Understanding and Designing Human Data Relations

Alex Bowyer

Table of Contents

# Frontmatter

This is the example frontmatter file. Use it for your abstract, dedications, acknowledgements etc.

# 1 Introduction

[Target 4,000 words]

## 1.1 Motivation & Problem

[Target 1,000 words]

## 1.2 Prior Work / Theoretical Frame

[Target 1,000 words]

## 1.3 The Research Question & My Approach

[Target 1,000 words]

*What role should people’s data play in their lives, and what capabilities do they need?*

Sub-questions:

* RQ1: How do people understand and relate to data?
* RQ2: What role do people have in the world of data?
* RQ3: What role does does play in people’s lives
* RQ4: What would an ideal world of humans and data look like?
* RQ5: What are the barriers and opportunities toward this ideal?

## 1.4 Contribution

[Target 500 words]

## 1.5 Thesis Structure

[Target 500 words]

## 1.6 Notes

# 2 Literature Review

## 2.1 Data-Centrism and the Need for Access

### 2.1.1 What is Data?

Data is an oft-used word that carries multiple meanings. In everyday speech, it might refer to mobile phone bandwidth, a filled application form or a collection of files. Even experts have a variety of definitions of data, as well as the related concepts of information and knowledge (Zins, [2015](#ref-zins2015)). In this study, we refer to data by its accepted definition as information or knowledge stored in a form suitable for computer processing. Wellisch expressed this as ‘the representation of concepts or other entities, fixed in or on a medium in a form suitable for communication, interpretation, or processing by human beings or by automated systems’ (Wellisch, [1996](#ref-wellisch1996)), which is a useful definition as it includes the fact that both humans and algorithms can use data, and that data is something that needs interpretation.

From a strict grammatical stance, ‘data’ is a plural of the singular ‘datum’ thus it is more correct to write ‘the data are correct’ - but this usage is rapidly declining from use (‘Data’, [no date](#ref-grammaristData)) and throughout this thesis I use the more widely adopted usage of treating data as a singular mass noun, as in ‘the data is correct’.

The concepts of ‘data’ and ‘information’ are closely related, so much so that they are often used interchangeably. Ackoff presented a model for distinguishing data, information, knowledge, understanding/intelligence and wisdom, in which he describes data as the physical symbols, effectively the 1’s and 0’s stored in a computer or the ink marks on a page, which becomes useful when humans or algorithms are able to deduce facts from those symbols to answer simple questions - at this point it becomes ‘information’. Being able to interpret deeper how and why questions allow information to become knowledge and understanding, towards the ultimate goal of wisdom (Ackoff, [1989](#ref-ackoff1989)). This is often represented as the DIKW pyramid (DIKW being shorthand for the data-information-knowledge-wisdom transformation that occurs as you move up through the layers), the origin of which is unknown (Wallace, [2007](#ref-wallace2007)). Figure 1 builds upon a representation by George Pór (Pór, [1997](#ref-por1997)) of the pyramid as a ‘wisdom curve’, showing how increasing meaning and value can be obtained from data as deeper questions can be asked of it. This theme of obtaining meaning and value from data is an important aspect of my research that I will refer back to.



REDRAW Figure 1: Making Data into Meaningful Information

This model that turning data into information can be thought of as using that data to answer questions is consistent with the idea that “information can be thought of as the resolution of uncertainty” (‘Information’, [no date](#ref-wikipediaInformation)). The exact origin of this definition is unknown but it is often attributed to mathematician Claude Shannon (Shannon, [1948](#ref-shannon1948)). Indeed from an etymological stance, one who is informed is one who has received knowledge or concepts as a result of what has been communicated to them. Thus we can consider that data is the material from which that information can be received. It follows also that data contains uncertainty that must be resolved in order for it to become meaningful information.

### 2.1.2 The Rise of Data-centrism

The earliest computer systems used data to store mathemical and scientific facts. Data processing allowed for previously manual operations to be performed with greater speed and accuracy, most famously the work of Alan Turing and the case of the Enigma code breakers during World War II (Hutton, [2012](#ref-hutton2012)). This work was the advent of general-purpose computing - machines that could be applied to any problem provided you could reduce that problem to data. Businesses over the following decades began to apply computers to myriad new problem areas in all different fields of work and life, and doing so began the encoding of information about people as data, be it for statistical purposes like censuses or research, or simply to enable the more efficient serving of customers by storing databases of customer records.

The personal computer revolution (‘The personal computer revolution’, [no date](#ref-britannicaPCrevolution)) of the late 1970s and 1980s put computers in every office and eventually every home too, and it soon became commonplace that each individual would have data stored about them in companies’ databases. In the subsequent years three factors have combined to accelerate this trend of storing data about people: i) labour costs have remained high and companies have sought ways to automate their businesses and to implement online services and call centres in place of in-person staff interaction, ii) computer processing and storage has become ever cheaper thanks to the advent of cloud computing, meaning that many business processes could be reduced to data processing tasks or entire businesses be moved online, and iii) the rise of smartphones and web-enabled devices have meant that the public are now ready and willing to conduct much of their daily business online through the web and apps. These factors have encouraged both commercial and civic providers to centralise their services and to ‘go digital’ to the greatest degree possible. In doing so they collect ever more data about people (now ‘service users’ or just ‘users’). Data is now seen as a resource which can be mined for value, and harnessed for profit and business efficiency - ‘the new oil’ (Toonders, [2014](#ref-toonders2014)). Zuboff, in her 2019 book on ‘surveillance capitalism’, characterises this new digital world as the collection of human behaviour data so that it can be used as free raw material and converted into profit through hyper-personalised advertising and targeting by software platforms (Zuboff, [2019](#ref-zuboff2019)). This philosophy is also known as ‘data-ism’ (Brooks, [2013](#ref-brooks2013)).

As a result of data-ism, the collection of data about people has become an inevitable part of modern life. We live ‘digital lives’ (‘Our Digital Lives’, [2018](#ref-ted2018)) where we each interact directly and indirectly with hundreds of digital systems every day - as you shop, socialise, or browse online; as you listen to music or watch TV; as you interact with governments or healthcare services; as you travel, and many more. Every one of those interactions indicates the presence of data about you stored in a company database. Every aspect of our lives involves the input, processing and output of data – either provided by, collected from, or generated about, us. And the digital data we create and consume (whether consciously or not - data sharing is often unwitting (Crabtree and Tolmie, [2018](#ref-tolmie2018))) has a direct influence on our lived experience - from decisions about what we are entitled to and what opportunities we will be offered, to the advertisements and content recommendations we are shown while we browse.

Unfortunately, the large-scale systems which collect data about us now function as ‘data traps’ (Abiteboul, André and Kaplan, [2015](#ref-abiteboul2015)) - where data about us is easily gathered but very hard to remove or even to access. This creates a lack of agency for the individuals living in this data-centric world. The World Economic Forum’s “Rethinking Personal Data” project recognised the critical role that data - specifically personal data - data created by and about people - now holds, and identified that ‘an asymmetry of power exists today […] created by an imbalance in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information’ (World Economic Forum, [2011](#ref-WEF2011), [2013](#ref-WEF2013), [2014b](#ref-WEF2014context), [2014a](#ref-WEF2014lens)).

### 2.1.3 Data Protection & GDPR

Since as early as 1973, the need to protect individuals’ rights over their data has been recognised (US Department of Health Education and Welfare, [1973](#ref-USDOHEW1973)). The 37-nation organisation OECD in 1980 stated that “the right of individuals to access and challenge personal data is […] the most important privacy protection safeguard” and issued recommendations that individuals should be given basic privacy rights, including the right to be informed whether data is stored about them, and the right to an intelligible copy of that data (Organisation for Economic Co-operation and Development, [1980](#ref-OECD1980)).

Over the subsequent decades, lawmakers began to enact laws to deliver these rights to individuals, notably the UK’s Data Protection Act 1984 (which set up an independent body, the Data Protection Registrar (now the Information Commissioner’s Office) with which organisations were required to register their usage of personal data), Ireland’s Data Protection Act 1988 (which introduced the concept of a ‘duty of care’ for data collectors - that they are expected to avoid causing damage or distress to data subjects), the EU’s Data Protection Directive in 1995 and the UK’s Data Protection Act in 1998. However, such laws were generally found to be ineffective - in 2002 Simon Davies, director of Privacy International said that the UK’s DPA was “almost useless in limiting the growth of surveillance” (Millar, [2002](#ref-millar2002)).

It was only in 2018, when the EU’s General Data Protection Regulation (GDPR) came into force, carrying with it significant designed-to-hurt fines for non-compliance (Kelly, [2020](#ref-kelly2020); Leprince-Ringuet, [2021](#ref-zdnet2021)), that individuals have been able to practically exercise their data rights to any meaningful degree (‘The GDPR: Does it Benefit Consumers in Any Practical Way?’, [2020](#ref-atebits2020)). The GDPR – which gives individuals key rights including rights to timely data access, explanation, erasure and correction (Information Commissioner’s Office, [2018](#ref-ico2018)) – can be seen as the first serious attempt to rebalance the aforementioned power imbalance over data between citizens and organisations and is generally regarded as a landmark piece of legislation and a strong template for individual data protection. Around the world, companies have overhauled their privacy policies and updated their business practices to comply with the GDPR and other similar legislation, such as Japan’s 2017 Act on the Protection of Personal Information, India’s 2019 Personal Data Protection Bill and the 2020 California Consumer Protection Act. In the USA, there has been no national privacy law yet, but the GDPR’s influence is being felt in court rulings (Hoofnagle, Sloot and Borgesius, [2019](#ref-hoofnagle2019)).

Also in 2018, the Cambridge Analytica scandal (‘Facebook–Cambridge Analytica Data Scandal’, [2014](#ref-wikipedia2018cambAna)) broke; the personal data of 87 million people, acquired from Facebook, was exploited with the apparent intent of influencing voting outcomes including the UK’s 2016 Brexit referendum and the USA’s 2017 election of Donald Trump. This combined with widespread public information campaigns about GDPR have led to a heightened awareness of personal data rights (European Union Agency for Fundamental Rights, [2020](#ref-EUAFR2020)) and at the time of writing in 2021, personal data protection laws and individual digital rights remain a rapidly evolving area.

From the GDPR and its antecedents, a number of key terms have been established which I will adopt in this thesis, specifically (Information Commissioner’s Office, [2014](#ref-ico2014); The European Parliament and the Council of the European Union, [2016](#ref-GDPR2016)):

* *Personal data* means any information relating to an identifiable natural person - one who can be identified directly or indirectly by reference to an identifier such as a name, identification number or location or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.
* The *data subject* is the identified individual, living or deceased, who the personal data relates to.
* A *data controller* is the legal entity (company, public authority, agency, individual or other body) which collects or stores personal data about an individual and determines the means and purposes for which it is processed. Liability for data protection compliance rests with the data controller.
* A *subject access request* is the right to a copy of your personal data.
* *Data portability* is the right to receive a copy of all stored data about you, not just that which you provided, in an accessible and machine-readable format such as a CSV file, so that you can transport it to another service or make use of it.

### 2.1.4 The Need for Practical and Effective Data Access

The World Economic Forum called in 2011 for a balanced ecosystem around personal data, and identified *transparency* as a key principle needed to achieve this: People need to know what data is captured, how it is captured, how it will be used and analysed and who has access to it. Additionally people must understand the *value* created by the use of their data and the way in which they are compensated for this (World Economic Forum, [2011](#ref-WEF2011)). It is almost impossible for people to assess that value, because they are unaware of most of their data (Spiekermann and Korunovska, [2017](#ref-spiekermann2017)). Having *awareness* of your personal data is a critical first step, so that people might assess “to what extent the bargain is fair” (Larsson, [2018](#ref-larsson2018)). In this regard, the GDPR can be seen as an important step in the right direction, as it requires data controllers to document their data practices and to provide data copies.

However, it is not sufficient simply to grant data subjects the technical or procedural capabilities to see the stored records about them. Access must be *effective*. Every individual must have the knowledge, skills and structures in place that enable them to achieve their objectives with their personal data data (Gurstein, [2003](#ref-gurstein2003)). Gurstein later identified seven aspects that are necessary for access to be effective (Gurstein, [2011](#ref-gurstein2011)) and to avoid a ‘data divide’ of those who can harness their data and those who cannot:

1. *Internet*: If data access is via Internet, then issues with affordability, bandwidth, network censorship, or disabilities limiting physical access to Internet devices or terminals would make access ineffective.
2. *Computers and software*: Sufficiently powerful computers must be available, for a sufficient amount of time, with sufficiently capable software to perform necessary interpretation or actions.
3. *Skills*: If technical skills or knowledge are required to use the software and/or to interpret, analyse or visualise the data, then access is ineffective.
4. *Content and formatting*: The data should be in an appropriate language and format to allow use at various levels of linguistic and computer literacy
5. *Sensemaking*: People need sufficient knowledge and skill to be able to interpret their data and extract meaningful information from it.
6. *Advocacy*: People need support and training to make use of their data and representation if they are to use it at a wider community level.
7. *Governance*: There must be financing and appropriate law or policy to support people’s desired usage of their data.

Unfortunately people’s ability to derive value from their data, or to assess its value is limited; it is an asset over which we have little control. Our existing data ‘resides in isolated silos kept apart by technical incompatibilities, semantic fuzziness, organizational barriers [and] privacy regulations’. This lack of effective data access is detrimental to trust, innovation and growth (Abiteboul, André and Kaplan, [2015](#ref-abiteboul2015)).

Beyond these operational concerns over effective access, there are practical limitations affecting people’s ability to make use of their data. Where people are given interfaces their data, access is typically via a list or feed combined with a search box. Studies have shown that people prefer to find information by *orienteering* rather than search - associatively traversing related datapoints (Teevan *et al.*, [2004](#ref-teevan2004); Karger and Jones, [2006](#ref-karger2006)). Having our documents distributed across multiple platforms, applications and devices makes interrogation and orienteering hard (Krishnan and Jones, [2005](#ref-krishnan2005)). Abowd and Mynatt highlight that in presenting information about people and their activities, everyday computing needs to address the facts that users activities rarely have a clear beginning or end, are often interrupted, are often concurrent with other activities; that *time* is an important factor in finding and interpreting information; and that associative modelling of information is more useful than hierarchical models, because future usage goals cannot always be anticipated (Abowd and Mynatt, [2000](#ref-abowd2000)). Recognising these needs, Krishnan and Jones identify that an effective information access system should support giving historical context, *finding trends and patterns*, time-based contextual retrieval, automatic structuring and multiple perspectives of the information (Krishnan and Jones, [2005](#ref-krishnan2005)). Shneiderman, in the context of considering the effectiveness of interactive information visualisations, identified the need to support seven types of information interaction: *overview*, *pan & zoom*, *focus* (context & distortion), *detail on demand*, *filter*, *relate*, *history* and *extract* (Shneiderman, [1996](#ref-shneiderman1996)). While any one of the capabilities mentioned in this paragraph does exist in data interfaces today, it is clear that no such general-purpose personal information access system exists with all or even most of those capabilities exists today. The development and state of the art in the field of Personal Information Management Systems is explored in section 2.2 below.

### 2.1.5 Research Gap: The Human Experience of Data

In this section, I have described the establishment of the data-centric world in which we live today, the imbalance this creates between data subjects and data controllers, and what can be viewed as nascent attempts by governments to redress that imbalance through the creation of new laws. I have also outlined where research thinking has exceeded the practical data capabilities we have today, in identifying many factors and capabilities that should be considered when it comes to giving people a meaningful relationship with their personal data.

To date, people’s relationship with their personal data and the information within it has barely been explored. What mental models to people have around data? What value does it carry to them and what meaningful place does it (or should it) hold in their life? What is it that makes data meaningful and what do people want from their data? What is it like to live in this data-centric world where your abilities over your data are limited by lack of access to data and a lack of suitable interfaces and technologies to properly manage your digital life? This is one aspect of the research gap this thesis will address - discovering the human experience of data.

## 2.2 Personal Data Interaction

### 2.2.1 Computers as General Purpose Information Tools

In the immediate aftermath of the second World War, Dr. Vannevar Bush wrote a landmark article for The Atlantic Monthly in which he envisioned a new scientific agenda for America and the world - to harness new general information-processing capabilities of computers to make the stored knowledge of mankind accessible and usable to all, for the betterment of society. He proposed the ‘Memex’, a device in which people would store their books, communications and records digitally so that it “might be consulted with exceeding speed and flexibility” - a personal filing system to serve as “an enlarged intimate supplement to his memory”. He emphasised the importance of allowing information to be stored in “associative chains of related materials” so that people would be able to retrieve information in the same way we think of it, traversing related items or ideas (Bush, [1945](#ref-bush1945)). During the next three decades, while computer systems were moving out of science labs and being established in workplaces as a means to automate and improve business processes, researchers began to look beyond usage in business and consider how computers might be used by ‘the common man’ to store one’s personal information in digital files (Nelson, [1965](#ref-nelson1965)), for interpersonal communication (Shannon, [1948](#ref-shannon1948)), to augment human intellect (Engelbart, [1962](#ref-engelbart1962)) and to model human thought (Simon and Newell, [1958](#ref-simon1958)).

Collectively, these constituted a recognition that computers could be considered a general purpose tool that anyone could use for their own purposes, and in the 1970s and 1980s the home computer revolution (‘The personal computer revolution’, [no date](#ref-britannicaPCrevolution)) seemed to place the potential power that “having reduced your affairs to software, software can take care of them for you” (Gelernter, [1994](#ref-gelernter1994)) into the hands of ordinary people.

### 2.2.2 Personal Information Management

Through the examination of people’s desk-based working practices, researchers began to understand how people handle information to inform the design of computer information systems. In 1983, Thomas Malone observed that categorisation is hard, and that any system must not only help the user to find information, but to remind the user of things to do. Computers could help through automatic classification, but should also allow both physical and logical “piles” of information to be arranged by the user (Malone, [1983](#ref-malone1983)). ‘Personal Information Management’ (PIM) was first mentioned in 1988 by Lansdale, who identified a need to design information management systems according how the psychology of the people who use them rather than to simulate office practices. By paying attention to how people categorise, recognise and recall information, and labelling information with appropriate attributes, allowing information to be retrieved by different properties (Lansdale, [1988](#ref-lansdale1988)). PIM includes both directly interacting with digital files, webpages and e-mails as well as ‘meta-activities’ such as finding, arranging, searching, browsing, re-finding, categorising, sensemaking, keeping and discarding personal information. William Jones summarised PIM as “the art of getting things done in our lives through information” (W. Jones, [2011a](#ref-jones2011)).

Driven in part by the pursuit of better “time management” in the late 20th century (characterised by PDAs, palmtops and electronic organisers) (Etzel, [1995](#ref-etzel1995)) and the focus on personal productivity in the early 2000s (characterised by ‘GTD’ self-help books and to-do list software) (Andrews, [2005](#ref-wired2005)) and the continuing challenge of overcoming information overload in an increasingly digital world, PIM has been a thriving field both in research and in practice, with a peak in activity around the mid ’00s. Since the 1990s numerous PIM system designs have emerged, each exhibiting some of the following six traits which I will now explain: Spatial, Semantic, Networked, Temporal, Contextual and Subjective.

*Spatial* PIM systems are based on the idea that people remember “where” they have put things and that this allows information to be quickly returned by associating it with a place (Negroponte and Bolt, [1978](#ref-bolt1978)), much as as people keep current information ‘in reach’ on a desk (Klein *et al.*, [2004](#ref-klein2004)). Spatial approaches recognise that *keeping* is a valuable activity in its own right, that informs sensemaking(Marshall and Jones, [2006](#ref-marshall2006)). Placed information also performs an important *reminding* function (Barreau, [1995](#ref-barreau1995); Barreau and Nardi, [1995](#ref-barreau1995a)).

Building on Bush’s ideas of “associative chains of related materials”, *networked* PIM systems focus on the relationships between data. HyperText, as conceived in 1965 (Nelson, [1965](#ref-nelson1965)) was designed to keep connections between information and allow the computer to understand what linked information *is*. The version of hypertext we use today is much weaker than Nelson’s HyperText or Berners-Lee’s Semantic Web and does not achieve these goals, as the inventors agree (Ross, [2005](#ref-ross2005); Nelson, [2006](#ref-nelson2006); Ziogas, [2020](#ref-ziogas2020)). In the absence of connected networks of personal information and with people collecting more information than they discard (Whittaker and Hirschberg, [2001](#ref-whittaker2001)), the 2000s saw software like Google Desktop Search (‘Google Desktop Search’, [2004](#ref-wikipedia2004googledesktop)) and Infovark (‘Infovark Company Profile’, [2007](#ref-crunchbase2007)) emerge to try and discover users’ data files and unify access to them, with limited impact (Bergman *et al.*, [2008](#ref-bergman2008)). Around this time, Microsoft invented WinFS, a system to re-invent the modern day operating system to be based upon relational structured data rather than file storage, but sadly it was never released (‘WinFS’, [no date](#ref-wikipedia2005winfs)). Paul Dourish et. al. proposed Placeless Documents, which relied on the idea of assigning user-specific properties to documents so that their could be arranged and recalled by their common properties rather than their location (Dourish *et al.*, [2000](#ref-dourish2000); Dourish, [2003](#ref-dourish2003)). *Metadata* – information about what the data *is* – is critical to information organisation (Foulonneau and Riley, [2008](#ref-foulonneau2008)). One of the more advanced networked PIM systems is the Networked Semantic Desktop, which recognises that critical metadata is lost when files are copied or emailed, and attempts to maintain metadata and traceability by integrating PIM with peer-to-peer (P2P) technology (Decker and Frank, [2004](#ref-decker2004)). Tags, which emerged as a means to organise data through systems like del.icio.us (‘Delicious’, [2003](#ref-wikipedia2003delicious)) and Flickr in the 2000s, are still widely used on social media and websites today, and are even available within macOS (Frost, [2019](#ref-frost2019)). Tags can be seen as a continuation of attempts to attach metadata to personal data to give it meaning, even though the dream of “folksonomies” has not been fully realised (Abbattista *et al.*, [2007](#ref-abbattista2007); Terdiman, [2008](#ref-terdiman2008)).

*Semantic* PIM systems, or “The Semantic Desktop” as it is often known, takes the idea of metadata even deeper and focuses on what the information means. The idea is to present an integrated view of a person’s stored knowledge by representing their documents, data and messages as URL-addressable semantic web resources (Sauermann, Bernardi and Dengel, [2005](#ref-sauermann2005)). The focus is on both the retrieval of documents and of facts (Schumacher, Sintek and Sauermann, [2008](#ref-schumacher2008)). This implicitly means that the computer must know more about what the data it stores represents, elevating it from number cruncher to something that holds a collection of information about the world. Hendler and Berners-Lee see semantic web technologies as the building blocks for a new age of *social machines*(Hendler and Berners-Lee, [2010](#ref-hendler2010)), machines that operate in society at an information level. This desire to give computers greater understanding of data has created emergent industries focused on using linguistics and statistics to perform content analysis, text mining and information extraction (Hotho, Nürnberger and Paaß, [2005](#ref-hotho2005)). It has even been proposed that AI might help computers to understand users’ mental models (Nadeem and Sauermann, [2007](#ref-nadeem2007)).

While folders have emerged as the dominant means to organise computer files and are effective because they allow you to arrange information according to its meaning to you (Bergman *et al.*, [2012](#ref-bergman2012); Bergman, [2013](#ref-bergman2013)), supporters of *temporal* PIM systems argue they are inadequate as an organising device. Freeman and Gelernter proposed Lifestreams, a PIM system based on the principled that storage should be transparent, archiving and compatibility should be automatic, and concise overviews of groups of related information should be available (Freeman and Gelernter, [1996](#ref-freeman1996)). Central to this system is the idea that personal data can most easily be navigated when viewed as a *timeline*, partly because almost all data can be associated to a specific time, but also because this maps onto the idea of relating personal information to human memory (Lansdale and Edmonds, [1992](#ref-lansdale1992)). TimeSpace provides another model of a PIM system that organises personal information by both time and the user’s own activities, to support interaction with a “continuously changing and evolving information space” (Krishnan and Jones, [2005](#ref-krishnan2005)). Time-based PIM approaches also coincide with a drive to move beyond files as a system of information storage. Gelernter believed we should not have to put effort into organising files, and argued somewhat prophetically that commercial factors have skewed personal data systems design away from the realities of human lives (Steinberg, [1997](#ref-steinberg1997)). In my own 2011 article “Why files need to die”, I mapped out how a personalised timeline could allow better personal information organisation and retrieval (Bowyer, [2011](#ref-bowyer2011)). Echoing this as well as Decker’s desire to maintain an information trail for every piece of information, Siân Lindley et. al., having called for time to become a subject of design research in its own right (Odom *et al.*, [2018](#ref-odom2018)), explored the concept of the *file biography*, a concept which allows the history of information to be kept as the file is used and changed. File biographies tell a story, and help to reconfigure our thinking away from mindsets around copying, deleting and sharing, to view information as fluid (Lindley *et al.*, [2018](#ref-lindley2018)). Moving into the world of online information collaboration, *activity streams* can also be seen as a recognition of the importance of tracking data as it changes, and offer new affordances (Hart-Davidson, Zachry and Spinuzzi, [2012](#ref-hartdavidson2012)).

In 1995, Barreau highlighted the importance of *context* to PIM; People need access to different information according to what they are doing (Barreau, [1995](#ref-barreau1995)) In 2000, Abowd and Mynatt’s vision for “everyday computing” highlighted the importance of paying attention to the user’s context in order to offer access to the most relevant information and features, and they suggest context can be identified by considering the “5 W’s” - who, where, what, when and why (Abowd and Mynatt, [2000](#ref-abowd2000)). Context-aware computing has emerged as a subdiscipline of research in its own right (Dey, [2001](#ref-dey2001)). Dourish identified that context is both a problem of representation, in that it is information that can be captured and represented, and of interaction, in that it is a relational property between objects or activities. He calls for *embodied interaction* - allowing users to create their own practices and meanings in the course of their PIM system interaction, noting that context is not objective and predetermined, it arises from the activity (Dourish, [2004](#ref-dourish2004)); you need different organisations of information in different contexts. This means that PIM systems need to support representing a given set of information in different ways (Lansdale and Edmonds, [1992](#ref-lansdale1992)) - but more that than, that different information should be shown according to the current context; different *perspectives* are needed to segment your life. TimeSpace uses ‘activity workspaces’ to achieve this (Krishnan and Jones, [2005](#ref-krishnan2005)), but Karger et. al.’s Haystack system refines the concept further, introducing the concept of *lenses*. Perspectives change which information records are included, whereas lenses allow you to focus on different attributes of what might be the same or different information (Karger *et al.*, [2005](#ref-karger2005)). Using a similar premise, Jilek’s “context spaces” system attempted a dynamically reorganising contextual sidebar, but is limited in flexibility as it uses rigid types for specific contexts. Lindley observes that different information abstractions are needed for different audiences, from which we can infer that in a multi-user system, no single arrangement of information will suffice because in the same context two people may have different needs (Lindley *et al.*, [2018](#ref-lindley2018)).

This is why the sixth trait of PIM systems is important: *subjectivity*. Information organisation cannot be handled in a deterministic, objective manner. Any PIM system must be tailored to, and adaptable by, the user. Shipman and Marshall found that forcing users into explicit information models or workflows is harmful to user experience, and that interactive systems have to address the challenge of being just explicit enough but still allowing for differences in individual mental models (Shipman and Marshall, [1999](#ref-shipman1999)). Bergman et. al. (Bergman, Beyth-Marom and Nachmias, [2003](#ref-bergman2003)) proposed three principles for subjective PIM, and their 2003 assertion that these principles are not currently well implemented in PIM systems remains true today:

1. the subjective classification principle - all related items should be classified together regardless of technological format
2. the subjective importance principle - the subjective importance of information should determine its degree of visual salience and accessibility
3. the subjective context principle - information should be retrieved and viewed by the user in the same context in which it was previously used

Teevan’s take on PIM subjectivity is important: “The user should feel in control of the information”. She argues that this can be done by “understanding what *conceptual anchors* the user creates and keeping them constant while the data changes.” (Teevan, [2001](#ref-teevan2001)). As with semantic PIM systems, we can see that a successful system (or at least, its designers) must understand a great deal about their users. The approach of *lifelogging* becomes relevant here: gathering as much data as possible, so that the maximum possible context, detail and understanding can be gained about that individual. This was most famously explored by Jim Gemmell and Gordon Bell in their MyLifeBits project (Gemmell, Bell and Lueder, [2006](#ref-gemmell2006); Bell and Gemmell, [2009](#ref-bell2009)) which tried to capture an entire life electronically, in pursuit of Vannevar Bush’s vision of augmenting human memory. When PIM is viewed through the lens of memory augmentation it is clear that the system design will necessarily have to be highly subjective. A PIM system is a *place* for your own personal information (W. Jones, [2011a](#ref-jones2011)).

### 2.2.3 Personal Informatics & The Quantified Self

Towards the end of the first decade of the 2000s, researchers and enthusiasts took PIM beyond task management and turned PIM thinking toward the self. Enabled by developments in fitness tracking and wearable sensors, it was now possible to gather physical activity data about yourself, such as step counts, heart rate (HR) or calories burned. In 2007, tech writers Kevin Kelly and Gary Wolf set out a vision for what they called *the Quantified Self*, that is, to achieve increased self-knowledge through self-tracking, not just of physical metrics but almost any aspect of your own life that could be numerically recorded in a computer (Kelly and Wolf, [2007](#ref-kelly2007)). The Quantified Self movement (QSM) is now a world-wide community of enthusiasts who have developed hundreds of tools and techniques for monitoring themselves through data for the purposes of self improvement, and also has a non-profit organisation aiming to ‘advance discovery through increasing access to data’ (‘About The Quantified Self’, [no date](#ref-quantifiedself)). Around 2009, researcher Ian Li began writing about what he called *personal informatics*, noting that it can be difficult to know ourselves due to incomplete self-knowledge, difficulties in monitor our own behaviours, and being too busy to introspect. He proposes that “Computers can help: They can store large amounts of data, analyse the data for patterns, visualise the data, and provide feedback at opportune times (Li, [2009](#ref-li2009)).” Just as QSM has gained traction with enthusists in the general public, so personal informatics has grown as an area of research, development and study in academic circles. While QSM and lifelogging focus slightly more on capturing data about oneself and personal informatics focuses slightly more on the mechanisms of collecting and reviewing self-tracking data, there is so much overlap that all three can all be considered the same field, which for convenience I will refer to by the shorthand *self informatics* (SI) throughout this thesis. SI can be seen as a distinct advancement from PIM because of its focus on *using* personal information for personal benefit. SI can be seen as the antithesis of corporate data-centric motives outlined in 2.1 - as here, data is gathered for the data subject’s benefit rather than that of the data-gathering orgnanisation.



Figure 2: Li et. al.’s Stage-based Model of Personal Informatics Systems (Li, Dey and Forlizzi, [2010](#ref-li2010))

Li, Dey and Forlizzi conducted participatory research with SI practitioners and identified five stages of personal informatics systems (which can be seen as refinement of William Jones’ list (W. Jones, [2011b](#ref-jones2011p72)) of the six activities involved in PIM). The five stages, illustrated in Figure 2, each of which can be driven by the user, the SI system or both, are:

* *preparation* - motivating oneself and deciding what to collect,
* *collection* - recording or capturing subjective and objective data manually or automatically,
* *integration* - combining, organising and transforming the data so that it can be interpreted as needed,
* *reflection* - reviewing, exploring, interrogating and considering the combined and analysed information, and
* *action* - where, armed with a new understanding of oneself, the data subject may change behaviours or set new goals.

Of these, reflection is perhaps the most important. Reflective learning (Boud, Keogh and Walker, [1985](#ref-boud1985)) has been recognised as a valuable means of knowledge acquisition and improvement in a variety of contexts including education (Dewey, [1938](#ref-dewey1938)), business (Beck *et al.*, [2001](#ref-beck2001)), and research (Lewin, [1946](#ref-lewin1946)). In the context of the wisdom curve (see Figure 1 above), reflection can be seen as asking questions of data in order to acquire knowledge about oneself. Knowledge about onself (a.k.a. self-insight (Hixon and Swann, [1993](#ref-hixon1993))) serves not only to satisfy curiosity (Li, Dey and Forlizzi, [2010](#ref-li2010)) but can improve self-control (O’Donoghue and Rabin, [2001](#ref-o2001)), increase self-awareness (Aslam *et al.*, [2016](#ref-aslam2016)) and enable positive behaviours such as saving energy (Seligman and Darley, [1976](#ref-seligman1976)).

Reflection can be facilitated in SI systems by enabling the tracking of subjective factors such as mood, health or activity, and can be triggered by means of notifications, or during more direct information exploration by the user as they recall or revisit experiences (Rivera-Pelayo *et al.*, [2012](#ref-rivera2012)). To aid interpretation of data by SI users, *contextualisation*, enhancing information with additional facts to ease its comprehension. This can include social, spatial or historical context, subjective or objective metadata or external sources of information (e.g. weather) (Rivera-Pelayo *et al.*, [2012](#ref-rivera2012)), or external devices (Dey, [2000](#ref-dey2000)). There are two phases of reflection, *discovery* and *maintenance*. During the initial discovery phase, typical questions that SI users ask concern the *history* of data changes, understanding the *context* of a datapoint, the *factors* that cause a pattern in data, and the identification of suitable *goals* to pursue. During the maintenance phase, questions asked concern *status* (how well you are doing at meeting your goals) and *discrepancies* (examining the difference between actual behaviour and desired behaviours). In order for a SI user to successfully reach this maintenance phase where they can continue to reflect upon their actions and adjust their goals, they must have been able to successfully navigate each of the 5 phases illustrated in Figure 2; if they have not collected the right data, they cannot integrate it, if they have not been able to integrate the collected data in a meaningful way, they cannot reflect upon it, and so on. Li et. al. framed this the *barriers cascade* (Li, Dey and Forlizzi, [2010](#ref-li2010)), and the pursuit of new ways to overcome these barriers has in effect been the major driving force behind all SI approaches; this is especially evident in the QSM (Choe *et al.*, [2014](#ref-choe2014)). Progress in easing the SI journey through the barriers cascade is being made: in 2011, Jones had noted that people often postpone or don’t have time for meta-level information management activities (W. Jones, [2011a](#ref-jones2011)), but by 2019 the increased automation around self-tracking and data collection was judged to have given people more free time and energy for reflection and managing their goals (Feng and Agosto, [2019](#ref-feng2019)).

### 2.2.4 The Emergence of Complex Digital Lives

[Target 600 words] - tolmie -sharing complexity, accountability, meaningful control - data traps, imbalance -

### 2.2.5 Research Gap: The Data Beyond The Individual

[Target 400 words]

## 2.3 Human-Centric Computing

[Target 2,500 words]

### 2.3.1 Human Computer Interaction Foundations

[Target 400 words]

### 2.3.2 Data Transcendence & Human Data Interaction

[Target 600 words]

### 2.3.3 People In Context: Human-Centred Design

[Target 600 words]

### 2.3.4 The Personal Data Economy

[Target 600 words]

### 2.3.5 Defining the Research Agenda for Human-Centricity in Practice

[Target 350 words]

## 2.4 Research Gap

[Target 500 words]

# 3 Approach & Methods

[Target X words]

* this PhD comprises a number of studies conducted between 2016 and 2021.
* there are three strands of enquiry:

## 3.1 1. The Human Perspective on Data Centricism

The objective of this part of the research is to fain a rich understanding of what data IS to people as they live their lives in the modern world.

Epistemiology:

* Here and throughout this thesis, we take a subjectivist perspective, looking more at what people think than trying to discover a single objective truth.
* As Peirce’s pragmatic maxim says, the meaning of anything in the world is understood through the conception of its practical effects, and theories that are more successful at controlling and predicting our world can be considered closer to the truth.
* We take an inductive research approach, finding patterns and forming theories to explain them

Therefore the key approach to this research objective is to be able to meaningfully engage with people about the subject of data, and conduct considered conversations to uncover their points of view.

(insert concentric circles diagram: Your Life > Your Data > )

Using a general approach of leading people into the subject of data using meaningful representations or starting with things that they already know (the devices, software and companies in their own life) I found effective ways to engage with participants and then get them talking about their data.

The interviews I conducted were divided into two strands: - exploring families’ data in a civic context, as it might be used in the provision of care or other state Services - exploring individual ‘digital life’ - the data that participants hold, use, or live with (or is held about them) by the companies they interact with.

(do these need little subsections as in 2 below?)

in the former case, I modelled data types and produced data cards and designed other activities to help stimulate discussion. In the latter, I used sketch interviewing as a technique. Then used standard qualitative interviewing as a technique (how much do I need to say about this?)

### 3.1.1 Data Cards and Card Sorting Techniques

Describe and show Family Civic Data cards used in 2017 study Describe the card-sorting techniques used - the riskiness vs who-should-control-it corkboard technique used in 2017 study

Describe why chose this - boundary objects (REF Star)/things to think with (REF Brandt and Messeter)

### 3.1.2 Family facts

Describe the facts-on-poles technique used in 2017 study.

### 3.1.3 Sketch Interviewing

Describe and show Sketch interviewing as used in the 2017 interviews and GDPR study.

### 3.1.4 Studies undertaken towards this research goal.

* 4 x 2 hr family ‘home visits’ (REF Mannay)
* 3 x 1hr ‘mobile phone/usefulness interviews’
* 5 x 1.5 hr ‘digital life sketch interviews (2017)’
* 11 x 1 hr ‘digital life interviews’ (2019)

## 3.2 2. The Lived Experience of A Data-Centric world

follow same pattern as above.

this time, importance of looking beyond the individual. looking at the way data exists and is used within relationships people have with organisations that hold data, within the data centric world.

additional Epistemiologies we draw on here are foucault’s thinking that meaning comes from discourse, and Deweyan pragmatism, that the purpose of research is to find solutions to real world problems in a manner that the solutions are generalisable.

here we move beyond people’s perspectives and thoughts to their actual lived experiences. our objective here is to understand the lived experience in a data centric world and how data can affect your life and relationships.

building on the prior research around attempts to model and interact with data in a more whole life, systemic way, we step up and look at data from the ‘life’ level - where we can see that, given the fact that most data is not directly held by individuals but by organisations - its role in relationships is key.

### 3.2.1 Early help

* explain the early help context and how it is a good setting to explore data relations especially w.r.t to empowerment

In order to fully explore this we conducted qualitative fieldwork with both sides of the early help relationship, exploring both current experiences of data use but also imagined future uses with different configurations that might be considered more human centric.

### 3.2.2 Card Sorting

### 3.2.3 Sentence Ranking

### 3.2.4 Ideation Grids

### 3.2.5 Group Poster Design

### 3.2.6 Storyboarding Cards

(participatory design) Quick explainer and reference to Golembewski

### 3.2.7 GDPR

* explain the GDPR context and how this is also a way to explore practical experience of real data processes

### 3.2.8 Privacy Policy interviews

as a starting point, having identified companies in people’s lives I start with the objective reality of what those companies say they collect, and use that as a way in to meaningful data conversations with participants.

### 3.2.9 View Your Data interviews

then to explore a real world process, I took participants on a journey of discovery of requesting, accessing and viewing their own held data.

### 3.2.10 Studies undertaken towards this research goal.

HDI in context of care study - 1 x 2h workshop with families (x participants) - 2 x 2h workshop with staff (x participants) - 1 x 2h workshop with staff and parents (x participants)

GDPR study - 10 x 1h+ privacy policy discussion - guiding through email etc - 10 x 2h view your data discussion

## 3.3 3. Uncovering and Exploring Design Constraints and Opportunities

Epistemiology: practical, pragmatism. What is possible. research through design.

This research goal starts from a different perspective, looking from the designer / technologists perspective, to gain an empathetic understanding of the current status quo of data systems and then applying human-centric system to reconsider the design of these systems. The goal is to uncover the barriers and challenges that must be overcome through design of systems and processes, and exploring the possible avenues of design that could be exploited

Here, research was conducted through embedding myself in projects aligned to the research objectives of building new data systems. Not empirical structured research but rather a process of acculturation to the world of building data systems and learning through participation and through research through design and through technical prototyping.

## 3.4 Research Through design

background on how this can be framed as research through design

## 3.5 Rapid prototyping

background on how prototyping is used to explore new ideas - pivoting etc.

### 3.5.1 Studies undertaken towards this research goal

two major projects:

* SILVER (ref early help as earlier)
* BBC R & D/PDS

two minor projects:

* web augmentation
* experimenting with using own data and getting usable data with GDPR

## 3.6 Structure of the following chapters.

Findings from the first research goal, The Human Perspective on Data Centricism, are written up in chapter 4, Understanding data

Findings from the second research goal, The Lived Experience of A Data-Centric world, are written up in chapter 5, understanding data relations.

Findings from the third research goal, Uncovering and Exploring Design Constraints and Opportunities are written up in chapter 6, understanding design constraints and possibilities.

In chapter 7, these findings are brought together to address the overall research question and suggest avenues for future research, design and development that can be applied to design better human data relations.

# 4 The Human Perspective on Data and Data-Centrism

[Target x words]

# 5 Living In A Data-Centric World

[Target x words]

# 6 Human-Centricity: Design Constraints and Opportunities

[Target x words]

# 7 Designing Better Human Data Relations

[Target x words]

# Bibliography

Abbattista, F. *et al.* (2007) ‘Shaping personal information spaces from collaborative tagging systems’, in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. (PART 3), pp. 728–735. doi: [10.1007/978-3-540-74829-8\_89](https://doi.org/10.1007/978-3-540-74829-8_89).

Abiteboul, S., André, B. and Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 5. ACM, pp. 32–35. doi: [10.1145/2670528](https://doi.org/10.1145/2670528).

‘About The Quantified Self’ (no date). Available at: <https://quantifiedself.com/about/what-is-quantified-self/> (Accessed: 22 March 2021).

Abowd, G. D. and Mynatt, E. D. (2000) ‘Charting Past, Present, and Future Research in Ubiquitous Computing’, *ACM Transactions on Computer-Human Interaction*, 7(1), pp. 29–58. doi: [10.1145/344949.344988](https://doi.org/10.1145/344949.344988).

Ackoff, R. L. (1989) ‘From data to wisdom’, *Journal of Applied Systems Analysis*, 16(1), pp. 3–9.

Andrews, R. (2005) ‘GTD : A New Cult for the Info Age’, *Wired*. Available at: <https://www.wired.com/2005/07/gtd-a-new-cult-for-the-info-age/>.

Aslam, H. *et al.* (2016) ‘Harnessing Smartphones as a Personal Informatics Tool towards Self-Awareness and Behavior Improvement’, *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*. IEEE, pp. 467–474. doi: [10.1109/DASC-PICom-DataCom-CyberSciTec.2016.92](https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.92).

Barreau, D. K. (1995) ‘Context as a factor in personal information management systems’, *Journal of the American Society for Information Science*, 46(5), pp. 327–339. doi: [10.1002/(SICI)1097-4571(199506)46:5<327::AID-ASI4>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-4571(199506)46:5%3C327::AID-ASI4%3E3.0.CO;2-C).

Barreau, D. and Nardi, B. A. (1995) ‘Finding and reminding’, *ACM SIGCHI Bulletin*, 27(3), pp. 39–43. doi: [10.1145/221296.221307](https://doi.org/10.1145/221296.221307).

Beck, K. *et al.* (2001) ‘The Agile Manifesto’. Available at: <http://agilemanifesto.org/>.

Bell, G. and Gemmell, J. (2009) *Total recall: how the E-memory revolution will change everything*. Dutton (09), pp. 47–5062–47–5062. doi: [10.5860/choice.47-5062](https://doi.org/10.5860/choice.47-5062).

Bergman, O. (2013) ‘The Effect of Folder Structure on Personal File Navigation’, *Journal of the American Society for Information Science and Technology*, 64(July), pp. 1852–1863. doi: [10.1002/asi](https://doi.org/10.1002/asi).

Bergman, O., Beyth-Marom, R. and Nachmias, R. (2003) ‘The user-subjective approach to personal information management systems’, *Journal of the American Society for Information Science and Technology*, 54(9), pp. 872–878. doi: [10.1002/asi.10283](https://doi.org/10.1002/asi.10283).

Bergman, O. *et al.* (2008) ‘Improved search engines and navigation preference in personal information management’, *ACM Transactions on Information Systems*, 26(4). doi: [10.1145/1402256.1402259](https://doi.org/10.1145/1402256.1402259).

Bergman, O. *et al.* (2012) ‘How do we find personal files?: The effect of OS, presentation & depth on file navigation’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2977–2980. doi: [10.1145/2207676.2208707](https://doi.org/10.1145/2207676.2208707).

Boud, D., Keogh, R. and Walker, D. (1985) *Reflection: Turning experience into learning*. Routledge.

Bowyer, A. (2011) ‘Why files need to die’. Available at: <http://radar.oreilly.com/2011/07/why-files-need-to-die.html>.

Brooks, D. (2013) ‘The Philosophy of Data’. Available at: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>.

Bush, V. (1945) ‘As we may think’, *The Atlantic Monthly*, 3(2), pp. 35–46. doi: [10.1145/227181.227186](https://doi.org/10.1145/227181.227186).

Choe, E. K. *et al.* (2014) ‘Understanding quantified-selfers’ practices in collecting and exploring personal data’, in *Proceedings of the 32nd annual acm conference on human factors in computing systems - chi ’14*. New York, New York, USA: ACM Press, pp. 1143–1152. doi: [10.1145/2556288.2557372](https://doi.org/10.1145/2556288.2557372).

Crabtree, A. and Tolmie, P. (2018) ‘The practical politics of sharing personal data’, in *Personal and Ubiquitous Computing*. Springer-Verlag (2), pp. 293–315. doi: [10.1007/s00779-017-1071-8](https://doi.org/10.1007/s00779-017-1071-8).

‘Data’ (no date). Grammarist. Available at: <https://grammarist.com/usage/data/>.

Decker, S. and Frank, M. (2004) ‘The Networked Semantic Desktop’, *WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web*. doi: [10.1108/eb057368](https://doi.org/10.1108/eb057368).

‘Delicious’ (2003). Available at: <https://en.wikipedia.org/wiki/Delicious_(website)>.

Dewey, J. (1938) ‘Experience and education’.

Dey, A. K. (2000) *Providing Architectural Support for Building Context-Aware Applications*. PhD thesis.

Dey, A. K. (2001) ‘Understanding and using context’, *Personal and ubiquitous computing*, pp. 4–7. Available at: [www.cc.gatech.edu/fce/contexttoolkit https://dl.acm.org/citation.cfm?id=593572%0Ahttp://dl.acm.org/citation.cfm?id=593572](www.cc.gatech.edu/fce/contexttoolkit%20https://dl.acm.org/citation.cfm?id=593572%0Ahttp://dl.acm.org/citation.cfm?id=593572).

Dourish, P. (2003) ‘The appropriation of interactive technologies: Some Lessons From Placeless Documents’, *Computer Supported Cooperative Work*, 12(4), pp. 465–490.

Dourish, P. (2004) ‘What we talk about when we talk about context’, *Personal and Ubiquitous Computing*, 8(1), pp. 19–30. doi: [10.1007/s00779-003-0253-8](https://doi.org/10.1007/s00779-003-0253-8).

Dourish, P. *et al.* (2000) ‘Extending document management systems with user-specific active properties’, *ACM Transactions on Information Systems*, 18(2), pp. 140–170. doi: [10.1145/348751.348758](https://doi.org/10.1145/348751.348758).

Engelbart, D. C. (1962) ‘Augmenting human intellect: A conceptual framework’. Menlo Park, CA, USA: Stanford Research Institute.

Etzel, B. (1995) ‘New strategy and techniques to cope with information overload’, in *IEE colloquium on information overload*. IEE (223), pp. 2–2. doi: [10.1049/ic:19951427](https://doi.org/10.1049/ic:19951427).

European Union Agency for Fundamental Rights (2020) ‘Your Rights Matter: Data Protection and Privacy 2020’, p. 20. doi: [10.2811/031862](https://doi.org/10.2811/031862).

‘Facebook–Cambridge Analytica Data Scandal’ (2014). Available at: <https://en.wikipedia.org/wiki/Facebook–Cambridge_Analytica_data_scandal>.

Feng, Y. and Agosto, D. E. (2019) ‘Revisiting personal information management through information practices with activity tracking technology’, *Journal of the Association for Information Science and Technology*, 70(12), pp. 1352–1367. doi: [10.1002/asi.24253](https://doi.org/10.1002/asi.24253).

Foulonneau, M. and Riley, J. (2008) *Metadata for digital resources : implementation, systems design and interoperability*. Chandos Pub, p. 203.

Freeman, E. and Gelernter, D. (1996) ‘Lifestreams: A Storage Model for Personal Data’, *SIGMOD Record (ACM Special Interest Group on Management of Data)*. Association for Computing Machinery (ACM), 25(1), pp. 80–86. doi: [10.1145/381854.381893](https://doi.org/10.1145/381854.381893).

Frost, A. (2019) ‘Forget Folders: The Best Ways to Organize Your Files with Tags and Labels’. Available at: <https://zapier.com/blog/how-to-use-tags-and-labels/>.

Gelernter, D. (1994) ‘The cyber-road not taken: Lost on the info-highway? Here’s some stuff that could really change your life.’, *The Washington Post*, 3, pp. 1–5. Available at: [https://www.nexis.com/results/enhdocview.do?docLinkInd=true&ersKey=23\_T28577898130&format=GNBFI&startDocNo=51&resultsUrlKey=0\_T28577898181&backKey=20\_T28577898182&csi=8075&docNo=61 https://www.nexis.com/results/enhdocview.do?docLinkInd=true&ersKey=23\_T280](https://www.nexis.com/results/enhdocview.do?docLinkInd=true&ersKey=23_T28577898130&format=GNBFI&startDocNo=51&resultsUrlKey=0_T28577898181&backKey=20_T28577898182&csi=8075&docNo=61%20https://www.nexis.com/results/enhdocview.do?docLinkInd=true&ersKey=23_T280).

Gemmell, J., Bell, G. and Lueder, R. (2006) ‘MyLifeBits: A personal database for everything’, *Communications of the ACM*, 49(1), pp. 88–95. doi: [10.1145/1107458.1107460](https://doi.org/10.1145/1107458.1107460).

‘Google Desktop Search’ (2004). Available at: <https://en.wikipedia.org/wiki/Google_Desktop>.

Gurstein, M. (2003) ‘Effective use: A community informatics strategy beyond the digital divide’, *First Monday*, 8(12). doi: [10.5210/fm.v0i0.1798](https://doi.org/10.5210/fm.v0i0.1798).

Gurstein, M. B. (2011) ‘Open data: Empowering the empowered or effective data use for everyone?’, *First Monday*. First Monday, 16(2). doi: [10.5210/fm.v16i2.3316](https://doi.org/10.5210/fm.v16i2.3316).

Hart-Davidson, W., Zachry, M. and Spinuzzi, C. (2012) ‘Activity streams: Building context to coordinate writing activity in collaborative teams’, in *SIGDOC’12 - proceedings of the 30th acm international conference on design of communication*. New York, New York, USA: ACM Press, pp. 279–287. doi: [10.1145/2379057.2379109](https://doi.org/10.1145/2379057.2379109).

Hendler, J. and Berners-Lee, T. (2010) ‘From the Semantic Web to social machines: A research challenge for AI on the World Wide Web’. doi: [10.1016/j.artint.2009.11.010](https://doi.org/10.1016/j.artint.2009.11.010).

Hixon, J. G. and Swann, W. B. (1993) ‘When Does Introspection Bear Fruit? Self-Reflection, Self-Insight, and Interpersonal Choices’, *Journal of Personality and Social Psychology*, 64(1), pp. 35–43. doi: [10.1037/0022-3514.64.1.35](https://doi.org/10.1037/0022-3514.64.1.35).

Hoofnagle, C. J., Sloot, B. van der and Borgesius, F. Z. (2019) ‘The European Union general data protection regulation: What it is and what it means’, *Information and Communications Technology Law*. Taylor & Francis, 28(1), pp. 65–98. doi: [10.1080/13600834.2019.1573501](https://doi.org/10.1080/13600834.2019.1573501).

Hotho, A., Nürnberger, A. and Paaß, G. (2005) ‘A brief survey of text mining.’, in *Ldv forum*. Citeseer (1), pp. 19–62.

Hutton, D. M. (2012) ‘Turing’s Cathedral: The Origins of the Digital Universe’. Emerald Group Publishing Limited.

‘Information’ (no date). Available at: <https://en.wikipedia.org/wiki/Information>.

Information Commissioner’s Office (2014) ‘Data controllers and data processors: what the difference is and what the governance implications are’, p. 20. Available at: <https://ico.org.uk/media/for-organisations/documents/1546/data-controllers-and-data-processors-dp-guidance.pdf%0Ahttps://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/>.

Information Commissioner’s Office (2018) ‘Your data matters - Your rights’. Available at: <https://ico.org.uk/your-data-matters/>.

‘Infovark Company Profile’ (2007). Available at: <https://www.crunchbase.com/organization/infovark>.

Jones, W. (2011a) ‘The Future of Personal Information Management Part I: Our Information, Always and Forever’.

Jones, W. (2011b) ‘The Future of Personal Information Management Part I: Our Information, Always and Forever’, p. 72.

Karger, D. R. *et al.* (2005) ‘Haystack: A customizable general-purpose information management tool for end users of semistructured data’, in *2nd biennial conference on innovative data systems research, cidr 2005*, pp. 13–27. Available at: [https://s3.amazonaws.com/academia.edu.documents/46870765/haystack.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1552482265&Signature=DUiWWHjrmBJbaWqlsCQcQj6POmg%3D&response-content-disposition=inline%3B filename%3DHaystack\_A\_customizable\_general-purpose](https://s3.amazonaws.com/academia.edu.documents/46870765/haystack.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1552482265&Signature=DUiWWHjrmBJbaWqlsCQcQj6POmg%3D&response-content-disposition=inline%3B%20filename%3DHaystack_A_customizable_general-purpose).

Karger, D. R. and Jones, W. (2006) ‘Data unification in personal information management’, *Communications of the ACM*, 49(1), p. 77. doi: [10.1145/1107458.1107496](https://doi.org/10.1145/1107458.1107496).

Kelly, K. and Wolf, G. (2007) ‘What is the quantified self’. Available at: <https://web.archive.org/web/20100507215130/http://www.kk.org/quantifiedself/2007/10/what-is-the-quantifiable-self.php>.

Kelly, R. (2020) ‘The Biggest ICO Fines Ever Issued’. Available at: <https://digit.fyi/data-protection-2020-the-biggest-fines-ever-issued-by-the-ico/>.

Klein, B. *et al.* (2004) ‘Enabling flow - A paradigm for document-centered personal information spaces’, in *Proceedings of the eighth iasted international conference on artificial intelligence and soft computing*, pp. 187–192. Available at: <https://www.semanticscholar.org/paper/Enabling-flow%3A-%7BA%7D-paradigm-for-document-centered-Klein-Agne/22be4a7b25e75de235e5d96bad6ab4ab4583daac>.

Krishnan, A. and Jones, S. (2005) ‘TimeSpace: Activity-based temporal visualisation of personal information spaces’, *Personal and Ubiquitous Computing*, 9(1), pp. 46–65. doi: [10.1007/s00779-004-0291-x](https://doi.org/10.1007/s00779-004-0291-x).

Lansdale, M. and Edmonds, E. (1992) ‘Using memory for events in the design of personal filing systems’, *International Journal of Man-Machine Studies*, 36(1), pp. 97–126. doi: [10.1016/0020-7373(92)90054-O](https://doi.org/10.1016/0020-7373(92)90054-O).

Lansdale, M. W. (1988) ‘The psychology of personal information management’, *Applied Ergonomics*, 19(March 1988), pp. 55–66. doi: [10.1016/0003-6870(88)90199-8](https://doi.org/10.1016/0003-6870(88)90199-8).

Larsson, S. (2018) ‘Algorithmic governance and the need for consumer empowerment in data-driven markets’, *Internet Policy Review*, 7(2). doi: [10.14763/2018.2.791](https://doi.org/10.14763/2018.2.791).

Leprince-Ringuet, D. (2021). Available at: <https://www.zdnet.com/article/gdpr-fines-increased-by-40-last-year-and-theyre-about-to-get-a-lot-bigger/>.

Lewin, K. (1946) ‘Action Research and Minority Problems’, *Journal of Social Issues*, 2(4), pp. 34–46. doi: [10.1111/j.1540-4560.1946.tb02295.x](https://doi.org/10.1111/j.1540-4560.1946.tb02295.x).

Li, I. (2009) ‘Designing Personal Informatics Applications and Tools that Facilitate Monitoring of Behaviors’, *Uist*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.232.8536>.

Li, I., Dey, A. and Forlizzi, J. (2010) ‘A stage-based model of personal informatics systems’, *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*. New York, New York, USA: ACM Press, p. 557. doi: [10.1145/1753326.1753409](https://doi.org/10.1145/1753326.1753409).

Lindley, S. E. *et al.* (2018) ‘Exploring new metaphors for a networked world through the file biography’, *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April, pp. 1–12. doi: [10.1145/3173574.3173692](https://doi.org/10.1145/3173574.3173692).

Malone, T. W. (1983) ‘How do people organize their desks?: Implications for the design of office information systems’, *ACM Transactions on Information Systems*, 1(1), pp. 99–112. doi: [10.1145/357423.357430](https://doi.org/10.1145/357423.357430).

Marshall, C. C. and Jones, W. (2006) ‘Keeping encountered information’, *Communications of the ACM*, 49(1), pp. 66–67. doi: [10.1145/1107458.1107493](https://doi.org/10.1145/1107458.1107493).

Millar, S. (2002) ‘UK singled out for criticism over protection of privacy’. Available at: <https://www.theguardian.com/technology/2002/sep/05/security.humanrights>.

Nadeem, D. and Sauermann, L. (2007) ‘From Philosophy and Mental-Models to Semantic Desktop Research: Theoretical Overview’.

Negroponte, N. and Bolt, R. A. (1978) *Spatial data management system*. MASSACHUSETTS INST OF TECH CAMBRIDGE ARCHITECTURE MACHINE GROUP.

Nelson, T. (2006) ‘Lost in hyperspace’, *New Scientist*, 191(2561). doi: [10.1002/elsc.200620112](https://doi.org/10.1002/elsc.200620112).

Nelson, T. H. (1965) ‘Complex information processing’, pp. 84–100. doi: [10.1145/800197.806036](https://doi.org/10.1145/800197.806036).

Odom, W. *et al.* (2018) ‘Time, Temporality, and Slowness’, pp. 383–386. doi: [10.1145/3197391.3197392](https://doi.org/10.1145/3197391.3197392).

O’Donoghue, T. and Rabin, M. (2001) ‘Choice and procrastination’, *The Quarterly Journal of Economics*. MIT Press, 116(1), pp. 121–160.

Organisation for Economic Co-operation and Development (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Available at: <https://www.oecd.org/digital/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm>.

‘Our Digital Lives’ (2018) in *TED talks*. TED. Available at: <https://www.ted.com/playlists/26/our_digital_lives>.

Pór, G. (1997) ‘Designing Knowledge Ecosystems for Communities of Practice’, in *Advancing organizational capability via knowledge management*.

Rivera-Pelayo, V. *et al.* (2012) ‘A framework for applying Quantified Self approaches to support reflective learning’, *Proceedings of the IADIS International Conference Mobile Learning 2012, ML 2012*, pp. 123–131.

Ross, G. (2005) ‘An introduction to Tim Berners-Lee’s Semantic Web’. Available at: [https://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/ http://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/](https://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/%20http://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/).

Sauermann, L., Bernardi, A. and Dengel, A. (2005) ‘Overview and outlook on the semantic desktop’, in *CEUR workshop proceedings*.

Schumacher, K., Sintek, M. and Sauermann, L. (2008) ‘Combining fact and document retrieval with spreading activation for semantic desktop search’, in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 569–583. doi: [10.1007/978-3-540-68234-9\_42](https://doi.org/10.1007/978-3-540-68234-9_42).

Seligman, C. and Darley, J. M. (1976) ‘Feedback as a means of decreasing residential energy consumption’, *Journal of Applied Psychology*, 62(4), pp. 363–368. doi: [10.1037/0021-9010.62.4.363](https://doi.org/10.1037/0021-9010.62.4.363).

Shannon, C. E. (1948) ‘A mathematical theory of communication’, *The Bell system technical journal*. Nokia Bell Labs, 27(3), pp. 379–423.

Shipman, F. M. (. and Marshall, C. C. (1999) ‘Formality Considered Harmful : Experiences , Emerging Themes , and Directions on the Use of Formal Representations in Interactive Systems’, pp. 333–352.

Shneiderman, B. (1996) *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualisations*.

Simon, H. A. and Newell, A. (1958) ‘Heuristic Problem Solving: The next advance in operations research’. doi: [10.1057/978-1-349-94848-2\_792-1](https://doi.org/10.1057/978-1-349-94848-2_792-1).

Spiekermann, S. and Korunovska, J. (2017) ‘Towards a value theory for personal data’, *Journal of Information Technology*, 32(1), pp. 62–84. doi: [10.1057/jit.2016.4](https://doi.org/10.1057/jit.2016.4).

Steinberg, S. G. (1997) ‘Lifestreams’, *Wired*. Available at: <https://www.wired.com/1997/02/lifestreams/>.

Teevan, J. *et al.* (2004) ‘The perfect search engine is not enough: A study of orienteering behavior in directed search’, in *Conference on human factors in computing systems - proceedings*, pp. 415–422. Available at: <http://people.csail.mit.edu/teevan/work/publications/papers/chi04.pdf>.

Teevan, J. B. (2001) ‘Displaying dynamic information’, in *Conference on human factors in computing systems - proceedings*, pp. 417–418. doi: [10.1145/634067.634311](https://doi.org/10.1145/634067.634311).

Terdiman, D. (2008) ‘Using tags to improve the Flickr experience’. Available at: <https://www.cnet.com/news/using-tags-to-improve-the-flickr-experience/>.

The European Parliament and the Council of the European Union (2016) ‘Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data’, pp. 16–32. Available at: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679%20https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES).

‘The GDPR: Does it Benefit Consumers in Any Practical Way?’ (2020). Atebits.com. Available at: <https://www.atebits.com/the-gdpr-does-it-benefit-consumers-in-any-practical-way/>.

‘The personal computer revolution’ (no date) in *Britannica*. Available at: <https://www.britannica.com/technology/computer/The-personal-computer-revolution>.

Toonders, J. (2014) ‘Data Is the New Oil of the Digital Economy’. Available at: [https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/ https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/%0Ahttp://www.wired.com/insights/2014/07/data-new-oil-digital-economy/](https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/%20https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/%0Ahttp://www.wired.com/insights/2014/07/data-new-oil-digital-economy/).

US Department of Health Education and Welfare (1973) ‘Records Computers and the Rights of Citizens’.

Wallace, D. P. (2007) *Knowledge management: Historical and cross-disciplinary themes*. Libraries unlimited.

Wellisch, H. H. (1996) *Abstracting, indexing, classification, thesaurus construction: A glossary*. American Society of Indexers.

Whittaker, S. and Hirschberg, J. (2001) ‘The Character, Value, and Management of Personal Paper Archives’, *ACM Transactions on Computer-Human Interaction*, 8(2), pp. 150–170. doi: [10.1145/376929.376932](https://doi.org/10.1145/376929.376932).

‘WinFS’ (no date). Available at: <https://en.wikipedia.org/wiki/WinFS>.

World Economic Forum (2011) *Personal data : The emergence of a new asset class*, pp. 1–40. Available at: [http://www3.weforum.org/docs/WEF\_ITTC\_PersonalDataNewAsset\_Report\_2011.pdf http://www.weforum.org/reports/personal-data-emergence-new-asset-class](http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf%20http://www.weforum.org/reports/personal-data-emergence-new-asset-class).

World Economic Forum (2013) ‘Unlocking the Value of Personal Data: From Collection to Usage Prepared in collaboration with The Boston Consulting Group Industry Agenda’, (February).

World Economic Forum (2014a) *Rethinking Personal Data : A New Lens for Strengthening Trust*. May, p. 35. Available at: <http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf>.

World Economic Forum (2014b) *Rethinking personal data: Trust and context in user-centred data ecosystems*. May, p. 35. Available at: <http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf>.

Zins, C. (2015) ‘What is the meaning of "data", "information", and "knowledge"?’, *Institute of Knowledge Sharing*, 3(1).

Ziogas, G. (2020) ‘The Inventor of the World Wide Web Says the Internet Is Broken’. Available at: <https://medium.com/digital-diplomacy/the-inventor-of-the-world-wide-web-says-the-internet-is-broken-fbce1c8bf6cf>.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile. Available at: <https://books.google.co.uk/books?id=W7ZEDgAAQBAJ>.