, in force 1995–2018) also included a right to obtain erasure from the **data controller** in relation to **personal data** that was incomplete, inaccurate or otherwise non-compliant with the Directive.

The right has, however, risen to recent prominence following the 2014 judgment of the Court of Justice of the European Union (‘CJEU’) in the case between Google Spain and the Spanish data protection regulator (the ‘AEPD’). Otherwise known as the *Costeja González* case, the complainant objected to the continued indexing of a 1998 newspaper article which mentioned his social security debts and associated real-estate auction. He complained to the AEPD that the social security proceedings were long resolved, no longer relevant and unnecessarily injurious to his **reputation**. The AEPD upheld his complaint, and the CJEU found in favour of his ‘derecho al olvido’ (i.e., his right to be forgotten).

Although the *Costeja González* case was decided before the GDPR came into force, many see its impact in the expansion of the right to erasure under Article 17 GDPR. Compared to the Directive, Article 17 GDPR is far more detailed and generous in its description of the circumstances in which a **data subject** may have a right to erasure; for example, when they withdraw their **consent**, or when the processing of their **data** is no longer necessary for its original purpose.

Further reading:

Aidinlis, S., 2020. The right to be forgotten as a fundamental right in the UK after Brexit *Communications Law* 25(2), 67-78, https://ssrn.com/abstract=3554625.

Sharma, S., 2019. Data subjects’ rights. *In*: Sharma, S., *Data privacy and GDPR handbook.* Hoboken: Wiley, 193–232.

See also: DELETION, DIGITAL FOOTPRINT ERASER, NECESSITY, REVOCATION, RIGHT TO DATA PROTECTION

Right to Be Informed

Many legal **jurisdiction**s provide individuals with a right to receive **information** which is relevant to them as (for example) consumers, patients and **data subject**s. It is a necessary adjunct to **informed consent** in many contexts.

For example, the US consumer bill of rights includes a right to be provided with sufficient **information** to make good purchasing decisions. In most cases such contextualised rights are **privacy** neutral. However, in some cases, they throw up privacy tensions. For example, the sixth amendment to the US constitution includes the right to know their accuser’s **identity** (as well as the charges and evidence against them); this is certainly in tension with the accuser’s privacy and in some cases may have significant implications for the accuser.

In many cases the doctrine manifests more as a duty on the discloser (such as a doctor or **data controller**) than as a right held by the data subject., For example, the **GDPR** does not refer to a 'right to be informed'. Instead, the **transparency** requirements in Articles 12–14 are framed as data controller *obligations*, rather than data subject rights. Nonetheless, these obligations are sometimes referred to collectively as the data subject's 'right to be informed'.

Further reading:

Information Commissioner's Office, 2023. *Right to be informed.* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-be-informed/.

Waller, S.W., Brady, J.G., Acosta, R.J., Fair, J. and Morse, J., 2011. Consumer protection in the United States: an overview. *European Journal of Consumer Law*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1000226.

See also: RIGHT OF ACCESS, RIGHT TO PRIVACY

Right to Be Let Alone

In their pioneering examination of American common and constitutional law, Warren and Brandeis claimed to have detected a **privacy** right taking the form of a right to be let alone, a right to be free from unwarranted **scrutiny** or **interference**. Their paper became seminal and began a long process of developing privacy protections in US law. It has been taken by some to be vague; the paper defines neither privacy nor the idea of being let alone, although the basic idea that Warren and Brandeis posit is that the individual has (some) rights to withdraw into **seclusion**, **solitude**, **intimacy** where their activities are not of **public interest**.

Prosser later synthesised the case law that flowed from Warren and Brandeis' article, reducing their 'right to be let alone' to a short and disconnected list of **privacy tort**s. This approach has been updated by Solove, who has pointed out that the right to privacy in US law is (as in many jurisdictions) composed of a heterogenous mix of constitutional rights, evidentiary privileges and statutory provisions.

Further reading:

Prosser, W.L., 1960. Privacy. *California Law Review*, 48, 383–423, https://lawcat.berkeley.edu/record/1109651?ln=en.

Solove, D.J., 2006. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477–564, https://scholarship.law.upenn.edu/penn_law_review/vol154/iss3/1/.

Warren, S.D. and Brandeis, L.D., 1890. The right to privacy. *Harvard Law Review*, 4, 193–220. https://doi.org/10.2307/1321160.

See also: ATTENTIONAL PRIVACY, INTRUSION UPON SECLUSION, RESERVE, RIGHT TO PRIVACY

Right to Correct

See: RECTIFICATION

Right to Data Portability

In 2018, the EU GDPR introduced a general right to portability of **personal data**. It gives individuals the right 'to receive the personal data concerning him or her, which he or she has provided to a **data controller**, in a structured, commonly used and machine-readable format and have the right to transmit those **data** to another controller'.

This only applies, however, in the context of digital **data processing**, and when the personal data are processed using the basis of **consent** or contractual purposes. When personal data are processed, for example, using the basis of **public interest** (as in the case of public authority **information** about citizens) this commercial transfer of information is not available.

In the United States, where healthcare is primarily provided via the free market, the right has a more longstanding (but more specific) lineage via the **HIPAA**.

Further reading:

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L. and Sanchez, I., 2018. The right to data portability in the GDPR: towards user-centric interoperability of digital services. *Computer Law & Security Review*, 34(2), 193–203, https://doi.org/10.1016/j.clsr.2017.10.003.

See also: DATA PORTABILITY, DATA TRANSFER, INTEROPERABILITY

Right to Data Protection

The European Union introduced a 'right to **data protection**' in its **Charter of Fundamental Rights**, which came into force in 2009. It establishes the protection of **personal data** as a human right: broader than the right to privacy in **information**, which only covers **identifiable data** that relate to an aspect of the **data subject** (or their lives) which is in some way ‘private’. The right to data protection is thus tied to the concept of *identification*, as opposed to a culturally determined distinction between public and private spheres of life or self.

Outside of EU law, the distinction between **privacy** and **data** protection can be less obvious. Some authors see data protection and privacy rights as overlapping in the subset of privacy referred to as **informational privacy**, which O’Neill and Diker Vanberg regard as a key aim of data protection law. Dove, however, delineates informational privacy from the right to data protection, characterising the former as a state of non-access for certain kinds of information, with data protection more concerned with the **fairness** of processing than with **availability**.

Further reading:

Diker Vanberg, A., 2021. Informational privacy post GDPR – end of the road or the start of a long journey? *The International Journal of Human Rights*, 25(1), 52–78, https://doi.org/10.1080/13642987.2020.1789109.

Dove, E.S., 2019. The EU General Data Protection Regulation: implications for international scientific research in the digital era. *Journal of Law, Medicine & Ethics*, 46(4), 1013-1030, https://doi.org/10.1177/1073110518822003.

O’Neill, O., 2013. Can data protection secure personal privacy? *In*: Terry Sheung-Hung Kaan and Calvin Wai-Loon Ho (eds) *Genetic privacy*: *an evaluation of the ethical and legal landscape*. London: Imperial College Press.

See also: CHARTER RIGHTS

Right to Deletion

See: ERASURE, RIGHT TO BE FORGOTTEN

Right to Object

Not to be confused with the (often temporary) **right to restriction** of processing under Article 18 **GDPR**, or the more permanent right to erasure under Article 17, Article 21 GDPR provides **data subject**s with a ‘right to object’ to the processing of **personal data**, where the **processing** is based, **legitimate interest**, **public interest** or is conducted for direct marketing purposes.

The right to object is not as widely debated as other GDPR **data** subject rights, but Esposito has made an interesting case for the use of the right to receive an 'impersonal price', by objecting to **data processing** for price **personalisation**. Tosoni has also suggested that Article 22 GDPR can also be used to establish a right to object to significant decisions based on automated processing.

Further reading:

Esposito, F., 2022. The GDPR enshrines the right to the impersonal price. *Computer Law & Security Review*, 45, 105660, https://doi.org/10.1016/j.clsr.2022.105660.

Tosoni, L., 2021. The right to object to automated individual decisions: resolving the ambiguity of Article 22 (1) of the General Data Protection Regulation. *International Data Privacy Law*, 11(2), 145–62, https://doi.org/10.1093/idpl/ipaa024.

See also: PRICE DISCRIMINATION, RIGHT TO DATA PROTECTION

Right to Privacy

See: HISTORY OF PRIVACY

Right to Rectification

See also: RECTIFICATION

Right to Restriction

Article 18 of the EU's **GDPR** introduced a right to restrict the **processing** of **personal data** where:

(a) the **data subject** contests the accuracy of the **data**, and the **data controller** needs time to verify the **accuracy**.
(b) the processing is unlawful, and they prefer restriction to **erasure**.
(c) the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims.
(d) they object to the processing, and the controller needs time to ascertain whether they have an overriding **legitimate interest** in the meantime.

Restriction thus limits the controller's ability to use **information**, but to an extent that falls short of erasure (i.e., the data should still be retrievable). Kuru and Beriain give an interesting account of how the right may (dis)function in the context of **genetic data** which identifies more than one biological family member.

Further reading:

Kuru, T. and de Miguel Beriain, I., 2022. Your genetic data is my genetic data: unveiling another enforcement issue of the GDPR. *Computer Law & Security Review*, 47, 105752, https://doi.org/10.1016/j.clsr.2022.105752.

See also: RIGHT TO OBJECT, RIGHT TO DATA PROTECTION

Risk

Risk is often quantified based on a calculation of the likelihood of an adverse event occurring under conditions of uncertainty, multiplied by the severity of the event's potential impact. The International Organization for Standardization defines it as the 'effect of uncertainty on objectives'. *Taking a risk* means pursuing a course of action despite consequent **exposure** to the possibility of adverse events.

Risk connotes an attempt to understand the potential for **harm** to occur, as well as the nature of that harm, and therefore to reason about it rationally.

A framework for risk allows **risk management**, the coordination or planning of activities to achieve a desired risk profile. Actors can be *risk averse*, if they wish to minimise their exposure to adverse events, or *risk accepting/risk tolerating*, if they are prepared to take risks to increase the probability of beneficial outcomes. *Insurance* is the pooling or distribution of risk across a group of voluntary parties, who together pay premiums to fund compensation in the event of adverse events; the rate of premium is related to the size of the risk undertaken and the costs of the risk occurring. A *black swan* event is a catastrophic event of low probability but highly serious consequences.

In the context of **privacy**, there is usually a risk that **information** could be divulged leading to a privacy **breach**. This is measured by a privacy **risk assessment**. Risk management is an approach to privacy management, by which the risk of a privacy breach is reduced to an acceptable level. By this approach, risk can never be reduced to zero, so the possibility of a breach will always remain. Associated with the risk of a privacy breach are business risk (the risk of loss of **reputation**, goodwill or money as a result of a **privacy** breach) and reputational risk (the loss of reputation as a result of information being made **public**).

Further reading:

Hopkin, P. and Thompson, C., 2022. *Fundamentals of risk management*: *understanding, evaluating and implementing effective enterprise risk management*, 6th edition. London: Kogan Page.

International Organization for Standardization (ISO), 2009. *Risk Management – Vocabulary*, www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en.

Rescher, N., 2022. *Risk theory*: *rational decision in the face of chance, uncertainty, and risk*. Cham: Springer, https://doi.org/10.1007/978-3-030-78502-4.

See also: PRIVACY RISK, PRIVACY IMPACT ASSESSMENT, PRIVACY INSURANCE, PRIVACY CONCERN, RESILIENCE, RISK TOLERANCE, RISK-UTILITY TRADE OFF

Risk Assessment

The calculation or evaluation of a risk (usually of an adverse event).

In **privacy**, accurate or precise risk assessment is notoriously difficult, in part because the harms are difficult to define and even more difficult to quantify. **Risk** assessment tends to focus on **confidentiality** rather than privacy, because the risk can be taken as the likelihood of a **breach**, irrespective of what **harm**s may occur as a result. In **statistical disclosure control** all risk assessment is couched in this way.

Further reading:

Duncan, G.T., Elliot, M. and Salazar-González, J.-J., 2011. *Statistical confidentiality*: *principles and practice*. New York: Springer, https://doi.org/10.1007/978-1-4419-7802-8.

See also: DISCLOSURE RISK, PRIVACY METRIC, PRIVACY RISK, RISK TOLERANCE

Risk Management

The coordination or planning of activities to achieve a desired risk profile.

See also: RISK

Risk Tolerance

A person's or organisation's **risk** tolerance (or *risk appetite*) is their willingness and ability to accept risk. *Risk-seeking* assumes high risk tolerance, preferring opportunities with higher expected outcome values, even if these include significant risk of loss or negligible gain. *Risk-aversion* assumes low risk tolerance, preferring certain outcomes of low value over uncertain outcomes with higher expected value but greater risk of loss or negligible gain.

In a **privacy** context, risk tolerance will be one driver for an individual's **privacy calculus**. For organisations processing **data**, their risk tolerance will affect how **business case**s are constructed for **data sharing** projects and will also determine the details of **information governance** policies.

Further reading:
Breakwell, G.M., 2014. *The psychology of risk*, 2nd edition. Cambridge: Cambridge University Press.

See also: PRIVACY RISK

Risk–Utility Trade-Off

A term which captures the idea that in general reducing **cybersecurity** risk also impacts the *bona fide* use of cybersystems. The term has been particularly employed in the field of **anonymisation**, where it is recognised that **data utility** will necessarily decrease as the protection put in place to reduce **disclosure risk** increases. An example of this is **differential privacy**, which has been criticised for damaging **data** disproportionately to provide its much-vaunted guarantee of **confidentiality**. It is arguable that the damage caused by overprotection outweighs that caused by actual **breach**es.

Further reading:
Cox, L.H., Karr, A.F. and Kinney, S.K., 2011. Risk-utility paradigms for statistical disclosure limitation: how to think, but not how to act. *International Statistical Review*, 79(2), 160–83, https://doi.org/10.1111/j.1751-5823.2011.00140.x.
Duncan, G.T., Keller-McNulty, S.A. and Stokes, S.L., 2001. Disclosure Risk vs. Data Utility: The R-U Confidentiality Map. National Institute of Statistical Sciences Technical Report 121. NISS, www.niss.org/sites/default/files/technical-reports/tr121.pdf.

See also: INFORMATION LOSS, PRIVACY GUARANTEE, PRIVACY RISK

Roe v Wade

The 2022 US Supreme Court decision in *Dobbs v Jackson Women's Health Organization* removed the federal protection for **abortion** established by the *Roe v Wade* judgment in 1973. The **privacy** implications of the *Dobbs* decision are manifold but can be summarised as bodily and informational.

On the level of **bodily privacy**, the new freedom for US states to re-criminalise abortion has obvious implications for the constitutional **right to be let alone**. While the **informational privacy** ramifications may be less immediately evident, numerous commentators have suggested that **HIPAA** does not protect identifiable patient **information** to the extent that would prevent government agencies using medical records to bring criminal charges against patients or healthcare providers, in states where terminations of pregnancy have been re-criminalised.

Further reading:

Spector-Bagdady, K. and Mello, M.M. 2022. Protecting the privacy of reproductive health information after the fall of Roe v Wade. *JAMA Health Forum*, 3(6), e222656–e222656. Https://doi.org/10.1001/jamahealthforum.2022.2656.

Role-Based Access Control

A **cybersecurity** model that restricts access to resources by system users depending on the user's role, responsibilities and duties within an organisation.

See also: ACCESS CONTROL

Rounding

A set of perturbative **statistical disclosure control methods** where a figure is rounded off to a defined base; it is most often applied to **tables of counts**. Commonly the base is 3, 5 or 10, but in principle any number could be used. The rounding creates uncertainty for an **adversary** as to what the exact counts are. Rounding – particularly to small bases – can be vulnerable to **differencing** and **subtraction attack**s.

See also: PERTURBATION

RSA Encryption

A **cryptographic** technology, created in 1977, used to protect online **communication**s. A **public key** and a **private key** are used in RSA **encryption** to encrypt and decode communications. The communication is encrypted using the public key, and it is decrypted using the private key. While the private key needs to be kept **secret**, the public key may be shared with anybody. The communication is safe during transmission since only the receiver has access to the private key.

Further reading:

Milanov, E., 2009. The RSA algorithm. *RSA laboratories*, 1–11, https://pdfdirectory.com/pdf/0702-the-rsa-algorithm.pdf.

Boneh, D., 1999. Twenty years of attacks on the RSA cryptosystem. *Notices of the AMS*, 46(2), 203–13, www.ams.org/notices/199902/boneh.pdf.

R-U Map

A plot of the trade-off between **disclosure risk** and **data utility** intending as a decision support tool.

Further reading:

Duncan, G.T., Keller-McNulty, S.A. and Stokes, S.L., 2001. *Disclosure Risk vs. Data Utility*: *The R-U Confidentiality Map*. National Institute of Statistical Sciences Technical Report 121. NISS, www.niss.org/sites/default/files/technicalreports/tr121.pdf.

See also: RISK–UTILITY TRADE-OFF

Rumour

A rumour is a narrative that has spread beyond the original interlocutors, despite lacking verification. Rumours circulate informally, and can often contain, deliberately or otherwise, misinformation. They often have sensational content that subjects would prefer to keep private or otherwise restrict. Because they usually circulate covertly, they can be hard to counter.

See also: GOSSIP, REPUTATION

S

Safe Data

One of the **five safes**, it implies that the **data** have been treated and/or **disclosure risk** assessed as part of the holistic process of managing **confidentiality risk**.

Safe Harbor

In 2000, the European Commission issued a decision declaring that the safe harbor principles provided adequate protection for **personal data** transferred from the EU to the US. The US/EU Safe Harbor framework thus provided a means for EU organisations to transfer personal data to US counterparts who had in some way self-certified as compliant with these safe harbor principles, as issued by the US Department of Commerce.

The principles expressly stated that adherence could be limited for the purposes of **national security**, **public interest**, law enforcement or **compliance** with US statute, case-law or government regulations. In the ***Schrems I*** judgment, the Court of Justice of the European Union ('CJEU') found that these exemptions were too broad to ensure the **privacy** of EU citizens' data and held that the Commission's 2000 **adequacy** decision was invalid.

The Safe Harbor framework was re-negotiated into the form of the **Privacy** Shield, but Max Schrems persisted and in *Schrems II* that too was found to be an insufficient basis for an adequacy decision. Personal data therefore cannot, at the time of writing, be transferred from the EU to the US without the use of additional safeguards such as **standard contractual clauses**. Calls have been made for more comprehensive **US privacy laws**, an independent US **data protection** regulator and ratification of **Convention 108+**, to bolster legal protection for personal data at the federal level and make a valid adequacy decision from the European Commission a more achievable goal.

Further reading:

Rotenberg, M., 2020. Schrems II, from Snowden to China: toward a new alignment on transatlantic data protection. *European Law Journal*, 26(1–2), 141–52, https://doi.org/10.1111/eulj.12370.

Schrems, M., 2016. The privacy shield is a soft update of the safe harbor. *European Data Protection Law Review*, 2, 148, https://doi.org/10.21552/EDPL/2016/2/4.

See also: BRUSSELS EFFECT, DATA PROTECTION PRINCIPLES

Safe Output

One of the five safes it implies that statistical outputs have been checked (**risk** assessed) before release as part of the holistic process of managing **confidentiality risk**.

See also: FIVE SAFES

Safe People

One of the five safes, it implies that the people accessing the **data** have been formally approved (and usually have gone through some training) before accessing the data as part of the holistic process of managing **confidentiality risk**.

See also: FIVE SAFES

Safe Projects

One of the five safes, it implies that any project that uses the **data** has been fully pre-specified and assessed through a formal project approvals process before it is allowed to proceed as part of the holistic process of managing **confidentiality risk**.

See also: FIVE SAFES

Safe Settings

One of the five safes, it implies that the environment in which the **data** are held and analysed is **secure** and that **data governance** processes are sound, as part of the holistic process of managing **confidentiality risk**. The term in fact predates the five safes, with, for example, Marsh et al. raising the alternative approaches of **safe data** vs safe settings as early as 1994.

Further reading:
Marsh, C., Dale, A. and Skinner, C., 1994. Safe data versus safe settings: access to microdata from the British census. *International Statistical Review*, 35–53, https://doi.org/10.2307/1403544.

See also: DATA ENVIRONMENT, FIVE SAFES

Safety

A state in which people, things or systems are shielded from damage, injury or **harm**. Personal safety is defined as bodily safety from attacks or accidents (such as protection from pollution or natural disasters). Many precautions and processes are used to guarantee safety. In the context of technology, safety includes methods to prevent harm and threats to devices, systems or **databases**.

Froomkin and Colangelo have argued that safety of all kinds is supported by **privacy**, that many aspects of US law are based implicitly on this insight and that technologies such as the **Internet of Things** and **network**ed autonomous vehicles are unsafe precisely because their privacy rules are inadequate. Examples where **locational privacy** enhances safety include witness protection programmes and protections from **doxxing** and stalking. **Informational privacy** can help protect people in sensitive or high-profile jobs, and also protect them from **identity theft**. **Spatial privacy** protects the safety of intimate relations. **Decisional privacy** protects individuals from government **intrusion** (in a high-profile example, the now overturned **Roe v Wade** Supreme Court decision protected the wellbeing of pregnant women). **Communication privacy** helps protect journalists and their sources, and also whistleblowers. **Associational privacy** allows people to seek safety in groups.

On the other hand, intuitively, safety and privacy may pull in opposite directions in some circumstances too. Ubiquitous **CCTV** might deter violent crime, while the safety of minors, dependants or the elderly infirm may be protected by monitoring by guardians or carers.

Further reading:

Froomkin, A.M. and Colangelo, Z., 2020. Privacy as safety. *Washington Law Review*, 95(1), 141–203, https://digitalcommons.law.uw.edu/wlr/vol95/iss1/6/.

See also: SECURITY

Salt

Salt is random **data** that is appended to a string before **encryption**. The addition of the salt ensures that the **encryption** is strong and protects against the issue of commonly used **password**s being guessed by an **adversary**.

Further reading:
Rosulek, M., 2021. Hash functions. *In: The Joy of Cryptography,* 203–13, https://joyofcryptography.com/pdf/chap11.pdf.

See also: CRYPTOGRAPHY

SAML

See: SECURITY ASSERTION MARKUP LANGUAGE

Sample Unique

A **data unit** within a sample **dataset** which is unique within that dataset on a given set of **key variable**s. Some **disclosure risk assessment** methods are based on the probability that a sample unique might be inferred to be a **population unique**.

See also: SPECIAL UNIQUE

Sample Unit

A **data unit** within a **dataset** which itself is a sample of some **population**.

Sampling

A form of **suppression**, commonly used to produce **microdata** files from a **census**, in which only a proportion of the original **data** records on a microdata file are shared. In the context of disclosure control, sampling means that an **adversary** could not have **response knowledge** as s/he could not be certain that any **population unit** was represented in the file.

Sampling Fraction

The proportion of the relevant **population** contained within a **dataset**. With simple random sampling, the sample fraction represents the proportion of **population unit**s that are selected in the sample. With more complex **sampling** methods, this is usually the ratio of the number of units in the sample to the number of units in the population from which the sample is selected. The lower the sampling fraction, the greater the protection afforded to a **dataset**, as an **adversary** becomes less able to infer with **confidence** that a **sample unique** is a **population unique**.

Sandbox

A **security** that separates a program from the rest of the system, enabling it to operate in a regulated environment without harming other **software** or **data**. Typically, a sandbox provides a constrained setting for a programme to run in, with restricted access to system resources. Software developers may be sure that bugs in new code being tested will not disrupt the system or other apps. Similarly, to stop **malware** from infiltrating a system, any malicious code that is executed can be kept separate from the rest of the system in a sandbox and may then be quickly and safely removed. Sandboxing is also utilised for **security** in software such as web browsers and email clients.

Further reading:
Rieck, K., Holz, T., Willems, C., Dussel, P. and Laskov, P., 2008. Learning and Classification of Malware Behavior. *In:* Zamboni, D. ed., *Detection of Intrusions and Malware, and Vulnerability Assessment*, Berlin: Springer, 108–25, https://doi.org/10.1007/978-3-540-70542-0_6.

See also: DATA ENVIRONMENT, NETWORK SECURITY, RESTRICTED ACCESS

Sandboxing

See: SANDBOX

Scenario Analysis

A form of threat analysis which is designed to specify the possible **attack vector**s for a **reidentification attack** or other vector for **statistical disclosure**. This enables a more sophisticated understanding of **risk** than simply identifying theoretical vulnerabilities in the **data** themselves.

In the approach developed by Elliot and Dale, there are 12 elements to a scenario: motivation, means, opportunity, **target variable**s, goals achievable by other means, effect of **data divergence**, **attack** type, **key variable**s, likelihood of attempt, likelihood of success, consequences of attempt, effect of variations in the **data situation**.

Further reading:
Elliot, M. and Dale, A., 1999. Scenarios of attack: the data intruder’s perspective on statistical disclosure risk. Netherlands Official Statistics, 14(Spring), 6–10, www.researchgate.net/publication/343963431_Scenarios_of_attack_the_data_intruder’s_perspective_on_statistical_disclosure_risk.

See also: THREAT ANALYSIS, VULNERABILITY

Schrems

Max Schrems is an Austrian **privacy** advocate who brought the cases commonly known as *Schrems I* and *Schrems II* before the Court of Justice of the European Union. These cases resulted in the toppling of the US–EU **Safe Harbor** in 2015, and the US–EU Privacy Shield in 2020.

Scraping

The process of automatically extracting **information** from human-readable output.

Web scraping (from websites) is used for **data collection**, market research, price monitoring and other purposes. Some webpages may have specific policies that deny collecting **data** in an automated way.

Screen scraping takes **information** from a visual display and converts its analogue form into digital. It is important to follow ethical guidelines when scraping.

Scrutiny

Scrutiny is a thorough examination or detailed observation of something. While it may only imply the focused attention of legal or scientific research, scrutiny in a public context may be **privacy**-threatening. Socially, scrutiny is often for the purpose of detecting past error or rule-breaking, or proactively to ensure that mistakes are not made or rules are not broken (for instance, in the workplace, or in a law enforcement context).

Scrutiny is usually targeted at a specific object, such as a person, organisation, document or process. **Transparency** is often couched as willingness to be scrutinised (to be 'open to scrutiny'). Where people's attitudes, contacts and behaviour are under scrutiny, then their **privacy** is clearly under threat. **Public figure**s face pressures from voters and the media to be open to scrutiny, making it hard to balance their privacy with the demands of democratic **transparency**.

See also: ATTENTIONAL PRIVACY

SDC

See: STATISTICAL DISCLOSURE CONTROL

SDL

See: STATISTICAL DISCLOSURE LIMITATION

SDLC

See: SOFTWARE DEVELOPMENT LIFECYCLE

Search

In general, searching is the process of looking for something. As such, it has an immediate negative effect on **privacy**, and the US government's ability to search its citizens and their property is limited by the Fourth Amendment to its Constitution against **unreasonable search**es and seizures.

More specifically, search has become a leading technology in the navigation of the **World Wide Web**, a general **information** retrieval (IR) technique for accessing relevant documents and **data** placed on the Web. Search is driven by a *query*, such as a question or a key word, which acts as input for an information retrieval **algorithm**. The output of a search is a list of webpages ranked in order of estimated relevance, also known as *search results*.

Search engines are **software** services that perform search. Typically a search engine will store information in a *search index* or *Web index* for fast retrieval, to produce real-time answers to queries. Text or machine-readable files are the easiest to index, whereas media such as audio and video rely on **metadata** and tags. The index is periodically updated by a *Web crawler*, a **bot** that traverses the Web to find new pages and resources to index. The parts of the Web that are not accessible to crawlers are collectively known as the **Deep Web**, which has a deliberately covert part called the **Dark Web**; the Deep Web therefore does not appear in search results.

Search engines affect privacy in several ways. First, they make it simple to find information about individuals, even if the information is outdated or false, as long as it has remained online; Web search was the original impetus for the right to be forgotten following *Google Spain v AEPD* (2014). Google has even been used to catch criminals. Whereas complex information environments facilitate privacy through **obscurity**, search reduces the complexity and makes information simpler and cheaper to find. However, information can remain somewhat obscure if the retrieved page does not appear in the first two or three pages of search results, since few search engine **user**s scroll down that far.

Second, search companies improve their engines' performance by gathering data about their users; this data can also be used to target advertisements to them, providing an income stream for what would otherwise be a free service. Search data is very valuable for advertisers, as it shows what is of interest to the user at the very point at which the advert is served. This is **personal data**, and may include query history, **location data** and clickthroughs. These can be very disclosive, and many people have had their query history brought up as evidence against them in criminal inquiries. Furthermore, many search companies provide other services too (for instance, Google provides email and video sharing), which can generate more data for the profile of the user.

Some search engines, called *anonymous* or *private* search engines, make a virtue of collecting and/or retaining as little personal data as possible. **Onion routing** is a means of disguising the **identity** of the searcher. **Do Not Track** measures are sometimes proposed, either in regulations, or at one point in a failed Web protocol, to allow people to opt out of data collection from search and other browsing behaviour.

Further reading:

Tene, O., 2008. What Google knows: privacy and Internet search engines. *Utah Law Review*, 4, 1433–92. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1021490.

See also: RIGHT TO BE FORGOTTEN, TARGETED ADVERTISING, TOR, WEB PROFILING

Search Engine

A **software** service that performs web **search**es.

Seclusion

Seclusion is a state of being removed from other people. It can apply to an individual, an intimate group or even a corporate group (as, for example, with medieval monks and nuns, whose vows included a life of seclusion).

Seclusion has been used in the literature as an important aspect of **privacy**. For instance, Anita Allen criticised views of privacy as primarily informational by citing people's desires for seclusion, among other things – a type of privacy that is non-informational. The **right to be let alone**, posited by Warren and Brandeis, aimed among other things to protect the seclusion of individuals or households.

Further reading:
Allen, A.L., 1988. *Uneasy access*: *privacy for women in a free society*. Totowa: Rowman & Littlefield.
Webb, D., 2007. *Privacy and solitude*. London: Hambledon Continuum.

See also: INFORMATIONAL PRIVACY, INTIMACY, INTRUSION UPON SECLUSION, SOLITUDE

Secondary Data

Usually, **data** that have been collected with the primary intention that they will be used by third parties. This is commonly used in the context of social research in respect **census** and survey data where an organisation such as a national statistical institute may collect data with the intent of making them available to researchers. These types of secondary use are normally consented – although censuses themselves are not consented.

The term secondary data can also sometimes be used to refer to any data that is undergoing a **secondary use**.

See also: INTENTIONAL DATA, PRIMARY DATA

Secondary Differentiation

A tactic adopted by an **adversary** engaged in a **reidentification attack** which allows the **adversary** to distinguish between multiple candidate linkages between **data unit**s and **population unit**s.

For the situation where multiple data units linked to a single population unit, this will involve the **adversary** identifying variables where the **data** units differ and then targeting resources at establishing the value of those variables for the population unit. For a single data unit matched against multiple population units, this involves identifying which of the population units matches the data unit on variables not included in the original **key variable** set.

Secondary Use

The reuse of **data** collected for one purpose for a different purpose.

Secondary use can reduce **data subject autonomy** and impact on **privacy** in ways that cannot necessarily be foreseen, and specifically raise the **risk** of going beyond the **data subject**s' original **consent** (where consent has been obtained), or their reasonable expectations of privacy. The EU's **GDPR** regulates secondary uses of **personal data** through the **purpose limitation** principle. This is based on a longstanding privacy principle, which Solove dates from a 1973 US Department of Health, Education and Welfare report into the dangers of State-aggregated **database**s being used for additional, unconsented purposes.

Further reading:

Solove, D.J., 2006. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477–564, https://scholarship.law.upenn.edu/penn_law_review/vol154/iss3/1/.

See also: SECONDARY DATA

Secrecy

See: SECRET

Secret

A secret is (usually) a piece of **information** that is allowed to circulate around an in-group, but non-members of the group are to be prevented from receiving it. The practice of keeping secrets is called *secrecy*. The in-group may be as small as a single person.

When secrets are information about individuals, then secrecy has a strong relation to **privacy**. However, not all secrets have this property. *State secrets* are information whose circulation governments wish to suppress; the highest level of secrecy is often referred to as *top secret*. *Trade secrets* are **intellectual property** of companies that have economic value precisely because they are not widely known to the **public**, including competitors, and which are not disclosed to outsiders (e.g., to employees or business partners) without legal restrictions in place, such as **confidentiality** agreements. Other forms of intellectual property, such as copyrighted material, can be published more widely, with reliance on property law to regulate their use (as opposed to patents or copyrighted intellectual property, which are disclosed with legal restrictions on their use by outsiders). Religious or esoteric rites are often secret or performed secretly.

In **cybersecurity** a secret refers to confidential **data** that is protected from unauthorised access. This can take the form of a **password**, an access token or cryptographic keys.

Further reading:
Horn, E., 2011. Logics of political secrecy. *Theory, Culture and Society*, 28(7–8), 103–22, https://doi.org/10.1177/0263276411424583.

See also: CRYPTOGRAPHY, PUBLICITY, SECURITY

Secret Ballot

In a democratic voting system, a **secret** ballot is such that no one can observe how individual voters voted in elections or referendums. The aim of this is to reduce the possibility of intimidation, bribery and illicit influence. However, some early liberals (John Stuart Mill, for instance) argued that the secret ballot took **privacy** too far, because a voter’s duty was to think of the **public interest**, not self-interest, and this could only be verified by a **public** vote. Typical methods of ensuring secrecy include voting on pieces of paper in screened voting booths, voting by machine and postal voting. Postal voting has its critics, however, as it means that the voting process itself is not overseen by any officials. One member of a household may be able to observe or influence other members' votes.

Further reading:

Teorell, J., Ziblatt, D. and Lehoucq, F., 2017. An introduction to special issue: the causes and consequences of secret ballot reform. *Comparative Political Studies*, 50(5), 531–44, https://doi.org/10.1177/0010414016641977.

See also: IDEOLOGICAL PRIVACY

Secret Sharing

Secret sharing is a method used in multi-party computation to ensure input **privacy**, that is, to enable machine learning from multiple **data** sources without the analyst having access to, or gaining knowledge of, the data sources being analysed.

In effect, important **information** is split between different intermediate analytical agents so that none of them has access to enough data to break **confidentiality**, but they jointly can compute the same output as an **algorithm** applied directly to all the input data.

Secret sharing can be more or less secure. For it to be totally secure, an **adversary** must be unable to infer anything about the input data from any incomplete combination of the shares; they must have access to all the shares to gain any information at all. At a lower level of **security**, it may be that the **adversary** can gain partial knowledge of the input data from combining a fraction of the shares. Secure secret sharing also requires that the intermediate analytical agents do not collude with each other.

Further reading:

Cramer, R., Damgård, I.B. and Nielsen, J.B., 2015. *Secure multiparty computation and secret sharing*. New York: Cambridge University Press.

Ricciato, F., Bujnowska, A., Wirthmann, A., Hahn, M. and Barredo-Capelot, E., 2019. A reflection on privacy and data confidentiality in Official Statistics. *Presented at 62nd ISI World Statistics Conference*, https://ec.europa.eu/eurostat/cros/content/reflection-privacy-and-data-confidentiality-official-statistics-0_en.

See also: INFORMATION SECURITY, SECURE MULTI-PARTY COMPUTATION, VERIFIABLE SECRET SHARING

Secure Communication

See: COMMUNICATION PRIVACY, COMMUNICATION

Secure Messaging

A server-based approach using **communication** platforms and protocols that prioritise the protection of sensitive **information** when exchanging messages. Strong **authentication** authenticates the **user**. Messages are transmitted from a message centre over an **SSL** connection and/or protected by equally secure measures. When the communication is established for the first time, an authentication using a message unlock code (MUC) is needed. Furthermore, encrypted messaging may be used by any email programs without requiring the installation of additional **software**.

Secure messaging is not perfect, **trust** in the service providers is required for the system to work and **risk**s remain; potential user behaviour risks including improper key management, concerns about government **surveillance** and **backdoor**s.

Further reading:

Unger, N., Dechand, S., Bonneau, J., Fahl, S., Perl, H., Goldberg, I. and Smith, M., 2015, May. SoK: secure messaging. *In: 2015 IEEE symposium on security and privacy*, IEEE, 232-249, https://doi.org/10.1109/SP.2015.22.

Kobeissi, N., Bhargavan, K. and Blanchet, B., 2017. Automated verification for secure messaging protocols and their implementations: a symbolic and computational approach. *In: IEEE European symposium on security and privacy*, 435–50, https://doi.org/10.1109/EuroSP.2017.38.

See also: ENCRYPTION, ENCRYPTION ALGORITHM, SECURE COMMUNICATION

Secure Multi-Party Computation

A cryptographic method, Secure Multi-Party Computation (SMPC) enables many parties to calculate a function or carry out a computation without disclosing their **secret** inputs to one another. Each participant in SMPC makes a private input, and the objective is to calculate a function of these inputs whilst maintaining **privacy**. SMPC uses **cryptographic protocol**s which let each party encrypt their private input so that only authorised parties may decode and utilise the **data**. Without disclosing any private **information** to one another, the parties then exchange their encrypted inputs and employ cryptographic methods to compute the required function on the encrypted data.

Further reading:

Chuan, Z., Shengnan, Z., Minghao, Z., Zhenxiang, C., ChongZhi, G., Hongwei, L. and Yuan, T., 2019. Secure Multi-Party Computation: theory, practice and applications. *Information Sciences*, 476, 357–72, https://doi.org/https://doi.org/10.1016/j.ins.2018.10.024.

See also: CRYPTOGRAPHY, INFORMATION SECURITY, INPUT PRIVACY

Secure Sockets Layer (SSL)

A web server and a client browser can connect securely using the Secure Sockets Layer (SSL) **cryptographic protocol**. To protect **data** sent between the server and client, SSL employs a mix of public-key and **symmetric key encryption**. The server displays its **digital certificate** to the client to authenticate its **identity** when a **user** connects to a website using SSL. **Data** is communicated securely between the server and client after the client has confirmed the **identity** of the server and a safe, encrypted connection has been established. The more recent **Transport Layer Security (TLS)** protocol, which offers comparable **encryption** and **authentication** procedures, has mostly supplanted SSL.

Further reading:

Kiljan, S., Simoens, K., Cock, D.D., Eekelen, M.V. and Vranken, H., 2016. A survey of authentication and communications security in online banking. *ACM Computing Surveys*, 49(4), 1–35, https://doi.org/10.1145/3002170.

Secure Web Gateway (SWG)

A **security** mechanism created to protect **user**s and organisations from web-based dangers including **malware** and **phishing**. Typically, an SWG mediates between the user’s computer and the **Internet**, monitoring Web traffic in real time and filtering out any hazardous **data**.

An SWG may incorporate multiple **security** capabilities such as URL filtering, content filtering**, anti-virus software** and anti-malware scanning. Advanced threat detection and response capabilities, such as **sandbox**ing, behavioural analysis and **machine learning**-based security, may also be offered by SWGs.

Further reading:

Akhawe, D., Barth, A., Lam, P.E., Mitchell, J. and Song, D., 2010. Towards a formal foundation of web security. *In*: *23rd IEEE Computer Security Foundations Symposium*, 290–304, https://doi.org/10.1109/CSF.2010.27.

Secure Web Platform

A Web-based program or service that has been built with **security-by-design** principles. **Security** features and controls will be included at every tier of the application stack, including the Web server, application server, **database**, and the **user** interface; features like **encryption**, **access control**, **authentication** and **authorisation**, **data** validation and audit logging are likely to be present. These measures are intended to guard against web-based attacks such as **SQL injection**, **cross-site scripting** and **cross-site request forgery**.

Further reading:

Akhawe, D., Barth, A., Lam, P.E., Mitchell, J. and Song, D., 2010. Towards a formal foundation of web security. *In*: *23rd IEEE Computer Security Foundations Symposium*, 290–304, https://doi.org/10.1109/CSF.2010.27.

See also: SECURE WEB GATEWAY

Security

Security is the protection against or reduction of vulnerabilities in socio-technical systems. Many security systems function to protect **privacy** either directly or indirectly. **Information security** involves holding **information** in such a way that unauthorised people cannot gain access to it or otherwise violate its **integrity**, while **cybersecurity** is a similar concept applied to computer systems. **Secure communication**s cannot be intercepted by eavesdroppers. *Home security* and *corporate **security*** describe systems for protecting private property. **National security** is an ecosystem of policies, systems and organisations for protecting sovereign states.

While both privacy and security are – all else being equal – good things which often support one another, their goals are not identical, and they can also work against each other. For instance, **personal data** that is excessive, collected without **consent**, outdated or used for arbitrary purposes may be securely held behind state-of-the-art **firewall**s, but still breaches privacy norms and **data protection** law. A welfare delivery system may be designed to secure the money or benefits to be disbursed against fraudulent claims, requiring privacy-invasive measures to monitor the behaviour of its clients. Ubiquitous **CCTV** cameras may deter crime, leading to greater security in public places. Conversely, much privacy policing is norm-based, and has little or no security input. A sign saying **Private Property** may be sufficient to protect land from **intrusion**, but nothing prevents an intruder simply strolling past it.

Further reading:

Anderson, R., 2020. Security engineering: a guide to building dependable distributed systems. Indianapolis: Wiley.

Baldwin, D.A., 2018. The concept of security. *In: National and International Security*, London: Routledge, 41–62, https://doi.org/10.4324/9781315184517.

See also: CIA TRIAD, CRYPTOGRAPHY, CYBERSECURITY, ENDPOINT SECURITY, GRADUATED SECURITY, HUMAN-CENTRED CYBERSECURITY, INFORMATION SECURITY, INTERNAL SECURITY TESTING, LAYERED SECURITY MODEL, NATIONAL SECURITY, NETWORK SECURITY, ONTOLOGICAL SECURITY, PROVABLE SECURITY, SECURITY ASSERTION MARKUP LANGUAGE, SECURITY AUDIT, SECURITY-BY-DESIGN, SECURITY INFORMATION MANAGEMENT, SECURITY PARAMETER, SECURITY POSTURE, SECURITY REQUIREMENT, SECURITY TOKEN, SECURITY-BY-OBSCURITY, SELF-CONTROL SECURITY, SEMANTIC SECURITY, TRANSPORT LAYER SECURITY, ZERO TRUST SECURITY

Security Assertion Markup Language (SAML)

An XML-based **standard** used to exchange **authentication** and **authorisation** information between parties including users, service providers and **identity providers**. **User**s may authenticate once with one system and then access many other systems without having to do so repeatedly thanks to SAML. A **security token** is a piece of **data** that includes details about the identity and privileges of a user. The service provider asks the **identity** provider for a security token whenever a user tries to access a system. A **security** token containing data about the user's identity and rights is then sent by the identity provider to the service provider.

For single sign-on (SSO) authentication across several **application**s and services, SAML is frequently utilised. **Federated identity** management, which enables users to access resources across several companies without requiring unique login **credentials** for each one, also uses SAML.

Further reading:
Hughes, J. and Maler, E., 2005. *Security assertion markup language (SAML) v2. 0 technical overview. OASIS SSTC Working Draft sstc-saml-tech-overview-2.0-draft-08*, 13, 12, https://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.html.

See also: IDENTITY MANAGEMENT, INFORMATION SECURITY

Security Audit

A **security** audit's objective is to assess an organisation's security **compliance** and spot any vulnerabilities that an **adversary** may use against it. It is presented as a report that summarises its findings, including any vulnerabilities found, and offers suggestions for fixing them. A rating or score indicating the degree of **risk** connected to the organisation's **security posture** may also be included in the audit. Organisations may use security audits as a key tool to ensure that their systems do not compromise **privacy** while still maintaining the **accuracy** and **accessibility** of their **data** and IT infrastructure.

Security-By-Design

A method involving creating software and systems from the beginning with **security** in mind. Instead of just adding security measures after the fact, security-by-design aims to build systems that are naturally safe and resistant to cyberattacks. Security considerations are incorporated into each stage of the **software development lifecycle**, from gathering requirements and designing a system through putting it into practice, testing it and deploying it. This may entail employing secure coding techniques, carrying out in-depth security testing and code reviews, and putting in place **encryption** and access restrictions to safeguard **sensitive data**.

See also: CYBERSECURITY, SECURE WEB PLATFORM

Security-By-Obscurity

A notion that relies on **secrecy** or **obscurity** rather than putting in place appropriate **security** measures to safeguard a system from attacks. The assumption that an **adversary** would be unable to compromise a system or obtain access to **information** merely because they are unaware of the precise specifics of how it is protected is usually seen as a poor approach to security. Using obscure or undocumented security measures, utilising rare or private security **protocol**s or disguising the location of critical **data** are a few examples of security-by-obscurity.

Further reading:
Mercuri, R.T. and Neumann, P.G., 2003. Security by obscurity. *Communications of the ACM*, 46(11), 160, https://doi.org/10.1145/948383.948413.

See also: SECURITY-BY-DESIGN

Security Information Management (SIM)

This process comprises gathering, collating, processing and analysing **security**-related information from relevant sources. Security Information Management aims to give enterprises a consolidated picture of their security. The **information** may be used to identify possible security problems, investigate security occurrences and pinpoint areas where an organisation's security architecture needs to be strengthened.

The most common SIM technologies are security information and event management (SIEM) systems, which gather and analyse security-related data, and security analytics engines, which employ **machine learning** and other advanced analytics techniques to find patterns and abnormalities in the data.

See also: INFORMATION SECURITY, SECURITY AUDIT

Security Parameter

A **security** parameter in **cryptography** is a numerical variable that both represents and determines how secure a cryptographic system is. A cryptographic algorithm's key size, block size, are frequently chosen based on the **security** parameter. For instance, the size of the modulus, which defines the size of the **public key** and **private key** used for **encryption** and **decryption**, serves as the security parameter in **RSA encryption**.

Security Posture

A composite term which describes the capacity of an organisation to deal with cyberthreats. This includes the protective controls that are in place to prevent or deter **attack**s, and the organisation’s ability to detect attacks when they occur. A **security** posture will be reflected throughout an organisation in policies and procedures, training of staff and the design of hardware and **software** systems and in the managerial prioritisation of security in business systems. An organisation’s security posture is one aspect of its **cyber resilience**.

Security Requirement

A defined criterion that specifies essential **security** features and capabilities of a system or **application**. Collectively, these specifications are used to ensure that the system or application satisfies the security requirements of its intended use. The requirements may be generated from security policies, legislation, regulations or industry best practice. Many security desiderata, including **confidentiality**, **integrity**, **availability**, **accountability** and non-repudiation, may be included in the scoping of the requirements.

See also: SECURITY PARAMETER

Security Token

Authenticating a **user**’s **identity** or granting access to a system or **application** can require the use of a **security** token, which can be either physical or digital. It is intended to offer another level of protection on top of a login and **password**. Security tokens might come in the form of a smart card, USB key or mobile application. They often have a microprocessor incorporated in them or some other kind of technology that creates a special code or **credential** that can be used to confirm the user’s identification. The code or credential normally has a time limit and is updated often, adding an extra degree of protection against illegal access.

In **two-factor authentication** systems, a user must provide both a password and a **security** token to access a system or application. As the user must hold the **security** token in addition to knowing the password, this offers a higher level of **security** than conventional **username** and password **authentication**.

Further reading:
Hallsteinsen, S. and Jorstad, I., 2007. Using the mobile phone as a security token for unified authentication. *In*: *2007 Second International Conference on Systems and Networks Communications*, 68, https://doi.org/10.1109/ICSNC.2007.82.

See also: IDENTIFICATION CARD

Self-Archiving

Self-archiving is the practice of an author, artist or creator putting their work online and usually giving **open access** to it. The term is most used in the context of peer-reviewed research papers but can include any type of **intellectual property**.

Further reading:
Harnad, S., 2001. The self-archiving initiative. *Nature*, 410, 1024–5, https://doi.org/10.1038/35074210.

Self-Control Security

A collection of guidelines meant to improve people's capacity to prevent **security** attacks, particularly those resulting from their own behaviour. Self-control security entails the adoption of mental and behavioural practices that reduce the likelihood of security problems. For example, this may involve using strong **password**s known only to the **user**, refraining from **risk**y online behaviour (such as downloading unfamiliar files or clicking on dubious links) and regularly updating **software**.

Further reading:
Kumru, C.S. and Thanopoulos, A.C., 2008. Social security and self-control preferences. *Journal of Economic Dynamics and Control*, 32(3), 757–78, https://doi.org/10.1016/j.jedc.2007.02.007.

See also: MENTAL CAPACITY

Self-Disclosure

An individual's voluntary sharing of **information** about themselves with an**other**, originating in – mostly Western – psychology literature, which particularly considers its importance in both **identity** formation and the management of relationships with others.

Self-disclosure can be viewed as an individual controlling their **privacy** by choosing who to reveal information about themselves to, a process which is subverted by **statistical disclosure** and other forms of **informational privacy breach**.

Social media has completely transformed the process of self-disclosure, making it both inherently more stylised and **public**.

Further reading:

Cozby, P.C., 1973. Self-disclosure: a literature review. *Psychological Bulletin*, 79(2), 73–91, https://psycnet.apa.org/doi/10.1037/h0033950.

Luo, M. and Hancock, J.T., 2020. Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current Opinion in Psychology*, 31, 110–15, https://doi.org/10.1016/j.copsyc.2019.08.019.

See also: AUTONOMY, IDENTITY MANAGEMENT, SELF

Self-Reflection

The introspective process of reflecting upon oneself, one's thoughts, feelings and actions, to promote self-awareness and personal development. Critically, this is a private endeavour, to promote **autonomy** and authenticity, which is challenged by any kind of unconsented observation. In its purest form self-reflection takes place in the person's thoughts, minimising the **risk** of observation. Self-reflection requires **psychological privacy** and this may be subverted by the development of **neurotechnology**, through which one's thoughts might in principle be read by **other**s.

Further reading:

Kupfer, J., 1987. Privacy, autonomy, and self-concept. *American Philosophical Quarterly*, 24(1), 81–9, www.jstor.org/stable/20014176.

Self-Sovereign Identity (SSI)

Identity management involves being able to single out or authenticate a specific individual. However, means to do this may require the categorisation of the individual in ways they do not like, by focusing on aspects of their identity that they may wish to suppress or de-emphasise. If a person relies on a large **identity provider**, this also creates the problem that the provider needs to gather a lot of sensitive identifying **information** centrally.

The alternative is that the person must manage their own identity, with many different codes and **password**s varying across all their service providers, leading to complexity and a fragmented online persona.

Self-sovereign identity (SSI) is the idea of allowing individuals to control the identifying information they provide. This facilitates a more unified experience, including, for instance, simpler digital payments. The technology most often suggested as appropriate for SSI is **blockchain**, where **public-key infrastructure** secures the information cryptographically. Identifying information is encrypted and stored on the blockchain, where its use can be monitored and controlled by the individual, who only gives it out when they **consent** to its use. **Personal data** does not need to be perpetually held by companies; rather, they can hold anonymised digital identifiers (DIDs) which can be authenticated via the transparent and immutable resources of the blockchain.

Further reading:

Preukschat, A. and Reed, D., eds, 2021. *Self-sovereign identity*. Shelter Island: Manning Publications.

Self, The

The self is a highly controversial topic, but broadly speaking can be thought of as the individual as revealed or experienced through its own reflective consciousness. Some, including Gavison, have argued that refusal of access to the self is definitive of **privacy**, although this has been denied by Solove as missing the point that privacy also involves allowing access to the self, for example in intimate relationships. Furthermore, theorists of postmodernism have gone so far as to deny that the individual’s self exists in any meaningful fashion, being fragmented and decentred. Whether or not this is true, there are people who suffer from disorders of the self (e.g., schizophrenics), and others who think there is no such thing (e.g., Buddhists), yet those of each group surely demand and deserve privacy.

Further reading:

Gallagher, S., ed., 2011. *The Oxford handbook of the self*. Oxford: Oxford University Press.

Gavison, R., 1980. Privacy and the limits of law. *Yale Law Journal*, 89(3), 421–71, https://doi.org/10.2307/795891.

Solove, D.I., 2008. *Understanding privacy*. Cambridge: MIT Press.

See also: INVIOLATE PERSONALITY

Semantic Security

Semantic **security** is a feature of cryptographic systems that ensures an opponent, even one with infinite computer capacity, cannot deduce anything significant about the **plaintext** message from the **ciphertext**. In other words, semantic security ensures that an **adversary** may only deduce what the ciphertext has previously revealed about the original **communication**. Block ciphers and stream ciphers are examples of randomised **encryption** techniques that are commonly used to establish semantic security.

Further reading:
Razzaq, A., Latif, K., Ahmad, H.F., Hur, A., Anwar, Z. and Bloodsworth, P.C., 2014. Semantic security against web application attacks. *Information Sciences*, 254, 19–38, https://doi.org/10.1016/j.ins.2013.08.00.

See also: INFORMATION SECURITY

Semi-Invasive BCI

A form of **Brain–Computer Interface** which involves the placement of electrodes and sensors onto the surface of the brain to detect and stimulate brain activity.

Sensitive Data

See: SENSITIVE VARIABLE, SENSITIVITY

Sensitive Variable

Information about a person more likely to cause **harm** if misused. There is no absolute criterion for what a sensitive variable is; the distinction between sensitive and non-sensitive can depend on the context. For example, one's religion might be considered as a sensitive variable in some countries and not so in others. The UK's **Data Protection** Act 1998 *did* use the term 'sensitive **personal data**', and the term sensitive data is still used more colloquially. The **GDPR** lists certain types of **data** as **special category data**, which require a higher degree of justification and safeguarding to be processed lawfully.

Note that **target variable**s are likely to be sensitive as an **adversary** will usually be targeting **information** that is interesting to them, and sensitivity and interestingness are related.

Sensitivity

A concept which indicates whether a piece of **data** has the potential to cause **harm** (most likely to the **data subject**). This is often though the revelation of personal characteristics considered to be more socially or politically delicate (e.g., ethnicity or health status).

A classification of data in some legal instruments to indicate types of data which might require additional **security** measures or pre-conditions for **data processing**. Under the EU **GDPR** these are referred to as **special category data**.

In the context of **machine learning**, sensitivity represents a model's capacity to accurately recognise positive examples. It is frequently used in binary classification situations where the desired outcome is the positive instance and everything else is the negative instance.

In **differential privacy**, sensitivity measures how much a function's output changes when its input changes. Sensitivity, in this case, is an important factor for determining the amount of noise to add to the output to achieve differential **privacy**.

Sometimes, 'sensitive' is used to mean either confidential and/or disclosive. This is best avoided as it lacks precision and causes term confusion.

See also: CONFIDENTIALITY, SENSITIVE VARIABLE

Sensor

A sensor is a device designed to detect or measure some feature or event in an environment, and to communicate its reading to an external system or agent. The output is often modelled by a mathematical *transfer function* which takes the possible range of detectable features as input. **Internet**-enabled sensors are a vital core technology underpinning the **Internet of Things**. Miniaturised sensors are particularly concerning for **privacy**.

Further reading:
Fraden, J., 2016. *Handbook of modern sensors*: *physics, designs, and applications*, 5th edition. Cham: Springer, https://doi.org/10.1007/978-3-319-19303-8.

See also: SURVEILLANCE

Serial Number

A **direct identifier**, either an integer or alphanumeric generated in a sequence so that each new instance is drawn from a series (either sequentially or randomly). This is commonly used for identification of instances of mass-produced material items. But serial numbers may also be used as a form of **pseudonymisation** of **personal data**.

As a pseudonym, serial numbers are trivially reversable to anyone who has access to the lookup table of serial numbers to meaningful identifiers and may be vulnerable to timestamp attacks (particularly if the draw is sequential). Some serial numbers are used in multiple transactions (e.g., UK National Insurance numbers and US Social Security numbers) and therefore may lead to **linkability** issues.

Service User Agreement

Like **Terms of Service**, a Service User Agreement is a common term for a legally enforceable contract between a provider and a **user** of a service. It may provide a means by which a commercial organisation communicates its collection and use of **personal data** to its customers. In a digital context, it may also contain restrictions on how a piece of **software** is used, and thus (arguably) restrict the **decisional privacy** of end-users.

Sessional Cookie

Small text files kept in a user's browser while they are online. They are frequently used to (i) preserve a **user**'s state of engagement with a website, enabling the website to remember user preferences, login **information** and other details while the user switches between pages; (ii) enhance performance; and (iii) simplify the user's interaction with the website. When a person signs out of the website or shuts their browser, sessional **cookie**s are immediately destroyed. They are therefore a safer alternative to permanent cookies, which may be used for **tracking** and can be retained for extended periods of time.

Shoulder Surfing

Shoulder surfing is the practice of gaining **information** (e.g., about passwords) by observing the behaviour of the **user** of a device (stereotypically by looking over their shoulder, so that the observer remains unobserved). Observation devices such as binoculars might be used, but shoulder surfing is characterised by its not requiring any technical knowledge or **hacking** skills.

Further reading:
Bošnjak, L. and Brumen, B., 2020. Shoulder surfing experiments: a systematic literature review. *Computers and Security*, 99, 102023, https://doi.org/10.1016/j.cose.2020.102023.

See also: ANALOGUE HOLE

SIM

See: SECURITY INFORMATION MANAGEMENT

Single Out

One understanding of **identifiable natural person** is that anyone who can be singled out – that is, differentiated from others within a **dataset** – can be identified. This is a view supported by the 2014 EU **Article 29 Working Party** guidance on **anonymisation** of **personal data**.

This view has been challenged in several publications (for example, Mourby and Mackey) and does not represent (for example) the draft guidance of the UK Information Commissioner's Office on **pseudonymisation**.

The difficulty with the 'singling out as identification' approach can be illustrated by considering a hypothetical **data**set containing just one **record**. Clearly that record will be singled out in the dataset, but equally clearly the fact that it is singled out is uninformative. How informative singling out will depend on the number of records in a dataset and the number of those records compared to the size of the reference **population**.

Technically, singling out is a synonym of the more long-standing term **uniqueness**. That concept has been subject has been given more nuanced treatment in the **statistical disclosure control** literature with, for example, the development of the notion of **special unique**ness.

Further reading:
Information Commissioner's Office, 2022. *Anonymisation, pseudonymisation and privacy enhancing technologies guidance*, https://ico.org.uk/media/about-the-ico/consultations/4019579/chapter-3-anonymisation-guidance.pdf.

Mourby, M. and Mackey, E., 2023. Pseudonyms, Profiles and Identity in the Digital Environment. *In*: van der Sloot, B. and van Schendel, S. eds, *The boundaries of data*: *technical, practical and regulatory perspectives*. Amsterdam: Amsterdam University Press.

See also: IDENTIFIED DATA, IDENTIFIABLE DATA

Single Sign-On

An approach to **authentication** often used in enterprise systems where once a **user** has passed authentication, they are able to access to all sub-systems for which they are authorised users. Often single sign-on is now combined with **multi-factor authentication**.

Singularity, The

Also referred to as the *technological singularity*, this is the hypothesis, usually attributed to Von Neumann, that technology in general, and **artificial intelligence** in particular, will surpass and then accelerate away from human intelligence, with unforeseeable consequences for human societies, **identity** and even survival. The hypothesis is hotly disputed, with some such as Ray Kurzweil going as far as predicting dates for its occurrence, and others such as Paul Allen remaining sceptical. An alternative framing of the hypothesis focuses on the merging of humans and their technological artefacts, with technologies such as **brain–computer interface**s being the forerunner of a human–computer symbiosis.

The impacts of the singularity on **privacy** are as difficult to predict as any other human concern, but it seems probable that they will be significant. Some speculations are that pervasive super-intelligent AI may effectively equate to omni-surveillance and that human beings directly connected to the Internet may subvert the meaning of individual identities and therefore privacy.

Further reading:

Allen, P. and Greaves, M., 2011. The singularity isn't near. *Technology Review*, 12, 7–8, https://www.technologyreview.com/2011/10/12/190773/paul-allen-the-singularity-isnt-near/.

Kurzweil, R., 2005. The singularity is near. *In*: Sandler, R., ed, *Ethics and emerging technologies.* London: Palgrave Macmillan, 393–406, https://doi.org/10.1057/9781137349088_26.

Schrader, D.E. and Ghosh, D., 2018. Proactively protecting against the singularity: ethical decision making in AI. *IEEE Security & Privacy*, 16(3), 56–63, https://doi.org/10.1109/MSP.2018.2701169.

Slander

See: DEFAMATION

Smart City

A city or urban area that uses advanced digital technologies to improve the quality of life of its habitants, increase the efficiency of operations and promote sustainable economic development. Smart cities often use **Internet of Things**, **communication network**s and other technologies to collect and analyse **data** on transportation systems, energy, resource consumption, traffic and other areas. Using this data, smart cities can implement solutions to solve problems and meet the needs of their habitants more efficiently.

Furthermore, smart cities can promote sustainability through the adoption of technologies such as renewable energy and **Artificial Intelligence** to manage resources. While a smart city offers undoubted advantages, they require an increasing amount of data collected by government and organisations that may raise **privacy** concerns. This constant **surveillance** can lead to paternalistic control, eroding individual privacy. Chinese smart city technology, for example, is heavily focused on surveillance, and is often branded 'safe city technology'. To counter charges of paternalism, **transparency** in data collection is desirable.

Controversy surrounds some smart city schemes. Alphabet (the parent company of Google) ran into trouble between 2018 and 2022 with its proposed plans for turning the Toronto waterfront into an advanced smart city. Public resistance fuelled by a campaign by privacy activists led to the plans being shelved. Relatively few smart city plans have been implemented at scale, and the hype may be disguising a more incremental approach to the use of technology to promote sustainability.

Further reading:

Carr, C., and Hesse, M., 2020. When Alphabet Inc. plans Toronto's waterfront: new post-political modes of urban governance. *Urban Planning*, 5(1), 69–83, www.cogitatiopress.com/urbanplanning/article/view/2519.

O'Hara, K., 2022. Digital modernity. *Foundations and Trends in Web Science*, 9(1–2), 1–254, http://dx.doi.org/10.1561/1800000031.

Zhang, K., Ni, J., Yang, K., Liang, X., Ren, J. and Shen, X.S., 2017. Security and privacy in smart city applications: challenges and solutions. *IEEE Communications Magazine*, 55(1), 122–9, https://doi.org/10.1109/MCOM.2017.1600267CM.

See also: SMART DEVICE

Smart Device

Smart devices are electronic devices that use advanced technologies to collect, process and use **data** to function in an interactive, automated and autonomous way. Smart devices are often equipped with sensors, **Internet** connectivity and **artificial intelligence algorithm**s that enable them to collect and analyse data about the environment and make decisions based on this data.

Data collected contains users' conversations, location, habits and other sensitive **information** which may be used for **targeted advertising** and **profiling**. How secure the devices are is rarely transparent and they may be open to **hacking**, leaving the user vulnerable. An important question is who the device ultimately works for: the owner, or the manufacturer/vendor; when the device is used for marketing, the answer may not be clear.

Further reading:

Dubois, D.J., Kolcun, R., Mandalari, A.M., Paracha, M.T., Choffnes, D. and Haddadi, H., 2020. When speakers are all ears: characterizing misactivations of IOT smart speakers. *Proceedings on Privacy Enhancing Technologies*, 2020(4), 255–76, https://doi.org/10.2478/popets-2020-0072.

Zheng, S., Apthorpe, N., Chetty, M. and Feamster, N., 2018. User perceptions of smart home IoT privacy. *In*: *Proceedings of the ACM on human–computer interaction*, 1–20, https://doi.org/10.1145/3274469.

See also: INTERNET OF THINGS

Smart Grid

A smart grid is an electricity grid which is configured with intelligent systems for such functions as smart metering, maximising the use of renewable power, conserving power and spreading its use away from peak periods, connecting smaller producers and mobile users (such as electric vehicles), decentralised control and allowing active input from consumers. The usual purposes of the smart grid are to make electricity usage more efficient, and particularly to aid the decarbonisation of the economy.

Privacy and **security** issues loom large in their design. The smart grid must be online for real-time distributed decision-making, and its computing infrastructure is usually envisaged as being in the cloud; hence it inherits the privacy and security concerns of **cloud computing**. As a decentralised **network** of heterogeneous devices and third-party service providers, coordinated through public **communication**s **network**s, it is vulnerable to **attack**. Furthermore, fine-grained **data** about household use of electricity can disclose sensitive matters. The requirement to be secure from theft of services implies **surveillance** of household usage, while the potential for many providers to demand the data (e.g., for billing) means **personal data** may be disseminated.

Further reading:
Liu, J., Xiao, Y., Li, S., Liang, W. and Chen, C.L.P., 2012. Cyber security and privacy issues in smart grids. *IEEE Communications Surveys & Tutorials*, 14(4), 981–97, https://doi.org/10.1109/SURV.2011.122111.00145.

See also: NETWORK SECURITY

Smart Implants

A device implanted into a human, which includes microprocessors for control, sensory processing, diagnosis, and so on. At present such implants are primarily used for medical reasons, to replace or augment a damaged body part. However, use cases of augmentation of healthy humans are likely extensions.

Because such devices can be connected to the **Internet** (and indeed users may benefit form that connectivity), some observers fear this technology is the gateway to the **Internet of people**.

See also: BRAIN–COMPUTER INTERFACE

Smart Meter

See: SMART GRID

Smishing

A **cybersecurity attack**, also known as SMS phishing, carried out via text message. A variant on **phishing**, victims are tricked into disclosing **identity information** to the **adversary**, enabling **identity theft**.

Further reading:
Yeboah-Boateng, E.O. and Amanor, P.M., 2014. Phishing, SMiShing & Vishing: an assessment of threats against mobile devices. *Journal of Emerging Trends in Computing and Information Sciences*, https://e-tarjome.com/storage/btn_uploaded/2020-09-12/1599891065_11216-etarjome%20English.pdf.

Mishra, S. and Soni, D., 2020. Smishing Detector: a security model to detect smishing through SMS content analysis and URL behavior analysis. *Future Generation Computer Systems*, 108, 803–15, www.sciencedirect.com/science/article/pii/S0167739X19318758.

SNARK

See: SUCCINCT NON-INTERACTIVE ZERO-KNOWLEDGE PROOF

Social Credit System

A social credit system is an extension of credit scoring, where **data** about someone's transactions and behaviour is used to create a measure of their general **trustworthiness** and/or pro-social attitudes. The measure could, in theory, be used to determine what services or privileges they should receive. The idea is controversial, because it may penalise behaviour that is legal, yet socially or politically disapproved of. Furthermore, if the data was crowdsourced, it might encourage citizens to snoop upon each other.

Many national and local governments have schemes that broadly meet this definition (for example, the UK's 2011 Troubled Families Programme), but the Chinese government has gone furthest in considering the possibility. Even there, there is no national system in place, but rather a thin patchwork of experimental local schemes of unproven efficacy.

Further reading:

Brusee, V., 2023. *Social credit*: *the warring states of China's emerging data empire*. Singapore: Palgrave Macmillan, https://doi.org/10.1007/978-981-99-2189-8.

Hayden, C. and Jenkins, C., 2014. 'Troubled Families' Programme in England: 'wicked problems' and policy-based evidence. *Policy Studies*, 35(6), 631–49, https://doi.org/10.1080/01442872.2014.971732.

See also: SURVEILLANCE, PROFILING, SOCIAL PROFILING

Social Engineering

Cybersecurity is often focused on technical and cryptographic means of protecting confidential **information**. However, cybersecurity systems are embedded within organisational contexts, which provide other **attack vector**s for an **adversary**. Social engineering is the use of psychological techniques to manipulate individuals into giving away important **security** details, such as passwords or answers to security questions.

A social engineering attack may, for example, involve the adversary posing as someone in authority, or an IT service provider, or an overworked and harassed secretary, and requesting enough information to grant access to a system. The adversary may befriend multiple employees and piece together a picture of how the security system works. Or they may send misleading messages (**phishing** or **spear phishing**), which may ask for verification of key information. The **attack** may be as simple as reading **password**s from sticky notes attached to a computer screen. If the target is lucrative, the adversary might invest a large amount of time in grooming or befriending a target.

Further reading:
Hadnagy, C., 2018. *Social engineering*: *the science of human hacking*, 2nd edition. Indianapolis: Wiley.

Social Genome

Closely related to the term **digital footprint**, an individual's social genome is the totality of all the **data** about them that is in the **public domain**.

Social Linked Data (SOLID)

Solid (SOcial LInked Data) is a project led by Tim Berners-Lee aimed at 're-decentralising' the **World Wide Web**. The Web was conceived by Berners-Lee, its inventor, as a decentralised **information** space, where permissionless **network**s and universalised addresses enabled free information flow. However, he has argued that, as it has grown, it has become less centralised, with more content and functions contained within 'walled gardens' such as **social network**s. These spaces are made more valuable as they grow by network effects, and, because it is hard to take **data** from them (as they eschew universal addresses and use proprietary ontologies and knowledge graphs to organise information), they are hard for **user**s to leave without sacrificing the quantity and quality of service they provide. This contains an implicit threat to **privacy** because users in effect pay for that service by consenting to its exploitation of the **personal data** that they create by using the services.

Berners-Lee's solution was the Solid project, launched in 2016. Solid's vision is to define a platform for decentralised apps using current World

Wide Web Corporation (W3C) standards. Apps are defined using linked data **standard**s and focus on the front-end service provision. Access to essential data is provided by APIs from **personal data stores** called Personal Online Datastores (PODs). A Solid user could have as many PODs as desired and would give access to apps to data stored in one or more of the PODs they owned, thereby retaining control of the data. Users could put their PODs on their own servers, or alternatively subscribe to servers that will host their PODs. Other essential platform services include **identity management**, standards for APIs, permission types, messaging, logins, and so on.

Establishing the Solid vision will require not only the infrastructure, but a flourishing ecosystem of apps to recreate the network effects that have made the walled garden model of **surveillance capitalism** so successful. To that end, Solid is accompanied by a commercial company called Inrupt, which provides developer tools, an enterprise Solid server, and other services.

Further reading:
Inrupt, 2023. Inrupt, www.inrupt.com/.
Solid, 2023. Solid. Cambridge, MA: Solid Project, https://solidproject.org/.

See also: PERSONAL INFORMATION MANAGEMENT SYSTEM

Social Network

A set of entities that are connected through social interaction.

The standard representation of a social **network** is as a web of dyadic ties between population units. Social network data is increasingly of interest to data scientists as network information may indicate dissemination patterns of **information** and influence.

Network **data** is invariably considered problematic from a **privacy** point of view. Information about who a **population unit** is connected to is likely to be both sensitive and disclosive, as even quite a small network fragment is likely to be **population unique**.

An online platform that allows users to connect with each other and share content such as posts, images and videos. Social networks are often used to maintain contact with friends and family, to discover new job opportunities and to promote personal or professional activities.

Some examples of social network platforms are Facebook, Instagram, Twitter and LinkedIn. Each platform has its own unique features and functionality, but all of them allow users to create a personal profile, follow other users’ updates and interact with them through comments, private messages or reactions to posts.

Social networks can present **risk**s to user privacy due to the amount of personal information users can share on these platforms. For example, in their basic profile, **user**s may share their name, age, location, photo and other personal information. Possibly even more important, in the process of engaging with others on the platform, each user will create posts expressing opinions, revealing their offline location and activity, their state of mind, who they are with and **other information that** can be used to track their online activities, create profiles and/or undermine their **autonomy**. Most social networks actively collect data about their interactions with users. This is typically used both to improve services to users, such as personalised **search**, and to contribute to the network's business model, such as better **targeted advertising**. Users tend to have **consent**ed to this by accepting the platform's **privacy policy**.

Specific **privacy risks** arising from engagement with social network platforms are **identity theft**, online activity **tracking**, **phishing** and unconsented sharing of **personal data**. A secondary issue is that as part of an individual's online presence, social network activity contributes significantly to each user's **digital footprint** and can therefore be leveraged as an **attack vector** by adversaries who wish to reidentify them in other data.

After several highly visible scandals, such as the Facebook–Cambridge Analytica scandal in 2018, social network platforms have become more rigorous in providing **privacy settings** options. However, in practice the major responsibility for protecting privacy still lies with users themselves.

Further reading:

Backstrom, L., Dwork, C. and Kleinberg, J., 2007. Wherefore art thou R3579X? Anonymized social networks, hidden patterns, and structural steganography. *In*: *Proceedings of the 16th international conference on World Wide Web*, 181–90, https://doi.org/10.1145/2043174.2043199.

Saeri, A.K., Ogilvie, C., La Macchia, S.T., Smith, J.R. and Louis, W.R., 2014. Predicting Facebook users' online privacy protection: risk, trust, norm focus theory, and the theory of planned behavior. *The Journal of Social Psychology*, 154(4), 352–69, https://thejsms.org/index.php/JSMS/article/view/693.

Mislove, A., Viswanath, B., Gummadi, K.P. and Druschel, P., 2010. You are who you know: inferring user profiles in online social networks. *In*: *Proceedings of the third ACM international conference on Web search and data mining,* https://doi.org/10.1145/1718487.1718519.

See also: BIG DATA, PRIVACY PARADOX, PROFILING, SOCIAL NETWORK ANALYSIS, SOCIAL PROFILING, SOCIAL STEGANOGRAPHY

Social Network Analysis

The mathematical or statistical analysis of **social network** data. Such **data** are typical represented in a graph in which the nodes are entities (individuals, organisations or other **population unit**s) and edges represent a (social) connection between entities.

Further reading:
Scott, J., 2024. *Social network analysis*. London: Bloomsbury Research Methods.

Social Profiling

Social profiling is a type of **profiling** that focuses specifically on the data generated by a person’s activities on **social media**. Social profiles are typically used to improve services to users, including not only desired services such as personalised **search**, but also those that contribute to **social network** business models, such as better **targeted advertising**.

Further reading:
Bilal, M., Gani, A., Lali, M.I.U., Marjani, M. and Malik, N., 2019. Social profiling: a review, taxonomy, and challenges. *Cyberpsychology, Behavior, and Social Networking*, 22(7), 433–50, https://doi.org/10.1089/cyber.2018.0670.

See also: PERSONALISED SERVICES

Social Steganography

See: STEGANOGRAPHY

Sock Puppet

See: FAKE PROFILE

Software

Software is a set of programs that instruct a computer or **smart device** how to perform a task.

Software for **privacy** refers to code to implement programs, applications or systems used to protect against and respond to **attack**s. Users should regularly evaluate and update their software to ensure that they are protected against the latest privacy and **security** threats. **Cybersecurity** software include **firewall**s, **anti-virus software** and software to enable **encryption**. Software is also used by adversaries (e.g., **virus**es, **trojan horse**s and other **malware** are all software).

Further reading:
Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S. and Balissa, A., 2018. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, 23, 259–89. https://doi.org/10.1007/s10664-017-9517-1.

See also: PRIVACY ENGINEERING, SECURITY-BY-DESIGN

Software Development Lifecycle (SDLC)

A systematic method of developing **software** which has several stages, such as planning, analysis, design, implementation, testing, deployment and maintenance. A systematic approach helps to ensure that the **application** complies with the appropriate **privacy** rules and laws.

SOLID

See: SOCIAL LINKED DATA

Solitude

Solitude is one of the four states of **privacy** discussed in Westin's *Privacy and Freedom*. It signifies separation from others, and freedom from observation. However, the Stoic tradition, later incorporated into medieval Christian thought, emphasised that physical solitude is not needed to withdraw into oneself for, as Emperor Marcus Aurelius put it, a quiet and untroubled retreat. Some types of solitude, for example that of Robinson Crusoe on his island, or the prisoner in solitary confinement, have negative rather than positive connotations. Furthermore, insofar as solitude requires one to be alone, it clearly rules out other desirable states of privacy, such as **intimacy**.

Further reading:

Webb, D., 2007. *Privacy and solitude*. London: Hambledon Continuum.
Westin, A.F., 1967. *Privacy and freedom*. New York: Ig Publishing.

See also: ATTENTIONAL PRIVACY, PHYSICAL PRIVACY, SPATIAL PRIVACY

Sousveillance

Sousveillance is a term invented by analogy with **surveillance**, to mean observation 'from below'. The usual direction of surveillance is of members of the **public** by those in authority, whereas sousveillance is the observation or recording of people in authority by members of the public, using mobile or wearable devices, to empower themselves and to hold authority to account.

Further reading:

Mann, S., Nolan, J. and Wellman, B., 2003. Sousveillance: inventing and using wearable computing devices for data collection in surveillance environments. *Surveillance and Society*, 1(3), 331–55, https://doi.org/10.24908/ss.v1i3.3344.

See also: ACCOUNTABILITY, TRANSPARENCY, RADICAL TRANSPARENCY, WEARABLE COMPUTING

Spam

Spam is an unwanted message, typically sent to many **user**s, usually designed to promote a service or product but can also be used to spread **malware** or enable a **phishing attack**. Spam can be spread by emails, text messages, **social media** or through **application**s.

To protect users against spam, software such as spam filters and **anti-virus software** are used to identify and block the unwanted messages prior to reaching the users. In addition, user **awareness** is important to deter them from opening attachments or clicking on unknown links in messages.

Further reading:

Jindal, N. and Liu, B., 2007. Review of spam detection. *In*: *Proceedings of the 16th international conference on World Wide Web*, 1189–90, https://doi.org/10.1145/1242572.1242759.

Spam Filter

See: SPAM

Spatial Cloaking

Statistical disclosure control applied to **location data**. Techniques include adding noise to the location and coarsening the geographical detail (e.g., changing point locations to postcodes or even larger geocodes).

Further reading:
Gruteser, M. and Grunwald, D., 2003. Anonymous usage of location-based services through spatial and temporal cloaking. *In*: *Proceedings of the 1st international conference on mobile systems, applications and services*, 31–42, www.usenix.org/publications/library/proceedings/mobisys03/tech/full_papers/gruteser/gruteser.pdf.

Spatial Privacy

When an individual has spatial **privacy**, they can deny **other**s physical access to their space. O’Hara suggests that there are two chief dimensions of spatial privacy. First, there is the protection of **personal space**, which is a mobile space positioned relative to the individual. Invasion of personal space would also include direct access to, or touching, the body. Second, there is the protection of locations which are associated with or claimed by the individual. These may include fixed points which are **private property** (the grounds of a house, for instance), but also points which are temporarily claimed by the individual, such as a restaurant table. Breaches of **attentional privacy** need not involve someone physically invading the space; they would also include contamination, such as littering on private property.

Further reading:
O’Hara, K., 2023. *The seven veils of privacy*: *how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

See also: ISOLATION, SOLITUDE

Spear Phishing

A form of **phishing attack** which is targeted at specific individuals or organisations. The email used to deliver the attack will appear to come from a trusted possibly known entity. Unlike blanket phishing, where the approach is to send out bulk emails in the hope of getting a few 'bites', spear phishing is a form of **social engineering** which is increasingly using more fine-grained **personalisation**. Having been subject to an **informational privacy breach** might make a **data subject** more likely to be targeted for spear phishing.

Special Category Data

Article 9.1 of the EU **GDPR** sets out nine 'special categories' of **personal data**, which can only be processed if the **data controller** can satisfy an additional condition listed in Article 9.2.

These types of data, which reveal **information** about a person's health, genetics, sexual orientation, religious or philosophical beliefs, and so on, map onto key aspects of fundamental human rights. Genetic **data**, for example, engages **genetic privacy**, while data revealing religious or philosophical beliefs impacts the right to **freedom of expression**. The additional safeguards required by the GDPR is thus a reflection of its respect for these fundamental human rights to **privacy** and free expression.

Further reading:
Information Commissioner's Office, 2023. *A guide to lawful basis,* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-GDPR/lawful-basis-for-processing/special-category-data/.

See also: LAWFUL BASIS

Special Unique

A **record** within a (sample) **dataset** that is unique within that dataset on a set of **key variable**s and is also unique on a subset of those key variables. Special uniques are regarded as riskier than other **sample unique**s (called **random unique**s) as they may appear unusual and be vulnerable to **spontaneous recognition** and **fishing attack**s.

Further reading:
Elliot, M.J., Manning, A.M. and Ford, R.W., 2002. A computational algorithm for handling the special uniques problem. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 493–509, https://doi.org/10.1142/S0218488502001600.

Speech Recognition

Technology that enables a program to process human speech into text; involving the use of algorithms to identify words spoken aloud and convert them into readable text, high quality versions can handle natural speech, different accents and multiple languages.

Speech recognition **software** works by breaking down spoken waveform into individual sounds, analysing each sound and transcribing them into text. Digital assistants use a form of speech recognition as a user-friendly input processing format; these are becoming increasingly sophisticated. The potential pervasiveness of speech recognition within **communication** systems and as part of managing our online interactions with organisations open the possibility of a new form of **personal data** capture.

Further reading:
Ma, Z., Liu, Y., Liu, X., Ma, J. and Li, F., 2019. Privacy-preserving outsourced speech recognition for smart IoT devices. *IEEE Internet of Things Journal*, 6(5), 8406–20, https://doi.org/10.1109/JIOT.2019.2917933.

See also: DATA CAPTURE, INTERNET OF THINGS, SMART DEVICE

Split Tunnelling

A networking technology that enables remote users to connect to both a private **network** and a public network at the same time. Only the traffic intended for the private network is routed over a secure VPN (**Virtual Private Network**) tunnel when split tunnelling is used; all other traffic is forwarded to the **Internet** directly. Split tunnelling can boost network efficiency and consume less bandwidth, but it also raises **security** issues. **Malware** or adversaries may be able to enter the private network over the open Internet connection if the remote user’s device is compromised.

See also: NETWORK SECURITY

Spontaneous Recognition

The act of reidentifying a **population unit** within a de-identified **dataset** simply by observing the dataset. Whereas most **disclosure** scenarios involve the actions of a motivated adversary, spontaneous recognition only requires a **data user**, and the presupposition is that the user has not deliberately attempted a **reidentification**. The prerequisites for a spontaneous recognition are that the individual **population unit** must have an unusual combination of a small number of attributes and must be known to the user. As an example, imagine that one's neighbour is a 16-year-old male widower. If, when exploring a dataset, one found a **record** with those attributes, one might assume – because it is such a rare combination of characteristics – that the record is that of the neighbour.

Spontaneous recognition is one of the residual risks left when **data** are stored in a **safe setting** in which access and use is highly controlled. Some, such as Ritchie, argue that the **risk** has been overstated.

Further reading:
Ritchie, F., 2017. *Spontaneous recognition*: *an unnecessary control on data access?* ECB Statistics Paper, www.econstor.eu/handle/10419/175534.

See also: SPECIAL UNIQUE

Spoofing Attack

An **attack** where an **adversary** impersonates another **user** or another device in a **network**. During this attack, the adversary manipulates the **IP address** of a packet or a message so that it appears to come from a trusted source.

An adversary can also spoof emails by manipulating the email address to look like a trusted sender. The Domain Name System can also be spoofed by redirecting users to malicious websites. In ARP spoofing, the adversary manipulates the Address Resolution **Protocol** (ARP) to redirect traffic to a different device.

Further reading:
van, D.M.J.R., Zubizarreta, X., Lukvcin, I., Rugamer, A. and Felber, W., 2018. Classification of spoofing attack types. *In*: *2018 European Navigation Conference*, 91–9, https://doi.org/10.1109/EURONAV.2018.8433227.

See also: DNS SERVER

Spyware

Spyware is created with the intention of gathering **data** from a computer system without the **user**’s knowledge or **consent**. Spyware may be downloaded and installed on a computer through email attachments, software downloads and the exploitation of holes in operating systems or Web browsers. Once activated, spyware may detect a user’s keystrokes, follow their online activity, gather sensitive **data** such as **password**s or credit card details and transfer it to a remote server without the user’s knowledge or consent. Moreover, spyware has the capacity to alter system settings, add new programs or ads and impair system performance.

Key loggers, **adware**, **tracking cookie**s and remote access **Trojan horse**s are examples of spyware. Given that spyware frequently runs in the background without the user’s **awareness** or **agreement**, spyware may be challenging to find and remove.

Further reading:

Egele, M., Scholte, T., Kirda, E. and Kruegel, C., 2008. A survey on automated dynamic malware-analysis techniques and tools. *ACM Computing. Surveys*, 44(2), 1–42, https://doi.org/10.1145/2089125.2089126.

Kirda, E., Kruegel, C., Banks, G., Vigna, G. and Kemmerer, R., 2006. Behavior-based spyware detection. *In*: *Usenix Security Symposium*, 694, http://usenix.org/events/sec06/tech/full_papers/kirda/kirda.pdf.

See also: MALWARE

SQL

See: STRUCTURED QUERY LANGUAGE

SQL Injection

SQL injection is a **vulnerability** that allows the execution of malicious SQL code on a website in which **database**s are used. The **adversary** injects the code into web forms or **search** bars which usually are created to accept standard text input from **user**s. When the injection is successful, the adversary can access and modify entries from the **data**base. To prevent SQL injection attacks, it is important to include **software** for authenticating user input.

Further reading:

Mavromoustakos, S., Patel, A., Chaudhary, K., Chokshi, P. and Patel, S., 2016. Causes and prevention of SQL injection attacks in Web applications. *In*: *Proceedings of the 4th International Conference on Information and Network Security*, 55–9. https://doi.org/10.1145/3026724.3026742.

Shar, L.K. and Tan, H.B.K., 2012. Defeating SQL injection. *Computer*, 46(3), 69–77. https://doi.org/10.1109/MC.2012.283.

See also: STRUCTURED QUERY LANGUAGE

SSI

See: SELF-SOVEREIGN IDENTITY

SSL

See: SECURE SOCKETS LAYER

Stakeholder

A stakeholder to an action is an entity who might be affected by an action.

In a **risk management** approach to **privacy**, **data protection** or **functional anonymisation**, all stakeholders must be considered in the **risk assessment**. Primary stakeholders might include the **data controller** (and their organisation) and the **data subject**s. Secondary stakeholders might include the **public** (where identification is in the **public interest**), professional bodies and **regulators** (with codes of conduct), interest groups (especially privacy campaigners), prospective third-party **data** users and the media. Secondary stakeholders may not be mentioned in legislation, but they will affect the cost–benefit and **risk** analyses.

Further reading:

Arfi, E., 2021. The basics (3/3): key stakeholders in data protection. *Medium.com*, https://medium.com/privacy-focused/the-basics-3-3-key-stakeholders-in-data-protection-ac1a6cd59a2f.

Stalking

The practice of harassing a **person** by repeated or continuous contact with them against their wishes. Commonly, this takes the form of **surveillance**, but it can also include **communication** (emailing, sending text messages or telephoning), or turning up at the same venue. In most cases, gathering **information** about the stalked person or interacting with them is secondary to the purpose of intimidation. Stalkers desire to break into the world of the stalked person, and hence take care that their actions are noticed. Sometimes, stalkers have had some relationship with their victims, including romantic ones that have been ended by the victim. However, there are other motivations, and numerous typologies of stalkers and their motives have been proposed. Mullen et al list five categories of stalker: rejected, seeking **intimacy**, incompetent, resentful, and predatory. These five categories reappear in most alternative classifications, with some authors adding additional types.

Further reading:

Chan, H.C.O. and Sheridan, L., 2020. *Psycho-criminological approaches to stalking behavior*: *an international perspective*. Hoboken, NJ: John Wiley.

Mullen, P.E., Pathé, M. and Purcell, R., 2001. Stalking: new constructions of human behaviour. *Australian & New Zealand Journal of Psychiatry*, 35(1), 9–16, https://doi.org/10.1046/j.1440-1614.2001.00849.x.

See also: CYBERSTALKING, EXTRINSIC PRIVACY, HARASSMENT, VOYEURISM

Standard

Technical standards are specifications of practices, products or services, developed by a recognised body that can be followed repeatedly in a range of contexts, but which are not compulsory. **Privacy** standards play an important role in **information security**. They provide guidance about best practice and legal **compliance**, aid **interoperability** between systems and enable privacy and **security** to be integrated into other **information** management practices. Standards, which are intended to be practical, may also help flesh out policy objectives such as **privacy-by-design**. Indeed, many standards have evolved from concerns about compliance with complex legislation.

Standards bodies have integrated into the legislative process. For instance, the European Union has a legal framework for standardisation which enables the European Commission to mandate or request European standardisation organisations to draft standards to meet stated policy objectives. As an example, in 2014 the Commission released a mandate for standards to be drafted for privacy management standards in accordance with the **Data Protection Directive**. Standards can also affect legislation: the **OECD Guidelines** of 1980 influenced a generation of **data protection** regulation.

The International Organization for Standardization (ISO) is the main global standards body, and since 1987 has had a Joint Technical Committee with the International Electrotechnical Commission to develop IT standards, called ISO/IEC JTC 1. Its Working Group 5 focuses on identity management and privacy technologies, and its statement of privacy principles is extremely influential. The National Institute of Standards and Technology (NIST) is the US's main body and has produced influential standards such as the **Advanced Encryption Standard** (AES).

The standards development process is usually long and painstaking and involves multiple stakeholders. The ISO's method includes a study period of several months to map the domain, from which a proposal is produced. If accepted, a series of working drafts is developed, published and tested against expert opinion. Once these have become stable, the process becomes less speculative and more editorial, and various senior technical committees feed in to produce a draft standard, until consensus is reached and the final standard published.

Further reading:

European Union Agency for Network and Information Security, 2018. *Guidance and gaps analysis for European standardisation*: *privacy standards in the information security context*. Brussels: ENISA, www.enisa.europa.eu/publications/guidance-and-gaps-analysis-for-european-standardisation.

ISO/IEC, 2011. *Information technology – security techniques – privacy framework*. ISO/IEC International Standard 29100, Geneva: ISO, https://standards.iso.org/ittf/PubliclyAvailableStandards/c045123_ISO_IEC_29100_2011.zip.

See also: DICOM STANDARD, ISO27001, ISO27002

Standard Contractual Clauses

Under Article 46.2(d) of the EU **GDPR**, one basis under which **personal data** may be transferred outside the European Economic Area is the implementation of standard **data protection** clauses, often called standard contractual clauses or 'SCCs'.

These contractual clauses must be drafted by a national data protection **regulator** (or **Supervisory Authority**) and approved by the European

Commission. The 2020 judgment in the ***Schrems II*** case suggested that unspecified 'supplementary measures' may be needed when contractual clauses are relied upon as a basis for transfer. Brad and Liddell, however, have defended the continuing sufficiency of SCCs without the need for further safeguards.

Further reading:

Bradford, L., Aboy, M. and Liddell, K., 2021. Standard contractual clauses for cross-border transfers of health data after Schrems II. *Journal of Law and the Biosciences*, 8(1), 1–36, https://doi.org/10.1093%2Fjlb%2Flsab007.

See also: DATA TRANSFER

Standard Model Clauses

See: STANDARD CONTRACTUAL CLAUSES

Static Key

A static key is a **secret encryption** key used in **cryptography** to continually encrypt and decode **data**, for example in **database** encryption, file **encryption** and **network communication** protocols. In symmetric encryption techniques, where the same key is used for both encryption and **decryption**, a static key is frequently employed. The usage of a static key, however, poses **security** issues since anybody who obtains the key can decrypt the **information** that was encrypted with it.

These issues can be ameliorated by employing powerful **encryption algorithm**s that can withstand **attack**s, keeping the key secure, and updating it often to ensure that any **breach** is contained. When two parties need to safely exchange static keys without disclosing them to third parties, key exchange protocols are often used.

See also: CRYPTOGRAPHIC KEY, MANUAL KEY TRANSPORT, NETWORK ENCRYPTION

Statistical Disclosure

A form of **confidentiality breach** that occurs when, through some form of statistical **matching**, an individual **data subject** is either identified within a de-identified **dataset** (**reidentification**) and/or **information** about them is revealed (**attribution/inference**).

Further reading:
Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K. and De Wolf, P.P., 2012. *Statistical disclosure control*. New York: Wiley.

See also: STATISTICAL DISCLOSURE CONTROL

Statistical Disclosure Control (SDC)

1. The practice of manipulating **data** to reduce **disclosure risk** and particularly **reidentification** risk. Common data manipulation techniques include **sampling**, variable **suppression**, **global recoding**, **microaggregation**, **record swapping**, **cell suppression**, **rounding**, **overimputation**, and **noise addition**. Some techniques, such as noise addition, are perturbative and will therefore affect the **analytical validity** of the data. Other techniques, such as global recoding, reduce the quantity of the data and therefore impact the **analytical completeness**.

   Disclosure control can either be applied to data themselves (input disclosure control), or to the analytical outputs, typically before **publication** (**output statistical disclosure control**).

   Most readily applied to structured data, SDC is a pragmatic approach that has been criticised by proponents of **formal privacy** models as providing no provable guarantee.

   Within the SDC community, Elliot et al criticise the standard SDC approach to **risk** assessment as immature because it relies on abstractions from the data without considering the context and because it contains no assessment of the impact of a **breach** just its likelihood.
2. An umbrella term for the research field which dates from the late 1980s and covers disclosure risk **assessment**, disclosure control methodologies and **data utility** impact measurement.

Further reading:
Elliot, M., Mackey, E. and O'Hara, K., 2020. *The Anonymisation Decision-Making Framework: European practitioners' guide, 2nd edition*. United Kingdom Anonymisation Network, https://ukanon.net/framework/.
Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K. and De Wolf, P.P., 2012. *Statistical disclosure control*. New York: Wiley.

See also: STATISTICAL DISCLOSURE, ANONYMISATION

Statistical Disclosure Limitation (SDL)

A synonym of **statistical disclosure control** favoured in North America.

Steganography

Steganography is the art of concealing a message within another dummy message. The **adversary** is intended to see the dummy but should not suspect the existence of the concealed message – an advantage over ordinary **cryptography**, where the encrypted message is visible, and therefore open for the adversary to try to decrypt. The concealed message may be made up of components of the dummy (e.g., every third word), or may be alongside it (e.g., written in invisible ink). In a digital context, a large file such as a video may conceal a simpler message interleaved with the original bits.

Social steganography is a variant described by Marwick and boyd in which people signal their moods on **social media** with references that would only be understood by the target group. In their research they found that teenagers would craft their Facebook posts using song lyrics chosen so that their parents interpreted their posts one way, and their friends another.

Further reading:

Katzenbeisser, S. and Petitcolas, F.A.P., eds, 2000. *Information hiding*: *techniques for steganography and digital watermarking*. Norwood, MA: Artech.

Marwick, A.E. and boyd, d., 2014. Networked privacy: how teenagers negotiate context in social media. *New Media and Society*, 16(7), 1051–67, https://doi.org/10.1177/1461444814543995.

Storage Limitation

The EU **GDPR** sets out six principles in Article 5 that must be satisfied for **personal data** to be processed lawfully. The fifth of these is the storage limitation principle.

As the name suggests, the principle requires that personal data be retained for no longer than is necessary to complete the purposes for which it was originally collected, or some compatible purpose. Personal data may be retained for 'longer periods' for the sake of scientific research. A retention policy is a common way for organisations to achieve **compliance** with the storage limitation principle.

Further reading:

Information Commissioner's Office, 2023. *A guide to the data protection principles*, https://ico.org.uk/media/for-organisations/uk-GDPR-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles-0-0.pdf.

See also: DATA PROTECTION PRINCIPLES, DATA RETENTION, DATA STORAGE, ERASURE, PURPOSE LIMITATION, RIGHT TO OBJECT

Streisand Effect

The Streisand effect concerns the unintended consequences of trying to suppress **information**. The act of **suppression** may draw more attention to the information than it would have received otherwise.

In 2003, the musical star Barbra Streisand sued an environmentalist project which documented coastal erosion in California with open access photographs of the coastline. One of their photographs happened to capture Streisand's clifftop mansion in Malibu. The environmentalists had not identified or labelled the mansion, but the case itself (which she lost) brought the photograph into public **awareness**. The photo had been downloaded a handful of times prior to the case but went viral afterwards.

In general, someone who litigates to protect their **privacy**, or against a **libel** or **slander**, **risk**s the Streisand effect by taking their grievance into the **public** arena.

Further reading:

Hagenbach, J. and Koessler, F., 2017. The Streisand effect: signalling and partial sophistication. *Journal of Economic Behavior and Organization*, 143, 1–8, https://doi.org/10.1016/j.jebo.2017.09.001.

See also: CENSORSHIP, OPEN ACCESS

Structural Zero

In a **table of counts** a structural zero denotes a cell count which is constrained to zero because of some secondary or tertiary factor. The most frequent occurring tertiary examples are in **data** about children, who cannot, for example, be married, hold a full-time job or drive a car because of legal constraints. Secondary structural zeroes derive from the definitions of the data structure itself; so, for example, we might allow a progression from married to divorced but not married to single, because single is defined within our data as meaning ‘never yet married’. Primary structural zeroes are zero counts which could theoretically be above zero, but empirical facts make them impossible (so for example scoring 1000 goals in a game of football is possible in that there is nothing in the rules that prevent it, but the reality of a game of football mean that it will never happen in practice). Primary structural zeroes tend not actually appear in data because they will be designed out.

Structural zeroes need to be accounted for when assessing **disclosure risk** as they do not cause **attribution disclosure** risk in the way that empirical zeroes do.

Structured Query Language (SQL)

Relational **database**s are most often managed and analysed using the Structured Query Language (SQL).

Further reading:

Jamison, D.C., 2003. Structured query language (SQL) fundamentals. *Current Protocols in Bioinformatics*, 9-2, https://doi.org/10.1002/0471250953.bi0902s00.

See also: BIG DATA, DATA MINING

Subject Access Request

See: DATA SUBJECT ACCESS REQUEST, RIGHT OF ACCESS

Subjective Harm

The intangible **harm**s of a **privacy** violation can be the most significant form of damage that ensues. Although this introduces an element of subjectivity in their quantification they should still be considered. The importance of subjective harm is reflected in the damages awarded for **breach** of privacy rights within the **European Court of Human Rights**. Subjective harms could include distress, damage to **reputation** and violation of moral **integrity** (i.e., of a person’s internal, emotional state and sense of personal **dignity** and self-worth).

Further reading:

van der Sloot, B., 2017. Where is the harm in a privacy violation? Calculating the damages afforded in privacy cases by the European Court of Human Rights. *JIPITEC*, 8(322), www.jipitec.eu/issues/jipitec-8-4-2017/4641.

Subliminal Advertising

In the United States, the perceived threat of subliminal advertising peaked in the 1950s, around the same time that **brainwashing** was also considered to be a psycho-political threat. The popular psychology, and associated moral panic, around subliminal messaging has been largely disproven, and replaced with concerns as to whether online **behavioural advertising** can ‘nudge’ people into a predictable course of consumption. Nudging is not necessarily conceived of as being subliminal, but rather as sufficiently pervasive to have a cumulative effect on an individual’s thought process in a way that erodes their **autonomy** and **decisional privacy**.

At least some **Internet** users have demonstrable concerns about the intrusive nature of personalised advertisements, hence the commercial availability of modern ad-blocking tools to help filter out potentially manipulative content.

Further reading:

Fullerton, R.A., 2010. ‘A virtual social H-bomb’: the late 1950s controversy over subliminal advertising. *Journal of Historical Research in Marketing*, 2(2), 166–73, https://doi.org/10.1108/17557501011042533.

See also: NUDGE THEORY

Subtraction Attack

A method of attacking aggregate **data** – in particular **tables of counts** – by removing known units (contributions) from the aggregate data to improve the inferences that can be made about units which are unknown or partially known. In tables of counts subtractions reduce the cell counts and if any cell count reaches zero then deterministic **attribution disclosure** becomes possible.

Smith and Elliot developed the **Subtraction-Attribution Probability** as a means for assessing **risk** of subtraction attacks. **Differential privacy** provides significant protection against such attacks.

Further reading:

Smith, D. and Elliot, M., 2008. A measure of disclosure risk for tables of counts. *Transactions on Data Privacy*, 1(1), 34–52, www.tdp.cat/issues/tdp.a003a08.pdf.

Succinct Non-Interactive Zero-Knowledge Proof (SNARK)

A succinct non-interactive **zero-knowledge** proof (SNARK) enables one party to demonstrate to another party that they are aware of a **secret** without disclosing any details about the secret itself. Because SNARKs are non-interactive, the prover can produce the evidence without interacting with the verifier. Moreover, because SNARKs are brief, the proof may be created and confirmed using little processing power.

Blockchain systems frequently employ SNARKs to demonstrate the legitimacy of transactions while hiding their contents, retaining **anonymity** in this way while still guaranteeing the chain’s **integrity.** Another use for SNARKs is in electronic voting systems, where it is crucial to guarantee the legitimacy of the vote without disclosing the voter’s **identity**.

SNARKs create the proof and check its correctness using advanced mathematical methods. A mix of public and **private key**s is often used to construct the proof, so that anybody having the associated **public key** may verify it.

Further reading:

Partala, J., Nguyen, T.H., and Pirttikangas, S., 2020. Non-interactive zero-knowledge for blockchain: a survey. *IEEE Access*, 8, 227945–61, https://doi.org/10.1109/ACCESS.2020.3046025.

See also: CRYPTOGRAPHIC KEY, ENCRYPTION

Super Cookie

Super **cookie**s or persistent cookies are saved on a **user**’s device and are highly challenging to remove or block. Unlike **sessional cookie**s, which are normally removed when a user clears their browsing history or shuts their browser, super cookies are designed to be more durable. They are frequently made by combining the use of Flash cookies, HTML5 local storage, and ETags.

Super cookies can be used to gather **information** about a user’s **browsing history** and other actions over a longer period of time, which has drawn criticism for their potential to be used for monitoring and **surveillance**, and the creation of richer and therefore more valuable user profiles.

Further reading:

Eckersley, P., 2010. How unique is your web browser? *In*: *Privacy enhancing technologies*: *10th international symposium, proceedings*, 10, 1–18, https://doi.org/10.1007/978-3-642-14527-8_1.

Kretschmer, M., Pennekamp, J. and Wehrle, K., 2021. Cookie banners and privacy policies: measuring the impact of the GDPR on the Web. *ACM Transactions on the Web*, 15(4), 1–42, https://doi.org/10.1145/3466722.

See also: PROFILING, SURVEILLANCE, WEB PROFILING

Supervisory Authority

Each country within the EU has a national **data protection authority** termed a **Supervisory Authority**. The **GDPR** defines a supervisory authority as an independent public authority, established by a Member State to monitor the application of the Regulation. The political independence of the Supervisory Authority from the government is paramount, which led to concern expressed by some commentators that the UK's post-Brexit reforms could undermine the autonomy of the **Information** Commissioner's Office.

Further reading:

Santatzoglou, S. and Tzanou, M., 2023. An (in) adequate data protection regime after Brexit? Bulk surveillance powers, national security and the future of EU–UK data transfers. *In*: Celeste, E. et al (eds), *Data protection and digital sovereignty post-Brexit*. Oxford: Hart Publishing.

See also: EUROPEAN DATA PROTECTION BOARD, REGULATORS

Suppression

A **statistical disclosure control** process where parts of the **data** are unavailable to the **data user**. All **metadata-level controls** (e.g., **sampling**, or **global recoding**) could be viewed as forms of suppression, but the term is more usually used to describe more targeted approaches such as **cell suppression**, the removal of outliers and **local suppression** of values within **microdata** records.

Further reading:

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K. and De Wolf, P.P., 2012. *Statistical disclosure control*. New York: Wiley.

See also: GLOBAL SUPPRESSION

Surname Attack

An **attack** on genomics **data** aimed at exploiting the patrilineal correlation between the **Y chromosome** and family names (in some cultures). The attack involves external **database**s such genealogy websites and, in most cases, will require additional **information** to be attached to the **genomics data** to have more than a marginal chance of success. It is effectively a form of **reconstruction attack** because surnames themselves are not unique and therefore do not on their own lead to **reidentification**. However, as such data expands in use and volume, surname attacks become more a more feasible **attack vector**.

Further reading:

Gymrek, M., McGuire, A.L., Golan, D., Halperin, E. and Erlich, Y., 2013. Identifying personal genomes by surname inference. *Science*, 339(6117), 321–4, https://doi.org/10.1126/science.1229566.

Samani, S.S., Elliot, M. and Brass, A., 2017. *Challenges in genomic privacy*: *an analysis of surname attacks in the population of Britain*. CMIST Working Paper 2017-03. Cathie Marsh Institute for Social Research, https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2017/Surname%20Analysis-Working%20Paper.pdf.

Surveillance

Surveillance refers to the monitoring of behaviour. Someone being monitored is referred to as being *under surveillance*, and the entity conducting the surveillance is called the *surveillant*. Recently, the verb *to surveil* has been coined (a back formation from the noun), with participles *surveilled* and *surveilling*; it is still non-standard, although appearing in some dictionaries.

The use of the term ‘surveillance’ brings with it some implications, although these need not always hold. Surveillance is usually of a particular thing – often a person or group of people, or it could be of a building, a car or a railway carriage. There is an assumption of indiscriminate observation – surveillance uncovers everything that a device or a surveillant can capture through a period of time. It generally takes place in a real-world context, to capture behaviour that is ‘natural’ or not posed (we would not usually say that a theatre audience puts the actors under surveillance). There is usually a reason for surveillance, such as crime prevention or evidence gathering. It is often covert, so that those under surveillance are unaware and will not modify their behaviour for the surveillant. An alternative model is to make the surveillance obvious, so that people self-censor, self-consciously acting within the rules (as for instance with a speed camera, which is often used to deter fast driving rather than to catch speeding drivers).

Surveillance can be performed by a human watcher, but mechanical devices, such as hidden cameras, cameras in drones, **CCTV**, bugs, **Internet** traffic monitors, telephone taps and reconnaissance satellites, are more usual (and usually cheaper). In a targeted surveillance operation, several of these devices can be used together, to create a multimodal picture of the target.

Government surveillance is usually regulated, but not in such a way as to alert the person under surveillance (often a suspected criminal, or foreign agent). The police or espionage agencies will apply to the courts for permission to surveil a target (which, if granted, is often time-limited), based on the evidence they already have of suspicious activity. In more authoritarian regimes, surveillance can be ubiquitous – notoriously in East Germany, where it was conducted by the Ministry for State Security or Stasi via an enormous network of informants. However, not all surveillance is malevolent – children or medical patients might be put under surveillance to prevent them coming to harm. Surveillance of traffic may be intended to prevent congestion. Surveillance of a **public** space may protect those in it from crime or violence.

Further reading:
Lyon, D., 2007. *Surveillance studies*: *an overview*. Cambridge: Polity Press.
Lyon, D., 2018. *The culture of surveillance*. Cambridge: Polity Press.

See also: ATTENTIONAL PRIVACY, BIG BROTHER, LIFELOGGING, PANOPTICON, SOUSVEILLANCE, SURVEILLANCE CAPITALISM

Surveillance Capitalism

A term coined by Zuboff to describe the business models of large technology companies who provide low-cost or free services in exchange for the use of **data** generated by their users. The services provided play several roles. First, they connect users in a **network**, which itself enhances the benefits they receive. Second, their use provides insight into the demands, requirements and desires of the users themselves (for instance, a **search** for a particular item reveals an interest in that item that is particularly salient at the time of search). Hence analysis of the data generated by users (their **data exhausts**) helps improve and personalise the user experience, increasing the value delivered to the users and helping grow the network. Third, the quality of the services provided brings more activity online, where it can generate more data (for instance, with digital payments superseding offline banking). Fourth, while users are not charged (much) for joining the network or using the services, the data their activities generate can be separately monetised by the company, most obviously in the creation of **targeted advertising**. Since this mode of capitalism was developed, in the 2010s, digital advertising has grown dramatically, to the detriment of other media, particularly newspapers.

The major pioneers of **surveillance** capitalism are Google and Facebook; a leading strategist of surveillance capitalism, who worked for both companies, is Sheryl Sandberg (Google 2001–8; Chief Operating Officer at Facebook/Meta Platforms 2008–22). Zuboff argues that, while these companies and others grew up to service and empower individuals and attract them to the online world, they thereby gained access to **information** about users' actions and preferences, not only allowing them to monetise the data but also giving them the tools for social prediction and even control. Many companies have surveillance capitalist models, for example via 'super-apps' which combine services in **e-commerce,** payments and banking, gaming, social **network**ing, news, entertainment, and so on, in a single platform. Such super-apps, including WeChat in China, Grab in Singapore and MyJio in India, are intended to gather rich data about users.

Zuboff also argues that this is an entirely new phase of capitalism, based on the transformation of human experience into behavioural information.

Further reading:

Zuboff, S., 2019. *The age of surveillance capitalism*: *the fight for a human future at the new frontier of power*. London: Profile.

See also: BEHAVIOURAL ADVERTISING, BIG DATA, SOCIAL NETWORK, TARGETED ADVERTISING

Swapping Key

In the **statistical disclosure control** technique **record swapping**, the swapping key is the set of variables used to match **record**s to select which records to swap.

Further reading:

Dalenius, T. and Reiss, S.P., 1982. Data-swapping: a technique for disclosure control. *Journal of Statistical Planning and Inference*, 6(1), 73–85, https://doi.org/10.1016/0378-3758(82)90058-1.

SWG

See: SECURE WEB GATEWAY

Symmetric Cryptography

See: CRYPTOGRAPHY

Symmetric Key Encryption

See: CRYPTOGRAPHIC KEY

Synthetic Data

See DATA SYNTHESIS

T

Table Redesign

A form of **statistical disclosure control** for **tabular data** where the table is reduced in size in order to aggregate cells with small cell counts (for **count data**) or a small number of contributors (for **magnitude data**).

Further reading:
Willenborg, L. and De Waal, T., 2012. *Elements of statistical disclosure control*. New York: Springer Science & Business Media, https://doi.org/10.1007/978-1-4613-0121-9.

Tabular Data

Aggregate **information** on entities presented in tables. **Data** in this form may be assumed to be intrinsically safe. However, although **reidentification attack**s are less likely to be meaningful than with **microdata**, **attribution** is still a possibility. Additional risks may also present themselves with multiple tables being released from a single underlying source **dataset** as these can be subject to **reconstruction attack**s.

See also: COUNT DATA

Tagging

To make digital **information** simpler to identify and manage, it is often helpful to tag it with certain keywords or **metadata**. This is frequently done on **social media** sites like Instagram, where users may tag their posts with hashtags to make them more visible and **search**able to other **user**s who are also interested in that subject.

Tagging has **privacy** consequences, as users effectively identify themselves and/or others when they tag images with names. Children can be at particular risk from enthusiastic parents. Tagging locations may disclose timestamped location data to an adversary and possibly details of their location, such as **security** measures, or the presence of expensive ornaments or pictures.

Further reading:
Besmer, A. and Richter, L.H., 2010. Moving beyond untagging: photo privacy in a tagged world. *In: Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 1563–72. https://doi.org/10.1145/1753326.1753560.

Target Dataset

A **dataset** in which an **adversary** attempts to identify **data subject**s.

See also: REIDENTIFICATION

Targeted Advertising

See: BEHAVIOURAL ADVERTISING

Target Variable

Used within **scenario analysis** to denote the **information** that an **adversary** wants to learn about a **population unit** or units.

TCB

See: TRUSTED COMPUTING BASE

T-Closeness

One of the ***k*-anonymity** family of **privacy model**s. *t*-closeness requires that the conditional distribution of Y given X (where X is a set of key variables) is no further than *t* from the unconditional distribution of Y. The distance between distributions is calculated using the Wasserstein (or Earth mover's) distance.

Further reading:
Li, N., Li, T. and Venkatasubramanian, S., 2007. T-closeness: privacy beyond k-anonymity and l-diversity, *In*: *23rd IEEE international conference on data engineering,* IEEE, 106–15, https://doi.org/10.1109/ICDE.2007.367856.

TCP

See: TRANSMISSION CONTROL PROTOCOL

Technical And Organisational Measures

See: APPROPRIATE TECHNICAL AND ORGANISATIONAL MEASURES

TEE

See: TRUSTED EXECUTION ENVIRONMENT

Telephone Tapping

The act of monitoring and/or recording telephone conversations without the **awareness** or **consent** of every party to the conversation. This can be achieved by intercepting the phone signal or by installing monitoring devices on the phone line or in the phone itself.

See also: SURVEILLANCE, COMMUNICATION PRIVACY, SECURE COMMUNICATION

Terms of Service

A common way for an organisation to present terms of a consumer contract to a potential customer is through terms of service. Once accepted, these terms form the basis of a binding legal contract. It is common practice for **data** collectors to require customers to accept a **privacy policy** as part of these standard terms of service. Technically, this is a misuse of the term ‘policy’, which is generally an internal document setting out **privacy** practices in an organisation and does not constitute a **consent form**.

Consent obtained through terms of service – which must be provided to receive said service – cannot be a legal basis for processing **personal data** under the EU’s **GDPR**, because this will not be seen as ‘freely given’ consent. An alternative **legal basis for processing**, such as legitimate interests or necessity for the performance of a contract, must be used instead.

See also: DATA PROCESSING, DATA PROTECTION, NECESSITY

Territorial Privacy

Territorial **privacy** is a combination of **spatial privacy** and **attentional privacy**. It is the absence of **intrusion** into one’s personal territory, either physically, or via **search** or **surveillance**. The US Constitution’s Fourth Amendment, protecting against **unreasonable search**es, is a legal protection against territorial privacy. In a digital age, territorial privacy might be involved in two ways. First, one’s online spaces may be seen as a type of territory which could be invaded by surveillance. Second, sensors and smart objects connected by the **Internet of Things** may be seen as invading one’s territory.

Further reading:

Könings, B., Schaub, F., Weber, M. and Kargl, F., 2010. Towards territorial privacy in smart environments. *In: Proceedings of the intelligent information privacy management symposium, AAAI spring symposium*, 113–18, www.aaai.org/ocs/index.php/SSS/SSS10/paper/viewPaper/1043.

Reynard, C.A., 1950. Freedom from unreasonable search and seizure: a second class constitutional right? *Indiana Law Journal*, 25(3), 259–313, www.repository.law.indiana.edu/ilj/vol25/iss3/1.

Territorial Scope

The EU’s **GDPR** has two dimensions to its scope: material and territorial. While **material scope** defines its subject matter, its territorial scope determines where the GDPR applies.

A significant innovation of the GDPR was its ‘extra-territorial effect’. This derives from Article 3 GDPR, which states that the Regulation applies:

When the **data controller** or processor responsible for the processing is established in the EU, even if the **data processing** takes place outside the EU.

When the **data subject**s are resident in the EU and are offered goods and services, or monitored in the EU, even if the responsible data controller or **data processor** is not established in the EU.

The GDPR also considers where the impact of the processing may be felt. Even when organisations do not have to follow comprehensive

privacy laws, for example in the US, they may still be caught by the GDPR through, for example, their use of **tracking cookie**s to monitor people in the EU who access their website.

Further reading:

Greze, B., 2019. The extra-territorial enforcement of the GDPR: a genuine issue and the quest for alternatives. *International Data Privacy Law*, 9(2), 109–28, https://doi.org/10.1093/idpl/ipz003.

Maelen, C.V., 2022. GDPR codes of conduct and their (extra)territorial features: a tale of two systems. *International Data Privacy Law*, 12(4), 297–315, https://doi.org/10.1093/idpl/ipac018.

See also: BRUSSELS EFFECT, JURISDICTION

Text Anonymisation

Text anonymisation is a form of **statistical disclosure control** (SDC) applied to text – either in free text fields in structured **dataset**s or in standalone documents. The process consists of two steps: first, identifying disclosive pieces of text, and then treating that text in some way, most commonly **redaction**, generalisation or **pseudonymisation**.

Note that compared to advanced SDC techniques, text anonymisation is a crude instrument and is generally recognised to be non-provable. Elements such as the context and the style in which something is written might provide background **information** to an **adversary**. For example, the use of a superficially innocuous phrase might identify the author of the text and that in turn might constrain the possible **identity** of an individual referred to.

Further reading:

Hassan, F., Domingo-Ferrer, J. and Soria-Comas, J., 2018. Anonymization of unstructured data via named-entity recognition. *In*: *Proceedings of modeling decisions for artificial intelligence*: *15th international conference*, Cham: Springer, 296–305. https://doi.org/10.1007/978-3-030-00202-2_24.

El Emam, K. and Arbuckle, L., 2013. *Anonymizing health data*: *case studies and methods to get you started*. Sebastopol, CA: O'Reilly.

See also: ANONYMISATION

TFA

See: MULTI-FACTOR AUTHENTICATION

The Onion Router

See: TOR

Therapeutic Alliance

A feature of a constructive relationship between therapist and client. It is recognised that a strong therapeutic alliance is a key determinant of whether therapy is successful. Critical to that is the client's **trust** of the therapist, underpinned by (among other things) a strong **confidentiality** assurance.

Further reading:

Horvath, A.O. and Luborsky, L., 1993. The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology*, 61(4), 561, https://psycnet.apa.org/doi/10.1037/0022-006X.61.4.561.

See also: CLIENT CONFIDENTIALITY, DUTY OF CONFIDENCE

Thermal Imaging

Thermal imaging is the practice of detecting with a thermal camera the infrared radiation emitted by an object. The result is a 'heat map' image, which (a) works even in the absence of visible light, and (b) enables the distinguishing of the hotter parts of the object from its cooler parts, or from the passive environment. As heat is often the result of activity (e.g., occupation of a house), thermal imaging may enable some inferences to be made about what is going on around or within the image.

It became prominent as a **privacy** issue with its use to gather evidence for the indoor cultivation of marijuana, which requires a relatively hot environment provided by high-intensity lamps. The US Supreme Court ruled in *Kyllo v United States* (2001) that the use of Forward Looking Infra-Red (FLIR) imaging was an **unreasonable search** in the absence of a warrant, as the police were exploring details of Kyllo's property that would ordinarily require physical **intrusion** to detect. On the other hand, in a completely opposite decision, the Supreme Court of Canada judged in *R v Tessling* (2004) that the use of FLIR did not constitute an unreasonable search, because the camera did not intrude into the suspect's home, but merely measured the radiation emanating into the public space from its walls.

Further reading:

Kerr, I. and McGill, J., 2007. Emanations, snoop dogs and reasonable expectations of privacy. *Criminal Law Quarterly*, 52(3–4), 392–432, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1302546.

See also: REASONABLE EXPECTATION OF PRIVACY, SURVEILLANCE

Third Party

In general, a third party is an actor that is involved in a situation but is not one of the principal participants and, thus, has a lesser interest. The term is used in a wide variety of different ways, dependent on the context.

The EU's **GDPR** most frequently discusses three kinds of actor: **data controller**s, **data processor**s and **data subject**s. One other category, mentioned only seven times in the text of the Regulation, is that of the 'third party'.

A third party is defined in Article 4.10 GDPR as any legal or **natural person** who is *not* a data controller, processor or subject but who is nevertheless authorised by a **data** controller (or processor) to access **personal data**. This means that they neither determine the means or purposes of the **data processing** (which would make them the controller) nor process the data according to the controller's written instructions (which would make them a processor). Neither can they be the subject of the **information** in question, as that would make them the data subject.

Other uses include trusted third party and third-party **tracker**.

Further reading:

Kollnig, K., Binns, R., Van Kleek, M., Lyngs, U., Zhao, J., Tinsman, C. and Shadbolt, N., 2021. Before and after GDPR: tracking in mobile apps. *Internet Policy Review*, 10(4), https://doi.org/10.48550/arXiv.2112.11117.

Foitzik, P., 2019. What you must know about third parties under the GDPR. *The Privacy Advisor*, https://iapp.org/news/a/what-you-must-know-about-third-parties-under-the-gdpr-ccpa/.

See also: TRUSTED THIRD PARTY

Third Party Doctrine

A US legal doctrine under which people who voluntarily give **information** to third parties have no reasonable expectation of **privacy** with respect to it.

Further reading:
Kerr, O.S., 2009. The case for the Third-Party Doctrine. *Michigan Law Review*, 107(4), 561–601, https://repository.law.umich.edu/mlr/vol107/iss4/1/.
Murphy, E., 2009. The case against the case for Third-Party Doctrine: a response to Epstein and Kerr. *Berkeley Technology Law Journal*, 24(3), 1239–53, https://lawcat.berkeley.edu/record/1122308/files/fulltext.pdf.

See also: REASONABLE EXPECTATION OF PRIVACY, THIRD PARTY

Thought Police

In George Orwell's classic novel *Nineteen Eighty-Four*, the Thought Police is the branch of law enforcement that ruthlessly detects, punishes and eliminates *thoughtcrime*, that is, the beliefs, unorthodox opinions, emotions, attachments and doubts that are not approved by **Big Brother**'s governing regime. By extension, the term is now often used metaphorically to describe those who profess outrage at deviations from accepted or prescribed thinking or language, especially on political subjects.

While Orwell's target was the way that social **interference** and sanction enforced conformity, the rise of **neurotechnology** opens the possibility that thoughts may be read directly, and therefore the potential for them to be policed.

Further reading:
Orwell, G., 1949. *Nineteen eighty-four*. London: Martin Secker & Warburg.

See also: CHILLING EFFECT, IDEOLOGICAL PRIVACY, PSYCHOLOGICAL PRIVACY

Threat Modelling

Threat modelling is an important component of **cybersecurity**, which helps organisations prevent and mitigate **security** and **privacy** risks and minimise security incidents by identifying potential vulnerabilities. This involves identifying assets, evaluating **risk**s and prioritising threats and conducting **security audit**s, possibly accompanied by **penetration testing** or other methods of **vulnerability** discovery.

Further reading:
Tatam, M., Shanmugam, B., Azam, S. and Kannoorpatti, K., 2021. A review of threat modelling approaches for APT-style attacks. *Heliyon*, 7(1), www.cell.com/heliyon/pdf/S2405-8440(21)00074-8.pdf.

See also: ATTACK, ATTACK SURFACE, DATA SITUATION AUDIT, VULNERABILITY MANAGEMENT

Threshold Rule

Also known as the *n*-rule, a rule for determining whether a statistical output is safe for release. The rule says that at least n **data-**units must have contributed to each element of an output. The rule tends to be applied either directly to **tables of counts** or indirectly in summary statistics such as means or correlations.

Further reading:
Griffiths, E., Greci, C., Kotrotsios, Y., Parker, S., Scott, J., Welpton, R., Wolters, A. and Woods, C., 2019. *Handbook on statistical disclosure control for outputs. Safe Data Access Professionals Working Group*. https://securedatagroup.files.wordpress.com/2019/10/sdc-handbook-v1.0.pdf.

See also: OUTPUT STATISTICAL DISCLOSURE CONTROL, SAFE OUTPUT

Time Bomb

A time bomb is a type of **malware** that is set to launch at a specified moment. A time bomb can be employed in a range of attacks, including destroying files, stealing **data** or rendering a specific machine or system unusable. It can be timed to go off at a certain date and time, after a predetermined number of system restarts, or following some predetermined **user** activity.

Further reading:
Crandall, J.R., Wassermann, G., De Oliveira, D.A., Su, Z., Wu, S.F. and Chong, F.T., 2006. Temporal search: detecting hidden malware timebombs with virtual machines. *ACM SIGOPS Operating Systems Review*, 40(5), 25–36, https://doi.org/10.1145/1168917.1168862.

See also: ATTACK SURFACE

Time Series

A time series is a special form of **longitudinal data** where a single measurement or observation is repeated and tagged chronologically. The **data** points can be recorded at regular intervals or episodically, and can represent any form of measurement, for example, temperature, stock prices, website traffic, transactions on a credit card or a person's weight.

To find patterns and trends, time series **data** may be examined using a variety of statistical and **machine learning** approaches, autoregression models, forecasting and anomaly detection.

Where time series **data** are about individual people, they may be challenging to anonymise though simple **statistical disclosure control**s, as the repeated observations quickly form a unique signature.

Further reading:
De Montjoye, Y.A., Hidalgo, C.A., Verleysen, M. and Blondel, V.D., 2013. Unique in the crowd: the privacy bounds of human mobility. *Scientific Reports*, 3(1), 1–5. https://doi.org/10.1038/srep01376.

See also: BIG DATA, TRAFFIC DATA

TIPS (Trust, Identity, Privacy, Security)

An acronym denoting **Trust**, **Identity**, **Privacy** and **Security**. It has become common to group these concepts together in various combinations as they are heavily interrelated, particularly in a digital context, and so have begun to define a research community. Recently, other letters have been added by some authors, with an additional P for protection and an additional S for **safety**.

Further reading:
Belanger, F., Hiller, J.S. and Smith, W.J., 2002. Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *The Journal of Strategic Information Systems*, 11(3–4), 245–70, https://doi.org/10.1016/S0963-8687(02)00018-5.

TLS

See: TRANSPORT LAYER SECURITY

Tokenisation

See: SECURITY TOKEN

Topcoding

A **statistical disclosure control** method in which values of an ordinal or continuous variable above a designated threshold are grouped to a single category. A typical example is age, where ages above, say, 94 are represented as 95+. The rationale for topcoding is that higher values of such variables are typically rarer and therefore intrinsically more disclosive.

TOR (The Onion Router)

An **open-source privacy-enhancing technology** that enables anonymous Web browsing. To prevent anybody from tracking the user's online activities or location, it operates by routing **Internet data** randomly across a complex **network** of servers run by volunteers all around the world. To preserve the **user**'s **anonymity**, each node in the TOR network only knows the **IP address** of the node that delivered the traffic to it and the IP address of the node to which it is sending the traffic to, and so the whole network needs to be monitored to establish the IP addresses of users and the Web resources they are accessing.

TOR is used to preserve the online **privacy** of Web users in authoritarian states, to avoid monitoring and to circumvent **censorship** rules, and by activists in democratic countries. It is also used for criminal **communication**s and underpins a good proportion of activity on the **dark web**.

Further reading:

Dingledine, R., Mathewson, N. and Syverson, P.F., 2004. Tor: The second-generation onion router. *In*: *USENIX security symposium*, 303–20, www.usenix.org/conference/13th-usenix-security-symposium/tor-second-generation-onion-router.

Jardine, E., 2015. The Dark Web dilemma: Tor, anonymity and online policing. *Global Commission on Internet Governance Paper Series*, 21, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2667711.

Burgess, M., 2022. How Tor is fighting – and beating – Russian censorship. *Wired*, 28 July, 2022, www.wired.com/story/tor-browser-russia-blocks/.

See also: ANONYMISING PROXY, LOCATION DATA, LOCATION TRACKING, TRAFFIC DATA

Tracing

The determination and recording of the past locations of an object or a person. Alternatively, the discovery of the current whereabouts of something (e.g., the police trying to trace the owner of a particular vehicle).

See also: LOCATION TRACKING, SURVEILLANCE, TRACKING

Tracker

A tracker is typically used to gather **data** about the online activity of a **user**. **Information** collected is likely to include **browsing history**, **search** activity, data indicating users' location, and so on, and is usually timestamped to give a timeline of the activity. Trackers are often used by advertisers to build a profile of a user, their interests and behaviour, which can then be used for multiple purposes, ranging from optimising user experience to **personalisation** of services and **behavioural advertising**. Much of the data gathered will be **personal data**.

Trackers may collect data without **informed consent**, or using **consent** given via a take-it-or-leave-it **privacy policy**, and users' data may be sold as a commodity to third-party companies. To protect their **privacy**, users can use **tracker blocker**s.

Further reading:

Mandalari, A.M., Dubois, D.J., Kolcun, R., Paracha, M.T., Haddadi, H. and Choffnes, D., 2021. Blocking without breaking: identification and mitigation of non-essential IoT traffic. *Proceedings on Privacy Enhancing Technologies*, 4, 369–88, https://doi.org/10.2478/popets-2021-0075.

See also: LOCATION DATA

Tracker Blocker

See: TRACKER

Tracking

The logging, often in real time, of the location and/or behaviour of an object or a **person**. In the physical world, this involves making a **record** of where the tracked object has been. Online, tracking usually involves logging the websites a person has visited, and the resources they have downloaded, to create a **browsing history**. Commercial organisations may wish to track their customers as they move through their website.

Further reading:

Harrison, G., Grant-Muller, S.M. and Hodgson, F.C., 2020. New and emerging data forms in transportation planning and policy: opportunities and challenges for ‘track and trace’ data. *Transportation Research Part C: Emerging Technologies*, 117, 102672, https://doi.org/10.1016/j.trc.2020.102672.

Munzert, S., Selb, P., Gohdes, A., Stoetzer, L.F. and Lowe, W., 2021. Tracking and promoting the usage of a COVID-19 contact tracing app. *Nature Human Behaviour*, 5(2), 247–55, https://doi.org/10.1038/s41562-020-01044-x.

Samarasinghe, N. and Mannan, M., 2019. Towards a global perspective on web tracking. *Computers and Security*, 87, 101569, https://doi.org/10.1016/j.cose.2019.101569.

See also: CROSS DEVICE TRACKING, CUSTOMER TRACKING, DO NOT TRACK (PROTOCOL), TRACKER, LOCATION TRACKING, SURVEILLANCE, TRACING

Traffic Data

Traffic **data** is a rich form of **metadata** which captures the volume and types of data communicated across a **network**, including the quantity of **data**, the number of packets sent and received and the origins and destinations of the traffic. Some elements of traffic data, such as **IP address**es, port numbers and timestamps, can provide **information** on how **user**s and devices behave on a network.

See also: INTERNET PROTOCOL

Transmission Control Protocol (TCP)

The Transmission Control Protocol (TCP) is an Internet protocol used for transmitting **data** over the **Internet** between a client and a server and sits alongside the **Internet Protocol** (IP) **network** layer protocol to make up the *Internet Protocol suite* (TCP/IP). It ensures that data is transmitted in a reliable way, in the correct order, and without errors by implementing congestion control and flow control. This is implemented by establishing a two-way connection between client and server, dividing data into packets (segments) and transmitting the packets with sequence numbers to allow the server to reassemble the data in the correct order. Where simpler messaging is possible, without error recovery, the **User** Datagram Protocol (UDP) is used instead of TCP. Between them, these two protocols are responsible for virtually all of the transport of data on the Internet.

The **Transport Layer Security** (TLS) protocol adds a layer of **security** on top of the TCP/IP transport protocols.

Further reading:
Bellovin, S.M., 1989. Security problems in the TCP/IP protocol suite. *SIGCOMM Computing Communications Review*, 19(2), 32–48, https://doi.org/10.1145/378444.378449.

See also: DATA FLOW

Transparency

Transparency spans broad legal, ethical and political principles, as well as more particular disclosure obligations under various **data** protection laws. One common theme across these phenomena is the underlying rationale that **interference** with an individual's **privacy** should be open to scrutiny (and, thus, potential challenge), rather than covert.

It is often asserted that informational transparency is an essential precondition for **accountability**, which in turn helps prevent corrupt practices among powerful actors. While the case for (e.g.) public **availability** of government spending data is relatively straightforward, the role of transparency in upholding personal privacy is less clear. The primary audience of the transparent **information** could be the affected **data subject**, a statutory regulator or a more heterogeneous ecosystem of interested players as part of systemic oversight.

The EU's **GDPR** focuses on obligations to provide details of **processing** to the individual data subjects. The individual is thus framed as the main beneficiary of informational transparency, which is in keeping with the individual-centred model of regulation also discussed in the context of **informed consent**. Those who argue for a more systemic model of accountability have suggested that details of personal data processing should be available to a broader set of players, who will have the time, motivation and expertise to scrutinise data controllers and hold them to account.

Another issue is that some transparency processes lead to confidentiality risks for individual data subjects. For example, the EU Clinical Trials Regulation requires drug developers to submit clinical study reports, which are then made publicly available effectively as **open data**. This allows claims about new drug safety and efficacy to be subject to wider **scrutiny**. However, the publicly available versions cannot contain any detailed information about individual study participants that would constitute their **personal data**, and as such a compromise must be struck between regulatory transparency and **data protection**.

Further reading:

Duncan, G.T., Jabine, T.B. and de Wolf, V.A. eds, 1993. *Private lives and public policies: Confidentiality and accessibility of government statistics*. Washington, DC: National Academies Press.

Vayena, E. and Blasimme, A., 2018. Health research with big data: time for systemic oversight. *The Journal of Law, Medicine and Ethics*, 46(1), https://doi.org/10.1177/1073110518766026.

See also: DATA PROTECTION PRINCIPLES, EXPLAINABLE AI

Transparency Notice

See: PRIVACY NOTICE

Transport Layer Security (TLS)

TLS is an Internet protocol designed to encrypt **data** to be transmitted between a client and a server using the **Transport Control Protocol**, ensuring **confidentiality** by **using cryptography** to implement **secure communications**. A TLS connection negotiates **encryption key**s and algorithms to be used for the **encryption** of the data. TLS helps mitigate **man-in-the-middle attack**s, by preventing an **adversary** from accessing the unencrypted data. TLS is a key component of the **Hypertext Transfer Protocol Secure**.

Further reading:
Rescorla, E., 2018. *The transport layer security (TLS) protocol version 1.3*. rfc8446, www.rfc-editor.org/rfc/rfc8446.

See also: DIGITAL CERTIFICATE, IP ADDRESS, SECURITY, DATA TRANSFER

Trapdoor

See: BACKDOOR

Trespass

Trespass covers a series of torts which defend people and their property from **intrusion**. It particularly includes trespass to land, that is, unauthorised intrusion onto land or property; trespass to the **person**, such as assault and battery, which restrict a person’s bodily **integrity**; and trespass to goods, that is, unauthorised use of someone’s **private property**.

Further reading:
Quinn, F., 2019. *Elliott & Quinn's tort law*, 12th edition. Harlow: Pearson Education.

See also: BODILY PRIVACY, PRIVACY TORT, PROPRIETARY PRIVACY, TERRITORIAL PRIVACY

Trojan Horse

A Trojan Horse (or *Trojan*) is a form of **malware** that poses as a trusted **application** to mislead **user**s into installing it on their device. Once installed, a Trojan will be able to carry out malicious tasks, such as stealing confidential **information**, editing or deleting files or opening a **backdoor** to give hackers access to the machine.

Further reading:
AbdAllah, E.G., Hassanein, H.S. and Zulkernine, M., 2015. A survey of security attacks in information-centric networking. *IEEE Communications Surveys & Tutorials*, 17(3), 1441–54. https://doi.org/10.1109/COMST.2015.2392629.

Trust

Trust is **confidence** that another **person** (or system) is *trustworthy*. In philosophy and social science, someone who trusts is a *trustor*, and a trusted person is a *trustee*. If someone is trustworthy, they must have the capabilities, willingness and incentives to act in the interests of the trustor, relative to commitments they have made to the trustor. Abstract systems and technologies can also be trusted, although this is usually expressed as trust in their *reliability*, that is, that they will meet their specifications.

The problem of trust is how, under conditions of uncertainty where future behaviour of the trustee can only be estimated, to ensure that all and only trustworthy people/systems are trusted. A trustee is usually trusted in some limited domain (for instance, trusted to supervise children but not trusted with administrative tasks). *Placing trust* in a trustee involves the trustor taking a **risk**, because they will rely on the trustee fulfilling their commitments.

An *untrustworthy* person is either unable, unwilling or not motivated to act in a trustor's interests. A would-be trustor *mistrusts* or *distrusts* a person/system if they believe that the person/system is untrustworthy. Mistrust/distrust is therefore not simply the absence of trust, but a positive judgment of untrustworthiness.

The failure of a trusted person to deliver their commitments is usually seen as fatal to trust. On a common model, trust is built up slowly as the trustee provides evidence of their trustworthiness to the trustor but can disappear immediately if the trustee fails. Empirically, this is not always the case, but **security** engineering assumes that failure to deliver security commitments is catastrophic.

Privacy and trust are often linked in the computing literature. **Data subject**s are seen as trusting others with their **personal data** – in other words, believing that the **data controller** is able, willing and incentivised to hold their **data** securely. The discipline of *trusted systems engineering* is rather misnamed since it is aimed at engineering *trustworthy* systems. It cannot be guaranteed that they will be trusted, since this depends on the external perspective of the trustor, not on the engineer's work.

Further reading:
Giddens, A., 1990. *The consequences of modernity*. Cambridge: Polity Press.
Hawley, K., 2019. *How to be trustworthy*. Oxford: Oxford University Press.
O'Hara, K., 2004. *Trust*: *from Socrates to spin*. Duxford: Icon Books.

See also: DATA ETHICS, DATA TRUST, FAIR INFORMATION PRACTICE PRINCIPLES, INFORMATION ETHICS, INFORMATION SECURITY, PRIVACY ENGINEERING

Trusted Computing Base (TCB)

The group of hardware, **software** and **firmware** elements of a computer or **network** known as the Trusted Computing Base (TCB) are essential for preserving the **security** and **integrity** of the system. The TCB implements security regulations, validating **user**s and protecting confidential **information** against unauthorised access, alteration or **deletion**. Therefore, the components of the TCB must managed and controlled to higher standards than other systems, as a **breach** of the TCB might jeopardise the entire system's security.

Further reading:
Marshall, D.A. and Michael, V.J., 1995. Trusted computing update. *Computers & Security*, 14(1), 57–68, https://doi.org/https://doi.org/10.1016/0167-4048(95)97026-7.

See also: ACCESS CONTROL, INFORMATION SECURITY, NETWORK SECURITY, TRUST

Trusted Execution Environment (TEE)

Execution environments with a heightened level of protection for important programs and **data**. TEEs often take the form of hardware implementations and offer a safe enclave where data and code may be stored and run separately from the rest of the system. Its **isolation** prevents unauthorised users or apps from accessing or tampering with the code or data.

TEEs are frequently employed in mobile devices such as smartphones to secure sensitive data such as payment **credentials** or **biometric data** or to provide secure **application** and workload execution in **cloud computing** environments. They usually offer a constrained range of **security** services, including key management, **digital signature**, **encryption** and **decryption**.

Further reading:
Sabt, M., Achemlal, M. and Bouabdallah, A., 2015. Trusted execution environment: what it is, and what it is not. *In*: *2015 IEEE Trustcom/BigDataSE/ISPA*, 57–64. https://doi.org/10.1109/Trustcom.2015.357.

See also: DATA ENVIRONMENT, SAFE SETTINGS, TRUST, TRUSTED RESEARCH ENVIRONMENT

Trusted Research Environment

A form of **data safe haven** that provides researchers with secure access to potential disclosive of **sensitive data**.

Mourby et al have argued that secure research environments can provide the **appropriate technical and organisational measures** needed to achieve **functional anonymisation**, at least from the perspective of a researcher accessing what could otherwise be **personal data** (for them).

Further reading:

NHS Digital, 2023. *Trusted Research Environment Service for England* https://digital.nhs.uk/coronavirus/coronavirus-data-services-updates/trusted-research-environment-service-for-england.

See also: DATA ENVIRONMENT, SAFE SETTINGS, TRUSTED EXECUTION ENVIRONMENT, TRUST, TRUSTED THIRD PARTY

Trusted Third Party

The term 'trusted third party' is used in both **cryptography** and in **information governance** more generally to refer to an entity who is trusted by both the discloser and recipient of **information**, who holds the **data** for the recipient to access on carefully managed terms or processes the data on behalf of the stakeholders of a larger data process. For example, trusted third parties can be used as a mechanism for linking together data which is held by different parties who cannot share data with each other.

A **data safe haven** can act as a trusted third party, particularly when one data owner makes information available to multiple recipients on a controlled access basis. An intermediary between two parties may be used to provide access to **personal data** in a secure environment, under controlled conditions. This is a means of minimising the **risk** of a **breach** of **confidentiality** or **data protection** law.

Note that a trusted third party is distinct from the term third party within the EU's **GDPR**, which means a party that is neither a **data controller**, nor a **data processor**, nor a **data subject**. In GDPR terms, a Trusted Third Party is likely to be a data processor, which uses data in accordance with the purposes and instructions of others.

See also: SAFE SETTINGS, DATA ENVIRONMENT, DATA GOVERNANCE, DATA INTERMEDIARY, DATA RECIPIENT, TRUST, TRUSTED RESEARCH ENVIRONMENT

Trustworthy Digital Identity

The reliable, secure verification of an entity's **identity** in an online context. It requires robust verification, **authentication** and **access control** processes to underpin online interactions, transactions and resources access. Trustworthiness is evidenced along multiple interrelated dimensions, covering **privacy**, **security**, **ethics**, **resilience** and **reliability**.

Further reading:
Samir, E., Wu, H., Azab, M., Xin, C. and Zhang, Q., 2021. DT-SSIM: a decentralized trustworthy self-sovereign identity management framework. *IEEE Internet of Things Journal*, 9(11), 7972–88, https://doi.org/10.1109/JIOT.2021.3112537.

See also: DIGITAL IDENTITY, IDENTITY MANAGEMENT, RELIABILITY, TRUST

Tunnel Encryption

See: VIRTUAL PRIVATE NETWORK

Two Factor Authentication (TFA)

See: MULTI-FACTOR AUTHENTICATION

U

Ubicomp

See: UBIQUITOUS COMPUTING

Ubiquitous Computing (UBICOMP)

The incorporation of computing technology into daily life with the intention of improving the imperceptibility of interactions with technology. Ubicomp amplifies the advantages of convenience technologies, increasing productivity and quality of life, but it also poses serious **privacy threat**s or, perhaps more accurately, it transforms the way we think about privacy.

Large volumes of **personal data** are often gathered and analysed as part of the intrinsic operation ubicomp systems through **sensor**s, **CCTV** and other cameras, and other embedded devices in buildings, workplaces and public spaces in the environment. Ubicomp does not distinguish between **sensitive** and mundane **information** and so it could be argued that public acceptance of Ubicomp goes hand in hand with implicit acceptance of indiscriminate **surveillance**.

Further reading:
Weiser, M., 1993. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7), 75–84, https://doi.org/10.1145/159544.159617.

See also: BIG DATA, INTERNET OF THINGS,WEB PROFILING

UK GDPR

Since its departure from the European Union, the United Kingdom has begun to refer to the UK General Data Protection Regulation (GDPR) in its domestic legislation; despite its name, this is essentially a shorthand for the **GDPR** provisions retained in national law and has no special status beyond ordinary UK statute.

Unambiguous Consent

Under Article 6 of the EU **GDPR**, **consent** must be unambiguous and **freely given** to constitute a **lawful basis** for **data processing**. Unambiguous consent is not the same as **explicit consent**, which is an additional requirement for **special category data** processed under Article 9 GDPR.

Even if consent does not need to be explicit to be unambiguous, it normally requires some positive action from the **data subject**. A pre-ticked box, for example, will not be sufficient for unambiguous consent.

Further reading:

Article 29 Working Party, 2017. *Guidelines on consent under Regulation 2016/679*, https://ec.europa.eu/newsroom/article29/items/623051.

Information Commissioner's Office, 2023. *What is valid consent?* https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/consent/what-is-valid-consent/.

Unicity

A concept related to **uniqueness** used by some in the computer science discipline to denote uniqueness on any combination of the variables that make up a **dataset**.

Uniform Resource Identifier (URI)

A Uniform Resource Identifier (URI) is a sequence of symbols that identifies a resource on the **World Wide Web**, providing a unique and persistent **identifier** that can be used across **application**s to allow concepts or objects to be referred to, and to be used to refer to the same thing in different **dataset**s. There are specific types of URI. The *Uniform Resource Locator* (URL) is the most familiar, being a means of locating a resource within a **network**'s file system (e.g. www.e-elgar.com/products/reference-and-dictionaries/). URLs are colloquially known as *Web addresses*. A *Uniform Resource Name* (URN) is a unique name for something without enabling it to be identified or located – it can therefore refer to something non-existent.

Unique and persistent resource identification was originally conceived by Tim Berners-Lee as a means of facilitating linking in hypertext. With the growth of the World Wide Web, and in particular the idea of *linked data*, URIs became a strategic requirement. If referents of data were identified using persistent URI schemes, then further **data** about those referents could be linked more easily, possibly even semi-automatically.

Hence the key privacy implication of URIs is that they facilitate **data linkage** about individuals.

Further reading:
Berners-Lee, T., Fielding, R. and Masinter, L., 2005. *Uniform Resource Identifier (URI): generic syntax*. Internet Society Network Working Group, https://datatracker.ietf.org/doc/html/rfc3986.

Unique Identifier

One or more pieces of **information** that, taken together, form a unit that can uniquely attributed to a **population unit**.

See also: DIRECT IDENTIFIER, INDIRECT IDENTIFIER

Uniqueness

The property of being unique on a given set of variables/**attribute**s. This is a fundamental concept in **disclosure risk** assessment. If a **population unit** is known to have a unique set of attributes and a **data unit** with those attributes is present in a **dataset**, then **reidentification** can occur. Uniqueness is also sometimes used to refer to the rate of such uniques in a dataset or a **population**.

Further reading:
Bethlehem, J.G., Keller, W.J. and Pannekoek, J., 1990. Disclosure control of microdata. *Journal of the American Statistical Association*, 85(409), 38–45, https://doi.org/10.2307/2289523.

See also: UNICITY

Unreasonable Search

The Fourth Amendment of the US Constitution protects its citizens from unreasonable searches and seizures by government agencies. Originally, these were understood as physical interference in property, papers or **person**s, but following the 1967 judgment *Katz v United States* it was extended to cover anywhere or anything about which someone has a **reasonable expectation of privacy** (in Katz's case, a call from a public telephone booth, but also including other **public** areas such as hotel rooms).

A **search** is a **reasonable search** if it is carried out with the subject's **consent**, or during an arrest or pursuit. It is also reasonable if the object of search is not held in private, for example if it is in plain view or out in the open (even if it is clearly indicated as **private property**), because the subject has no reasonable expectation of privacy in such cases.

If the search is not reasonable in Fourth Amendment terms, a warrant from a qualified judge or magistrate, justified by a probable cause, containing a clear description of what is to be searched or seized, is needed.

Kim has argued that 'reasonableness' is the source of most of the term's controversy. For example, an individual's right to privacy could be outweighed by the **public interest** in **security** from crime, but this calculation may be different in the case of stop-and-search, which may cause significant social friction with, for example, ethnic minority communities.

Further reading:
Kim, J.Y., 2022. What is an unreasonable search? *Oregon Law Review*, 101(1), 95–136, https://scholarsbank.uoregon.edu/xmlui/handle/1794/27923.

Untraceability

An object is untraceable if it cannot be found, **singled out** or accessed by an interested observer. The evidence that the observer already possesses of the object's **identity**, history, whereabouts, transactions, outputs, behaviour or movements is insufficient in that case to link it to any known context in the present. For example, email is untraceable if its sender cannot be **identified**; a stolen laptop is untraceable if its location (or that of its thief) cannot be discovered; a payment is untraceable if either the source account or the payee is not identifiable.

Untraceability is analogous in many ways to **anonymity**, but not identical. The identity of a criminal may be known to the observer, so they are not anonymous, but they remain untraceable in terms of location, and so cannot be arrested. Conversely, the Unknown Warrior is traceable, since his body lies in a tomb in Westminster Abbey, but he is anonymous, having been taken unidentified from the field of battle.

Further reading:
Chaum, D.L., 1981. Untraceable electronic mail, return addresses and digital pseudonyms. *Communications of the ACM*, 24(2), 84–90, https://doi.org/10.1145/358549.358563.

See also: SECURE COMMUNICATION, SECURE MESSAGING, TRACKING

U-Probability

One of the two key parameters of **probabilistic record linkage**, the probability of a pair of **record**s being of different **population unit**s (i.e., a non-match), given that they have the same value for a given variable. This can often be estimated from the **data** themselves, or from external **information**, as it will approximate to the proportion of the **population** in question that have the relevant characteristic.

Further reading:
Fellegi, I.P. and Sunter, A.B., 1969. A theory for record linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210, http://dx.doi.org/10.1080/01621459.1969.10501049.

See also: FALSE POSITIVE, *m*-PROBABILITY

URI

See: UNIFORM RESOURCE IDENTIFIER

User

A person who uses a computer, **network**, service or digital artefact. A service user often has an account on that service and is **identified** to the **system** by a **unique username**.

The term *end user* refers to the ultimate human users of some **software** system or **dataset**. They may have little or no expertise or understanding of the system, and consequently it needs to be designed in order to facilitate use. This may sometimes undermine system **security**, as **access control**s cannot be too stringent for the likely expertise of the end user, and the system may also need to be accessible on a range of devices.

Data users are usually the analysts or end users of a particular dataset. **Privacy risk**s increase with the number of data users.

Further reading:
Adams, A. and Sasse, M.A., 1999. Users are not the enemy. *Communications of the ACM*, 42(12), 40–6, https://doi.org/10.1145/322796.322806.

See also: AUTHENTICATION, DATA ENVIRONMENT, RISK–UTILITY TRADE-OFF

User-Centred Design

An approach to product and service design that emphasises consideration of **user** needs, wants and experiences throughout the design and development phases. This will often involve direct engagement with users themselves to test out product ideas or to refine designs of prototypes.

In principle, user-centred design can lead to enhanced consideration of **security** and **privacy**, although this is not definitional. User-centred design can incorporate processes that enable comprehension of the user's **privacy concern**s and expectations, so that knowledge can be used to develop a design that is open, **safe** and considerate of their privacy interests. This approach ties user-centred design to **privacy-by-design** concepts, such as **data minimisation**, acceptability and **transparency**.

Further reading:

Abras, C., Maloney-Krichmar, D. and Preece, J., 2004. User-centered design. *In*: Bainbridge, W., ed., *Encyclopedia of human-computer interaction*, Thousand Oaks: Sage Publications, 37(4), 445–56.

Ayalon, O. and Toch, E., 2021. User-centered privacy-by-design: evaluating the appropriateness of design prototypes. *International Journal of Human-Computer Studies*, 154, 102641, https://doi.org/10.1016/j.ijhcs.2021.102641.

See also: DATA-PROTECTION-BY-DESIGN, PROFILING, SECURITY-BY-DESIGN, USER MODELLING

User Modelling

A term arising from the human computer interaction community in the late 20th-century literature, which has now been mostly replaced by the synonymous term **profiling**.

Further reading:

Webb, G.I., Pazzani, M.J. and Billsus, D., 2001. Machine learning for user modeling. *User Modeling and User-Adapted Interaction*, 11, 19–29, https://doi.org/10.1023/A:1011117102175.

See also: PRIVACY-RELATED INTERACTION, USER-CENTRED DESIGN

Username

A local **unique identifier** which enables a **user** to access a **system** or service. Usually used in combination with a **password** and possibly additional means of **authentication**.

US Privacy Laws

The United States of America has been relatively slow to pass general **data protection** legislation at the federal level. Individual states, most notably California, have passed laws which regulate **personal data** irrespective of the type of **information**, or the actors involved. At the national level, the Genetic Information Non-discrimination Act of 2008 was devoted explicitly to privacy, but only **genetic privacy**, to address concerns about the power of **genomic**s **data** following the Human Genome Project.

The American Data Privacy and Protection Act represents a significant step towards general data protection in the US. However, its future is uncertain at the time of writing, as it will require the bi-partisan support in the US Senate that previous efforts have lacked. As with genetic data in 2008, the political hook on which the ADPPA is hung is (in part) the fears of the **privacy threat**s posed by **artificial intelligence**, and the inadequacy of current laws to regulate **algorithmic** opacity.

The current draft of the ADPPA only protects the data of American residents, and as such it is not anticipated that it will make it significantly easier to share personal data from the European Union (because the information of EU residents will not be captured). The Trans-Atlantic Data Privacy Framework will therefore remain a more promising avenue for the United States to secure **adequacy** status from the European Commission.

Further reading:

Kaufmann, J., Hilgert, F. and Wohlthat, R., 2022. The proposed American Data Privacy and Protection Act in comparison with GDPR: does the current US bill of the ADPPA converge towards the 'gold standard' concepts under the EU GDPR – or not? *Computer Law Review International* 23(5), 146–52, https://doi.org/10.9785/cri-2022-230505.

See also: SCHREMS

Utility First

The traditional approach to **statistical disclosure control** where the desired properties of the **data** for the intended processing are defined and instantiated first and the **disclosure risk** is only assessed once those are defined.

See also: PRIVACY FIRST

V

Value–Action Gap

See: ATTITUDE–BEHAVIOUR GAP

Value of Data

In a commercial context, **data** has value for the firms that collect it, whether about customers or products. The value may be set by commoditising the data and selling it to others. Data will also have value internally. It can be used to match consumers with products and services, to retain consumers or increase their satisfaction rate, to provide insight into patterns of consumer behaviour or the trends in the markets or to optimise business processes.

Realising this value is not always easy, especially in fast-moving markets. Many firms struggle to extract all the value, or to produce a business model for the strategic use of data. Consequently, they may collect too much **information**, on the off-chance that it might be useful in future, or may collect the wrong information, or find it difficult to integrate it into the company's workflow. To extract value effectively, **data curation** needs to be effective, so that it can be held securely and brought in to use in a timely fashion.

An additional point worth noting is that careless handling of **personal data** carries with it large business **risk**; a privacy scandal may, through loss of **reputation**, result in the data being of negative value to the firm.

Further reading:
Günther, W.A., Mehrizi, M.H.R., Huysman, M. and Feldberg, F., 2017. Debating big data: a literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191–209, https://doi.org/10.1016/j.jsis.2017.07.003.

See also: BIG DATA, COMMODIFICATION, CUSTOMER RELATIONSHIP MANAGEMENT, CUSTOMER TRACKING, DATA STEWARD, E-COMMERCE, ECONOMICS OF PRIVACY, TRACKING, VALUE OF PRIVACY

Value of Privacy

The value of privacy can be described and even measured in two different ways: (i) psycho-socially (including culturally, ethically and politically), and (ii) economically.

Along the first dimension, the value of privacy consists in the importance of the roles it plays in personal psychology and social structures. There is much debate on this topic, and many different viewpoints. One important question is whether privacy has an intrinsic value, or whether its value is contingent on the contribution(s) it makes to other important value(s). If privacy is intrinsically valuable, then we must value and protect it for its own sake, independently of its other social effects.

If its value is contingent or functional, then a further question arises as to what function(s) it performs. It may have social functional value, for instance, enabling valuable social activities such as the practice of democracy, or the operation of associations within civil society. Or its functional value may be to the individual, for example supporting human **dignity**, the construction of **identity**, **intimacy**, or **autonomy** and freedom. Its value may arise from a combination of any of these for the individual and society. The answers to these questions will determine how and whether a right to privacy ought to be claimed and supported, and how we should judge a **conflict of rights** with privacy (e.g., the right to **freedom of expression**).

Another view is that the value of privacy is dependent on some other factor, so that it *reduces* to the value of that factor (for instance, human dignity). On this reductionist view, the value of privacy is wholly derivative. Finally, we may question whether (or when) the value of privacy is positive or negative. This may depend on circumstances. Different cultures may value privacy to different degrees; less individualistic societies may be more prone to see privacy as a hindrance to social stability, for instance.

Along the second dimension, the economic value of privacy to an individual can be detected via the price they are prepared to pay for its provision. To measure this accurately will require efficient marketplaces (e.g., for **privacy-enhancing technology**) where the relevant choices can be made, and **privacy preference**s revealed.

Further reading:

O'Hara, K., 2023. *The seven veils of privacy: how our debates about privacy conceal its nature*. Manchester: Manchester University Press.

Posner, R.A., 1981. The economics of privacy. *American Economic Review*, 71(2), 405–9, www.jstor.org/stable/1815754.

See also: ECONOMICS OF PRIVACY, INFORMATION ETHICS, CULTURAL VARIATION OF PRIVACY

Value Sensitive Design (VSD)

Value sensitive design (VSD) is a style of technology design intended to incorporate human values into design, beyond those functions and constraints that must be included in the design for it to work. VSD must in general assume some context of use for a technology, and consider the **stakeholder**s, both passive and active, involved in its use. Modelling the artefact involves not only considering how it achieves its intended function, but also how it will promote or hinder the interests of the stakeholders – interests that may go beyond the function of the technology itself.

With respect to **privacy**, the most obvious application of VSD is in ensuring that a technology is as protective of privacy as possible. It is therefore a potential methodological approach to **privacy-by-design** and **data-protection-by-design**.

Further reading:
Friedman, B., 1996. Value-sensitive design. *Interactions*, 3(6), 17–23, https://doi.org/10.1145/242485.242493.

Veil

A veil is a piece of gauze or cloth that covers the face, hair or head of a **person**. It need not be intended to conceal **identity** and is often designed to prevent someone being the object of **scrutiny** or attention, to assert a religious or cultural identity or to conceal expression (e.g., to ensure the **dignity** of a grieving person in a **public** setting). It is a common form of dress in many societies and cultures. It often has a gendered role, for instance isolating women from male attention (e.g., married or 'respectable' women, or perhaps younger women on order to preserve modesty and 'purity'). In a few countries, veiling for women is compulsory. More rarely, cultures favour the veiling of men (such as the Tuaregs of the Saharan region). Veiling often has a religious basis, but may also indicate social rank, removing high status people from the view of the many.

Metaphorically, veiling is used to convey something being **mask**ed or otherwise partially concealed from scrutiny. A new product may be *unveiled* to the public. In **privacy** terms, practices such as **transparency** or journalistic investigation are often described as *lifting the veil* on some previously hidden issue.

Further reading:

El Guindi, F., 1999. *Veil: modesty, privacy and resistance*. Oxford: Berg.
Murphy, R.F., 1964. Social distance and the veil. *American Anthropologist*, 66(6 pt.1), 1257–74, https://doi.org/10.1525/aa.1964.66.6.02a00020.

See also: FEMINIST CRITIQUE OF PRIVACY, RESERVE, SECLUSION

Verifiable Secret Sharing (VSS)

A group of participants can safely exchange a **secret** value, such as a **private key**, among themselves using the cryptographic approach known as verifiable secret sharing (VSS). The secret is divided into several individually unintelligible shares and distributed across the participants. The main characteristic of VSS is that any subset of participants may verify the original secret value by combining their shares, but any collection of participants' shares that does not contain all of them does not reveal the secret. VSS is therefore a type of **secret sharing** that is secure against cheating participants.

Further reading:

Chandramouli, A., Choudhury, A. and Patra A., 2022. A survey on perfectly secure verifiable secret-sharing. *ACM Computing Surveys*, 54(11s), 1–36, https://doi.org/10.1145/3512344.
Feldman, P., 1987. A practical scheme for non-interactive verifiable secret sharing. *In*: *28th Annual Symposium on Foundations of Computer Science (sfcs 1987)*, 427–38, https://doi.org/10.1109/SFCS.1987.4.

See also: AUTHENTICATION, CRYPTOGRAPHIC KEY, MUTUAL AUTHENTICATION

Vicarious Liability

A tort law doctrine under which an employer can be held responsible for **negligence** or wrongdoing of their employees. This could include neglectful use, or deliberate misuse, of the **personal data** of others. Countries across the world have developed different rules in this regard. In South Africa, for example, employers will be the 'responsible person' regarding their employees' **breach**es of the Protection of Personal Information Act of 2013. Under Canadian case-law, an employer can be held vicariously liable for a breach of **common law** rights to privacy where they have materially contributed to the **risk** of a privacy violation.

In the UK, the Supreme Court has confirmed that liability under the (then) Data Protection Act 1998 can be attributed to employers when their employees' breach has a sufficiently close connection to their ordinary course of employment. This precedent should still apply under the new Data Protection Act 2018. Even when an employee has acted maliciously, an employer can be held liable for any poor **data governance** the disgruntled employee has exploited in their actions.

Outside of tort law, statutory **data protection** laws in the UK and EU focus on **data controller**s, rather than individuals. This means that a company (for example) is the legal **person** responsible for **compliance**, not its individual employees (to the extent that the latter are processing personal data in the course of their employment). As such, while data protection regimes do not often use the term 'vicarious liability', the definition of terms such as 'data controller' attributes liability, by default, to organisations rather than employees.

Further reading:

Bascerano, E.G. and Millard, D., 2016. Employers' statutory vicarious liability in terms of the Protection of Personal Information Act. *Potchefstroom Electronic Law Journal*, 19(1), 1–38.

Flett, E., Wilson, J. and Gover, R., 2020. Morrisons off the hook as employers welcome clarity on vicarious liability for data breach: *WW Morrison Supermarkets plc v Various Claimants* [2020] UKSC 12. *Journal of Intellectual Property Law & Practice*, 15(7), 504–6, https://doi.org/10.1093/jiplp/jpaa084.

von Tigerstrom, B., 2018. Direct and vicarious liability for tort claims involving violation of privacy. *Canadian Bar Review*, 96(3), 539–64, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3319845.

See also: DATA BREACH, DATA PROCESSING, PRIVACY TORT

Virality

Virality is the property of online **information** or content which diffuses speedily across the Internet to many **user**s in a short space of time. Such content is often sensational or quirky. Many **social network** platforms are designed so that content can easily be recommended or passed on by people and so *go viral*. It chiefly affects privacy if the viral information **breach**es it; in that case, it will be very difficult to restore the affected individual's **reputation**.

Further reading:
Sampson, T.D., 2012. *Virality: contagion theory in the age of networks*. Minneapolis: University of Minnesota Press.

See also: GOSSIP, INFORMATIONAL PRIVACY, INTERNET OF PEOPLE, RUMOUR

Virtual Machine (VM)

Multiple operating systems can operate on a single physical machine thanks to a virtual machine (VM), which is a **software** simulation of a real computer **system**. It essentially provides the functionality of one computer system while running as a program within another.

Virtual machines may be used for a variety of tasks, including the creation, testing and deployment of software as well as the execution of programs that need a specific (perhaps obsolete) operating system.

Further reading:
Smith, J. and Nair, R., 2005. *Virtual machines: versatile platforms for systems and processes*. Amsterdam: Elsevier.

Virtual Private Network (VPN)

Using a Virtual Private Network (VPN), two or more devices may connect over the **Internet**, allowing **secure communication** and **secure messaging**. By establishing a virtual tunnel between the devices, VPNs enable private and secure **data transfer**, as if the connected device were entirely within the private **network**, whereas in reality it is using the **public** data transport system to access it. It is as if the **boundary** of the private network has been shifted outward.

A **user**'s device establishes a secure connection to a VPN server, which serves as a **proxy**, when they connect to a VPN. The VPN server **encrypt**s the **data** and transfers it to its destination, receiving the user's Internet traffic. Since the data is encrypted and **secured** by the VPN, it is impossible for an external **adversary** to track the user's online activity.

People and organisations frequently use VPNs to safeguard their online privacy and security, evade Internet **censorship**, and access **information** that might be restricted in their **jurisdiction**. They may also be used to securely connect to remote networks.

Further reading:

Khanvilkar, S. and Khokhar, A., 2004. Virtual private networks: an overview with performance evaluation. *IEEE Communications Magazine*, 42(10), 146–54, https://doi.org/10.1109/MCOM.2004.1341273.

Venkateswaran, R., 2001. Virtual private networks. *IEEE Potentials*, 20(1), 11–15, https://doi.org/10.1109/45.913204.

See also: INFORMATION SECURITY, REMOTE ACCESS, RESTRICTED ACCESS, TRAFFIC DATA

Virus

A class of **malware** created to replicate and propagate from one computer to another.

A computer virus may damage a computer by corrupting or destroying files, reducing system performance and stealing confidential **data**. While some viruses are made to propagate swiftly and inflict **harm** as soon as they infect a computer, others are designed to lay dormant until activated by a certain event. Utilising **anti-virus software** and keeping it updated is the usual way to prevent infections by known viruses or those whose behaviour follows a known pattern.

Further reading:

Han, X. and Tan, Q., 2010. Dynamical behavior of computer virus on Internet. *Applied Mathematics and Computation*, 217(6), 2520–6, https://doi.org/10.1016/j.amc.2010.07.064.

Yang, L.-X. and Yang, X., 2014. A new epidemic model of computer viruses. *Communications in Nonlinear Science and Numerical Simulation*, 19(6), 1935–44, https://doi.org/10.1016/j.cnsns.2013.09.038.

See also: TIME BOMB, WORM

Vishing

See: PHISHING, SMISHING

Vital Interests

Under Article 9 of the EU **GDPR**, **special category data** (i.e., sensitive **personal data** relating to traits such as health, ethnicity or sexual orientation) can only be processed if it satisfies one of ten potential conditions. Acting in the vital interests of the **data subject**, or another **natural person**, is one such condition if the data subject is legally incapable of giving **consent**.

The GDPR recitals make it clear that a vital interest means a **risk** to life or physical **integrity**, for example in the context of epidemics, emergency medical care or natural disasters.

Further reading:

Gazi, T., 2020. Data to the rescue: how humanitarian aid NGOs should collect information based on the GDPR. *Journal of International Humanitarian Action*, 5(1), 1–7.

Information Commissioner's Office, n.d. *Vital interests*, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/vital-interests/.

See also: DATA PROTECTION, MENTAL CAPACITY

VM

See: VIRTUAL MACHINE

Voice over Internet Protocol (VoIP)

Voice over Internet Protocol (VoIP) technology enables users to make voice calls over the **Internet** through a broadband connection. VoIP translates voice signals into digital **data** packets that may be sent using the Internet data transport infrastructure and received as an audio stream by the receiver.

VoIP depends for its quality on low packet loss and packet latency; that is, the data needs to get to its destination very quickly, otherwise the **user** experience of the audio or video is unacceptable. This means that **security** measures cannot interfere too much with the **data flow**. For instance, a lot of VoIP service providers do not employ **end-to-end encryption** to **secure communication**s, making them **vulnerable** to interception. As an additional **privacy threat**, call **metadata** may also be gathered and utilised for monitoring and **surveillance**.

Further reading:

Goode, B., 2002. Voice over internet protocol (VoIP). *Proceedings of the IEEE*, 90(9), 1495–1517, https://doi.org/10.1109/JPROC.2002.802005.

See also: COMMUNICATION, COMMUNICATION PRIVACY, DATA QUALITY, ENCRYPTION, HACKING, MAN-IN-THE-MIDDLE ATTACK, PACKET SNIFFING

VoIP

See: VOICE OVER INTERNET PROTOCOL

Voyeurism

Voyeurism is the practice of watching other people's **private** behaviour for pleasure or through compulsion, usually without the knowledge or **consent** of the **person** watched. It is commonly associated with the observation of sexual behaviour and other revealing activities such as undressing, sunbathing or bathing, and **intrusion** into **gendered spaces**. However, it is not exclusively a sexual impulse, and has been implicated in such activities as watching reality television or **eavesdropping** on arguments.

Further reading:
Hopkins, T.A., Green, B.A., Carnes, P.J. and Campling, S., 2016. Varieties of intrusion: exhibitionism and voyeurism. *Sexual Addiction and Compulsivity*, 23(1), 4–33, https://doi.org/10.1080/10720162.2015.1095138.

See also: ATTENTIONAL PRIVACY, INTRUSION UPON SECLUSION, SECLUSION, SURVEILLANCE

VPN

See: VIRTUAL PRIVATE NETWORK

VSD

See: VALUE SENSITIVE DESIGN

VSS

See: VERIFIABLE SECRET SHARING

Vulnerability

A weakness or disadvantage that leads an entity to be at **risk** of adverse outcomes.

When applied to people in a **privacy** context, a vulnerability suggests that they are open to exploitation by malicious actors. When applied to systems, a vulnerability refers to a gap or flaw in a system that can be exploited by an **adversary** to compromise the **CIA Triad** (**confidentiality**, **integrity** and **availability**). There is therefore the potential for unauthorised access, retrieval of users' **personal information** or the misconfiguration of a system, if the vulnerability is known, or if it is of a common type.

System vulnerabilities can be triggered by poor **software** design choices, overlooking **security** measures or failures to update software to take account of current **threat**s. They can be minimised by security administrators, using **encryption** and **authentication**, and regular **security audit**s. Vulnerabilities in commercial software are managed using **patches** – software updates distributed regularly to users by vendors as a service.

Further reading:
Syed, R., 2020. Cybersecurity vulnerability management: a conceptual ontology and cyber intelligence alert system. *Information & Management*, 57(6), 103334, https://doi.org/10.1016/j.im.2020.103334.

See also: AUTHORISATION, CYBERSECURITY, SECURITY-BY-DESIGN, VULNERABILITY MANAGEMENT

Vulnerability Management

The process of locating, assessing, prioritising and addressing security vulnerabilities in computer systems. By proactively detecting and correcting **security** flaws before an **adversary** can exploit them, **vulnerability** management seeks to lower the **risk** of cyber-attacks and **data breach**es.

Management frameworks consist of several steps, including discovery, **risk assessment**, vulnerability assessment, **remediation**, verification and reporting.

Further reading:
Syed, R., 2020. Cybersecurity vulnerability management: a conceptual ontology and cyber intelligence alert system. *Information & Management*, 57(6), 103334, https://doi.org/10.1016/j.im.2020.103334.

See also: CYBER RESILIENCE, CYBERSECURITY

Vulnerable Population

A set of people sharing one or more characteristics that means that they have reduced capacity to make informed decisions or are susceptible to exploitation. Examples include children and adults with learning disabilities.

Further reading:
Waisel, D.B., 2013. Vulnerable populations in healthcare. *Current Opinion in Anaesthesiology*, 26(2), 186–92, https://doi.org/10.1097/ACO.0b013e32835e8c17.

W

Wearable Computing

Wearable computing devices, or *wearables*, are **Internet**-enabled devices worn by individuals either in their daily lives or during events (such as sporting events, sleep or combat). They include smart watches and fitness devices worn on the wrist, devices strapped to the head and body, items such as spectacles or earpieces, jewellery, and actual clothes made of fabrics with computational devices embedded within them. Some wearables can even be attached directly to the skin like a tattoo (*e-skin*).

Typical functions of wearables include acting as **sensor**s (perhaps of medical parameters of the body, but also of the environment, perhaps measuring air pollution), providing **information** to the wearer (e.g., with **augmented reality** glasses or headsets), compensating for disabilities, behaviour **tracking**, drug delivery, epidemiology or enabling an organisation to gain a global vision of an environment using its members as a collaborative swarm of human sensors. They might even be a fashion statement (clothes that change colours or display varying images). Healthcare and well-being are currently the most common applications.

The information collected by wearables is strongly disclosive of the wearer's behaviour, health and **location**, and may also be a **privacy threat** to others in the environment (for instance, smart glasses could be used for **stalking**). They can therefore be a clear threat to privacy, especially if the information they collect is not stored or transmitted securely. As wearables are necessarily very light and with limitations on battery power, the amount of **software** or hardware devoted to security they can incorporate is limited.

Further reading:

Iqbal, S.M.A., Mahgoub, I., Du, E., Leavitt, M.A. and Asghar, W., 2021. Advances in healthcare wearable devices. *npj Flexible Electronics*, 5, article no.9, https://doi.org/10.1038/s41528-021-00107-x.

Xue, Y., 2019. A review on intelligent wearables: uses and risks. *Human Behavior and Emerging Technologies*, 1(4), 287–94, https://doi.org/10.1002/hbe2.173.

See also: ELECTRONIC HEALTH RECORD, HARASSMENT, INFORMATION SECURITY, INTERNET OF THINGS

Wearable Tech

See: WEARABLE COMPUTING

Web 2.0

Web 2.0 (pronounced ‘Web two’) is a generic name for the generation of the **World Wide Web** that became prominent in the period 2000–5. Whereas the earlier Web had a more broadcast flavour, with most participants being passive readers of webpages, the protocols that characterised Web 2.0 enabled **user**s to generate their own content using accessible tools.

Initially, the **privacy concern**s of Web 2.0 were focused on the temptation to overshare in blogs or social networks. However, these were eventually dwarfed by the possibilities of analysis of **inferred data** at scale (**big data**) created about Web 2.0 participants, over and above the **declared data** they volunteered. Consequently, privacy debates at the high point of Web 2.0 had a very different flavour from those after about 2012, when the power of ‘big tech’ became apparent.

Further reading:

Caviglione, L. and Coccoli, M., 2011. Privacy problems with Web 2.0. *Computer Fraud and Security*, 2011(10), 16–19, https://doi.org/10.1016/S1361-3723(11)70104-X.

Child, J.T., Haridakis, P.M. and Petronio, S., 2012. Blogging privacy rule orientations, privacy management, and content deletion practices: the variability of online privacy management activity at different stages of social media use. *Computers in Human Behavior*, 28(5), 1859–72, https://doi.org/10.1016/j.chb.2012.05.004.

See also: DELETION, SOCIAL NETWORK

Web 3.0

Web 3.0 is a generic name for a vision of the future **World Wide Web** based on **decentralisation of the Web** and the use of **blockchain** data for data storage and also to transfer value via tokens or a **cryptocurrency**. It is a reaction to the perceived centralisation of the Web around large digital platforms, particularly **social media** companies and **search engine**s, exploiting **network** effects among their **user**s.

Privacy has long been perceived as a problem on the Web, partly because **data** is used routinely to power and **personalised services**, partly because data-gathering is opaque to the **data subject** and partly because the centralisation of data holdings by platforms is vulnerable to **attack** by **cybercriminal**s. The technology envisaged for Web 3.0, based on **distributed ledger** technology, would in theory enhance privacy by making **data ownership** and use of data **transparent** on the blockchain. Value could also be transferred to data subjects whose data is used, using tokens defined on the blockchain, allowing compensation for the use of data. Data on the blockchain would be **encrypted**, and if data subjects had the appropriate **cryptographic key** they could control access. **Individual**s could control their own identities, and identify themselves using the minimum **information**, if **self-sovereign identity** were implemented on the blockchain. Blockchain is also perceived to be more secure than current data storage technology.

Further reading:

Litwack, S., 2018. Is a decentralized 'web 3.0' the answer to our privacy concerns? *IAPP Privacy Tech*, https://iapp.org/news/a/is-a-decentralized-web-3-0-the-answer-to-our-privacy-concerns/.

See also: ACCESS CONTROL, DATA IN USE, DATA TRANSFER, CENTRALISED GOVERNANCE, PRIVACY AS CONTROL, SOCIAL NETWORK, VALUE OF DATA, VALUE OF PRIVACY

Web Beacon

A web beacon is a tiny graphic image or piece of code that is inserted into a webpage or email. It is sometimes referred to as a *tracking pixel*. **Tracker**s and advertisers frequently use web beacons to track **user** behaviour and collect **information** about how users engage with their content.

The image or code in a web beacon triggers a request to a remote server when a user opens a page or email that contains one. The remote server logs the request and records information about the user's activity, including the time and date of the request, the user's **IP address**, the kind of browser and device they are using and other information about their behaviour.

Users can deactivate picture-loading in their email client, install ad-blocking software, or select **privacy setting**s in their web browser to block **third-party cookie**s and prevent tracking across multiple websites, if they wish to protect themselves from web beacons and other tracking technologies.

Further reading:
Bouguettaya, A. and Eltoweissy, M., 2003. Privacy on the Web: facts, challenges, and solutions. *IEEE Security & Privacy*, 1(6), 40–9, https://doi.org/10.1109/MSECP.2003.1253567.

See also: PROFILING, WEB PROFILING

Web Bug

A web bug is a particular kind of **web beacon** that is made to be invisible to the **user**.

Web of Trust

One challenge with **public-key infrastructure (PKI)** is to ensure that the asserted link between an agent and their **public key** is authentic. A centralised solution to this is to have **certification authoritie**s to hold **database**s of links. However, in open systems, or low-trust contexts, it may not be desirable to rely on a centralised **trusted third party**, and instead a decentralised **trust** model may be used.

In an open system such as Pretty Good **Privacy** (PGP), decentralised trust is disseminated around a web of trust. Each successful interaction will help an agent build trust with those with which it has transacted. When such a recipient has decrypted a message from a sender, it is in a position to confirm the link between the sender and the public key. It may then use that knowledge to support the sender’s claim to own the key, for example by adding its own **digital signature** to an open certificate, or by responding to queries from others about ownership of the key. Those wishing to authenticate an agent’s ownership of a public key can choose to trust such evidence to a certain degree, for example depending on the **reputation** of the provider of the evidence. As more links are authenticated across the network, the decentralised web of trust grows stronger.

Further reading:
Caronni, G., 2000. Walking the web of trust. *In: Proceedings IEEE 9th international workshops on enabling technologies: infrastructure for collaborative enterprises (WET ICE 2000)*, https://doi.org/10.1109/ENABL.2000.883720.

See also: AUTHENTICATION, CERTIFICATION, PUBLIC-KEY CRYPTOGRAPHY

Web Profiling

Web profiling is the process of gathering and examining **information** about a **user**'s online behaviour and activities, including their **search** history, **social media** interactions, **browser history** and other digital traces. While it is only one type of **profiling**, Web profiling is the most commonly used, as more **information** can be gathered cheaply than with other types. It is mostly used for **behavioural advertising**.

Numerous methods, including **tracking cookie**s, **browser fingerprinting** and **cross-device tracking**, are used for Web profiling.

Further reading:
Datta, A., Datta, A., Makagon, J., Mulligan, D.K. and Tschantz, M.C., 2018. Discrimination in online advertising: a multidisciplinary inquiry. *In: Conference on Fairness*, *Accountability and Transparency*, 20–34, https://proceedings.mlr.press/v81/datta18a.

See also: DIGITAL FINGERPRINTING, DIGITAL FOOTPRINT, DIGITAL IDENTITY, SOCIAL PROFILING

Web Skimming Attack

In a Web skimming attack, hackers inject malicious code into a website's HTML payment form to collect **personal data** such as credit card numbers and **password**s.

Hackers generally acquire access to a form by taking advantage of coding flaws such as **SQL injection** or **cross-site scripting**. Upon gaining entry, they place JavaScript code to record users' payment **information** as it is entered and transfer it to a remote server under their control.

Further reading:
Ahmed, A.A. and Al Dabbagh, N.B., 2023. Web attacks and defenses. *Journal of Education & Science*, 32(2), 114–27, https://doi.org/10.33899/edusj.2023.137855.1319.
Shar, L.K. and Tan, H.B.K., 2012. Defeating SQL injection. *Computer*, 46(3), 69–77, https://doi.org/10.1109/MC.2012.283.

Whistleblowing

Whistleblowing is the practice of revealing confidential **information** whose revelation the *whistleblower* believes is in the **public interest**. A common type of whistleblower is an employee who disapproves of some unethical or criminal practice of the employer.

Whistleblowing has a complex relationship with privacy.

First, it almost always involves a **breach of confidence**, and so is always connected with an unauthorised flow of information.

Second, many privacy scandals have been revealed by whistleblowers. Edward Snowden's exposé of the National Security Agency's bulk **surveillance** programmes is perhaps the most prominent; another example is that of former Facebook product manager Frances Haugen, who released thousands of documents about its working practices to journalists and the Securities and Exchange Commission.

Third, whistleblowing engages the public interest; if there is public interest in the release of confidential information, that can be a defence for the whistleblower against legal action (even if the relationship was protected by a **non-disclosure agreement**). Many **jurisdiction**s contain specific protections for whistleblowers.

And fourth, whistleblowers themselves are likely to suffer in career terms (even if they are protected from legal action). Hence, methods to encourage whistleblowing or protect whistleblowers focus on **anonymity**, such as whistleblowing hotlines. Whistleblowers themselves may also avail themselves of anonymous communications, **encrypted** messages, **onion routing** or content-sharing sites such as WikiLeaks.

Further reading:

Ceva, E. and Bocchiola, M., 2019. *Is whistleblowing a duty?* Cambridge: Polity Press.

Greenwald, G., 2014. *No place to hide: Edward Snowden, the NSA and the surveillance state*. London: Hamish Hamilton.

See also: NATIONAL SECURITY

White Box Testing

A **software** testing approach that involves evaluating a system's internal architecture and operation (as opposed to treating it as a *black box* and looking only at inputs and outputs). Developers that have access to the source code scrutinise it and the system's underlying architecture to ensure it is operating effectively and correctly and to find any flaws or vulnerabilities.

Further reading:

Khan, M.E. and Khan, F., 2012. A comparative study of white box, black box and grey box testing techniques. *International Journal of Advanced Computer Science and Applications*, 3(6), https://doi.org/10.14569/IJACSA.2012.030603.

See also: INTERNAL SECURITY TESTING, VULNERABILITY

White Hat Attack

An **attack** on an organisation's systems and/or **data** designed to inform the organisation about where the **cybersecurity** vulnerabilities are and how to correct them. It is distinguished from a **grey hat attack** because the attack will have been **authorised** by the organisation. It will often be carried out by an employee of the organisation or a consultant. Where the personnel carrying out a white hat attack are part of an ongoing defensive function, they are often called a **red team**.

Further reading:

Gandhi, F., Pansaniya, D. and Naik, S., 2022. Ethical hacking: types of hackers, cyber attacks and security. *International Research Journal of Innovations in Engineering and Technology*, 6(1), 28–32, www.proquest.com/docview/2634513488?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals.

Morgan, G. and Gordijn, B., 2020. A care-based stakeholder approach to ethics of cybersecurity in business. *In*: Christen, M., Gordijn, B. and Loi, M., eds, *The ethics of cybersecurity*, 119–38, https://doi.org/10.1007/978-3-030-29053-5_6.

See also: BLACK HAT ATTACK, ETHICAL HACKING, HACKING, PENETRATION TEST, VULNERABILITY

Wiretapping

The US term for **telephone tapping**.

World Wide Web (WWW)

The World Wide Web (WWW, or the *Web*) is an **application** that uses the **Internet Protocol** suite to transfer documents and other digital resources between computers. The intended effect, revolutionary in its day (it was first proposed in 1989 by Tim Berners-Lee), is to make it appear that the downloaded documents and resources are sitting on the **user**'s computer or device, rather than on a remote server. It was envisaged as a hypertext system, so that links could be created between different documents. It was also radically decentralised; it is an open system governed by open standards, so no permission is needed to join it, or to add documents or pages to it, or to make links from a document one is editing to anywhere else on the Web.

The key technological aspects of the Web are the use of **uniform resource locator**s **(URLs)** as means of both identifying and locating digital resources, the *Hypertext Markup Language* (HTML) to render documents and webpages for the particular Web browser on the user’s device, and originally the **Hypertext Transfer Protocol (**HTTP) governing the transfer of **information** between server and client. In the intervening years, there have been improvements to the transfer protocol (or the *transport layer* of the Web). HTTPS is a secure version of HTTP, which enabled the growth of the Web as a site for banking, payments and **e-commerce** after 2000. QUIC is another transport protocol gaining popularity, as it deals more easily with errors, data loss, congestion control, latency and other data transport issues.

The Web was particularly responsible for the growth of the **Internet**, by acting as the platform on which increasingly interesting content could be accessed, whether news, entertainment or commerce. Hence it has had many unintended consequences, from misinformation to **cybercrime** to new **privacy** threats (as well as many benefits). Berners-Lee himself has suggested that the Web’s malign developments include its becoming more centralised, as well as its effects on privacy. To that end, he proposed the Solid (**Social Linked Data**) project to help rectify these and re-establish the original Web vision.

Further reading:

Berners-Lee, T. and Fischetti, T., 1999. *Weaving the Web: the original design and ultimate destiny of the World Wide Web*. New York: HarperCollins.

Mansour, E., Sambra, A.V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., Aboulnaga, A. and Berners-Lee, T., 2016. A demonstration of the solid platform for social web applications. *In: Proceedings of the 25th International Conference Companion on World Wide Web*, 223–6, https://doi.org/10.1145/2872518.2890529.

See also: DARK WEB, DECENTRALISATION OF THE WEB, HYPERTEXT TRANSFER PROTOCOL SECURE, STANDARD, WEB 2.0, WEB 3.0

Worm

A worm is a kind of **malware** that infects a network and replicates within it. Where **virus**es attach to existing programs, worms are self-contained programs that need no host and can spread and infect other systems via the **network** configuration.

Further reading:

Weaver, N., Paxson, V., Staniford, S. and Cunningham, R., 2003. A taxonomy of computer worms. *In: Proceedings of the 2003 ACM workshop on rapid malcode*, 11–18, https://doi.org/10.1145/948187.948190.

WWW

See: WORLD WIDE WEB

X

XAI

See: EXPLAINABLE AI

XSS

See: CROSS-SITE SCRIPTING

X Variable

See: EXPLANATORY VARIABLE

Y

Y Chromosome

The chromosome for maleness. Because it is inherited via the paternal line, it is culturally correlated with surnames and therefore can on its own be disclosive.

See also: SURNAME ATTACK

Y Variable

See: RESPONSE VARIABLE

Z

Zero Day Attack

Zero day attacks on vulnerable **software** get their name from the fact that they happen before the creator of the software is aware of the **vulnerability**, giving them no time to prepare or provide a **patch** to correct the issue.

In a zero day attack, hackers take advantage of a vulnerability, unknown to a system's creator or managers, to infiltrate the system without **authorisation**, steal **confidential** data, deploy **malware** or engage in other nefarious deeds. The creator's ignorance makes it difficult to deploy the appropriate **security** measures, such as **firewall**s and **anti-virus software**, to identify and stop this kind of attack. Once the vulnerability is discovered, a patch can be developed and distributed to **user**s, and once it has been applied generally, the **risk** of the **attack** becomes negligible.

Further reading:

Bilge, L. and Dumitracs, T., 2012. Before we knew it: an empirical study of zero-day attacks in the real world. *In: Proceedings of the 2012 ACM conference on computer and communications security*, 833–44, https://doi.org/10.1145/2382196.2382284.

See also: CYBERSECURITY, PATCH MANAGEMENT, VIRUS

Zero Knowledge

Zero knowledge refers to a type of cryptographic **algorithm** that enables one party to demonstrate to another party that they are aware of a certain piece of **information** or **secret** without actually disclosing the information or secret itself.

In a zero knowledge proof, the party showing their knowledge, known as the *prover*, can show the other party, known as the *verifier*, that they have the knowledge or **credentials** without the verifier learning anything more. This is not necessarily a strict mathematical proof; most zero knowledge proofs aim to show that the probability that the prover is dishonest is extremely low. They often take the form of a dialogue between prover and verifier. The verifier queries the prover, whose responses are cumulatively sufficient to establish possession of the knowledge to a high enough probability.

Further reading:
Blum, M., Feldman, P. and Micali, S., 2019. Non-interactive zero-knowledge and its applications. *In*: Goldreich, O., ed., *Providing sound foundations for cryptography: on the work of Shafi Goldwasser and Silvio Micali*, ACM, 329–49, https://doi.org/10.1145/3335741.3335757.

See also: INFORMATION SECURITY

Zero Knowledge Proof

See: ZERO KNOWLEDGE

Zero Trust Security

Under zero trust models or architectures, access requests (for example, to a **network** or elements of it) are assumed to be untrustworthy until proven otherwise. Each attempt to access a resource is subject to an established policy for **authentication**, **authorisation** and verification. This contrasts with models of **trust** that are satisfied at the perimeter and assumed for all connections inside.

Access controls are based on policies that specify the **least privilege** required for people, devices and programs to access resources. Controls will be realised in an array of **network security** and **endpoint security** tools, techniques and practices, including – but not restricted to – **firewall**s, **intrusion detection system**s, **intrusion prevention system**s and **user** and entity behaviour analytics.

Further reading:
Mehraj, S. and Banday, M.T., 2020. Establishing a zero trust strategy in cloud computing environment. *In: 2020 International Conference on Computer Communication and Informatics (ICCCI)*, 1–6, https://doi.org/10.1109/ICCCI48352.2020.9104214.
Bertino, E., 2021. Zero trust architecture: does it help? *IEEE Security & Privacy*, 19(5), 95–6, https://doi.org/10.1109/MSEC.2021.3091195.

See also: RESTRICTED ACCESS