



The University of  
**Nottingham**

UNITED KINGDOM · CHINA · MALAYSIA

Mortier, Richard and Haddadi, Hamed and Henderson, Tristan and McAuley, Derek and Crowcroft, Jon and Crabtree, Andy (2016) Human-data interaction. In: The encyclopedia of human-computer interaction. Interaction Design Foundation. ISBN 9788792964007

**Access from the University of Nottingham repository:**

<http://eprints.nottingham.ac.uk/38638/1/Human-Data%20Interaction.pdf>

**Copyright and reuse:**

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution No Derivatives licence and may be reused according to the conditions of the licence. For more details see: <http://creativecommons.org/licenses/by-nd/2.5/>

**A note on versions:**

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact [eprints@nottingham.ac.uk](mailto:eprints@nottingham.ac.uk)

# 41. Human-Data Interaction

by **Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, Jon Crowcroft and Andy Crabtree**

## 41.1 Introduction

We have moved from a world where computing is siloed and specialised, to a world where computing is ubiquitous and everyday. In many, if not most, parts of the world, networked computing is now mundane as both foreground (e.g., smartphones, tablets) and background (e.g., road traffic management, financial systems) technologies. This has permitted, and continues to permit, new gloss on existing interactions (e.g., online banking) as well as distinctively new interactions (e.g., massively scalable distributed real-time mobile gaming). An effect of this increasing pervasiveness of networked computation in our environments and our lives is that *data* are also now ubiquitous: in many places, much of society is rapidly becoming “data driven”.

Many of the devices we use, the networks through which they connect – not just the Internet but also alternative technologies such as fixed and cellular telephone networks – and the interactions we experience with these technologies (e.g., use of credit cards, driving on public highways, online shopping) generate considerable trails of data. These data are created both consciously *by us* – whether volunteered via, e.g., our Online Social Network (OSN) profiles, or observed as with our online shopping behaviour (World Economic Forum 2011) – and they are inferred and created *about us* by others – not just other people but, increasingly, machines and algorithms, too.



Author/Copyright holder: tara hunt. Copyright terms and licence: CC BY-SA 2.0

*We create data trails both consciously and unconsciously. Most of us are very self-conscious about what we post on our [Facebook](#) and [Twitter](#) accounts – but only a minority of people are aware that they leave a very detailed data trail in many other ways, e.g., when browsing online or walking in their cities carrying their smartphones. Increasingly, machines and algorithms are tracking every step we take – both online and offline – and we are very rarely “warned” or notified about this surveillance.*

### **41.1.1 The Evolution of Human-Computer Interaction**

We observe that Human-Computer Interaction (HCI) has grown out of and been traditionally focused on the interactions between humans and computers as artefacts, i.e., devices to be interacted with. As described by Jonathan Grudin (1990a, 1990b), a Principal Researcher at Microsoft Research in the field of HCI, the focus of work in HCI has varied from psychology (Card, Moran, and Newell 1983) to hardware to software to interface, and subsequently deeper into the organisation. This trend, moving the focus outward from the relatively simple view of an operator using a piece of hardware, continued with consideration of the richness of the inter-relationships between users

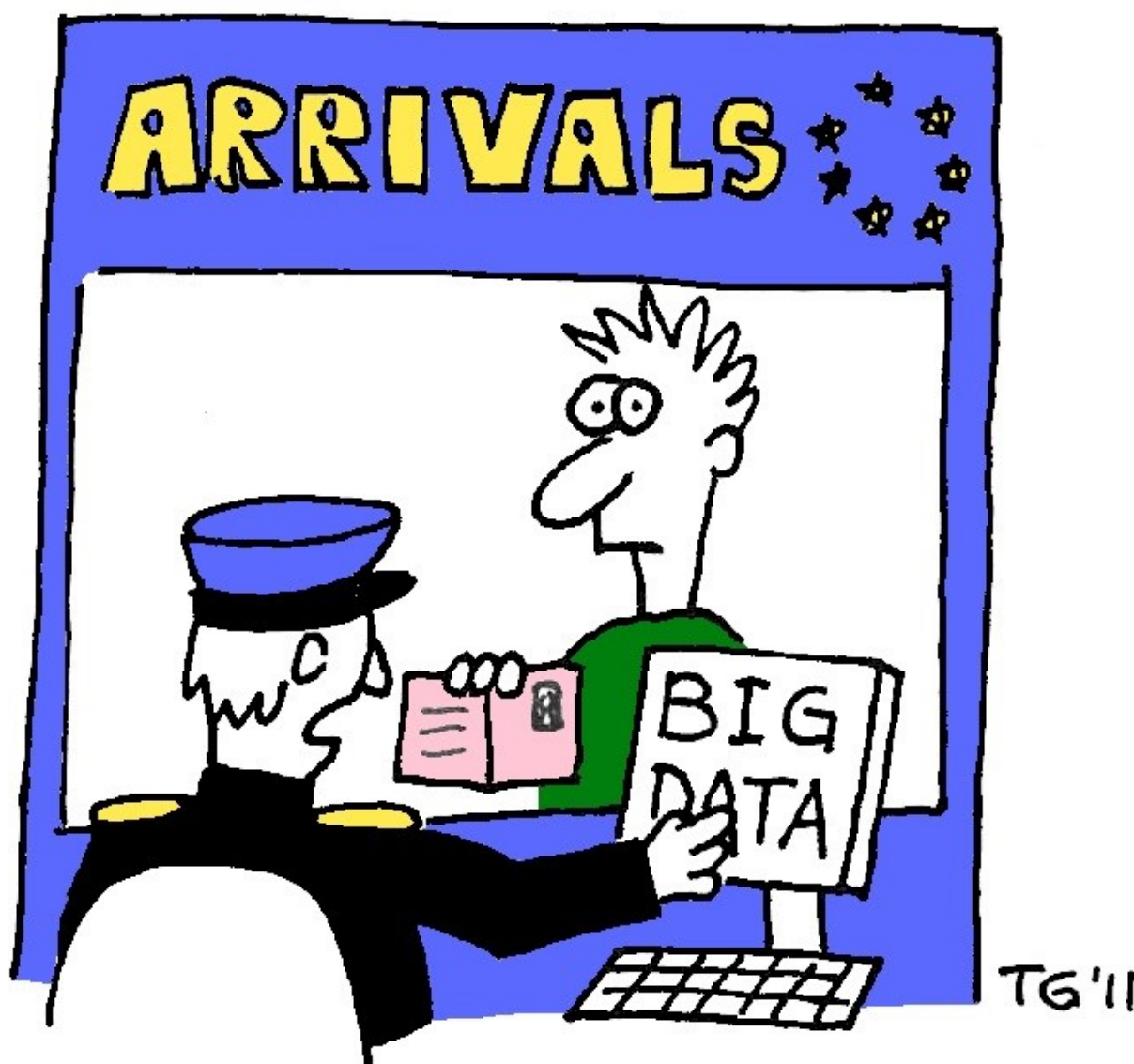
and computer systems as those systems have pervaded organisations and become networked, and thus the need to “explode the interface”, e.g., Bowers and Rodden (1993), professors at Newcastle and Nottingham Universities respectively.

### The evolution of Human-Computer Interaction:

- Psychology
- Hardware
- Software
- Multiple networked components
- Focus of interaction between users and computer systems as one heads deeper into the networked organisation.

*Author/Copyright holder: Jorge Gonzalez. Copyright terms and licence: CC BY-SA 2.0*

We believe that the continuing and accelerating trend towards truly ubiquitous and pervasiv computing points to a need to emphasise another facet of the very general topic of how people interact with computer systems: *how people should interact with data*. That is, not so much the need for us to interact directly with large quantities of data (still a relatively rare occupation), but the need for us all to have some understanding of the ways in which our behaviours, the data they generate, and the algorithms which process these data increasingly shape our lives. A complex ecosystem, often collaborative but sometimes combative (Brown 2014), is forming around companies and individuals engaging in the use of thee data. The nascent, multi-disciplinary field of Human-Data Interaction (HDI) responds to this by placing the human at the centre of these data flows, and it is concerned with providing mechanisms for people to interact explicitly with these systems and data.



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

*Author/Copyright holder: Thierry Gregorius. Copyright terms and licence: CC BY 2.0*

*We think that it's crucial to understand 1) how our behaviours, 2) how the data our behaviours generate, and 3) how the algorithms which process these data increasingly shape our lives. Human-Data Interaction (HDI) places the human at the centre of these data flows, and HDI provides mechanisms which can help the individual and groups of people to interact explicitly with these systems and data.*

In this article we will next go into more detail as to why HDI deserves to be named as a

distinct problematic (§2) before defining just what it is we might mean by HDI (§3). We will then give our story of the development of HDI to its state by the mid-2010s, starting with Dataware, an early technical attempt to enable HDI (§4). We follow this with a deeper discussion of what exactly the “I” in HDI might mean – how interaction is to be construed and constructed in HDI – and a recent second attempt at starting to define a technical platform to support HDI with that understanding in mind (§5 and §6 respectively). We conclude with a brief discussion of some exciting areas of work occurring in the second half of the 2010s that we identify (§7), though there are no doubt many more! Finally, after summarising (§8), we give a few indications of where to go to learn more (§9).

## 41.2 Why Do We Need HDI?

### 41.2.1 Life Goes On: We Still Need Privacy



*Author/Copyright holder: Sean MacEntee. Copyright terms and licence: CC BY 2.0*

*Privacy is not an outdated model. We need it more than ever.*

“One thing should be clear, even though we live in a world in which we share personal information more freely than in the past, we must reject the conclusion that privacy is an outmoded value ...we need it now more than ever.”

– Barack Obama, President of the USA (US Consumer Privacy Bill of Rights 2012)

Privacy has long remained a topic of widespread societal interest and debate as digital

technologies generate and trade in personal data on an unprecedented scale. Government and Industry proclaim the social and economic benefits to be had from personal data, against a counterpoint of a steady flow of scare stories detailing misuse and abuse of personal data. Industry efforts to quell anxiety proffer encryption as the panacea to public concerns, which in turn becomes a matter of concern to those charged with state security. Use of encryption in this way also glosses, hides or at least renders opaque a key threat to consumer or user privacy: the ability to “listen in” and stop devices “saying” too much about us. As Professor of Computer Science and Law at Stanford University Keith Winstein (2015) puts it,

“Manufacturers are shipping devices as sealed-off products that will speak, encrypted, only with the manufacturer’s servers over the Internet. Encryption is a great way to protect against eavesdropping from bad guys. But when it stops the devices’ actual owners from listening in to make sure the device isn’t tattling on them, the effect is anti-consumer.”  
– Keith Winstein

Many Internet businesses rely on extensive, rich data collected about their users, whether to target advertising effectively or as a product for sale to other parties. The powerful network externalities that exist in rich data collected about a large set of users make it difficult for truly competitive markets to form. We can see a concrete example in the increasing range and reach of the information collected about us by third-party websites, a space dominated by a handful of players, including [Google](#), Yahoo, Rubicon Project, Facebook and Microsoft (Falahrastegar et al. 2014, 2016). This dominance has a detrimental effect on the wider ecosystem: online service vendors find themselves at the whim of large platform and Application Programming Interface (API) providers, hampering [innovation](#) and distorting markets.

### 41.2.2 The Paradox of Privacy: The More We Reveal, the More Privacy We Desire

Personal data management is considered an intensely personal matter however: e.g., professor of Informatics Paul Dourish (2004) argues that individual attitudes towards personal data and privacy are very complex and context dependent. Studies have shown that the more people disclose on [social media](#), the more privacy they say they

desire, e.g., Taddicken and Jers (2011), of the Universities of Hamburg and Hohenheim respectively. This paradox implies dissatisfaction about what participants received in return for exposing so much about themselves online and yet, “they continued to participate because they were afraid of being left out or judged by others as unplugged and unengaged losers”. This example also indicates the inherently social nature of much “personal” data: as Andy Crabtree, Professor of Computer Science at the University of Nottingham, and Richard Mortier, University Lecturer in the Cambridge University Computer Laboratory (2015) note, it is impractical to withdraw from all online activity just to protect one’s privacy.

Context sensitivity, opacity of [data collection](#) and drawn inferences, trade of personal data between third parties and data aggregators, and recent data leaks and privacy infringements all motivate means to engage with and control our personal data portfolios. However, technical constraints that ignore the [interests](#) of advertisers and analytics providers, and so remove or diminish revenues supporting “free” services and applications, will fail (Vallina-Rodriguez et al. 2012; Leontiadis et al. 2012).

### **41.2.3 The Internet of Things Reshaped the Nature of Data Collection: From Active to Passive**

The Internet of Things (IoT) further complicates the situation, reshaping the nature of data collection from an active feature of human-computer interaction to a passive one in which devices seamlessly communicate personal data to one another across computer networks. Insofar as encryption is seen as the panacea to privacy concerns – and it is not: consumer data remains open to the kinds of industry abuses that we are *all* becoming increasingly familiar with – this gives rise to “walled gardens” in which personal data is distributed to the cloud before it is made available to end-users. Open IoT platforms, such as Samsung’s ARTIK, do not circumvent the problem either: they are only open to developers. This is not an IoT specific objection. However, IoT throws it into sharp relief: while security is clearly an important part of the privacy equation, it is equally clear that more is required.

### **41.2.4 Reclaiming Humanity: Active Players not Passive Victims of the Digital Economy**

There is need in particular to put the end-user into the flow of personal data; to make the parties about whom personal data is generated into active rather than passive participants in its distribution and use. The need to support personal data management



is reflected in a broad range of legal, policy and industry initiatives, e.g., Europe's General Data Protection Directive (European Parliament 2014), the USA's Consumer Privacy Bill of Rights (US Consumer Privacy Bill of Rights 2012) and Japan's revision of policies concerning the use of personal data (Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society 2014).

Here, issues of trust, accountability, and user empowerment are paramount. They speak not only to the obligations of data controllers – the parties who are responsible for processing personal data and ensuring compliance with regulation and law – but seek to shift the locus of agency and control towards the consumer in an effort to transform the user from a passive “data subject” into an active participant in the processing of personal data. That is, into someone who can exercise control and manage their data and privacy, and thus become an active player or participant in – rather than a passive victim of – the emerging data economy.

Having discussed why HDI is a topic that should concern us, we now turn to a more detailed discussion of just what it is that we might mean when we use the term HDI.

## 41.3 Just What is HDI?

As with most academic ventures, you might anticipate that answering the above question will not be straightforward. We believe that the richness of conceptions of data, reflected in its general definition, lead to a broad definition of HDI, e.g., the definition of data from the Oxford English Dictionary (2014):

- As a count noun: an item of information; a datum; a set of data.
- As a mass noun:
  - Related items of (chiefly numerical) information considered collectively, typically obtained by scientific work and used for reference, analysis, or calculation.
  - Computing. Quantities, characters, or symbols on which operations are performed by a computer, considered collectively. Also (in non-technical contexts): information in digital form.

When compounded with other nouns, however, it becomes more interesting:

- **Data trail: An electronic record of the transactions or activities** of a particular person, organisation, etc. Now esp. with reference to a person's financial

transactions, telephone and Internet usage, etc.

- **Data smog: A confusing mass of information**, esp. from the Internet, in which the erroneous, trivial, or irrelevant cannot be easily or efficiently separated from what is of genuine value or interest (often in figurative context); obfuscation generated by this; cf. [information overload](#).
- **Big Data**: Computing (also with capital initials) **data of a very large size**, typically to the extent that its manipulation and management present significant logistical challenges; (also) the branch of computing involving such data.

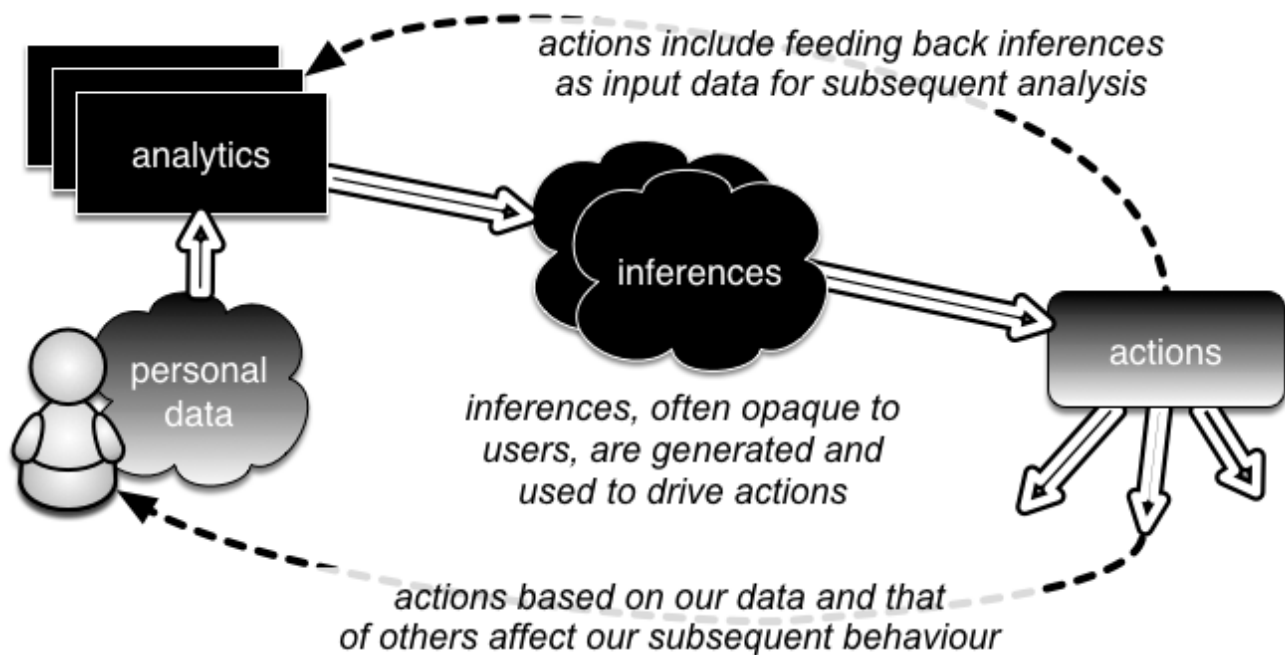
In many ways, it is the interplay between the last three definitions that we believe gives rise to the need for a broader conception of HDI: when the data trails of individuals' private behaviour are coalesced and analysed as big data; and where the results of that analysis, whether or not correct, are fed back into the data associated with an individual. Data, particularly personal data, can be seen as a *boundary object* (Star and Griesemer 1989; Star 2010), reflected in the many ways different communities refer to and think of data. For example, to [contrast](#) with big data we see data trails referred to as *small data* (Estrin 2013) where " $N = me$ ", pertaining to each of us as individuals. We see yet other terms used in other fields: *participatory data* (Shilton 2012) in health, *microdata* (Kum et al. 2014) in population informatics, and *digital footprint* (Madden et al. 2007) in the digital economy.

Looking at the literature shows several meanings already attached to the term; as of 2016, we are aware of at least five distinct versions:

- HDI is about federating disparate personal data sources and enabling [user control](#) over the use of "my data" (McAuley, Mortier, and Goulding 2011).
- HDI is about human manipulation, analysis, and sense-making of large, unstructured, and complex datasets (Elmqvist 2011).
- HDI is about processes of collaboration with data and the development of communication tools that enable interaction (Kee et al. 2012).
- HDI is about delivering personalised, context-aware, and understandable data from big datasets (Cafaro 2012).
- HDI is about providing access and understandings of data that is about individuals and how it affects them (Mashhadi, Kawsar, and Acer 2014).

While distinct, there is a connecting thread running through the different versions of HDI that suggests:

- That there is a great deal of digital data about, so much so that it might be seen as the next frontier for computing and society alike (Pentland 2012).
- That HDI is very much configured around large amounts of “personal data”, whether in terms of delivering personalised experiences or in terms of it being about individuals.
- That interaction covers a range of interrelated topics from data analytics to data [tailoring](#), and enabling access, control, and collaboration.



### we lack **legibility**

*means to inspect and reflect on "our" data, to understand its collection and processing*

### we lack **agency**

*means to manage "our" data and access to it, enabling us to act effectively in these systems as we see fit*

### we lack **negotiability**

*means to navigate data's social aspects in collaboration with others and their policies*

*Author/Copyright holder: Richard Mortier. Copyright terms and licence: CC BY-NC-ND*

*Figure 1: Data flows in the Human-Data Interaction model. We generate data which is analysed to produce inferences. These inferences in turn are fed back, affecting our behaviour and becoming themselves the subject of further analysis.*

On reading the literature, such as it is in the mid-2010s, this is still a fledgling field! – we come to the view that HDI is not about data per se then, not even digital data, but is very much centred on digital data pertaining to people and digital data that may be considered to be “personal” in nature. As McAuley (Professor of Digital Economy, Horizon Digital Economy Research), Mortier, and Goulding (Researcher at Horizon

Digital Economy Research) (2011) and Haddadi, Lecturer at Queen Mary University, et al. (2013) put it respectively,

“Modern life involves each of us in the creation and management of data. Data about us is either created and managed by us (e.g., our address books, email accounts), or by others (e.g., our health records, bank transactions, loyalty card activity). Some may even be created by and about us, but be managed by others (e.g., government tax records).”

“An ecosystem, often collaborative but sometimes combative, is forming around companies and individuals engaging in use of personal data.”

At the heart of HDI lies three core principles: legibility, agency and negotiability, set out by Mortier et al. (2014):

- **Legibility.** Premised on the **recognition** that interactions with data flows and data processes are often opaque, legibility is concerned with *making data and analytic algorithms both transparent and comprehensible to users.*
- **Agency.** The *means to manage “our” data and access to it*, agency enables us to *act effectively* in these systems, as and when we see fit. This not only includes the ability to opt-in or opt-out of data collection and processing but also the broader ability to engage with data collection, storage and use, and to understand and modify data and the inferences drawn from it.
- **Negotiability.** The *means to navigate data’s social aspects*, negotiability supports interaction between other data subjects and their policies. This enables the *ongoing engagement of users so that they can withdraw from data processing* either completely or in part, and can derive value from data harvesting for themselves.

### 41.3.1 Legibility – Enabling Data Subjects to Understand Data Concerning Them

Our interactions with online data systems are often opaque to us: there are few online analogues of physical world artefacts such as the mandatory signage required in locations covered by CCTV (closed-circuit television), where recordings are made and

monitored, typically for surveillance and security purposes. We argue that it is not enough simply to make these processes transparent: they are often technical and complex, and the implications of the data collected and processed are incomprehensible. Rather, we believe that they must be made *legible*, able to be understood by the people they concern. This is a prerequisite for our ability to exercise agency consciously in situations where personal data is being collected and processed. Already recognised in specific contexts such as consent and withdrawal (Coles-Kemp and Zabihi 2010), the need for data to be more legible is becoming pervasive as society becomes more data-driven.



Author/Copyright holder: Jorge Gonzalez. Copyright terms and licence: CC BY-SA 2.0

*It is not enough simply to make online data system processes visible or transparent. Their technical complexity and the manifold implications of the data collected and processed must also be made legible so that the people whom they concern can understand them.*

Data created about us are often less well-understood by us. For instance, third-party website tracking, when combined with recommender systems and data-mining algorithms can create new data from inferences, such as advertising preferences (Vallina-Rodriguez et al. 2012). Credit-scoring companies and “customer science” companies collect and mine shopping and transaction data to both predict and enable behaviours. Not *all* such data uses are strictly commercial, however. For instance,

personal data can be used to generate data for new crowdsourced applications such as traffic reports or optimised bus routes (Berlengerio et al. 2013). But new tools for informing people about their data, and the practices used around these data, are essential.

Data created by us arise from our interaction with numerous sensors and technologies, from what are now mundane technologies such as OSNs and websites. The richness and variety of such data, however, is continually increasing, particularly with the growing interest in lifelogging and the “Quantified Self” (Choe et al. 2014). For example, devices and sensors with which we explicitly interact when monitoring our health (e.g., continuous blood glucose monitoring, smart asthma inhalers, bathroom scales that track our weight, or smartphone apps that monitor our sleep patterns). Such devices can create “people-centric” sensor trails (Campbell et al. 2008). Related advances in portable medical sensors, affordable personal genomics screening, and other tools for mobile health diagnosis will generate new personal medical datasets (Kumar et al. 2013).

Legibility entails several features. First, we need to become aware that data is being collected, relatively straightforward to achieve as with, e.g., recent European legislation requiring that websites make clear to users when the site deposits browser cookies. The second, more complex, requirement is that we become aware of the data themselves and their implications. A data-centric view of the world requires that we pay attention to the correctness (in an objective knowledge sense) of data. In contrast, a human-centric view requires that systems allow for different *but equally valid* viewpoints of data. Similarly, interpretations of data may vary significantly over time, hence (for example) the recent Court of Justice of the European Union (2014) “right-to-be-forgotten” where public data about individuals can be removed from search engine results so that the distant past is not kept fresh in people’s minds, mirroring in some ways the natural human behaviour of forgetting once topical information.

Simply providing visualisations of data is a starting point, and a well-studied topic within HCI. However, designer with MetroMile, Chloe Fan (2013) observes that even this can pose problems due to the scale of data involved as Quantified Self app developers have found when presenting the large, detailed, rich data collected about aspects of a single individual, from physical activity to sleep [patterns](#) and diet. Zaslavsky, Principal Research Scientist at CSIRO, Perera, Research Associate with The Open University, and Georgakopoulos, Professor with RMIT, Australia (2012), note similar problems arise

with data that are inherent ambiguous such as those collected about communities through Internet-of-Things technologies. However, the potential for data visualisation to reveal aspects of the incentive models associated with the processing of data, and even the details of the processing algorithms themselves, may present more problematic challenges in a commercial [environment](#). One possible avenue is to engage with artists in attempting to make these very abstract concepts (data, algorithm, inference) legible to users (Jacobs et al. 2013, 2016).

### 41.3.2 Agency – The Capacity to Act for Ourselves within Data Systems

Empowering us to become aware of the fact and implications of the collection of our personal data is a beneficial first step. However, putting people at the heart of these data processing systems requires more: we require *agency*, the capacity to act for ourselves within these systems. In 2016, the right to be informed when personal data are collected was enshrined in legislation such as the European General Data Protection Directive. But as the intimacy, ubiquity, and importance of the personal data collected about us grows, we require a broader ability to engage with its collection, storage, and use to understand and modify raw data and the inferences drawn from them.

This is more than simply the ability to provide informed consent, though even that is often not achieved (or was as of the mid-2010s) (Ioannidis 2013; Luger, Moran, and Rodden 2013; Luger and Rodden 2013). The data collection process may have inherent biases due to contextual dependencies, temporal and other sampling biases, and simply misunderstood semantics. Inferences drawn from our personal data could be wrong, whether due to flawed algorithms, incomplete data or the way our attitudes and preferences change over time. User-centric controls are required, not only for consent but for the revocation of collected personal data (Whitley 2009).

In addition to a richer and more robust dialogue between regulators and the industry, we believe that enabling these requires stakeholders, including researchers, regulators, technologists, and industry, to establish qualitative and quantitative techniques for understanding and informing activity around human data. A survey of 1,464 UK consumers said that 94% believed that they should be able to control information collected about them (Bartlett 2012). It is worth noting that providing such abilities might also bring benefits to data collection and processing organisations as well: the

same survey reported that 65% of respondents said that they would share additional data with organisations “if they were open and clear about how the data would be used and if I could give or withdraw permission”.

Note that we do not suggest all users must become continuously engaged in the collection, management and processing of their personal data. Extensive work in the context of privacy and personal data has demonstrated such features as the privacy paradox, whereby privacy only becomes a concern after a violation (Barnes 2006), and we might reasonably anticipate that many people will not often need or desire the capacity to act within these data-collection and -processing systems. However, many will from time to time, and some enthusiasts may do so more frequently. We claim that they must be supported in doing so.

Evidence suggests that mechanisms for expressing data management, such as privacy policies, are difficult both to **design** (Trudeau, Sinclair, and Smith 2009) and to interpret (Leon et al. 2012), and so supporting users acting more broadly may prove a significant challenge. The interplay between data collectors and third-party data users introduces new challenges, both to the privacy of personal data and to the understanding of this privacy: How can we accurately measure the effects of personal data collection when the effects of this collection may span multiple entities and multiple time periods? If we cannot measure these effects, then it will be hard to convince people that they should be concerned, or that they should adopt privacy mechanisms such as differential privacy (Dwork 2006), privacy-preserving profiling and advertising schemes (Haddadi, Hui, and Brown 2010; Guha et al. 2009), or metaphors to simplify the configuration of such systems (Adams, Intwala, and Kapadia 2010; Kapadia et al. 2007).

It is also worth noting that not all activities associated with processing of personal data are harmful, and so granting users agency in these systems need not have only negative effects. Recommender systems (Ricci et al. 2010) can provide a useful function, saving us time and effort. Live traffic updates through services such as Google Maps assist us in avoiding traffic jams. Public health initiatives are often based on the aggregation of large quantities of highly personal data. The opportunity for data subjects to engage with data systems may enable them to correct and improve the data held and the inferences drawn, improving the overall quality and utility of the applications using our personal data.

### **41.3.3 Negotiability – The Ability for People to Re-evaluate**



## Their Decisions as Contexts Change

Legibility and agency are important, but we further need to allow people to re-evaluate their decisions as contexts change, externally (e.g., people and data crossing jurisdictional boundaries) and internally (e.g., [feedback](#) and control mechanisms have been shown to affect data-sharing behaviour (Patil et al. 2014). We term this *negotiability*.

Much debate around the use of personal data has assumed that data are considered a “good” that can be traded and from which economic value should be extracted (Organisation for Economic Co-operation and Development 2013). Although we agree that it may well be possible to enable an ecosystem using economic value models for utilisation of personal data and marketplaces (Aperjis and Huberman 2012), we believe that power in the system is—as of 2016—disproportionately in favour of the data aggregators that act as brokers and mediators for users, causing the apparent downward trajectory of economic value in the information age (Lanier 2013).

Effectively redressing this balance requires research to understand the *contextual integrity* (Nissenbaum 2004) of uses of our personal data, and how this impacts services and new uses of our data both for research and business (Shilton et al. 2009). Contextual effects mean that data connected with people cannot realistically be considered *neutral* or *value-free*, leading to problems with applying concepts such as the *data-driven society* or *Big Data* to individuals. Expecting people to be able to self-manage their personal, private data may be inappropriate given increased data collection (Solove 2013), and so legal and regulatory frameworks may need revisiting and readdressing (Westby 2011).

Some of these issues are already being faced by researchers carrying out experiments that use personal data. [Experiment](#) design requires careful consideration of the types of data to be used and the ways in which appropriate consent to use data can be obtained (Brown, Brown, and Korff 2010). Sharing of research data is becoming popular, and even mandated, as a mechanism for ensuring good science and the dissemination of good science (Callaghan et al. 2012). As a result, issues such as the privacy and ethics issues of sharing – and not sharing (Huberman 2012) – data are increasingly being discussed (O’Rourke et al. 2006).

Much of our presentation has focused on issues surrounding specifically personal data. The power of open data, open knowledge, and open innovation are also being

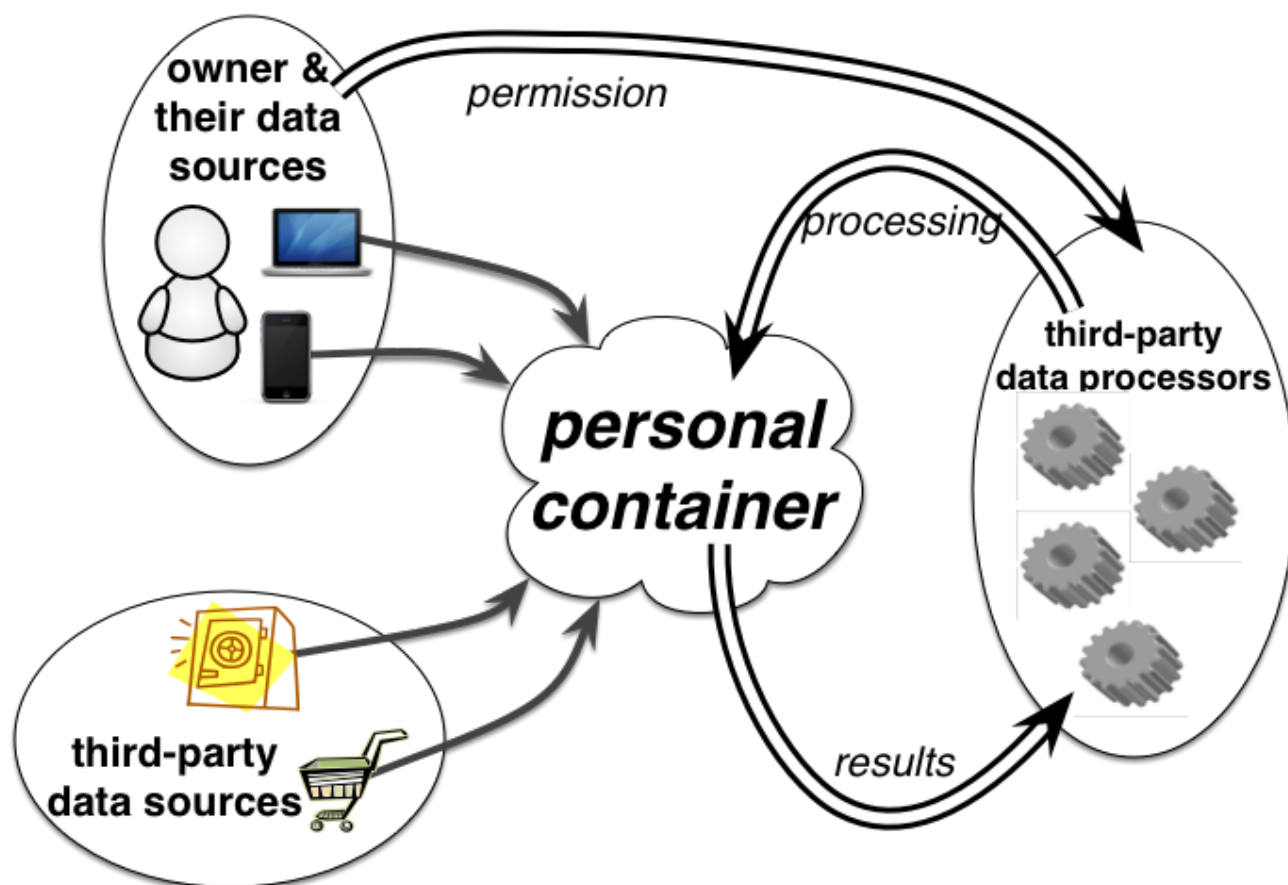
widely advocated by a number of independent organisations such as the [The Open Data Institute](#). The objective of these efforts is to free individuals and the Web from echo chambers and filter bubbles (Pariser 2011), empowering them through transparent access and audit of governments and various organisations. The underlying belief is that publishing data will help make it participatory and accessible, leading to innovation and thus benefit to the populace. Releasing data to the public, however, needs care and foresight into usage, correlation, and reputational side effects. For example, availability of crime data about a specific neighbourhood may end up reinforcing that area as a crime hub. Individuals hidden in previously anonymized, delinked personal data may become identifiable through application of newly available data (Ohm 2010). As a result, HDI needs to take into account not only personal data, but also current *and future* data.

Finally, as we build infrastructures and interfaces that enable users to understand and engage with data processing systems, we must consider how these will shape and be shaped by the ways that we reason about our data. The kinds of analogies we build and use in this reasoning will be informed by cultural and contextual differences and similarities and, in turn, will inform how we use, release, and distribute personal data in different communities and cultures.

Having discussed just what we might mean by HDI, we now turn the clock back to an early exploration of technical matters that informed the development of HDI. This provides a basis for the direction in which HDI has moved and for its current trajectory.

## 41.4 Dataware: HDI v0

The Dataware model of McAuley, Mortier, and Goulding (2011) was a very early foray into providing a particular instantiation of what later became core HDI concepts. The model is based on three fundamental types of interacting entity, depicted in Figure 2: the *owner* (or *user* or *subject*), by or about whom data is created; the *data sources*, which generate and collate data; and the *data processors*, which wish to make use of the user's data in some way.

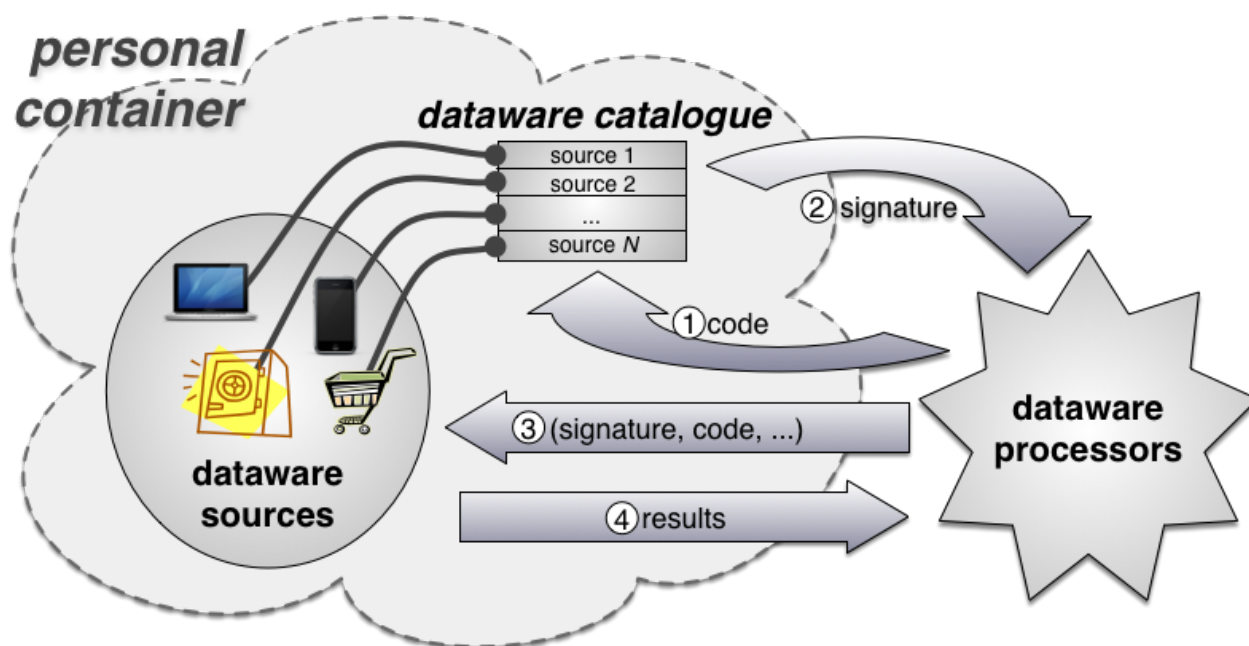


Author/Copyright holder: Richard Mortier. Copyright terms and licence: CC BY-NC-ND

Figure 2: Actors within the Dataware model: owner (or user or subject), sources, and processors, interaction among whom is mediated through the owner's personal container.

To assist the owner in managing the relationship between these entities, this model posits that the underlying technology will provide them with a *personal container* – a forerunner of the Databox, discussed later (§6) – that will enable them to oversee and manage access to their data sources and processing of their data by various data consumers. This is a logical, primarily cloud-hosted, entity formed as a distributed computing system, with the software envisaged to support it consisting of a set of APIs providing access to data held by data sources. *Data processors* would write code to use these APIs, and then distribute that code to the data sources which would take responsibility for executing it, returning results as directed by the data processor. The final and key piece of infrastructure envisaged is a *catalogue*, within which an owner would register all their data sources, and to which processors would submit requests

for metadata about the sources available, as well as requests to process data in specified ways.



Author/Copyright holder: Richard Mortier. Copyright terms and licence: CC BY-NC-ND

*Figure 3: Workflow in the Dataware architecture: requests are made of the owner's catalogue, which grants permission by signing the request. When the signed request is presented for processing at a source, the source can validate it has permission to be run.*

From a user's point of view, interaction with this model shown in Figure 2 works as shown in Figure 3: processors desiring access to one or more datasets within the catalogue present a request for access along with information about the request (minimally, a representation of the processing to be carried out); the user permits (or denies) the request, which is indicated by the catalogue returning some form of token to the processor representing granted permission; the processor subsequently presents the request (the processing to be carried out) and the token to the data sources it covers; finally, the data sources return the results of the processing as directed in the request to the data consumer. The model assumes that the catalogue and the data sources it references are governed by the user, including logging and auditing the uses made of data so that the user can retrospectively inspect what has been done, when, by whom and to what end. The model also permits a user to operate multiple catalogues, independent of each other, thereby providing a means to control

the problems of linking accounts across different sources. Interactions between such catalogues are not considered an explicit feature of the system.

Never realised to deployment, Dataware nevertheless explored some key interactional and technical issues that arise in HDI:

- The need for common, or at least self-describing, data formats, and associated means for source discovery.
- The need to support federation between an owner's many data sources, as many existing sources will remain within distinct organisations, and the associated need for identity mechanisms.
- The need for protocols that support not only resource discovery but also negotiation of permission to process data, ability for users to control the environment in which data processing executes, giving them complete control over exactly what is exfiltrated to the requesting data processor.

It is to a deeper exploration of the interactional issues that HDI throws up that we next turn.

## 41.5 Interaction: The “I” in HDI

We can thus consider dataware an attempt to build a digital infrastructure that supports [human data interaction](#) by surfacing a user's personal data sources, and what third parties would do with them or have done with them. It construes the “I” in HDI as an accountable transaction between the parties to it, configured in terms of request, permission, and audit. This certainly could transform the situation (as it stands in the mid-2010s), characterised by the largely unaccountable use of personal data by third parties. However, it does not describe how such transactions will be accountably conducted in [practice](#): on what accountable matters will requests, permissions and audits turn upon? It is towards unpacking what we mean by the accountable matters of human data interaction that we now turn.

### 41.5.1 Data as a Boundary Object

As previously noted, HDI is about digital data pertaining to people, that is considered personal in nature: it is an object-embedded-in-human-relationships, the view of data transactions within those relationships can be elaborated through the notion of *boundary objects*. To wit, HDI turns upon a “mutual modus operandi”, which involves

“communications” and “translations” that order the “flow” of information through “networks” of participants.

However, the question then arises: is the interactional arrangement Dataware provides, of request-permission-audit, sufficiently coherent to make HDI into a mundane infrastructure? On closer inspection, the answer is “no”: communications are not truly mutual. Instead, third parties—not data subjects—drive them, with interaction being done to the “users” rather than something done by them. Even where they do have the ability to refuse or remove permissions, users are dealing with one-way traffic: the Dataware modus operandi is asymmetrical. This then begs the question of what a symmetrical relationship might look like, e.g., how might users drive data sharing by (for example) actively seeking out data processors?

The situation is further complicated by the inherently cognitive character of the Dataware model: it is based on “my data” and on data “about me”, ignoring the *N*-dimensional character of human data that arises as data often relates not so much to “me” or “you” but to “us”. With this, the coherence of the “my data” model starts to break down and break down in challenging ways. It is not just a matter of handling what, for example, “you” posted on “my” Facebook page, but of handling the media we produce and consume *together*. Thus, the unit of data is not always “mine” but frequently “ours”. How is “our data” to be handled? How is social data to be catalogued and governed? The individuated model of “my” data breaks down when we start to think of how “our” data is owned, controlled and managed. It’s not enough to assign, e.g., an individual in a household to “house keep” it as can be done with day-to-day management of the home network (Tolmie et al. 2007). A host of relational issues are wrapped up in any such endeavour: the age of members of “our” cohort will shape ownership and control, as will the personal situations that members find themselves in. Who, for example, will own and control “our” children’s personal data? And what about elderly, infirm or temporally incapacitated members of “our” cohort?

Take, for example, a young child’s personal data – who owns it and who controls it? It cannot be assumed that the *same* person exercises ownership and control. Ownership may well reside with the person to whom the data applies, as it were. However, control in such a situation may well be delegated to another (e.g., a parent), thereby reflecting the organised practices of personal data handling (take, for example, a young child’s health records or bank details). The same does not apply to teenagers, however. As they develop their independence we might well expect, again in line with current

organised practices of human data interaction, that they will assume control over their own data along with a great many other aspects of their lives. Even so, this may be a phased rather than a sharp transition. The same may apply, in reverse, to an elderly member of the cohort who wishes to hand over the running of her affairs to someone else. Situated within a lively social context, and accompanied by differing relational rights and obligations, ownership and control cannot be permanently fixed and tied to an individual, as the Dataware model presumes. Instead, it will change over time with respect to a host of evolving relationships and contingencies.

In the real world, data sharing is “recipient designed” – i.e., shaped by people with respect to the relationship they have with the parties implicated in the act of sharing. What you tell people of how much you smoke or drink or what foods you eat or how much you weigh, for example, very much depends upon to *whom* you are doing the telling. Doctors know all too well, for example, that such matters are grossly underestimated when told to them. The same applies more generally; not that we grossly underestimate things but that we are selective in what we divulge about our personal lives, using the “selectivity” with respect to our relationship to the other parties involved. HDI views the recipient as the processor, which presents a particular request for computation to be carried out to the data source after it has been granted permission. While this holds true, the issue is to enable the user to design permission with respect to just *what* of the data is *available* to the processor, and to others within a particular cohort, too. Recipient design draws our attention to the need to support human judgement, decision-making and intervention in the course of HDI.

These subtleties of HDI in the social world indicate a need to develop a much more encompassing and dynamic model of human-data *interaction*. That would include possibilities for users to refuse or remove permissions to access data, and to redact data, both internally within a cohort (whether a family or some other grouping of people) and externally in our interactions with third parties. These problems, by no means exhaustive of the challenges confronting efforts to build digital infrastructures supporting HDI, suggest that there is a strong sense in which we need to factor “articulation work” into HDI.

## 41.5.2 Articulation Work in HDI

Articulation work speaks to the coordinate character of human action, to the gearing in of individual courses of action with one another. Kjeld Schmidt, a doctor of sociology

and professor of work, technology and organization at Copenhagen Business School, draws on ethnographic data to highlight several generic features of action and interaction upon which coordination turns (Schmidt 1994). These include “maintaining reciprocal awareness” of salient activities within a cooperative ensemble; “directing attention” towards the current state of cooperative activities; “assigning tasks” to members of the ensemble; and “handing over” aspects of the work for others to pick up and work on themselves. These general properties of coordinate action appear concretely in situated practices that create and sustain a “common field of work”, whether coordinating “walking” in the company of others or the “sharing” of personal data with processors. The common field of work in HDI is the catalogue of data sources that users generate. Data “sharing” is organised around the catalogue and is ostensibly coordinated through the interactional arrangement request-permission-audit. This is an insufficient arrangement when seen from the perspective of cooperative work, however, for reasons that Schmidt points out:

“...in order to be able to conceptualise and specify the support requirements of cooperative work we need to make a fundamental analytical distinction between (a) cooperative work activities in relation to the state of the field of work and mediated by changes to the state of the field of work, and (b) activities that arise from the fact that the work requires and involves multiple agents whose individual activities need to be coordinated, scheduled, meshed, integrated, etc. – in short: articulated.”  
– Kjeld Schmidt

Requests, permissions and audit logs are mechanisms of coordination within the field of work itself, but they do not articulate the field of work. They order the flow of information between users and third parties, but the flow itself stands in need of articulation. What, for example, occasions a request being made and being made in such a way for it to seem “reasonable” to a user? Consider the expectations we might ordinarily entertain and the potential responses that might attach to requests from strangers, for example. Add to the mix how we might ordinarily react to requests regarding our personal data from strangers, and it soon becomes clear that making a request is a non-trivial matter; that it requires articulation.



Thus, a key design challenge in HDI is not only one of developing appropriate mechanisms to coordinate the flow of information within the field of work, but of articulating and thus coordinating the work that makes flow *possible* as well. No such stipulation exists in HDI: neither the request nor audit function provide adequate support and with it insight into the cooperative arrangement of work between users and third parties or the status of data processing within that arrangement. Cooperative work in HDI effectively occurs within a black box. A user cannot tell then from either the request or the audit such things as where in the arrangement of work the processing of data has reached, who is doing what with it, what's going to happen next, if there are problems or issues of concern, and so on. The articulation of work is limited to who wants the data for what purposes and reviewing such information. There is then very little support within HDI as it stands for the ongoing management of relationships between the various actors implicated in personal data sharing. Again, it is hard to see on what basis HDI could become a stable socio-technical infrastructure in everyday life *without* such mechanisms.

A key challenge thus becomes one of creating computational mechanisms of interaction that build the elemental objects of articulation work into HDI to make salient dimensions of distributed action accountable to users, thereby enabling them to manage and coordinate interaction. In saying this, we are not saying that we should blindly follow prior stipulations of salient features (though it does seem that some will hold), but that we need to develop a *much better understanding* of what needs to be *articulated* with respect to personal data sharing and the cooperative work arrangements implicated in it.

The same applies to the field of work itself. Schmidt points out that the distributed activities of a cooperative work arrangement are articulated with respect to objects within the field of work itself (e.g., data sources within the catalogue). A key issue here revolves around the 'conceptual structures and resources' that order the field of work, enabling members of a cooperative ensemble to make sense of it and act upon it. Again the question of interactional adequacy arises when we ask what conceptual structures HDI provides? It's not that it doesn't provide any, but the terms on which it does so are problematic from an interactional perspective.

Take, for example, the Dataware catalogue. It is conceptually ordered in terms of 'tables' that render data sources intelligible in terms of accounts, applications, installs, and services, etc. The problem in this is that the conceptual structure of HDI as

instantiated in Dataware is rendered in terms of the underlying technology, rather than in terms of what is being done through that technology, such as the processing of biological data as part of a healthcare regime. The problem thus involves ordering the field of work such that it reflects the work-being-done, or the work-to-be-done, rather than the underlying technical components of that work. It is hard to see, then, how users can articulate their distributed activities with respect to objects in the field of work when those objects (data sources) lack legibility or intelligibility to the broader populace in contrast to computer scientists and software engineers. Other, more 'user friendly' – and more pointedly, data-relevant and service-specific – conceptual structures and resources are required.

### 41.5.3 Interactional Challenges in Articulating HDI

Articulating both the field of work and the cooperative arrangements of work implicated in HDI stand as two key challenges confronting HDI. We have seen that a mutual *modus operandi* is not in place and that the user whose data is being purposed by others does *not* have reciprocal opportunities for discovery. We have seen that data is not only 'mine' but 'ours' and thus is social in character. We have seen that ownership and control are not isomorphic and that the life world drives the dynamics of these aspects of interaction. We have seen that data sharing is recipient designed. And we have seen that, in short, the conceptual structures and resources ordering the field of work lack legibility, intelligibility, and accountability. Each of these problems is an inherent feature of the field of work in HDI and presents challenges to its ongoing articulation.

#### 41.5.3.1 User-driven Discovery

What exactly should be made discoverable, and what kinds of control can users exercise over the process of discovery? These issues prospectively turn upon the articulation of metadata about a user's personal data sources, ranging (for example) from nothing more than articulating where a user's catalogue or catalogues can be contacted to more detailed information concerning a catalogue's contents. The demands of articulation work place further requirements on this process. Even if users are willing to publish metadata about their data, some means of understanding *who* is interested in discovering it may well be needed so as to build trust into the process. This could involve providing analytics into *which* processors are interested, when, how often, etc. Such analytics might provide users with resources enabling them to decide what of their data to expose or hide, though discovery may also turn in important

respects upon other aspects of access control including defining pre-specified policies on who can and can't discover their data.

The issue of how users might drive the discovery process (finding data processors for themselves, whether for personal, financial or social purposes) is more problematic. We will soon discuss early thoughts on how this might be addressed (§6), and turn upon making discovery of data processors much like discovering new apps in app stores. Users are familiar with and make a conscious choice to visit app stores, where they are provided with rich metadata about apps and app authors that shapes their decision-making. Data processors could be 'vetted', much like apps in the iTunes Store, and progressively more detailed information about processing could be provided, much like app permissions in the Google Play Store. In addition, the social aspects of app stores also play an important role in the discovery process: user ratings and [social networking](#) links help build the trust between users and service providers that is essential in the discovery and adoption of new technologies.

### 41.5.3.2 From My Data to Our Data

The social challenges of data ownership and control make it necessary to consider how users can collate and collaboratively manage individual and collective data sources. Individuals will need resources that enable them to control their own personal data sources as well as resources that allow them to delegate control of data sources and catalogues to others such that (for example) "I" can assign control of "my" data sources to "you". How ownership and control relationships are represented within and between catalogues, and what mechanisms will be needed to provide adequate support for their ongoing articulation, is an open matter. Even so, transparency/awareness will be an important matter to consider along with rights management. The creation and curation of collective data sources is an equally challenging matter. Although this may appear trivial – for example, energy consumption data might relate as it does now to the household rather than specific individuals, with no complex identity and management issues involved – *purposing* such data is anything but a trivial matter. Who has the right to view and share such data? Who can edit it or revoke its use? Who actually owns and controls it? One view might be to default to the bill payer, but not all collective data sources are necessarily premised on contractual relationships. Add to the mix a world in which personal data harvesting becomes increasingly associated with the things with which we mundanely interact, and the possibility of opening up both collective and individual behaviours to

unprecedented scrutiny through data analytics becomes a real and problematic prospect. The inherent tension between individual and collective data will require the development of group management mechanisms that support negotiated data collection, analysis and sharing amongst a cohort.

### 41.5.3.3 The Legibility of Data Sources

Both the individual and negotiated production, analysis and sharing of personal data turn upon data sources being legible to users. For users to have agency – that is, the ability to exercise control – within an HDI system in any meaningful way, data sources must provide a minimum level of legibility as to what data they contain, what inferences might be drawn from that data, how that data can be linked to other data, and so on. Without some means to present this critical information, preferably in some form that can be standardised, users will find it hard even to begin to understand the implications of decisions they may make and permissions they give for processing of their data.

As part of this, it is key that users are not only able to visualise and inspect the data held by a source, but that they can also visualise and thus understand just what a data processor wants to take from a source or collection of sources and why – that just what is being ‘shared’ is transparently accountable to users, which may also involve making external data sources (e.g., consumer trends data) visible so that users understand just what is being handed over. Coupled to this is the need to enable recipient design by users. There are two distinct aspects to this. One revolves around enabling users to edit data, redacting aspects of the data they do not wish to make available to others both within a cohort and outside of it. The other revolves around controlling the presentation of data to processors when the accuracy of data needs to be guaranteed (e.g., energy consumption readings).

In summary, the challenges of articulating personal data within HDI are not settled matters. Rather, they open a number of thematic areas for further investigation, elaboration and support:

- *Personal data discovery*, including meta-data publication, consumer analytics, **discoverability** policies, identity mechanisms, and app store models supporting discovery of data processors.
- *Personal data ownership and control*, including group management of data sources, negotiation, delegation and transparency/awareness mechanisms, and rights management.

- *Personal data legibility*, including visualisation of what processors would take from data sources and visualisations that help users make sense of data usage, and recipient design to support data editing and data presentation.
- *Personal data tracking*, including real time articulation of data sharing processes (e.g., current status reports and aggregated outputs), and data tracking (e.g., subsequent consumer processing or data transfer).

Each of these themes stand in need of interdisciplinary investigation and elaboration. This includes ethnographic studies of current practices of individuals and groups around personal data creation and curation, co-designed interventions to understand future possibilities, and the engineering of appropriate models, tools and techniques to deliver the required technologies to support the complex processes involved in HDI and mesh the articulation of personal data with the organised practices of everyday life. What this amounts to in many respects is a call to the broader HCI community to engage with the study and design of boring things – infrastructures – for personal data is embedded within them: in health infrastructures, communication infrastructures, financial infrastructures, consumption infrastructures, energy infrastructures, media infrastructures, etc. It is a call to study and build HDI around the unremarkable ways in which personal data is produced and used within the manifold infrastructures of everyday life. Doing so, we might understand how personal data is accountably traded within human relationships and thereby develop actionable insights into what is involved in articulating those relationships in the future.

Thus, the analysis that discussion of the Dataware proposal generated, followed by the engagement of a more interactional lens on the problem, led to considerable development of our conception of HDI. In turn, this has generated considerable refinement of what a technical platform in support of HDI might be – we discuss one such proposal next.

## 41.6 Databox: HDI v1

Dataware focused on a computational model for processing of personal data – by moving code to data, the problems associated with release of data to third parties could be avoided. However, it failed to consider in any detail the numerous interactional challenges identified through consideration of the HCI literature and the concepts of boundary object and articulation work, discussed in the preceding section. Informed by that consideration, our current work related to HDI is concerned with development of

infrastructure technology to provide for HDI in supporting individuals (in the first instance) in management of their personal data. This effort refines the initial concept of a cloud-hosted, online Personal Container into a *Databox* (Haddadi et al. 2015). Your Databox is a physical device, supported by associated services, that enables you to coordinate the collection of your personal data, and to selectively and transiently make those data available for specific purposes. Different models are supported that will enable you to match your data to such purposes, from registration with privacy-preserving data discovery services so that data processors can find your Databox and request from you access to data it holds, to app stores in which you can search for data processing applications that you wish to provide with access to your data via your Databox. Its physicality offers a range of affordances that purely virtual approaches cannot, such as located, physical interactions based on its position and the user's proximity.

It is worth noting that we do not envisage Databoxes entirely replacing dedicated, application-specific services such as Facebook and Gmail. Such sites that provide value will continue receiving personal data to process in exchange for the services they offer. Nor is the Databox oriented solely to privacy and prevention of activities involving personal data. Rather, it is explicitly intended to enable new applications that combine data from many silos to draw inferences presently unavailable. By redressing the extreme asymmetries in power relationships in the current personal data ecosystem, the Databox opens up a range of market and social approaches to how we conceive of, manage, cross-correlate and exploit "our" data to improve "our" lives. What features must a Databox provide to achieve these aims? We answer in four parts: it must be a *trusted platform* providing facilities for *data management* for the data subjects as well as enabling *controlled access* by other parties wishing to use their data, while *supporting incentives* for all parties.

### 41.6.1 Trusted Platform

Your Databox coordinates, indexes, secures and manages data about you and generated by you. Such data can remain in many locations, but it is the Databox that holds the index and delegates the means to access that data. It must thus be highly trusted: the range of data at its disposal is potentially far more intrusive – as well as more useful – when compared to data available to traditional data silos. Thus, although privacy is not the primary goal of the Databox, there are clear requirements on the implementation of the Databox to protect privacy (Haddadi, Hui, and Brown 2010).

Trust in the platform requires strong security, reliable behaviour and consistent availability. All of the Databox's actions and behaviours must be supported by pervasive logging with associated tools so that users and (potentially) third-party auditors can build trust that the system is operating as expected and, should something unforeseen happen, the results can at least be tracked. We envisage such a platform as having a physical component, perhaps in the form-factor of an augmented home broadband router, under the direct physical control of the individual. Thus, while making use of and collating data from remote cloud services, it would also manage data that the individual would not consider releasing to any remote cloud platform.

### 41.6.2 Data Management

A Databox must provide means for users to reflect upon the data it contains, enabling informed decision-making, particularly about whether to delegate access to others. As part of these interactions, and to support trust in the platform, users must be able to edit and delete data via their Databox as a way to handle the inevitable cases where bad data is discovered to have been inferred and distributed. This may require means for the Databox to indicate this to third parties. Similarly, it may be appropriate for some data not to exhibit the usual digital tendency of a perfect record. Means to enable the Databox automatically to forget data that are no longer relevant or have become untrue may increase trust in the platform by users (Mayer-Schonberger 2009), though determining those characteristics automatically may be difficult. Even if data has previously been used, it may still need to be "put beyond use" (Brown and Laurie 2000). Concepts such as the European Union's "Right to be Forgotten" require adherence to agreed protocols and other forms of cooperation, by third-party services and data aggregators. The Databox can be used as a central point for negotiating such data access and release rights.

### 41.6.3 Controlled Access

Users must have fine-grained control over the data made available to third parties. At the very least, the Databox must be selectively queryable, though more complex possibilities include supporting privacy-preserving data analytics techniques, such as differential privacy (Dwork 2006) and homomorphic encryption (Naehrig, Lauter, and Vaikuntanathan 2011). A key feature of the Databox is its support for revocation of previously granted access. In systems where grant of access means that data can be copied elsewhere, it is effectively impossible to revoke access to the data accessed. In contrast, a Databox can grant access to process data locally *without* allowing copies to

be taken of raw data unless that is explicitly part of the request. Subsequent access can thus easily be revoked (McAuley, Mortier, and Goulding 2011). A challenge is then to enable users to make informed decisions concerning the impact of releasing a given datum as this requires an understanding of the possible future information-states of all third parties that might access the newly released datum. One way to simplify this is to release data only after careful and irreversible aggregation of results to a degree that de-anonymisation becomes impossible. More complex decisions will require an on-going dialogue between the user and their Databox, to assist in understanding the impact of their decisions and even learning from those decisions to inform future behaviour.

#### **41.6.4 Supporting Incentives**

A consequence of the controlled access envisioned above is that users may deny third-party services access to data. The Databox thus must enable services alternate means to charge the user: those who wish to pay through access to their data may do so, while those who do not may pay through more traditional financial means. One possible expression of this would be to enable the Databox to make payments, tracing them alongside data flows to and from different third-party services made available via some form of app store. Commercial incentives include having the Databox act as a gateway to personal data currently in other silos, and as an exposure reduction mechanism for commercial organisations. This removes their need to be directly responsible for personal data, with all the legal costs and constraints that entails, instead giving control over to the data subject. This is particularly relevant for international organisations that must be aware of many legal frameworks. A simple analogy is online stores' use of payment services (e.g., PayPal, Google Wallet) to avoid the overhead of Payment Card Infrastructure compliance.

This, then, is where HDI stands in the mid-2010s: a nascent field with some exciting possibilities for both technical development and human study. Next, we outline just a few of these.

### **41.7 Future Directions: What's Next?**

The principles of HDI underscore the need to develop a user-centric platform for personal data processing in the 21st century. While in its infancy, it is increasingly clear that HDI poses a broad range of challenges that are only now beginning to be



elucidated e.g., Crabtree and Mortier (2015). Many arise from engineering decisions taken early in the life of the Internet, where many features were eschewed in favour of building something that worked (Clark 1995). Thus, application (and higher) layer data flows are not a thing with which the Internet is concerned. The focus was and is on moving data packets between network interfaces and supporting delivery of those packets to the correct application. It is hard to envision completely redesigning the entire basis of the Internet at this late stage. However, a number of discrete challenges are key to putting HDI's principles into practice.

### **41.7.1 Accountability**

The potential efficacy of HDI fundamentally turns upon opening the Internet up as it were and making it accountable to users. What we mean by this is that at the network layer, the Internet only really supports accounting to the extent required for settlement between Internet Service Providers (ISPs), such as counting the number of bytes exchanged over particular network interfaces to enable usage-based billing. With the kinds of intimate data the IoT is envisioned to make available, this low-level "bits and bytes" accounting will be completely inadequate. It will be necessary to surface what data devices generate, how that data is recorded and processed, by whom, where it flows to, etc. This metadata must be made visible to users to enable legibility, agency and negotiability without infringing users' privacy.

### **41.7.2 Personal Infrastructures**

The advent and growth of the IoT, coupled with the lack of facility for easily managing ensembles of network-connected devices (at least, as it stands in the mid-2010s), increases the likelihood that we will suffer harm by leaking intimate information. There is need to complement the opening up of the Internet with the development of personal infrastructures that enable users to manage the flow of data.

One possible approach might be to provide smarter home hubs that support a range of interfaces and control points developed for specific purposes. Another is to support users in building their own infrastructure to a far greater extent than is possible today. Instead of relying on others (e.g., ISPs) to provide, configure and manage infrastructure to support users, we might seek to make it straightforward for users to create their own infrastructure services, configuring and managing facilities such as firewalling, virtual private networks, DNS and other services.

### 41.7.3 Resilience

Resilience is a key ingredient in the mix between the Internet, personal infrastructures, and IoT applications in critical domains, such as health and well-being or smart-device energy management. In short, we might ask what happens to such applications when the Internet goes down (e.g., when the local access router dies or there is a problem at the local exchange)? There is a critical need to build resilience into IoT infrastructures if we are to rely upon applications in critical domains.

One possible solution is to build IoT infrastructure into the local [physical environment](#) – e.g., into the fabric of the home – to provide the necessary fallback. This might be complemented by formal modelling techniques to enable the “in house” management of complex networked systems of “dumb” devices. That, in turn, raises the challenge of how users are to understand such techniques and interact with them to ensure quality of service and the ongoing protection of privacy in the face of contingency.

### 41.7.4 Identity

As Peter Steiner put it in a cartoon in *The New Yorker* (1993), “On the Internet, nobody knows you’re a dog”. Identity touches all aspects of HDI and requires that meaningful statements can be made about just who has access to a user’s data. The Internet, being concerned with moving packets between network interfaces, provides no inherent support for higher-level expressions of identity. Application layer means of supporting identity do exist – e.g., TLS client certificates and PGP public keys – but they are *very* complex to manage. Specific challenges here include how to ensure the availability of the necessary “secrets” (keys, certificates) on all devices that may be used to access relevant data; how to support the management of data corresponding to multiple identities held by a user; and how to handle the revocation of access.



Author/Copyright holder: Peter Steiner. Copyright terms and licence: Fair Use.

*"On the Internet, nobody knows you're a dog" is an adage which began as a cartoon caption by Peter Steiner and published by The New Yorker on July 5, 1993.*

## 41.7.5 Dynamics

Devices generating data change context as they are shared between individuals, and individuals change context as they move around in space and time. Applications and services will come and go as well. Enabling users to be aware of and to manage the dynamics of ongoing data processing – who or what has access to which data, for which purposes, etc. – is a critical challenge to the sustained harvesting of personal data. That ongoing data harvesting will be dynamic and will potentially implicate multiple parties (users and data consumers) also raises the challenge of understanding the dialogues that are needed to sustain it; particularly the “work” these dialogues need to support and how they should be framed, implemented and maintained.

## 41.7.6 Collaboration

Systems developed to support personal data management typically focus on the individual. But personal data rarely concerns just a single person. It is far more common for sources of personal data to conflate information about multiple individuals, who may have different views as to how personal it is. For example, smart metering data gives a household’s energy consumption in aggregate, and different

household members may want that data to be shared with data consumers at different levels of granularity. Supporting the collaborative management and use of personal data is another critical ingredient in the mix, all of which trades on making the data and data processing legible and putting the mechanisms in place that enable users to exercise agency and negotiability *locally* amongst their own cohorts as well as *globally*.

## 41.8 The Take Away



*“Remember when, on the Internet,  
nobody knew who you were?”*

*Author/Copyright holder: Peter Steiner. Copyright terms and licence: Fair Use.*

*“Remember when, on the Internet, nobody knew who you were?” is a play by Kaamran Hafeez on the famous Steiner cartoon, also published in The New Yorker, on February 16, 2015.*

So, in such a complex and emerging field, what should you take away? The cartoon above gives one key takeaway: the simple fact that we *do* live in a complex, increasingly data-driven world, and this is the case whether or not we understand or care. The aim of HDI as a research agenda is to bring this fact to the fore, to provoke

engagement from many parties to address the challenges we believe this raises. We hope that the framing of these debates as Human-Data Interaction, and the core principles we claim are at the heart of HDI, will assist and encourage researchers in many fields – including Computer Science, Law, Sociology, Statistics, Machine Learning among many others – to engage with the challenges and opportunities posed by our collective data driven future.

## 41.9 Where to Learn More?

As a nascent field, HDI is still very much under development – there are no books! However, there is a growing community of people interested in pushing forward its development, at:

<http://hdiresearch.org/>, and there are a number of ad hoc workshops and other activities occurring under various banners, e.g., in the UK, the [Alan Turing Institute](#) and the [IT as a Utility Network+](#).

There have also been press articles which garnered some interest in their comments sections, giving some small sampling of public responses to privacy and HDI, e.g.,

- Murphy 2014: <http://www.nytimes.com/2014/10/05/sundayreview/we-...>
- MIT Technology Review 2015: <http://www.technologyreview.com/view/533901/the-emerging-...>
- Kellingley 2015: <https://www.interaction-design.org/literature/arti...> human-data-interaction-hdi-the-new-information-frontier
- Naughton 2015: <http://www.theguardian.com/technology/2015/feb/01/control-personal-data-databoxend-user-agreement>.

## 41.10 Acknowledgements

This article grows out of work funded by several agencies including RCUK grants Horizon Digital Economy Research (EP/G065802/1), Privacy By Design: Building Accountability into the Internet of Things (EP/M001636/1), CREATE (AH/K000179/1), Databox (EP/N028260/1) and IT as a Utility Network+ (EP/K003569/1); and the EU FP7 User Centric Networking grant No. 611001. As well as thanking the HDI community (<http://hdiresearch.org>) for their ongoing engagement and input, we particularly thank Kuan Hon, Yvonne Rogers, Elizabeth Churchill, Ian Brown, Laura James, Tom Rodden, members of the QMUL Cognitive Science research group, and attendees at the IT-as-a-

Utility Network+ Human-Data Interaction workshop (October 2nd, 2013) for their input.

## 41.11 References

- Adams, Emily K., Mehool Intwala, and Apu Kapadia. 2010. "MeD-Lights: a usable metaphor for patient controlled access to electronic health records." In Proceedings of the 1st ACM International Health Informatics Symposium, 800–808. IHI '10. Arlington, Virginia, USA: ACM. isbn: 978-1-4503-0030-8. doi:10.1145/1882992.1883112.
- Aperjis, Christina, and Bernardo A. Huberman. 2012. "A Market for Unbiased Private Data: Paying Individuals According to their Privacy Attitudes." First Monday 17, nos. 5-7 (May). doi:10.5210/fm.v17i5.4013.
- Barnes, Susan B. 2006. "A privacy paradox: Social [networking](#) in the United States." First Monday 11, no. 9 (September 4). doi:10.5210/fm.v11i9.1394.
- Bartlett, Jamie. 2012. The Data Dialogue. London, UK: Demos, September 14. isbn: 978-1-909037-16-8.
- Berlingerio, Michele, Francesco Calabrese, Giusy Lorenzo, Rahul Nair, Fabio Pinelli, and Marco Luca Sbodio. 2013. "AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data." In Machine Learning and Knowledge Discovery in Databases, edited by Hendrik Blockeel, Kristian Kersting, Siegfried Nijssen, and Filip elezn, 8190:663–666. Lecture Notes in Computer Science. Berlin, Germany: Springer. doi:10.1007/978-3-642-40994-3
- Bowers, John, and Tom Rodden. 1993. "Exploding the Interface: Experiences of a CSCW Network." In Proceedings of the INTERACT '93 and CHI '93 Conference on [Human Factors](#) in Computing Systems, 255–262. CHI '93. Amsterdam, The Netherlands: ACM. isbn: 0-89791-575-5. doi:10.1145/ 169059.169205.
- Brown, I., and B. Laurie. 2000. "Security against compelled disclosure." In Proc. IEEE ACSAC, 2–10. December. doi:10.1109/ACSAC.2000.898852.
- Brown, Ian. 2014. "The Economics of Privacy, Data Protection and Surveillance." In Handbook on the Economics of the Internet, edited by M. Latzer and J.M. Bauer. Cheltenham, UK: Edward Elgar Publishing.
- Brown, Ian, Lindsey Brown, and Douwe Korff. 2010. "Using NHS Patient Data for Research Without Consent." Law, Innovation and Technology 2, no. 2 (December): 219–258. issn: 1757-9961. doi:10.5235/175799610794046186.
- Cafaro, Francesco. 2012. "Using embodied allegories to design gesture suites for human-data interaction." In Proceedings of the 2012 ACM Conference on Ubiquitous

Computing, 560–563. New York, NY, USA: ACM. isbn: 978-14503-1224-0.  
doi:10.1145/2370216.2370309.

- Callaghan, Sarah, Steve Donegan, Sam Pepler, Mark Thorley, Nathan Cunningham, Peter Kirsch, Linda Ault, et al. 2012. "Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres." *International Journal of Digital Curation* 7, no. 1 (March 10): 107–113. issn: 1746-8256. doi:10.2218/ijdc.v7i1.218.
- Campbell, Andrew T., Shane B. Eisenman, Nicholas D. Lane, Emiliano Miluzzo, Ronald A. Peterson, Hong Lu, Xiao Zheng, Mirco Musolesi, Kristf Fodor, and Gahng-Seop Ahn. 2008. "The Rise of People-Centric Sensing," *IEEE Internet Computing* 12, no. 4 (July): 12–21. issn: 1089-7801. doi:10.1109/ mic.2008.90.
- Card, Stuart K., Thomas P. Moran, and Allen Newell. 1983. *The psychology of human-computer interaction*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, February. isbn: 0898598591.
- Choe, Eun K., Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. "Understanding quantified-selfers' practices in collecting and exploring personal data." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1143–1152. Toronto, ON, Canada: ACM Press. isbn: 9781450324731. doi:10.1145/2556288.2557372.
- Clark, David D. 1995. "The Design Philosophy of the DARPA Internet Protocols." *SIGCOMM Comput. Commun. Rev. (New York, NY, USA)* 25, no. 1 (January): 102–111. issn: 0146-4833. doi:10.1145/205447.205458.
- Coles-Kemp, Lizzie, and Elahe K. Zabihi. 2010. "On-line privacy and consent: a dialogue, not a monologue." In *Proceedings of the 2010 workshop on New security paradigms*, 95–106. NSPW '10. Concord, MA, USA: ACM, September. isbn: 978-1-4503-0415-3. doi:10.1145/1900546.1900560.
- Court of Justice of the European Union. 2014. An internet search engine operator is responsible for the processing that it carries out of personal data which appear on web pages published by third parties. Judgment in Case C-131/12, May 13.
- Crabtree, Andy, and Richard Mortier. 2015. "Human Data Interaction: Historical Lessons from Social Studies and CSCW." In *Proceedings of European Conference on Computer Supported Cooperative Work (ECSCW)*. Oslo, Norway, September.
- Dourish, Paul. 2004. "What We Talk About when We Talk About Context." *Personal Ubiquitous Comput. (London, UK, UK)* 8, no. 1 (February): 19– 30. issn: 1617-4909. doi:10.1007/s00779-003-0253-8.
- Dwork, Cynthia. 2006. "Differential Privacy." In *Automata, Languages and*

Programming, edited by Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, 1–12. Berlin, Germany: Springer Berlin / Heidelberg. doi:10.1007/11787006 1.

- Elmqvist, Niklas. 2011. “Embodied Human-Data Interaction.” In Proceedings of the CHI Workshop on Embodied Interaction: Theory and Practice in HCI, 104–107. May.
- Estrin, Deborah. 2013. small data, N=me, Digital Traces. Talk presented at TEDMED 2013, Washington, DC, USA, April.
- European Parliament. 2014. Legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).  
<http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P7TA-2014-0212&language=EN>, March.
- Falahrastegar, Marjan, Hamed Haddadi, Steve Uhlig, and Richard Mortier. 2014. “The Rise of Panopticons: Examining Region-Specific Third-Party Web Tracking.” In Proc. Traffic Monitoring and Analysis, edited by Alberto Dainotti, Anirban Mahanti, and Steve Uhlig, 8406:104–114. Lecture Notes in Computer Science. Also as arXiv preprint arXiv:1409.1066. Springer Berlin Heidelberg, April. isbn: 978-3-642-54998-4. doi:10.1007/978-3-64254999-1 9.
- 2016. “Tracking Personal Identifiers Across the Web.” In Proceedings of Passive and Active Measurement (PAM).
- Fan, Chloe. 2013. “The Future of [Data Visualization](#) in Personal Informatics Tools.” In Personal Informatics in the Wild: Hacking Habits for Health & Happiness CHI 2013 Workshops. ACM.
- Grudin, Jonathan. 1990a. “Interface.” In Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work, 269–278. CSCW '90. Los Angeles, California, USA: ACM. isbn: 0-89791-402-3. doi:10.1145/99332. 99360.
- Grudin, Jonathan. 1990b. “The Computer Reaches out: The Historical Continuity of Interface Design.” In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 261–268. CHI '90. Seattle, Washington, USA: ACM. isbn: 0-201-50932-6. doi:10.1145/97243.97284.
- Guha, Saikat, Alexey Reznichenko, Kevin Tang, Hamed Haddadi, and Paul Francis. 2009. “Serving Ads from localhost for Performance, Privacy, and Profit.” In Proceedings of the Eighth ACM Workshop on Hot Topics in Networks (HotNets-VIII). New York City, NY, USA.
- Hamed Haddadi, Heidi Howard, Amir Chaudhry, Jon Crowcroft, Anil Madhavapeddy,



Derek McAuley, Richard Mortier, "Personal Data: Thinking Inside the Box", The 5th decennial Aarhus conference (Aarhus 2015), August 2015

- Haddadi, Hamed, Pan Hui, and Ian Brown. 2010. "MobiAd: private and scalable mobile advertising." In Proceedings of the fifth ACM International Workshop on Mobility in the Evolving Internet Architecture, 33–38. MobiArch '10. Chicago, Illinois, USA: ACM. isbn: 978-1-4503-0143-5. doi:10.1145/1859983.1859993.
- Haddadi, Hamed, Richard Mortier, Derek McAuley, and Jon Crowcroft. 2013. Human-data interaction. Technical report UCAM-CL-TR-837. Computer Laboratory, University of Cambridge, June.
- Huberman, Bernardo A. 2012. "Sociology of science: Big data deserve a bigger audience." Nature 482, no. 7385 (February 16): 308. issn: 1476-4687. doi:10.1038/482308d.
- Ioannidis, John P. A. 2013. "Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research." American Journal of Bioethics 13, no. 4 (March 20): 40–42. doi:10.1080/15265161.2013.768864.
- Jacobs, Rachel, Steve Benford, Mark Selby, Michael Golembewski, Dominic Price, and Gabriella Giannachi. 2013. A conversation between trees: what data feels like in the forest. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 129-138. DOI=<http://dx.doi.org/10.1145/2470654.2470673>
- Jacobs, Rachel, Steve Benford, Ewa Luger, and Candice Howarth. 2016. The Prediction Machine: Performing Scientific and Artistic Process. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16). ACM, New York, NY, USA, 497-508. DOI: <http://dx.doi.org/10.1145/2901790.2901825>
- Kapadia, Apu, Tristan Henderson, Jeffrey J. Fielding, and David Kotz. 2007. "Virtual Walls: Protecting Digital Privacy in Pervasive Environments." Chap. 10 in Proceedings of the 5th International Conference on Pervasive Computing, edited by Anthony LaMarca, Marc Langheinrich, and Khai N. Truong, 4480:162–179. Lecture Notes in Computer Science. Toronto, ON, Canada: Springer Berlin / Heidelberg, May. isbn: 978-3-540-72036-2. doi:10.1007/978-3-540-72037-9 10.
- Kee, Kerk F., Larry D. Browning, Dawna I. Ballard, and Emily B. Cicchini. 2012. "Sociomaterial processes, long term planning, and infrastructure funding: Towards effective collaboration and collaboration tools for visual and data analytics." In Presented at the NSF sponsored Science of Interaction for Data and Visual Analytics Workshop. Austin, TX, March.
- Kellingley, Nick. 2015. Human Data Interaction (HDI): The New Information Frontier.

<https://www.interaction-design.org/literature/article/human-datainteraction-hdi-the-new-information-frontier>, November.

- Kum, Hye-Chung, Ashok Krishnamurthy, Ashwin Machanavajjhala, and Stanley C. Ahalt. 2014. "Social Genome: Putting Big Data to Work for Population Informatics." *Computer* 47, no. 1 (January): 56–63. issn: 0018-9162. doi:10.1109/mc.2013.405.
- Kumar, Santosh, Wendy Nilsen, Misha Pavel, and Mani Srivastava. 2013. "Mobile Health: Revolutionizing Healthcare Through Transdisciplinary Research." *Computer* 46, no. 1 (January): 28–35. issn: 0018-9162. doi:10.1109/mc.2012.392.
- Lanier, Jaron. 2013. *Who Owns The Future?* New York, NY, USA: Simon & Schuster.
- Leon, Pedro G., Justin Cranshaw, Lorrie F. Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. 2012. "What do online behavioral advertising privacy disclosures communicate to users?" In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society*, 19–30. WPES '12. Raleigh, North Carolina, USA: ACM. isbn: 978-1-4503-1663-7. doi:10.1145/2381966.2381970.
- Leontiadis, Ilias, Christos Efstratiou, Marco Picone, and Cecilia Mascolo. 2012. "Don'T Kill My Ads!: Balancing Privacy in an Ad-supported Mobile Application Market." In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, 2:1–2:6. HotMobile '12. San Diego, California: ACM. isbn: 978-1-4503-1207-3. doi:10.1145/2162081.2162084.
- Luger, Ewa, Stuart Moran, and Tom Rodden. 2013. "Consent for All: Revealing the Hidden Complexity of Terms and Conditions." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2687–2696. CHI '13. Paris, France: ACM. isbn: 978-1-4503-1899-0. doi:10.1145/2470654.2481371.
- Luger, Ewa, and Tom Rodden. 2013. "An Informed View on Consent for UbiComp." In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 529–538. UbiComp '13. Zurich, Switzerland: ACM. isbn: 978-1-4503-1770-2. doi:10.1145/2493432.2493446.
- Madden, Mary, Susannah Fox, Aaron Smith, and Jessica Vitak. 2007. *Digital Footprints: Online identity management and search in the age of transparency*. PEW Internet & American Life Project. Retrieved Feb. 23, 2014 from <http://www.pewinternet.org/files/old-media//Files/R...> PIP Digital Footprints.pdf.pdf. 1615 L ST., NW – Suite 700 Washington, D.C. 20036: PEW Internet, December.
- Mashhadi, Afra, Fahim Kawsar, and Utku G. Acer. 2014. "Human Data Interaction in IoT: The ownership aspect." In *IEEE World Forum on Internet of Things (WF-IoT)*, 159–162. March. doi:10.1109/WF-IoT.2014.6803139.
- Mayer-Schonberger, V. 2009. *Delete: The Virtue of Forgetting in the Digital Age*.

Princeton University Press. isbn: 9781400831289.

- McAuley, Derek, Richard Mortier, and James Goulding. 2011. "The Dataware Manifesto." In Proceedings of the 3rd IEEE International Conference on Communication Systems and Networks (COMSNETS). Invited paper. Bangalore, India, January.
- MIT Technology Review. 2015. The Emerging Science of Human-Data Interaction. <http://www.technologyreview.com/view/533901/the-emerging-scienceof-human-data-interaction/>, January.
- Mortier, Richard, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2014. "Human-Data Interaction: The Human Face of the Data-Driven Society." <http://dx.doi.org/10.2139/ssrn.2508051>, SSRN (October).
- Murphy, Kate. 2014. "We Want Privacy, but Can't Stop Sharing." <http://www.nytimes.com/2014/10/05/sunday-review/we-want-privacy-but-cant-stopsharing.html>, New York Times (October).
- Naehrig, Michael, Kristin Lauter, and Vinod Vaikuntanathan. 2011. "Can Homomorphic Encryption Be Practical?" In Proc. ACM Cloud Computing Security Workshop, 113–124. Chicago, Illinois, USA. isbn: 978-1-4503-1004-8. doi:10.1145/2046660.2046682.
- Naughton, John. 2015. "Fightback against internet giants' stranglehold on personal data starts here." <http://www.theguardian.com/technology/2015/feb/01/control-personal-data-databox-end-user-agreement>, The Guardian (February).
- Nissenbaum, Helen F. 2004. "Privacy as Contextual Integrity." *Washington Law Review* 79, no. 1 (February): 119–157.
- Ohm, Paul. 2010. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." <http://uclalawreview.org/pdf/57-6-3.pdf>, *UCLA Law Review* 57:1701–1778.
- Organisation for Economic Co-operation and Development. 2013. Exploring the Economics of Personal Data - A Survey of Methodologies for Measuring Monetary Value. Technical report, OECD Digital Economy Papers 220. OECD, April 2. doi:10.1787/5k486qtxldmq-en.
- O'Rourke, JoAnne M., Stephen Roehrig, Steven G. Heeringa, Beth G. Reed, William C. Birdsall, Margaret Overcashier, and Kelly Zidar. 2006. "Solving Problems of Disclosure Risk While Retaining Key Analytic Uses of Publicly Released Microdata." *Journal of Empirical Research on Human Research Ethics* 1, no. 3 (September): 63–84. issn: 1556-2646. doi:10.1525/jer.2006.1.3.63.
- Oxford English Dictionary. 2014. <http://www.oed.com/view/Entry/296948>, February.

- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York, NY, USA: Penguin Press, May. isbn: 1594203008.
- Patil, Sameer, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2014. "Reflection or Action?: How Feedback and Control Affect Location Sharing Decisions." In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 101–110. Toronto, ON, Canada, April. doi:10.1145/2556288.2557121.
- Pentland, Alex. 2012. "Reinventing society in the wake of Big Data." <http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data>, Edge (August).
- Ricci, Francesco, Lior Rokach, Bracha Shapira, and Paul B. Kantor. 2010. *Recommender Systems Handbook*. 1st. New York, NY, USA: Springer-Verlag New York, Inc. isbn: 0387858199, 9780387858197.
- Schmidt, Kjeld. 1994. *Social Mechanisms of Interaction*. Technical report COMIC Deliverable 3.2. ISBN 0-901800-55-4. Esprit Basic Research Action 6225.
- Shilton, Katie. 2012. "Participatory personal data: An emerging research challenge for the information sciences." *Journal of the American Society for Information Science and Technology* 63, no. 10 (October): 1905–1915. issn: 1532-2882. doi:10.1002/asi.22655.
- Shilton, Katie, Jeff Burke, Deborah Estrin, Ramesh Govindan, Mark Hansen, Jerry Kang, and Min Mun. 2009. "Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing." In *Proceedings of the 37th Research Conference on Communication, Information and Internet Policy (TPRC)*. Arlington, VA, USA, September.
- Solove, Daniel J. 2013. "Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126, no. 7 (May): 1880–1903.
- Star, Susan Leigh. 2010. "This is Not a Boundary Object: Reflections on the Origin of a Concept." *Science, Technology & Human Values* 35 (5): 601–617. doi:10.1177/0162243910377624. eprint: <http://sth.sagepub.com/content/35/5/601.full.pdf+html>.
- Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." *Social Studies of Science* 19 (3): 387–420. doi:10.1177/030631289019003001.
- Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society. 2014. *Policy Outline of the Institutional Revision for Utilization of Personal Data*. [http://japan.kantei.go.jp/policy/it/20140715\\_2.pdf](http://japan.kantei.go.jp/policy/it/20140715_2.pdf).

- Taddicken, Monika, and Cornelia Jers. 2011. "The Uses of Privacy Online: Trading a Loss of Privacy for Social Web Gratifications?" In *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web*, 1st ed., edited by Sabine Trepte and Leonard Reinecke, 143–156. Springer-Verlag Berlin Heidelberg. isbn: 978-3-642-21520-9. doi: 10.1007/978-3-642-21521-6\_11.
- The Open Data Institute. <http://theodi.org/>.
- Tolmie, P., A. Crabtree, T. Rodden, C. Greenhalgh, and S. Benford. 2007. "Making the home network at home: digital housekeeping." In *Proceedings of ECSCW*, 331–350. Limerick, Ireland: Springer.
- Trudeau, Stephanie, Sara Sinclair, and Sean W. Smith. 2009. "The effects of introspection on creating privacy policy." In *WPES '09: Proceedings of the 8th ACM workshop on Privacy in the electronic society*, 1–10. Chicago, IL, USA: ACM, November. isbn: 978-1-60558-783-7. doi:10.1145/1655188.1655190.
- US Consumer Privacy Bill of Rights. 2012. *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy*. <https://www.whitehouse.gov/sites/default/files/privacy-final.pdf>, February.
- Vallina-Rodriguez, Narseo, Jay Shah, Alessandro Finamore, Yan Grunenberger, Konstantina Papagiannaki, Hamed Haddadi, and Jon Crowcroft. 2012. "Breaking for commercials: characterizing mobile advertising." In *Proceedings of the 2012 ACM Internet Measurement Conference*, 343–356. Boston, MA, USA: ACM, November. isbn: 978-1-4503-1705-4. doi:10.1145/2398776.2398812.
- Westby, Jody R. 2011. "Legal issues associated with data collection & sharing." In *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, 97–102. Salzburg, Austria, April. isbn: 978-1-4503-0768-0. doi:10.1145/1978672.1978684.
- Whitley, Edgar A. 2009. "Informational privacy, consent and the "control" of personal data." *Information Security Technical Report* 14, no. 3 (August): 154–159. issn: 13634127. doi:10.1016/j.istr.2009.10.001.
- Winstein, Keith. 2015. "Introducing the right to eavesdrop on your things." <http://www.politico.com/agenda/story/2015/06/internet-of-things-privacyconcerns-000107>, *The Agenda Magazine* (July).
- World Economic Forum. 2011. *Personal Data: The Emergence of a New Asset Class*. <http://www.weforum.org/reports/personal-data-emergence-newasset-class>. In collaboration with Bain & Company, January.
- Zaslavsky, Arkady, Charith Perera, and Dimitrios Georgakopoulos. 2012. "Sensing as

a Service and Big Data.” In Proceedings of the International Conference on Advances in Cloud Computing (ACC). Bangalore, India, July.

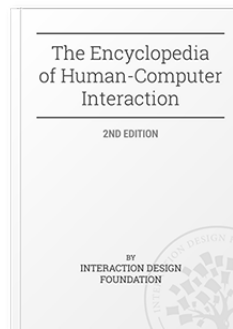
[← Emotion and website design](#)

[Design for All >](#)

Learnt something new? Share with your friends:



## 41.12 References



The Encyclopedia of Human-Computer Interaction, 2nd Ed.

BY MADS SOEGAARD AND RIKKE FRIIS DAM (EDS.)

### READ OFFLINE

 [PDF version](#)



## TABLE OF CONTENTS

[View chapter TOC](#) ▾

### AUTHOR(S)



**Richard Mortier**

Dr Richard Mortier is a University Lecturer in the Computer Laboratory at the University of Cambridge. He works in the Networks and Operating Systems group within the Systems Research Group, and he is interested in computing infrastructures, specifically networked systems. Before coming to Cambridge in 2014 Dr Mortier was a Horizon Transitional Fellow in the Horizon Institute, based in the School of Computer Science at the University of Nottingham. His research spans a range of topics, all with a networked systems angle. He has worked on topics from distributed system performance monitoring and debugging, to Internet routing protocols, to real-time media platform design and implementation. He has worked in a variety of roles, from high-level platform architect, to designer and implementer of complex networked systems, to website designer and builder. He has also consulted and worked for a broad range ... companies, including startups and corporates in both the IIS and IUK.

[View profile](#)



**Hamed Haddadi**

Dr Hamed Haddadi is a Senior Lecturer (Associate Professor) in Digital Media at School of Electronic Engineering and Computer Science, Queen Mary University of London. He was a postdoctoral researcher at Max Planck Institute for Software Systems in Germany and a postdoctoral research fellow at the Department of Pharmacology, University of Cambridge and The Royal Veterinary College, University of London. Additionally, Hamed has spent time working and collaborating with Intel Research, Microsoft Research, AT&T Research, Telefonica, and Sony Europe. He studied for his BEng/MSc/PhD at University College London and the University of Cambridge. He is currently serving as the Information Services Director for the ACM SIGCOMM Executive Committee.

[View profile](#)



**Tristan Henderson**

Dr. Tristan Henderson has been a Lecturer in Computer Science in the University of St Andrews since December 2006 and is now a senior lecturer. In 2011 he helped cofound SACHI and his research revolves around measuring communications networks and their users. His PhD looked at measuring networked game servers and game players, but more recently he has been investigating new methodologies for measuring mobile, wireless and online social

networks and their users. He attempts to use the results of this network measurement to build more usable networks. Specifically networks which support the functionality that users desire. For instance we can use network measurement to improve security and privacy so that people can use pervasive sensor networks. Or we can measure network behaviour and use this to inform the design of future networks. He is one of the Workshop Organisers for...

[View profile](#)



## Derek McAuley

Professor Derek McAuley is a Professor of Digital Economy in the School of Computer Science and Director of Horizon at the University of Nottingham. After his PhD and lectureship in the Computer Laboratory at the University of Cambridge he moved to a chair in Department of Computer Science at the University of Glasgow. He returned to Cambridge in July 1997, to help found the Cambridge Microsoft Research facility, moving on to found the Intel lablet in Cambridge in July 2002. He has worked for Intel from August 2005 until August 2006. founded two startups, XenSource (now Citrix) and Netronome. He also Chief Innovation Officer (in residence) at the Digital Catapult during its startup phase in 2013/2014.

[View profile](#)



## Jon Crowcroft

Jon Crowcroft is the Marconi Professor of Networked Systems in the Computer Laboratory, of the University of Cambridge. Prior to that he was professor of networked systems at UCL in the Computer Science Department. He has supervised over 45 PhD students and over 150 Masters students. He is a Fellow of the ACM, a Fellow of the British Computer Society and a Fellow of the IEE and a Fellow of the Royal Academy of Engineering, as well as a Fellow of the IEEE. He was a member of the IAB 96-02, and went to the first 50 IETF meetings; was general chair for the ACM SIGCOMM 95-99; is recipient of Sigcomm Award in 2009. He is the Principle Investigator in the Computer Lab for the EU Social Networks project, the EPSRC funded Horizon Digital Economy project, hubbed at Nottingham, the EPSRC funded project on federated sensor nets project FRESNEL, in collaboration with Oxford; and a new 5-year project... towards a Carbon Neutral Internet with Leeds. Jon's research interests include Communications, Multimedia and

[View profile](#)



## Andy Crabtree

A sociologist by background, Andy started working with computer scientists and software engineers when he did his PhD with John Hughes at Lancaster. John pioneered the use of ethnography in systems design and Andy has been working on shaping computing around the social for the best part of 20 years now. He has published widely, including a couple of textbooks on ethnography for design. He was awarded an RCUK Academic Fellowship in 2006, which



solidified his work at Nottingham in the Mixed Reality Laboratory, an interdisciplinary research group with an international reputation for innovative work at the interface between computing and society.

[View profile](#)