Understanding and Improving Human Data Relations

Alex Bowyer

2 Literature Review

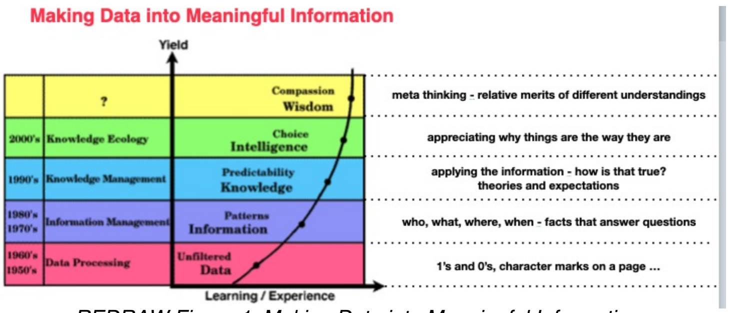
2.1 Data-centrism and the Need for Access

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). 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), 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) 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). 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). Figure 1 builds upon a representation by George Pór (Pór, 1997) 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). The exact origin of this definition is unknown but it is often attributed to mathematician Claude Shannon (Shannon, 1948). 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 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 Personal Data & the Rise of Data-centrism

The earliest computer systems used data to store mathematical 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). 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) 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). 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). This philosophy is also known as ‘data-ism’ (Brooks, 2013) and the analysis and exploitation of such data at scale is known as ‘big data’ (Neef, 2015).

As a result of data-ism, the collection of data about people has become an inevitable part of modern life. We live ‘digital lives’ (Various Authors, 2018) 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)) 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.

In 2017, the average American Internet user had 150 online user accounts with different providers (Caruthers, 2018). Data for the UK show the number of service and supply relationships each individual has to manage increasing from around 45 in 1997 to around 250 in 2020 (Henderson and Group, 2020). As the amount of personal data relating to each of us has increased, the need for individuals to be able to manage this has grown, but unfortunately, the large-scale systems which collect data about us now function as ‘data traps’ (Abiteboul, André and Kaplan, 2015) - 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” (Hoffman, 2011, 2013, 2014b, 2014a).

2.1.3 Legislating to Protect Personal Data: the Story so far

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). 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).

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 (ICO)) 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), and research commissioned but the ICO in 2008 stated that the European Data Protection Directive was “out of date, bureaucratic and excessively prescriptive” (McCullagh, 2009).

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; Leprince-Ringuet, 2021). This legislation has finally given individuals the practical means to exercise their data rights ('The GDPR: Does it Benefit Consumers in Any Practical Way?', 2020) against a backdrop of massive personal data use across society that had previously rendered data access requests impractical (Cormack, 2016). The GDPR – which gives individuals key rights including rights to timely data access, explanation, erasure and correction (Information Commissioner's Office, 2018) – 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).

Following the Snowden revelations (Gellman, 2013) in 2013, attention and concern over personal data use has grown year on year. In 2018, the Cambridge Analytica scandal ('Facebook–Cambridge Analytica Data Scandal', 2014; Chang, 2018) 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) 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; The European Parliament and the Council of the European Union, 2016):

- *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 Personal 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 (Hoffman, 2011). It is almost impossible for people to assess that value, because they are unaware of most of their data (Spiekermann and Korunovska, 2017). 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). 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 (Gurstein, 2003). Gurstein later identified seven aspects that are necessary for access to be effective (Gurstein, 2011) 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 for the layperson.
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*: Information presentation should be as clear as possible so that people can 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).

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; Karger and Jones, 2006). Having our documents distributed across multiple platforms, applications and devices makes interrogation and orienteering hard (Krishnan and Jones, 2005). 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). 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). 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). While any one of the capabilities mentioned in this paragraph does exist in at least some 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 Personal 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). 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), for interpersonal communication (Shannon, 1948), to augment human intellect (Engelbart, 1962) and to model human thought (Simon and Newell, 1958).

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) seemed to place the potential power that "having reduced your affairs to software, software can take care of them for you" (Gelernter, 1994) 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 also 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). *Personal Information Management (PIM)* was first mentioned in 1988 by Mark Lansdale, who identified a need to design information management systems according to the psychology of the people who use them rather than by simulating office practices. By paying attention to how people categorise, recognise and recall information, and labelling information with appropriate attributes, information can be retrieved by different properties (Lansdale, 1988). 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).

Driven in part by the pursuit of better "time management" in the late 20th century (characterised by PDAs, palmtops and electronic organisers) (Etzel, 1995) and the focus on personal productivity in the early 2000s (characterised by 'GTD' (Getting Things Done) self-help books and to-do list software) (Andrews, 2005) 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), much as as people keep current information 'in reach' on a desk (Klein *et al.*, 2004). Spatial approaches recognise that *keeping* is a valuable activity in its own right, that informs sensemaking (Marshall and Jones, 2006). Placed information also performs an important *reminding* function (Barreau, 1995; Barreau and Nardi, 1995).

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) 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; Nelson, 2006; Ziogas, 2020). In the absence of connected networks of personal information and with people collecting more information than they discard (Whittaker and Hirschberg, 2001), the 2000s saw software like Google Desktop Search ('Google Desktop Search', 2004) and Infovark ('Infovark Company Profile', 2007) emerge to try and discover users' data files and unify access to them, with limited impact (Bergman *et al.*, 2008). 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). 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; Dourish, 2003). *Metadata* – information about what the data *is* – is critical to information organisation (Foulonneau and Riley, 2008). 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). Tags, which emerged as a means to organise data through systems like del.icio.us ('Delicious', 2003) and Flickr in the 2000s, are still widely used on social media and websites today, and are even available within macOS (Frost, 2019). 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; Terdiman, 2008).

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). The focus is on both the retrieval of documents and of facts (Schumacher, Sintek and Sauermann, 2008). 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), 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). It has even been proposed that AI might help computers to understand users' mental models (Nadeem and Sauermann, 2007).

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; Bergman, 2013), 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). 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). 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). 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). 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). 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), 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). 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).

In 1995, Barreau highlighted the importance of *context* to PIM; People need access to different information according to what they are doing (Barreau, 1995) In 2000, Abowd and Mynatt 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). *Context-aware computing* (Abowd *et al.*, 1999; Eliasson, Cerratto Pargman and Ramberg, 2009) has subsequently emerged as a sub-discipline of research in its own right (Dey, 2001) (see also section 2.3.2). 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); 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) - 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), 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). 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 (Jilek *et al.*, 2018). 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).

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). Bergman et. al. (Bergman, Beyth-Marom and Nachmias, 2003) 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). With semantic PIM systems, we can see that a successful system (or at least, its designers) must understand a great deal about their users.

2.2.3 Personal Informatics & The Quantified Self

In the late '00s, researchers and enthusiasts took PIM beyond task management and turned PIM thinking toward the self. In pursuit of Bush's vision of augmenting human memory, Jim Gemmell and Gordon Bell in their MyLifeBits project at Microsoft (Gemmell, Bell and Lueder, 2006; Bell and Gemmell, 2009) tried to capture an entire life electronically. This became known as *lifelogging*: gathering as much data as possible, so that the maximum possible context, detail and understanding can be gained about that individual. 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 such as step counts, heart rates or calories burned, but almost any aspect of your own life that could be numerically recorded in a computer (Kelly and Wolf, 2007). The Quantified Self movement (QSM) is now a world-wide community of enthusiasts who have developed hundreds of tools and techniques for self-tracking/lifelogging and 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). 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)." 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 integrating and reviewing self-tracking data, there is so much overlap that all three can 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 organisation.

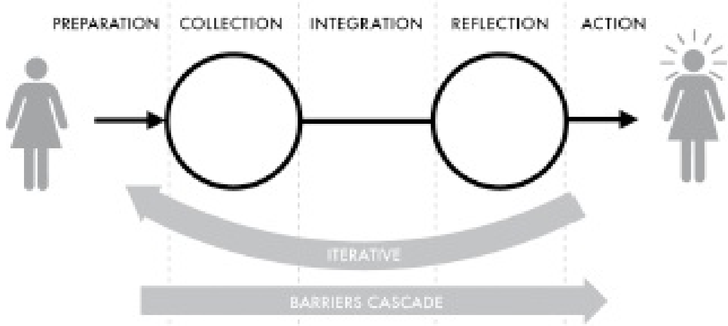


Figure 2: Li et. al.'s Stage-based Model of Personal Informatics Systems (Li, Dey and Forlizzi, 2010)

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) 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, as the capacity to gain new insight is the motivating reason to engage in SI. Reflective learning (Boud, Keogh and Walker, 1985) has been recognised as a valuable means of knowledge acquisition and improvement in a variety of contexts including education (Dewey, 1938), business (Beck *et al.*, 2001), and research (Lewin, 1946). 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 oneself (a.k.a. self-insight (Hixon and Swann, 1993)) serves not only to satisfy curiosity (Li, Dey and Forlizzi, 2010) but can improve self-control (O'Donoghue and Rabin, 2001), increase self-awareness (Aslam *et al.*, 2016) and enable positive behaviours such as saving energy (Seligman and Darley, 1976).

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). 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), or external devices (Dey, 2000). 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, these goals frame the questions asked, which 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), and the pursuit of new ways to overcome these barriers has in effect been the major problem space for all SI approaches; this is especially evident in the QSM (Choe *et al.*, 2014). While effortless SI is not yet a reality and many barriers still exist, 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), 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).

2.2.4 The Emergence of Complex Digital Lives

As described in 2.1.2 above, the rise of data-centrism has meant that every aspect of our lives now involves digital service providers and products which process personal data. Smartphones put computers in everyone's pockets, and cheap cloud computing and an open web allowed every organisation to serve the population digitally through apps and websites. In 2010, broadband access was declared a legal right in Finland ('Finland: Broadband Access Made Legal Right In Landmark Law', 2010), and in 2011, the UK Supreme Court declared that Internet access was an "essential part of everyday living" and denial of Internet access for criminals such as sex offenders was ruled unlawful (Roche, 2011; Wagner, 2012). Everyone now required access to information and online digital services. "The boundary between real life and online [had] disappeared" (Burkeman, 2011). The promise that whatever you want to do "there's an app for that" had become true (Apple, 2009). During the late '00s and throughout the 2010s data-centric companies disrupted almost every industry: Amazon (shopping & books), Uber (taxis), Netflix (movie rental), Spotify (music), AirBNB (accommodation), Google (email, news & advertising), Facebook (social networking & advertising), Paypal/Revolut/Monzo (banking), match/Tinder (dating), Steam (computer games), Just Eat (takeaways), and many more (Levine, 2011; Carter, 2015). As a result, we now produce rich data trails simply by going about our daily lives, and this has become "the driving force for value creation" online (Symons *et al.*, 2017). More recently as we start the 2020s, the trend has accelerated, with the COVID-19 pandemic necessitating the move of both information work and social activities to online using platforms such as Zoom, Google Docs and miro (O'Donnell, 2020).

Throughout the transition to this information economy, the computing industry has delivered revolutionary new capabilities, but with every provider offering their own apps and websites, the information landscape has become hugely challenging for people to manage; information overload is now a serious problem that has been linked to increased anxiety, impaired critical thinking, exhaustion, and loss of willpower and focus (Hemp, 2009; Tunikova, 2018; Fu *et al.*, 2020). Our personal information is fragmented and a unified interface is needed: "We must launch multiple applications and perform numerous repetitive searches for relevant information, to say nothing of deciding which applications to look in (Karger and Jones, 2006)." In the silo-ed world of today's Internet, this has only got worse. Bergman's subjective principles (see above) imply that our data should be able to move and be referenced freely, but it cannot. Our ability to share and connect data is limited (Crabtree and Tolmie, 2018). Our data is trapped (Abiteboul, André and Kaplan, 2015), not only because it is held by organisations without giving us effective access, but also by various practical means such as format incompatibilities, device restrictions, paywalls, and a lack of data portability. We need to free our data, as I expand upon (Bowyer, 2018).

It is clear that general-purpose computing has yet to provide people with the tools to manage their complex digital lives. There have been attempts to create general purpose interfaces for personal data, typically based around a timeline, such as AllOfMe.com ('AllofMe Company Profile', 2007; 'AllofMe.com Teaser Clip', 2008) in 2008 and myTimeline a decade later ('myTimeline', 2018); however none of these products have reached public availability. To date the closest market-successful tool that people have for general purpose information handling is Facebook, given that it can store personal information, handle asynchronous and instant messaging, news, photo sharing, some retail functionality, brand interaction & support, calendaring and event management, and group discussions. However, it is a closed system with no capability for customisation; none of its content is available outside the network and external content cannot be linked or interacted with except by import; as such it cannot be considered a PIM system. Its own Timeline feature, promoted at launch in 2011 as "the story of your life" and "a new way to express who you are" (Siegler, 2011) has been retired, along with many other tools designed to make information easier to manage such as personal news feeds and friend lists (Perez, 2018), a reminder that Facebook exists primarily to serve its advertisers, rather than the general public, as per the often-repeated saying "if you're not paying for it, you are the product". The most promising area for the development of interfaces for managing digital lives is the emerging "personal data locker" space, explored more in 2.3.4 below, which offer the promise of "a place for personal data", as Jones imagined PIM should be (W. Jones, 2011a), though at time of writing these are still quite limited. As Abiteboul noted in 2015, "everyone should be able to manage their personal data with a personal information management system" (Abiteboul, André and Kaplan, 2015), but as of yet, in any meaningful or holistic way, they cannot, because no general-purpose personal information management system for modern day digital lives exists.

2.2.5 Research Gap: The Data Beyond The Individual

In this section, I have detailed the ways in which personal information management systems have developed, and shown that they have not kept pace with the ever-more-complex needs of the Information Age. Most PIM systems treat data as a static resource to be filed and accessed much like you would a file in a 1970s office. Most digital services operate in isolation from each other, without any shared representation or co-operative understanding of an individual's personal information. Where personal data access is provided, it is limited in usage to the delivery of the specific service on offer, it is treated as a property asset and the data is not *participatory*. As Katie Shilton writes, "Much of the social impact of participatory personal data will depend on how data are captured and organized; who has access; whether individuals consent and participate; and how (or whether) data are curated and preserved (Shilton, 2011)." We need "fundamental changes in the way we represent and manipulate data" (Karger and Jones, 2006); we need holistic representations of data that can be subjectively meaningful and which allow for the constant change and evolution of data over time.

Of particular importance is that we recognise that people exist in an interconnected world of relationships - with other individuals, and with organisations, and that the role of data within those relationships needs to be examined. When your data is held by others, managing personal information is not just a matter of arranging your own bookshelves, but rather a multi-party negotiation over representation, ownership, access and consent. Data is a shared resource with multiple users, and only a few researchers have begun to look at people's interactions with data in this context (for example, activity streams (Hart-Davidson, Zachry and Spinuzzi, 2012), social sensemaking (Puussaar, Clear and Wright, 2017), and decentralised file storage (Zichichi, Ferretti and D'Angelo, 2020)). There has been negligible research into the role data plays within human relationships.

This is the second research gap that my thesis aims to address - to look at personal data holistically in the context of your life. How does the holding of personal data by third parties affect people's ability to function in modern life? Do people have meaningful control over their personal data in this multi-party landscape? What practical problems do data-holding organisations current practices cause for people? What role should data take in our complex digital lives?

2.3 Practical Human-centred Design

2.3.1 Human Computer Interaction Foundations

Up until the 1980s, the only reasons to consider the relationship between a human and the computer they were using were ergonomics, comfort and efficiency. People were shielded from the complexities of the machines they were using–the machine did the work and the human was just the *operator*. In the 1990s, the "first wave" of what is now known as *Human-Computer Interaction* (HCI) recognised humans as *actors* operating in groups, who had tasks to perform either using or assisted by technology (Bannon, 1995). People were now *users* of technology. Design thinking shifted from machine-centric to *user-centric design* (UCD), motivated by the goal of helping the user to do their tasks better. In the personal computer revolution of the 1990s, people began to work in complex and varied multi-user situations, and observation and understanding of a user's working environment provided empathy that enabled better design. There was a recognition that people use computers differently in different contexts. In the 2000s, as smartphones, broadband and Web 2.0 brought computing into every aspect of our lives, HCI's third wave looked beyond the workplace to consider users as unique humans with emotions and culture; design became about *experiences* (Bødker, 2006) which could span work, mobile and home domains. Computers were no longer just for work. This created a "chaos of multiplicity for HCI in terms of use technologies, use situations, methods and concepts" (Bødker, 2015); designers would now need to "embrace people's whole lives" (Bødker, 2006). The blueprint for how this could be achieved was to be found in Mark Weiser's seminal 1991 Scientific American article "The Computer for the 21st century", in which he envisioned a world where data could be accessed across many different devices, such that interfaces and interactions could be designed around the user's data needs in specific contexts. He recognised the need to put humans, not machines, at the centre of data interaction, and that in order to achieve "calm computing", technology would need to "disappear into the background" of our lives (Weiser, 1991; Weiser and Brown, 1996).

2.3.2 Data Transcendence, Context and Human-Data Interaction

Weiser's vision was significant because it recognised the need for data to transcend the confines of a single machine; to satisfy human needs in different contexts, data needs to be *pervasive* (Saha and Mukherjee, 2003; Krishnan, 2010). From a technical perspective, Weiser's vision has largely been realised, with today's smartphones, tablets and digital whiteboards / smart TVs corresponding directly to his imagined "tabs", "pads" and "boards" respectively. *Ubiquitous computing* now allows environments, vehicles and wearable computing to collect data via sensors – the "Internet of Things" (IoT), which enables context-aware computing (Abowd *et al.*, 1999; Eliasson, Cerratto Pargman and Ramberg, 2009). But what of the interaction perspective? As an answer to this question, the concept of *Human-Data Interaction* (HDI) emerged. This sub-discipline of HCI outlines the vision that the human needs to have a direct, explicit relationship with their own data (Mortier *et al.*, 2013, 2014), and that personal data should be considered an entity in its own right; people do not just need to interact with systems, but with the data itself. This can be seen as an echo of previous calls throughout the decades for a new relationship with our stored knowledge (Bush, 1945; Lansdale, 1988; Rogers, 2006; Hendler and Berners-Lee, 2010; W. Jones, 2011a).

Mortier et. al. laid out three tenets of HDI: Individuals need to have *agency* over how their data is used within the system, the data needs to be *legible* (i.e. understandable) to us, and we need *negotiability* - the ability to flexibly adapt and make use of the data. HDI has remained a small but important research niche within HCI, and many researchers continue to explore this field today ('Human Data Interaction Project at the Data to AI Lab, MIT', 2015; 'HDI Network Plus, University of Glasgow', 2018; 'HDI Lab, Heerlen', 2020; BBC R&D, 2017), as does this thesis. In order to understand what HDI might mean in practice we can look to Gregory Abowd's 2012 paper which aims to update Weiser's vision. In it, Abowd emphasises the importance of programming for *environments*, building a complete experience for the individual that considers not just the 2D screen they are using, but the entire surroundings and context of their environment. He imagines a hybrid, conjoined experience between people, devices, sensors and the cloud where data storage and processing need not be constrained to the input and output devices we use (Abowd, 2012) and crucially, that the individual within this *"everyday computing"* experience is harnessing technology for their own ends, not just being aided to complete a predetermined task (Abowd and Mynatt, 2000) – in essence they are able to program their own environment.

2.3.3 Human-Centred Design: A Sociotechnical Challenge

Abowd's vision is a helpful reference point to remind us how far from true human-data interaction we are today. As described above, data is trapped, and very few computing interactions today are designed as a situated experience. Some TV streaming services show a good example of an interaction whose design has taken into account context: instead of typing in long email addresses and passwords, difficult on a TV remote, you can visit a short link from a smartphone or PC where you are already authenticated. But even though there are pockets of research around contextual experiences (for example the work around second screening (T. Jones, 2011; Zúñiga, Garcia-Perdomo and McGregor, 2015)), in general most design work today still focuses on a single interaction surface. In order for technology to disappear into the background so that we might live in a calm, engaged manner, as outlined by Weiser and expanded upon by Yvonne Rogers (Rogers, 2006), a more humane interface is needed (Raskin, 2000), one which designs for the whole person. Judging the success of a user interaction can no longer be done by assessing task-completion efficiency (Abowd and Mynatt, 2000) but should consider the holistic needs of the individual at that moment in time.

Yet in the 2010s, there was a growing recognition that the world had lurched severely away from such goals. The design of information-consumption interfaces was having a detrimental effect upon people, not just in terms of the psychological impacts of information overload as detailed above in section 2.2.4, but also in terms of the impact on users' attention. This would become known as "the attention economy" (Simon, 1971; Croll, 2009; Cogran and Kinsley, 2012; Brynjolfsson and Oh, 2012). Social media technologies like infinite scrolling and smartphone notifications had created "a culture of perpetual distraction" (Timely, 2020) which "hijacks people's minds" (Harris, 2016). As Zeynep Tufekci put it in her TED talk, "we are creating a dystopia just to make people click on ads" (Tufekci, 2017). In 2013, Tristan Harris released a presentation calling on the tech industry to respect users' attention and minimize distraction (Harris, 2013a), which lead to the creation of the Center for Humane Technology (Harris, 2013b), a central group in this new movement to design for positive human values and to practice *value-sensitive design* (Friedman and Hendry, 2019). This focus beyond just supporting data interaction to understanding and enhancing the individual's lived experience can be seen as a central guiding tenet of *Human-centred design*.

We can see from the above that the design of human-centred personal data interaction is not purely a matter of designing better user interfaces, nor even of designing for the user's physical environment, but in fact a design challenge that exists at the sociotechnical (Bunge, 1999; Murton, 2011) level – it must take into account the social relationships of the individual (as detailed in 2.2.6) as well as the power imbalance that exists between data holders and data subjects (as detailed in 2.1.2). Andy Crabtree recognised the sociotechnical nature of the HDI challenge in his 2016 paper with Mortier on 'The Shifting Locus of Agency and Control' and highlighted particular aspects of this multi-party challenge around personal data, specifically being able to ensure the *privacy* of your data as well as the *accountability* data subjects require over data-processing algorithms and data-handling organisations (Crabtree and Mortier, 2016). These goals are now actively pursued through research into *privacy by design* (Cavoukian, 2010) and *Critical Algorithm Studies* (Gillespie and Seaver, 2016) respectively. In his subsequent work with Peter Tolmie, Crabtree focused on the particular HDI challenges around data-sharing, which must also be designed for (echoing Lindley's work on file biographies mentioned earlier) (Crabtree and Tolmie, 2018). These areas of pursuing a human-centric agenda within a sociotechnical context continue to be areas of active research today, as seen in projects such as Nesta's DECODE (Symons *et al.*, 2017), which focuses on individual empowerment, and UKRI's not-equal.tech (Crivellaro *et al.*, 2019), which focuses on *data justice* (Taylor, 2017).

2.3.4 The Emergent Human-Centred Personal Data Ecosystem

During the 2010s, while many were focused on the utility of PIM systems (as described in 2.2.2 above, and hereafter referred to as *'traditional PIM'*), some researchers, thought leaders and strategists were developing ideas that can be seen as the first socio-technical designs for personal data interaction. One of the earliest was Doc Searls, who launched a project called ProjectVRM with colleagues at Harvard University around 2008. He envisioned a model he called *Vendor Relationship Management (VRM)* which can be seen as the inverse of Customer Relationship Management (CRM) where organisations use data to profile and learn more about their customers and get their attention (Searls, 2008). In essence, the vision (expanded in his 2012 book (Searls, 2012)) was to combat the attention economy by turning the world of commerce inside-out; individuals would publish tightly controlled personal data about themselves and their needs, and retailers could respond to these individuals with product offers, from which (s)he would then select.

Drawing together the work of Searls and various other innovators in the VRM space, David Siegel catalogued an emerging vision of a personal data interface that could realise VRM thinking, which he called a *Personal Data Locker* (Siegel, 2009, 2010). (Equivalent terms *Personal Data Store*, *Personal Data Vault (PDV)* and *Personal Data Services* are also common). He described *'Pull-centric Computing'*, where information is 'pulled' at your request rather than being pushed upon you. The WEF's Rethinking Personal Data project (mentioned earlier) describes the potential for individuals to have their own *personal data ecosystem (PDE)* of *"commercial entities, acting as trusted intermediaries, exchanging assets on behalf of the individual, following a clear set of principles and legally binding contracts"*. They describe the PDV as being the technical means to place the individual *at the centre of that ecosystem*: the PDV provider would be *"an intermediary collecting user data and giving third parties access to this data in line with individual users' specifications"* (Hoffman, 2010). A 2010 report by nonprofit Mydex helps to contextualise the PDV, explaining that the PDV is a service to the individual that positions *"individuals as information managers"* at the *"epicenter of a new ecosystem of PIM services"* and that it will not just give access to data but *"transform relationships between individuals and organisations"* (Mydex CIC, 2010). In my view, this is what substantially differentiates the PDE from traditional PIM systems - it is a response to the sociotechnical need outlined in the previous section. A 2012 report from Ontario's Information Privacy Commissioner notes that the PDE collides with traditional concepts of ownership when it comes to data, that the PDE needs to *"provide a collection of tools and initiatives aimed at facilitating individual control over personal information"* wherever it is located; this is another way in which PIM within PDE can be differentiated from traditional PIM (Cavoukian, 2012).

It was against this landscape that *Personal Information Management Services* (PIMS[1]) became a business area in its own right, the basis for a *personal data economy*. PIMS is attempting to create a market for *"tools that help individuals gather, manage and use personal information to make better decisions and manage their lives better"*, with a potential market value (in the UK) of £16.5 billion, more than the automotive and pharmacetical industries (Ctrl-Shift, 2014). In 2016, a global network and non-profit initiative called MyData was founded, bringing together researchers, companies and public sector agencies in the PDE space, in pursuit of a *"fair, sustainable and prosperous digital society, where the sharing of personal data is based on trust, and relationships between individuals and organisations are balanced"* (MyData.org, 2018). An important aspect of MyData is its aim to combine companies' needs for data with individuals' digital human rights. Through analysis of principles of PIMS, VRM and other related spaces ('MyData Comparison of Principles document', 2017), the MyData declaration was produced, outlining a detailed vision for the PDE space to *"empower individuals with their personal data, thus helping them and their communities develop knowledge, make informed decisions, and interact more consciously and efficiently with each other as well as with organisations."* (MyData, 2017) MyData now has over 700 parties involved worldwide and provides a focal point to the PDE community.

The MyData declaration identifies data controllers' transparency with data and data-handling practices as an essential means for individuals to gain agency and accountability, and puts forward the idea that the individual should be the point of integration of their own personal data ecosystem; in other words, *"everything goes through me"*; this is the embodiment of the human-centric ideal of individual empowerment but will also be a good way for data controllers to ensure awareness, accuracy and consent. They also introduce the idea of a *personal data operator* (also known as a *data trust*) which is a key part of the personal data ecosystem - a trusted third party which stores or transfers data on behalf of the data subject, but does not use it themselves. An example operator is digi.me, which has developed a PDV with a *'private sharing'* model that allows users to allow subsets of their data to be used by external organisations or apps with strictly controlled parameters (Firth, 2019). The MyData/PDE space is very active currently, with many emerging businesses and startups having appeared in the last two to three years. Citizen.me ('Our Values', no date) is another company with a similar positioning. Other operators such as UBDI ('Whose data is it anyway?', 2019) and datacy ('About Us', no date) are positioned under a different business model which aims to help individuals take control of their personal data for profit. Open Humans has a PDV optimised to allow people to share their data for the benefit of research (Price Ball, no date). Ethi is a PDV platform focused on providing individuals with deep insights from their data, and tools to more easily delete personal data from data-holding organisations (Jelly, 2021). Many of these operators are building for a human-centric world that does not yet fully exist, and Geneva-based firm Hestia.AI / HestiaLabs also operates in this space, with a different focus: helping people to understand the existing data ecosystems that make use of their data (Dehaye, 2018), to build towards MyData ideals from the status quo.

2.3.5 Research Gap: Defining the Research Agenda for Achieving Human-Centricity in Practice

In this section, I have shown how the emergent human-centric personal data ecosystem has developed from its roots in HCI, ubicomp and HDI. The call for designs and sociotechnical systems that empower individuals with their personal data arise from the power imbalance (Hoffman, 2014a) that has emerged as a result of the datafication of modern life. In the third wave of HCI (Bødker, 2015), user interface design's main consideration was "what does *the user* want to do?". Over the last decade, catalysed by the shift by the explosion of Internet culture and the shift from self-install software products to massive-scale cloud-based Internet services, there has been a gradual but perceptible shift away from the tenet that the user's needs should come first: the designs of commercial and civic web applications now more reflect the question (considered from the provider's perspective) "What do *we* want the user to do?". Users (people) and their individual needs have been left behind. The MyData community have clearly outlined the goals to address this problem, but much of the focus at present is on technology questions of how to build better PDVs and better PIM interfaces, or on indentifying an effective business model that will facilitate the transition to a PDE, which is a necessary but distracting question. My research is situated at the bleeding edge of this emerging human-centric personal data ecosystem and being non-commercial, is able to take a more purist human-centric stance. After uncovering the human experience of personal data (as detailed in 2.1.5) and the lived experience of personal data usage within people's wider digital life and relationships (2.2.5), I will seek to address a third research gap - to understand the technical, legal, policy, economic and social realities of the PDE landscape itself, sufficient to inform the design of PDE processes and systems. Thinking of the barriers cascade in the SI space (Li, Dey and Forlizzi, 2010), what barriers exist that inhibit the building or adoption of PDE human-centric technologies? What opportunities might make it easier to overcome these barriers and to catalyse progress toward the human-centric agenda as envisioned in the MyData declaration? What are the key challenges faced when we attempt to build human-centric technologies in today's world? By applying learnings about human experiences and attitudes to the data-centric world to the practice of PDE design & development, can we more clearly map the road ahead and define a research agenda for the next step of tackling the PDE challenge?

By adopting both a participatory design and technical strategist's standpoint throughout this thesis, building on the theoretical foundations of effective data access, information management and human-centric data interaction, I aim to progress PDE / MyData thinking, using methods detailed in the next chapter, in pursuit of my primary research question, which is:

"What role should people's data play in their lives, what capabilities do they need, and how could these ideals be achieved?"

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.

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.

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

'About Us' (no date). datacy. Available at: https://datacy.com/personal-about (Accessed: 31 March 2021).

Abowd, G. D. (2012) 'What next, ubicomp?: celebrating an intellectual disappearing act', in *Proceedings of the 2012 acm conference on ubiquitous computing*. New York, New York, USA: ACM Press, pp. 31–40. doi: http://dx.doi.org/10.1145/2370216.2370222.

Abowd, G. D. *et al.* (1999) 'Towards a better understanding of context and context-awareness', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 304–307. doi: 10.1007/3-540-48157-5_29.

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.

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

'AllofMe Company Profile' (2007). Available at: https://www.crunchbase.com/organization/allofme (Accessed: 23 March 2021).

'AllofMe.com Teaser Clip' (2008). YouTube. Available at: https://www.youtube.com/watch?v=JWyqt4WL6xE (Accessed: 21 March 2021).

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/.

Apple (2009) 'iPhone 3G Commercial: "There's an app for that"'. YouTube. Available at: https://www.youtube.com/watch?v=mFlITzqRBWY.

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.

Bannon, L. J. (1995) 'From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in System Design', *Readings in Human–Computer Interaction*, pp. 205–214. doi: 10.1016/b978-0-08-051574-8.50024-8.

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.

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

BBC R&D (2017) 'Human Data Interaction - BBC R&D'. Available at: https://www.bbc.co.uk/rd/projects/human-data-interaction.

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.

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.

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.

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.

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.

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.

Bowyer, A. (2018) 'Free Data Interfaces: Taking Human- Data Interaction to the Next Level', *CHI Workshops 2018*. Available at: https://eprints.ncl.ac.uk/273825.

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

Brynjolfsson, E. and Oh, J. H. (2012) 'The attention economy: Measuring the value of free digital services on the internet', *International Conference on Information Systems, ICIS 2012*, 4, pp. 3243–3261.

Bunge, M. (1999) *Social Science Under Debate: A Philosophical Perspective*. University of Toronto Press. Available at: https://books.google.co.uk/books?id=-MLjZzJLbpkC.

Burkeman, O. (2011) 'SXSW 2011: The internet is over'. Available at: https://www.theguardian.com/technology/2011/mar/15/sxsw-2011-internet-online (Accessed: 23 March 2021).

Bush, V. (1945) 'As we may think', *The Atlantic Monthly*, 3(2), pp. 35–46. doi: 10.1145/227181.227186.

Bødker, S. (2006) 'When second wave HCI meets third wave challenges', *ACM International Conference Proceeding Series*, 189(October), pp. 1–8. doi: 10.1145/1182475.1182476.

Bødker, S. (2015) 'Third-wave HCI, 10 years later—participation and sharing', *Interactions*, 22(5), pp. 24–31. doi: 10.1145/2804405.

Carter, J. (2015) 'Who are the digital disruptors redefining entire industries?' Available at: https://www.techradar.com/uk/news/world-of-tech/who-are-the-digital-disruptors-redefining-entire-industries-1298171 (Accessed: 23 March 2021).

Caruthers, M. (2018) 'World Password Day: How to Improve Your Passwords'. Available at: https://blog.dashlane.com/world-password-day/ (Accessed: 5 May 2021).

Cavoukian, A. (2010) 'Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D', *Identity in the Information Society*, 3(2), pp. 247–251. doi: 10.1007/s12394-010-0062-y.

Cavoukian, A. (2012) 'Privacy by Design and the Emerging Personal Data Ecosystem', (October), pp. 1–39.

Chang, A. (2018) 'The Facebook and Cambridge Analytica scandal, explained with a simple diagram - Vox'. Available at: https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram.

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.

Cogran, P. and Kinsley, S. (2012) 'Paying Attention: towards a critique of the attention economy', *Cultural Machine*, 13.

Cormack, A. (2016) 'Is the Subject Access Right Now Too Great a Threat to Privacy?', *European Data Protection Law Review*, 2(1), pp. 15–27. doi: 10.21552/edpl/2016/1/5.

Crabtree, A. and Mortier, R. (2016) 'Personal Data, Privacy and the Internet of Things: The Shifting Locus of Agency and Control', *SSRN Electronic Journal*, pp. 1–20. doi: 10.2139/ssrn.2874312.

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.

Crivellaro, C. *et al.* (2019) 'Not-equal: Democratizing research in digital innovation for social justice', *Interactions*, 26(2), pp. 70–73. doi: 10.1145/3301655.

Croll, A. (2009) 'The Three Economies of Online Currency'. Available at: https://solveforinteresting.com/the-three-currencies-of-the-online-economy/.

Ctrl-Shift (2014) 'Personal Information Management Services: An analysis of an emerging market'. Nesta, p. 38. Available at: https://www.nesta.org.uk/report/personal-information-management-services-an-analysis-of-an-emerging-market/.

'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.

Dehaye, P.-O. (2018) 'Post-hearing questions by Senator Blumenthal to Mark Zuckerberg'. Available at: https://wiki.personaldata.io/wiki/Item:Q1800.

'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: http://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.

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.

Eliasson, J., Cerratto Pargman, T. and Ramberg, R. (2009) 'Embodied interaction or context-aware computing? An integrated approach to design', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Springer, Berlin, Heidelberg (PART 1), pp. 606–615. doi: 10.1007/978-3-642-02574-7_68.

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.

European Union Agency for Fundamental Rights (2020) 'Your Rights Matter: Data Protection and Privacy 2020', p. 20. doi: 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.

'Finland: Broadband Access Made Legal Right In Landmark Law' (2010). Available at: https://www.huffpost.com/entry/finland-broadband-access_n_320481 (Accessed: 23 March 2021).

Firth, E. (2019) 'Personal data has value in so many different ways'. digi.me. Available at: https://blog.digi.me/2019/09/04/personal-data-has-so-much-more-value-than-pure-cash/.

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.

Friedman, B. and Hendry, D. G. (2019) *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press (The mit press). Available at: https://books.google.co.uk/books?id=8ZiWDwAAQBAJ.

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/.

Fu, S. *et al.* (2020) ‘Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload’, *Information Processing and Management*, 57(6). doi: 10.1016/j.ipm.2020.102307.

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.

Gellman, B. (2013) ‘Edward Snowden, after months of NSA revelations, says his mission’s accomplished’, *The Washington Post*, 23. Available at: http://www.washingtonpost.com/world/national-security/edward-snowden-after-months-of-nsa-revelations-says-his-missions-accomplished/2013/12/23/49fc36de-6c1c-11e3-a523-fe73f0ff6b8d_story.html%5Cnhttp://www.washingtonpost.com/world/national-security/edward-.

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.

Gillespie, T. and Seaver, N. (2016) ‘Critical Algorithm Studies - A Reading List’. Available at: https://socialmediacollective.org/reading-lists/critical-algorithm-studies/.

‘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.

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.

Harris, T. (2013a) ‘A Call to Minimize Distraction Respect Users’ Attention’. Available at: http://www.minimizedistraction.com/.

Harris, T. (2013b) ‘Who We Are: Center for Humane Technology (CHT)’. Available at: https://www.humanetech.com/who-we-are.

Harris, T. (2016) ‘How Technology Hijacks People’s Minds — from a Magician and Google’s Design Ethicist’. Available at: https://www.tristanharris.com/2016/05/how-technology-hijacks-peoples-minds-from-a-magician-and-googles-design-ethicist/ (Accessed: 22 March 2019).

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.

‘HDI Lab, Heerlen’ (2020). Available at: https://hdilab.com/.

‘HDI Network Plus, University of Glasgow’ (2018). Available at: https://hdi-network.org/.

Hemp, P. (2009) ‘Death by Information Overload’. Available at: https://hbr.org/2009/09/death-by-information-overload (Accessed: 23 March 2021).

Henderson, I. and Group, B.-s. W. (2020) 'Customer — Supplier Engagement Framework Explained', pp. 1–7. Available at: https://me2ba.org/wp-content/uploads/2020/09/customer-supplier-engagement-framework-updated-9-28.pdf.

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.

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.

Hoffman, W. (2010) 'Rethinking Personal Data'. Available at: https://web.archive.org/web/20110220013300/http://www.weforum.org/issues/rethinking-personal-data.

Hoffman, W. (2011) *Personal data : The emergence of a new asset class*. World Economic Forum, pp. 1–40. Available at: http://www.weforum.org/reports/personal-data-emergence-new-asset-class.

Hoffman, W. (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.

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

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

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.

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

'Human Data Interaction Project at the Data to AI Lab, MIT' (2015). Available at: https://hdi-dai.lids.mit.edu/.

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/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.

Jelly, M. (2021) 'The Mission'. ethi.me. Available at: https://www.ethi.me/the-mission (Accessed: 31 March 2021).

Jilek, C. *et al.* (2018) 'Context spaces as the cornerstone of a near-transparent and self-reorganizing semantic desktop', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11155 LNCS, pp. 89–94. doi: 10.1007/978-3-319-98192-5_17.

Jones, T. (2011) 'Designing for second screens : The Autumnwatch Companion'. Available at: https://www.bbc.co.uk/blogs/researchanddevelopment/2011/04/the-autumnwatch-companion---de.shtml.

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.

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.

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. (2010) 'Pervasive Personal Information Spaces'. University of Waikato. Available at: https://researchcommons.waikato.ac.nz/handle/10289/4590.

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.

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.

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.

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.

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/.

Levine, R. (2011) 'How the internet has all but destroyed the market for films, music and newspapers'. Available at: https://www.theguardian.com/media/2011/aug/14/robert-levine-digital-free-ride (Accessed: 23 March 2021).

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.

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.

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.

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.

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

McCullagh, K. (2009) 'Protecting 'privacy' through control of 'personal' data processing: A flawed approach', *International Review of Law, Computers and Technology*, 23(1-2), pp. 13–24. doi: 10.1080/13600860902742562.

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

Mortier, R. *et al.* (2013) 'Challenges & opportunities in human-data interaction', *University of Cambridge, Computer Laboratory*. Citeseer. doi: 10.5210/fm.v17i5.4013.

Mortier, R. *et al.* (2014) 'Human-data interaction: The human face of the data-driven society', *Available at SSRN 2508051*. doi: 10.2139/ssrn.2508051.

Murton, D. (2011) 'A Brief History of the Evolution of Social Technology'. Available at: https://www.scottmonty.com/2011/04/brief-history-of-evolution-of-social.html.

MyData (2017) 'Declaration - MyData.org'. Available at: https://mydata.org/declaration/ (Accessed: 8 November 2019).

'MyData Comparison of Principles document' (2017). Available at: http://bit.ly/pd-principles.

MyData.org (2018) 'MyData - Who we are'. Available at: https://mydata.org/about/.

Mydex CIC (2010) 'The Case for Personal Information Empowerment : The rise of the personal data store', *World*, pp. 1–44.

'myTimeline' (2018). Available at: https://www.timelineinc.com/ (Accessed: 23 March 2021).

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

Neef, D. (2015) *Digital exhaust: what everyone should know about big data, digitization and digitally driven innovation*. Pearson Education.

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.

Nelson, T. H. (1965) 'Complex information processing', pp. 84–100. doi: 10.1145/800197.806036.

Odom, W. *et al.* (2018) 'Time, Temporality, and Slowness', pp. 383–386. doi: 10.1145/3197391.3197392.

O'Donnell, B. (2020) 'Zoom, the office and the future: What will work look like after coronavirus?' Available at: https://eu.usatoday.com/story/tech/columnist/2020/09/07/zoom-work-from-home-future-office-after-coronavirus/5680284002/.

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

'Our Values' (no date). Available at: https://www.citizenme.com/about/our-values (Accessed: 31 March 2021).

Perez, S. (2018) 'Facebook is shutting down Friend List Feeds'. Available at: https://techcrunch.com/2018/08/09/facebook-is-shutting-down-friend-list-feeds-today/.

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

Price Ball, M. (no date) 'About Us'. Available at: https://www.openhumans.org/about/ (Accessed: 31 March 2021).

Puussaar, A., Clear, A. K. and Wright, P. (2017) 'Enhancing Personal Informatics Through Social Sensemaking', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. Association for Computing Machinery, 2017-May, pp. 6936–6942. doi: 10.1145/3025453.3025804.

Raskin, J. (2000) *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.

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.

Roche, M. (2011) 'Full internet ban for sex offenders ruled unlawful'. Available at: https://ukhumanrightsblog.com/2011/08/12/full-internet-ban-for-sex-offenders-ruled-unlawful/ (Accessed: 23 March 2021).

Rogers, Y. (2006) 'Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences', *LNCS*, 4206, pp. 404–421. Available at: http://www.inf.ufg.br/$\sim$vagner/courses/mobilecomputing/docs/papers/03-Rogers_Ubicomp06.pdf.

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/.

Saha, D. and Mukherjee, A. (2003) 'Pervasive computing: A paradigm for the 21st century'. IEEE. doi: 10.1109/MC.2003.1185214.

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.

Searls, D. (2008) 'The Intention Economy: What Happens When Customers Get Real Power'. Available at: https://web.archive.org/web/20101226073246/http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2009_

Searls, D. (2012) *The intention economy: when customers take charge*. Harvard Business Press (04), pp. 50–2168–50–2168. doi: 10.5860/choice.50-2168.

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.

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

Shilton, K. (2011) ‘Participatory Personal Data: An Emerging Research Challenge for the Information Sciences’, *Advances in Information Science*.

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*.

Siegel, D. (2009) *Pull: The power of the semantic web to transform your business*. Penguin.

Siegel, D. (2010) ‘Personal Data Locker Vision Video’. Available at: https://vimeo.com/14061238.

Siegler, M. G. (2011) ‘Facebook Unveils Timeline: The Story Of Your Life On A Single Page’. Available at: https://techcrunch.com/2011/09/22/facebook-timeline/ (Accessed: 21 March 2021).

Simon, H. A. (1971) ‘Designing Organizations for an Information-Rich World’, *Computers, Communication, and the Public Interest.*, pp. 37–72.

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.

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.

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

Symons, T. *et al.* (2017) ‘Me, my data and I: The future of the personal data economy’, *DECODE (DEecentralised Citizen Owned Data Ecosystems) Report*, (732546), p. 88. Available at: https://media.nesta.org.uk/documents/decode-02.pdf.

Taylor, L. (2017) ‘What is data justice? The case for connecting digital rights and freedoms globally’, *Big Data and Society*, 4(2). doi: 10.1177/2053951717736335.

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.

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.

‘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.

Timely (2020) ‘The attention economy: what it is, what it’s doing to you’. Available at:

https://memory.ai/timely-blog/the-attention-economy.

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/.

Tufekci, Z. (2017) ‘We’re building a dystopia just to make people click on ads’. TED. Available at: https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads.

Tunikova, O. (2018) ‘Are We Consuming Too Much Information?’ Available at: https://medium.com/@tunikova_k/are-we-consuming-too-much-information-b68f62500089 (Accessed: 23 March 2021).

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

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

Wagner, A. (2012) ‘Is internet access a human right?’ Available at: https://www.theguardian.com/law/2012/jan/11/is-internet-access-a-human-right (Accessed: 23 March 2021).

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

Weiser, M. (1991) ‘The computer for the 21st century’, *Scientific American*, 265(3), pp. 94–105. doi: 10.1145/329124.329126.

Weiser, M. and Brown, J. S. (1996) ‘The coming age of calm technology’, *Beyond Calculation: The Next Fifty Years of Computing*. Available at: http://www.teco.edu/lehre/ubiq/ubiq2000-1/calmtechnology.htm http://link.springer.com/content/pdf/10.1007/978-1-4612-0685-9_6.pdf%5Cnpapers2://publication/uuid/F86D6ECE-A71E-4D20-A47B-9AF86A84923D.

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.

‘Whose data is it anyway?’ (2019). 04: UBDI. Available at: https://www.ubdi.com/blog/whose-data-is-it-anyway (Accessed: 31 March 2021).

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

Zichichi, M., Ferretti, S. and D’Angelo, G. (2020) ‘On the Efficiency of Decentralized File Storage for Personal Information Management Systems’. Available at: http://arxiv.org/abs/2007.03505.

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.

Zúñiga, H. de, Garcia-Perdomo, V. and McGregor, S. C. (2015) ‘What is second screening? Exploring motivations of second screen use and its effect on online political participation’, *Journal of communication*. Oxford University Press, 65(5), pp. 793–815.

1. The usage of the abbreviation PIMS here is not to be confused with its earlier use to refer to *'Personal Information Management Systems'* in traditional PIM terminology.↩