FAST GUIDE #01

RESEARCH DATA: the basics

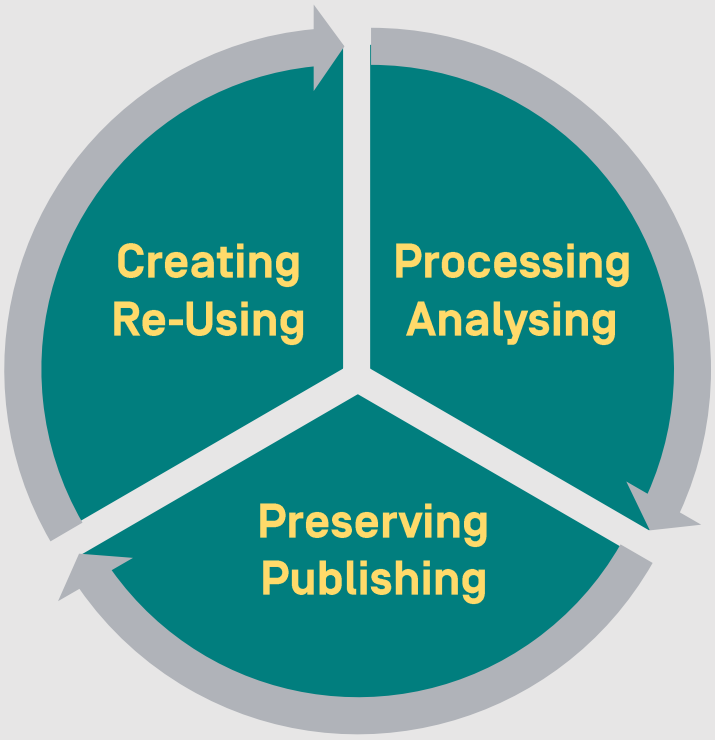
EPFL Library
Research Data Management
FAST GUIDES

RESEARCH DATA DEFINITIONS

- ✓ Material generated or collected during the course of conducting research [1].
- ✓ Factual records used as primary sources for scientific research, commonly accepted in the scientific community as necessary to validate research findings [2].
- ✓ Information collected, observed, or created, for purposes of analysis to produce original research results [3].
- ✓ Any information in binary digital form derived from the research process [4].

RESEARCH DATA LIFECYCLE

1. **Creating / Re-using**: planning data collection, locating existing data sources; producing, collecting or documenting data.
2. **Processing / Analyzing**: validating, cleaning, transforming data; creating metadata; using, creating analysis tools; interpreting the data.
3. **Preserving / Publishing**: reviewing the data; getting data into a format suitable for preservation; depositing data and metadata in archive / repository; promoting data re-use.



RESEARCH DATA TYPES

- **Observational Data**: data captured in-situ, can't be recaptured, recreated or replaced.
  Examples: Sensor readings, sensory (human) observations, survey results, interview notes, transcripts
- **Experimental Data**: data collected under controlled conditions, in situ or laboratory-based, should be reproducible, but can be expensive. Examples: gene sequences, chromatograms, spectroscopy, microscopy
- **Simulation Data**: result from using a model to study the behaviour and performance of an actual or theoretical system, models and metadata, where the input can be more important than output data.
  Examples: climate models, economic models, biogeochemical models
- **Derived/Compiled Data**: reproducible, but can be very expensive.
  Examples: derived variables, compiled database, 3D models
- **Reference or canonical Data**: static or organic collection [peer-reviewed] datasets, most probably published and/or curated. Examples: gene sequence databanks, chemical structures, census data, spatial data portals[5]

Raw Data

Raw data refer to data that have not been changed since acquisition, eg. a real-time GPS-encoded navigation file, and the initial time-series file of temperature values from a heat probe.

Processed Data/Active Data

Editing, cleaning or modifying the raw data results in processed data, eg. raw multibeam data files can be processed to remove outliers and to correct sound velocity errors[6].

Credits and sources

[1] https://www.ed.ac.uk/information-services/research-support/research-data-service
[2] https://www.oecd.org/sti/sci-tech/38500813.pdf
[3] http://www.ed.ac.uk/information-services/research-support/data-management
[4] https://www.degruyter.com/view/product/430793
[5] http://guides.library.stonybrook.edu/research-data-services/types
[6] http://www.marine-geo.org/help/data_FAQ.php

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #02

FAIR DATA PRINCIPLES[1]

EPFL Library
Research Data Management
FAST GUIDES

F — FINDABLE

*Data and metadata are **easy to find** by both humans and computers.*

- **F1** [Meta]data are assigned a globally unique and persistent identifier.
- **F2** Data are described with rich metadata.
- **F3** Metadata clearly and explicitly include the identifier of the data they describe.
- **F4** [Meta]data are registered or indexed in a searchable resource.

A — ACCESSIBLE

*Humans and computers can **readily access** or download datasets.*

- **A1** [Meta]data are retrievable by their identifier using a standardized communication protocol:
  - **A1.1** the protocol is open, free and universally implementable;
  - **A1.2** the protocol allows for an authentication and authorization procedure where necessary.
- **A2** Metadata are accessible, even when the data are no longer available.

I — INTEROPERABLE

*Data from different datasets are **prepared to be combined** or exchanged.*

- **I1** [Meta]data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- **I2** [Meta]data use vocabularies that follow FAIR principles.
- **I3** [Meta]data include qualified references to other [meta]data.

R — REUSABLE

*Published data can be **easily combined** or **replicated** in future research.*

- **R1** [Meta]data are richly described with a plurality of accurate and relevant attributes:
  - **R1.1** [meta]data are released with a clear and accessible data usage license;
  - **R1.2** [meta]data are associated with detailed provenance;
  - **R1.3** [meta]data meet domain-relevant community standards.

DESCRIBE

Describe provenance, usage and organization of data with standardized **metadata** [DataCite, RDA standards, DublinCore]. Make metadata available **even if** data are not.

OPEN

Open your data using standardized **licenses** [ex. Creative Commons]. **Limitations** may apply to the openness [ex. embargo]. Disclose files in **open formats**, even alongside proprietary formats.

LINK

Use persistent **identifiers** for datasets [ex. DOI, HANDL, URN] and tag all the metadata with the **same** identifiers. **Cross-link** datasets with linked-data standards [RDF].

PUBLISH

Deposit datasets in data **repositories**, favoring services with user-friendly **interfaces**.

"Data shoud be as open as possible, as closed as necessary."

Carlos Moedas
EU Commissioner

How FAIR are your data?
Take the FAIR **self-assessment test**[2]

Did you know?

40% of researchers are aware of the existence of FAIR principles[3]

20-50% increased citation for articles linked to associated data[4]

Credits and sources

[1] FAIR principles: go-fair.org/fair-principles
[2] FAIR self-assessment tool: ands-nectar-rds.org.au/fair-tool
[3] State of Open Data 2018: figshare.com/blog/State_of_Open_Data_2018/440
[4] Open Data Citation Advantage: sparceurope.org/open-data-citation-advantage

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #03

COST of Research Data Management

EPFL Library
Research Data Management
FAST GUIDES

RESEARCH DATA MANAGEMENT (RDM) activities to consider for cost estimation

DATA MANAGEMENT PLAN
Writing and continuous revision of a DMP

COLLECTION
Databases and software, data formatting and organization, data transfer

ACTIVE MANAGEMENT
Electronic Lab Notebook (ELN), Laboratory Information Management System (SLIMS), data sharing platform

DOCUMENTATION
Data description and metadata, documentation and transcription

STORAGE/BACK-UP
Data back-up, data storage

ACCESS AND CONTROL
Access control, data security, personal data protection

SHARING
Anonymization, copyright assessment, data cleaning, data publishing

ARCHIVING
Data preparation, long-term preservation data repository

Costs are cumulative and increase in time

- do not overcomplicate your processes
- do not adopt too many tools

Did you know?

RDM costs can be eligible for funding applications.

SWISS NATIONAL SCIENCE FOUNDATION
Data generated must be publicly accessible in non-commercial repositories provided there are no legal, ethical, copyright or other issues. The SNSF may allocate up to CHF 10,000 for Open Research Data activities.

ERC/H2020
Costs related to Open Access to research data (APC, RDM, curation and storage costs ...) are eligible for reimbursement during the project.

+5% **Expected RDM cost on the total project expenditure spent on properly managing and stewarding data[1].**

Cost reduction expected for projects tackling the issues of poor data quality, redundant data, and lost data[2]. **-15%**

TOOLS / RESOURCES

- **RESEARCH OFFICE BUDGET TEMPLATES**
  The RDM costs are already listed in the budget templates available on the EPFL-ReO website[3].
- **COST CALCULATOR**
  Try out the EPFL Library's online tool to calculate the storage costs for your research project[4].
- **QUESTIONS TO CONSIDER**
  An overview of possible costs per research activity is presented by the Utrecht University[5].
- **CHRONOS**
  Timekeeping for research projects, to justify eligible personnel costs to the funding bodies[6].

Credits and sources

[1] ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf (p.17) / [2] www.usgs.gov/products/data-and-tools/data-management/value-data-management / [3] research-office.epfl.ch/research-funding / [4] rdmepfl.github.io/costcalc / [5] www.uu.nl/en/research/research-data-management/guides/costs-of-data-management / [6] www.epfl.ch/research/services/manage-projects/chronos /

[7] Icons: https://www.flaticon.com/packs/business-seo

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #04

FILE FORMATS

EPFL Library
Research Data Management
FAST GUIDES

Definition

A **file format** is a standard way to encode data for storage in a computer file. It specifies how bits are used to encode information in a digital storage medium. File formats may be either proprietary or free and may be either unpublished or open[1].

When listing out the data formats you will be using, make sure to include:

- The necessary software to view the data (e.g. SPSS v.3; Microsoft Excel 97-2003).
- Information about version control.
- If data are stored in one format during collection and analysis and then transferred to another format for preservation: list out features that may be lost in data conversion such as system specific labels.

When selecting file formats for archiving, the formats should ideally be:

- Non-proprietary, unencrypted, uncompressed, commonly used by the research community.
- Compliant to an open, documented standard: interoperable among diverse platforms and applications, fully published and available royalty-free, fully and independently implementable by multiple software providers on multiple platforms without any intellectual property[2].

File formats extensions for reusability/preservation:

Type of data	APPROPRIATE	ACCEPTABLE	NOT SUITABLE
Tabular data with extensive metadata	.csv - .hdf5	.txt - .html - .tex - .por	
Tabular data with minimal metadata	.csv - .tab - .ods - SQL	.xml if appropriate DTD - .xlsx	.xls - .xlsb
Textual data	.pdf - .txt - .odt - .odm - .tex - .md -.htm - .xml	.pptx - .pdf with embedded forms - .rtf	.doc - .ppt
Code	.m - .R - .py - .iypnb - .rstudio - .rmd – NetCDF	.sdd	.mat - .rdata
Digital image data	.tif - .png - .svg - .jpeg	jpg - .jp2 - .tif - .tiff - .pdf - .gif - .bmp	.indd - .ait - .psd
Digital audio data	.flac - .wav - .ogg	.mp3 - .mp4 - .aif	
Digital video data	.mp4 - .mj2 - .avi - .mkv	.ogm - .webm	.wmv - .mov
Geospatial data	NetCDF, tabular GIS attribute data, .shp - .shx - .dbf - .prj - .sbx - .sbn - PostGIS - .tif - .tfw - GeoJSON	.mdb - .mif	
CAD/vector and raster data	.x3d - .x3dv - .x3db - PDF3D .pdf	.dwg - .dxf	
Generic data	.xml - .json - .rdf		

For further information: List of EPFL Recommended File Formats[3]

Credits and sources

[1] https://en.wikipedia.org/wiki/File_format

[2] https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats

[3] https://researchdata.epfl.ch/wp-content/uploads/2018/05/Recommended_DataFormats_-2018_03_05_Final.pdf

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #05

METADATA

EPFL Library
Research Data Management
FAST GUIDES

"Metadata is structured information associated with an object for purposes of discovery, description, use, management, and preservation"

(National Information Standards Organization, 2008)

METADATA IS UBIQUITOUS AND PROLIFERATIVE

METADATA IS EMBEDDED OR SUPPLEMENTAL

METADATA RESULT FROM AUTOMATIC OR MANUAL INPUT

INTEROPERABILITY IS BASED ON METADATA

5 METADATA FAMILIES

- **Technical metadata** (ex. version of producing device)
- **Use metadata** (ex. number of downloads)
- **Administrative metadata** (ex. publishing date, rights and licenses)
- **Descriptive metadata** (ex. title, author, keywords)
- **Preservation metadata** (ex. last checksum date)

From Excel to databases and semantic web knowledge bases, the more metadata you have, the better **data management system** you need.

FAIR data, good quality linked (open)data, mainly relies on rich, detailed, qualified, shared, standardized metadata.

HOW TO?

1. **Be systematic, adopt rules, use controlled values**
2. **Describe your data completely and consistently**
3. **Use standards**

Metadata and metadata standards creation, adoption and maintenance is a JOINT EFFORT within and between interest-based communities.

TOOLS TO BUILD YOUR OWN STRONG METADATA

FORMAT, TECHNICAL, INTERCHANGE STANDARDS : exif, IPTC, instrumentation specific standards...

VALUE NORMS, STANDARDS AND REFERENCES : ISO 8601, ISO 639-1, ISO 3166-1, thesaurii, vocabularies, lists of authorities...

CONTENT MODELS AND STANDARDS : ISA (Investigation-Study-Assay) framework, Force11 Software citation principles

STRUCTURE STANDARDS AND SCHEMAS : INSPIRE, SDMX, Darwin Core, Dublin Core, PROV model, Datacite

More resources

http://www.dcc.ac.uk/resources/metadata-standards/list

http://rd-alliance.github.io/metadata-directory/standards

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #06

CODE

EPFL Library
Research Data Management
FAST GUIDES

When working with code, good practices are also needed. In particular the publication of code is needed in order to understand, reuse and repeat the operation.

TIPS AND TRICKS FOR A BETTER EFFICIENCY IN CODE MANAGEMENT

VERSIONING

Versioning systems are powerful tools for code management. The most used is **Git**, it's free and open :

- It allows to **track changes** and to undo changes if needed. You can manage easily different versions of your code
- Connected to a repository your code and its modifications are **automatically backuped**
- You can also **work in team** easily on the same code

SHARING

In order to **share your code and make it visible**, repositories provide various services like version management system, wikis, task management and issues tracking, one of the most known is **Github**.

EPFL provides c4science.ch for code versionning. Data are stored in Switzerland.

EPFL provides also gitlab.epfl.ch (open-source github) but backup is not guaranteed.

DESCRIBING

README documentation is a really important part of coding. It allows you to **explain your code**, for you and others. You should add rich metadata and documentation (README, LICENSE, comments on code...) on any publication of the code.

Some tools like sphinx-doc.org and doxygen.nl can help you by going through your code and generating a preformated documentation.

LICENSING

It is important to explain **how your code can be used** by others (and related restrictions).

You have at least three options :

- Open source licenses (permissive as MIT or GPL)
- Academic licenses (restrict commercial usage)
- Commercial licenses (reserve commercial usage)

PUBLISHING

Don't forget to **generate a DOI** to uniquely identify a version of your software and to easily cite it.

Most code repository (like Zenodo or c4science) generate a DOI for your deposit.

TIP : Github provides an integration with Zenodo.

PRESERVING

Preservation is important for keeping your work secure and also for scientific validation.

C4science is a solution to **preserve your code** for the long term. If you are using another code repository, you can **always make a copy on c4science for preservation**.

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #07

ELN - Electronic Lab Notebook

EPFL Library
Research Data Management
FAST GUIDES

An ELN allows new capabilities compare to paper notebook :

- A **better knowledge transmission** internally and externally
- Increase the preservation by **automatic backup** and by storing everything on the same location
- An **uniformization of the work** by proposition template and sharing between members

When considering an ELN implementation in your lab, make sur to answer the following questions:

PRACTICAL

Are the storage method and location adequate for me [cloud based]?

- If your ELN is cloud based you might want to consider where your data are hosted and who can have access to it.

Can I have a connected computer where I need to use the notebook?

Do I work with pattern and my does my ELN support it?

Do I need support [hotline...]?

Do I need some specific tools?

INTERFACE

Do I find the interface suitable for me?

Is it compatible with mobile devices?

Do I need a sample/laboratory management?

INTEROPERABILITY

Can I import my previous notes?

- You might want to check the import option and if there is an API.

Can I export my data in an open way?

- You might want to check the export option and if there is an API.

What are the export formats?

Do I have data volume limitation?

- You might want to check the ELN business plan and the allowed storage for data.

IMPORT EXPORT

Is the ELN compatible with software I'm using to generate data?

- This might help you to import the data you generate.

Can I use my cloud software [Google drive, Mendeley ...]?

- This might help you for integrating the services you are using.

Can I use repositories?

- Like Zenodo [zenodo.org], figshare [https://figshare.com/], C4science[https://c4science.ch/]
- This might help you to publish your data.

ELN @EPFL

- SLIMS : Commercial solution proposed by the School. It integrates a sample management and different services for biologists. https://sv-it.epfl.ch/page-120709-fr-html/lims/
- ELN : Chemistry Notebook, developped by Luc Patiny https://eln.epfl.ch/
- Others : The EPFL Library Research Data team can help you to implement a different ELN.

Additional information can be find here

researchdata.epfl.ch and https://datamanagement.hms.harvard.edu/electronic-lab-notebooks

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #08

PERSONAL DATA management

EPFL Library
Research Data Management
FAST GUIDES

PERSONAL DATA

Personal data, art. 3a FADP:

"all information relating to an identified or identifiable person" (ex: name, date of birth, address, pictures, videos, IP address, GPS coordinates, biometric/genomic data, etc.)

SENSITIVE DATA

Sensitive data, art. 9.1 GDPR:

"data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"

THE SWISS FEDERAL ACT ON DATA PROTECTION *(FADP)*

The FADP applies to every research project conducted in Switzerland. Additional laws are enacted for the research involving human beings (Human Research Act)1.

The following rules and items are required for every research project (non exhaustive):

- **Hash** the identifiers if the project purposes can be reached without them
- **Data collected on internet**[3] are still **submitted to restrictions** (art. 22), even if the subjects published them
- Data collection and processing must follow the following principle: good faith, lawfulness, proportionality, exactitude, security
- **Pseudomisation:** restricted access right to the **pseudomisation key** must be implemented. Besides, the **risk of reidentification** must be assessed
- Anonymized data received from a **third party** still require the subject to be informed of this new use
- Research conducted on **human being** must comply with the Human Research Act
- Legal consent for person under 18 years is required to collect their data

THE EU GENERAL DATA PROTECTION REGULATION *(GDPR)*

"This Regulation applies to the processing of personal data of data subjects who are in the [European] Union" (art. 3).
Several derogations are available in the case of scientific and historical research (art. 89).
Example of a GDPR summary[2].

The additional following items are **mandatory to comply with the GDPR** (non exhaustive) :

- A description of how the following principles will be implemented (art. 5): Lawfulness, Data Minimization, Accuracy, Storage Limitation, Integrity, Transparency, Privacy-by-design, Confidentiality and Accountability
- If the data processing and storage are **outsourced**, documentation about the GDPR compliance of the external services is required
- **Inform the subjects** about their rights to modify their data, restrict the use of their data and withdraw their participation (chapter 3)
- **Privacy by design** (data protection as a priority, data minimization, pseudomization, etc.)
- A **Data Protection Impact Assessment** (DPIA) if the project may result in a high risk. High risk project may involve data processed on a large scale, innovative use of the data, sensitive data, vulnerable subjects, data transfers outside of the EU, etc.

Credits and sources

[1] https://www.admin.ch/opc/en/classified-compilation/20061313/index.html

[2] https://gdprexplained.eu/

[3] https://www.edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/Internet_und_Computer/services-en-ligne/medias-sociaux.html

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #09

DATA MASKING

EPFL Library
Research Data Management
FAST GUIDES

ADVANTAGES

WHY IT'S WORTH

- Complies with law
- Makes data sharable
- Prevents data misuse
- Makes data publishable

APPLICABILITY

TESTS ON HUMANS / SENSITIVE DATA

- Name, identification number, location data, online identifier, etc.
- Factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity

TECHNIQUES

PSEUDONYMIZATION

REVERSIBLE
(FOR WORKING DATA)

REPLACING
Replace data by identifiers. The key is stored separately and securely.

ENCRYPTING
Encrypt the data and store the key securely. Appropriate for long-term preservation, not for data publishing.

ANONYMIZATION

IRREVERSIBLE
(FOR PUBLISHED DATA)

GENERALIZING
Diminish granularity by generalizing the variables. Appropriate for data too specific or unique records.

SHUFFLING
Shuffle data over one / several columns without compromising their utility.

FAKING
Prevent the identification of specific records, adding fake data while preserving correlations.

REMOVING
Suppress data or part of the outlier records. Appropriate for processing identifiers.

3RD PARTY DATA

Using commercial datasets or collaborationg in a joint research? Then, define a **contract** for data sharing or publication, and distinguish between research **authors** and data **owners**.

HINT

Mitigate the identification risk, but preserve the data utility for research.

SOME TOOLS

TO MASK IDENTITY OR ASSESS IDENTIFICATION RISKS

- ARX Data Anonymization Tool [Java][1]
- Amnesia [online][2]
- ARGUS [Java][3]
- sdcMicro [R][4]
- Differential privacy queries [SQL][5]
- Faker [Python][6]

SUPPORT AND LAWS

- Human Research Ethics Committee[7]
- Federal Act on Data Protection[8]
- Human Research Act[9]
- GDPR[10]

Credits and sources

[1] arx.deidentifier.org / [2] amnesia.openaire.eu / [3] qosient.com/argus/anonymization.shtml / [4] cran.r-project.org/web/packages/sdcMicro/index.html
[5] github.com/uber/sql-differential-privacy / [6] faker.readthedocs.io/en/master / [7] research-office.epfl.ch/ethical-legal-review/epfl-hrec
[8] admin.ch/opc/en/classified-compilation/19920153/index.html / [9] admin.ch/opc/en/classified-compilation/19920153/index.html
[10] eur-lex.europa.eu/eli/reg/2016/679/oj / [Icons] cran.r-project.org/web/packages/sdcMicro/index.html

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #10

STORAGE, PUBLICATION AND PRESERVATION

EPFL Library
Research Data Management
FAST GUIDES

RESEARCH DATA

- Raw Data
- Processed Data
- Metadata
- Codes / Algorithms
- Virtual machines

STORAGE

- NAS
- Cloud solutions
- Local servers
- Shared databases
- ELN / LIMS
- Data management system

PUBLISHING

- Data papers
- Journals servers
- Data repositories
- Preprints
- Data citation mechanisms

PRESERVATION

- Data repositories
- Cold data
- Post-processed, curated data
- Archive-ready format converted files
- Certified, standardized Archival Management System

STAKEHOLDERS

- Teams
- Institutions
- Funders
- Research partners
- Private partners
- Research and scientific IT services providers

Publishing and deposit conditions

- Data ownership
- Stakeholders consent
- Compliance with protection laws
- Ensuring data integrity
- Providing appropriate metadata
- Clarifying reuse licensing
- Setting up embargoes and sampling rules, if needed

Preserving criteria

- Historical and scientific data value
- Data quality and uniqueness
- Reliability of sources
- Data preparation cost
- Repository and maintenance cost
- Deposit responsibility

How long to preserve?

- At least 10 years for the SNSF
- Evaluate preserving criteria
- Mind the retention and disposal schedules
- Stick to administrative and legal stakeholders requirements

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHEQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FAST GUIDE #11

DMP - Data Management Plan

EPFL Library
Research Data Management
FAST GUIDES

WHY A DMP?

COMPLIANCY Requested by research funders (public or private), a DMP enhances research reproducibility and the use of public funds.

TRANSPARENCY Usually published when the funding period ends, a DMP completes the research results with the information on data, software, protocols, sources, etc.

FORECAST To anticipate costs (materials and software) and identify risks (eg. data loss, incompatible formats, security). DMPs allow institutions to better allocate services.

STREAMLINE To reduce risks of data loss and the efforts of reverse engineering for new collaborators. A DMP boosts data reuse in the lab and outside.

Target the reproducibility of research results!
Anticipate questions about data in your projects.

WHAT'S IN A DMP?

DESCRIPTION Data types, formats, size.

COLLECTION Sources, experiments, analysis, simulations.

CURATION Metadata, naming, datasets structures.

STORAGE Active data, sharing tools, preservation.

RISKS Access rights, anonymization, ethics assessment.

PUBLICATION Data licenses, data repositories, IP.

COSTS For RDM: refer to Fast Guide #03.

Not just administrative hurdle!
Use your DMP as a reference tool during the data life-cycle.

WHEN A DMP?

IDEALLY At the conception of your research project.

USUALLY When requesting funds.

REALLY ASAP, but it is never too late.

The DMP is a living document!
Keep it up-to-date throughout the project.

A DMP IS...

... a written document describing how data of a research project is managed during the life-cycle.

FUNDERS REQUIRING A DMP

- SNSF
- H2020 (ERC, FET, MSCA, ...)
- EPFL (some internal projects)
- AXA Research Fund
- U.S. Federal Grants
- Wellcome Trust
- Ligue vaudoise contre le cancer

DOWNLOAD DMP TEMPLATES

- **SNSF DMP** [1]
  A template based on SNSF Open Research Data Policy, with added guiding examples.
- **ERC DMP** [2]
  A template based on the FAIR principles, with added guiding examples.
- **MSCA DMP** [3]
  This DMP form is suggested (not mandatory) for the Marie Skłodowska-Curie actions' applicants.
- **NCCR RDM STRATEGY** [4]
  UPCOMING: EPFL Library RDM team works on such a template right now.

[1] SNSF DMP template: researchdata.epfl.ch/wp-content/uploads/EPFL_Library_SNSF_DMP_Template.odt - [2] ERC DMP template: researchdata.epfl.ch/wp-content/uploads/EPFL_Library_ERC_DMP_Template.odt - [3] MSCA DMP template: ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template - [4] NCCR RDM Strategy template: contact the EPFL Research Data Library TEAM for info and support - [ICONS] flaticon.com/packs/essential-set-2

Contact and info researchdata.epfl.ch researchdata@epfl.ch

BIBLIOTHÈQUE EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE