

Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Discussion II: Designing and Pursuing Better Human Data Relations	2
1.1 Introduction & Background	2
1.1.1 Peripheral Research & Design Settings	3
1.1.2 Attribution of Insights	4
1.2 Defining and Refining ‘Human Data Relations’ (HDR)	5
1.2.1 The Role of Personal Data	6
1.2.2 Human Data Interaction or Human Information Interaction?	7
1.2.3 The Two Distinct Motivations for Human Data Relations	8
1.2.4 Four Objectives for Human Data Relations	10
1.2.5 Better Human Data Relations as a Recursive Public	11
1.3 The landscape of opportunity: Obstacles to better Human Data Relations, and how we might overcome them	12
1.3.1 Obstacles to the HDR Objective of Data Awareness & Understanding	14
1.3.2 Obstacles to the HDR Objective of Data Usability	16
1.3.3 Obstacles to the HDR Objective of Ecosystem Awareness & Understanding	20
1.3.4 Obstacles to the HDR Objective of Ecosystem Negotiability	26
1.3.5 Obstacles to the HDR Objective of Effective, Commercially Viable and Desirable Systems	32
1.4 The landscape of opportunity: Four approaches to improving Human Data Relations	40
1.4.1 A Primer on Theories of Change (ToC)	40
1.4.2 Approach 1 to Improving HDR: Discovery-Driven Activism	42
1.4.3 Approach 2 to Improving HDR: Building the Human-centric Future	48
1.4.4 Approach 3 to Improving HDR: Defending Autonomy and Nurturing the Information Landscape	64
1.4.5 Approach 4 to Improving HDR: Winning Hearts and Minds: Teaching, Championing and Selling the Vision	69

1.5 Thesis Conclusion	71
Bibliography	71

1 Discussion II: Designing and Pursuing Better Human Data Relations

“Civilizations advance not by the technology they know about, but by the technology they don’t have to know about.” – Anonymous proverb

1.1 Introduction & Background

Through the Case Studies (Chapter 4 & 5) and the discussion in Chapter 6, a clear understanding of what people want from direct and indirect data relations (RQ1 & RQ2) has been established. In this chapter, we turn our attention from theory to practice, from what is needed to *what is possible*. Specifically, this chapter will return to the overall research question and consider “*How might [better Human Data Relations] be achieved?*”, and answer this question by describing practical approaches for future research, innovation and policy that are either novel or already emergent.

This chapter is deliberately broad and open ended. It does not pretend to be complete or definitive in its interpretation of the outlook for HDR. It is not a roadmap, but rather a snapshot of ongoing work, identified challenges and known opportunities, forming an **anthology of reference material**, based on my research and design experience from my six years working to understand and advance HDR.

The rationale here is that it will be valuable for anyone working in the HDR space to have a good high-level understanding of the landscape as well as specific ideas to work with; the goal is to boost and strengthen any such activities so that they might benefit from the insights gained.

The shape this chapter takes is to consider the the six HDR wants that this thesis has uncovered [Chapter 6] as a basis for defining **objectives** for the HDR landscape, then to to illustrate what specific **obstacles** and **opportunities** are relevant when attempting to pursue those objectives, as well as to highlight specific **designerly insights** that are relevant.

There are many aspects to the wide-reaching objective of better HDR in practice: technical, design, commercial, legal, moral, social and political and this chapter does not cover them all, nor is it formal empirical research. Instead, detail is provided in the form of **real world practical designs and insights** from four industrial and academic research projects I was part of during the same timeframe as the empirical research, as well as from the work of other innovators and activists. This detail is contextualised relative to existing literature and the thesis’ earlier contributions.

In section 7.1.1 the peripheral R & D activities I undertook are described; forming the primary point of reference for this chapter, as this peripheral work has informed and allowed me to build upon the core HDR understanding from the empirical research, and much of the work has often aligned well to the six data wants [Chapter 6]. Often the work has exposed evolving areas where different actors are trying to bring about better HDR.

In section 7.1.2, I explain some important context about the nature of the ideas presented in this chapter and how to attribute them fairly.

In section 7.2, I formalise and expand the Human Data Relations concept. Additional insights into how people relate to data are identified, as well an important dichotomy of two distinct drivers that motivate people's needs for better relations with their data. I also conceptualise those who would pursue better HDR as *HDR reformers* and reflect on the researcher-turned-activist stance that drives this chapter, to conceptualise this community as a *recursive public*.

Section 7.3 and 7.4 form the main body of this chapter, beginning with a focus on the main obstacles that one must face in pursuit of the HDR objectives identified. Specific insights that can inform research, design and activist thinking are shared in inset boxes throughout 7.3 and 7.4.

7.4 is solution-focused, considering the nuts-and-bolts of *how* we might, using these insights begin to tackle the obstacles of 7.3 in pursuit of the HDR objectives. In 7.4.1 I introduce the concept of *Theories of Change (ToC)*, which are used as a framing device for structuring the four approaches to pursuing change that follow in 7.4.2 to 7.4.5. Each for these four approach sections is modelled as a different possible trajectories for change. Within each of these four trajectories, specific named opportunities are described or referenced in varying detail¹.

Section 7.5 concludes the thesis, summarising the insights presented in 7.3 and the change trajectories presented in 7.4, reflecting on my journey as a researcher and summarising the thesis' contributions as a whole.

1.1.1 Peripheral Research & Design Settings

[TODO Move 3.4.3 etc. to here and remove all refs to 3.4.3]

The majority of examples and learnings shared in this chapter come from my participation as an expert researcher and designer in two industrial research projects:

1. **BBC R&D's Cornmarket Project**, which explored through user experience design, technical prototyping and participatory research, how

¹Some of the challenges and opportunities herein are described in greater detail than others, corresponding only to my proximity and depth of engagement with those ideas rather than their relative merit, complexity or impact potential. Given the broad aim of mapping out a new field, I consider that it is more useful to introduce a range of applicable ideas even if some are only lightly detailed, rather than to detail just a few.

individuals might interact with data through a Personal Data Store interface (see 3.4.3.3)

2. **Sitra/Hestia.ai's #digipower Project**, a successor to Case Study Two, in which European politicians examined companies' data practices through exercising data rights and conducting technical audits (see 3.4.3.4) [TODO add links to both the page, and the sitra report, and (Bowyer, Pidoux, *et al.*, 2022; Pidoux *et al.*, 2022)]

In addition, my participation as an interface designer and front-end software developer in the following two academic research projects contributes secondarily to this chapter:

3. **Connected Health Cities (CHC)'s SILVER Project**, where I, along with a backend developer and a team of researchers, developed a prototype health data viewing interface for Early Help support workers (see 3.4.3.1).
4. **Digital Economy Research Centre (DERC)'s Healthy Eating Web Augmentation Project**, which explored the use of web augmentation techniques to modify the user interface of takeaway service Just Eat to insert health information, in support of healthy eating (see 3.4.3.2).

1.1.2 Attribution of Insights

While this thesis is my own original work, and many ideas presented in this chapter are fully original, some of the specific details, theories and ideas presented in this chapter arose or were developed or augmented through my close collaboration, discussion and ideation with other researchers, including:

- Jasmine Cox, Suzanne Clarke, Tim Broom, Rhianne Jones, Alex Ballantyne and others at BBC R&D;
- Paul-Olivier Dehaye, Jessica Pidoux, Francois Quellec at Hestia.ai;
- Stuart Wheater of Arjuna Technologies and Kyle Montague of Open Lab during the SILVER project; and
- Louis Goffe of Open Lab on the DERC Healthy Eating project
- earlier innovation work with Alistair Croll at Rednod, Montréal, Canada (circa 2011) and with Megan Beynon at IBM Hursley, UK (circa 2006).

Due to these collaborations and the ongoing and parallel nature of many of these projects to my PhD research, it is impossible to precisely delineate the origin of each idea or insight. In practice, ideas from my developing thesis and own thinking informed the projects' trajectories and thinking, and vice-versa. These ideas would not have emerged in this form without my participation, so they are not the sole intellectual property of others, but equally I would not have reached the same conclusions alone, so the ideas are not solely my own either. All diagrams and illustrations were produced by me, except where specified, and the overall synthesis and framing presented in this chapter is my own original work. Where this chapter includes material from the four projects, that material is either already public, or permission has been obtained from the corresponding project teams.

1.2 Defining and Refining ‘Human Data Relations’ (HDR)

Chapter 6 established six ‘wants’ that people have in their relationships with data: **visible, understandable and usable data; process transparency, individual oversight and decision-making involvement.**

The major contribution of this thesis, beyond the detail and evidence for these wants conveyed in chapters 4 to 6, is to synthesise these findings and conceptualise what people want from data holders into a clearly defined field for future research and innovation. Repurposing the concepts of ‘human-technology relations’ and later ‘human-data relations’ which have been the subject of some study in the contexts of philosophy, embodied interaction and the performing arts (Ihde, 1990; Hogan, 2012; Windeyer, 2021), I have chosen to name this field “**Human Data Relations**”, or HDR for short. I propose this field as a successor to Mortier *et al.*’s Human Data Interaction (HDI) (Mortier *et al.*, 2014). HDR builds upon HDI but is wider, broader and more sociotechnical; HDR encompasses all aspects of the ways in which people and organisations can and should relate to data, not just interaction with data itself. Through a greater focus on relationships and ecosystems and approaches that target today’s practical data-centric power-imbalanced reality, it can provide a more effective research agenda for the world of the 2020s. The field’s definition draws upon three distinct connotations or readings of its name:

Human Data Relations - A Definition

The field of human data relations encompasses all the ways in which humans and human organisations relate to, and with, data, specifically:

1. **Human-Data Relations:** the direct interaction of users with data to understand and use it, similar to HDI, and in service of the **direct data wants** [6.1] of visible, understandable and useable data.
 2. **Human “Data Relations”:** the relationships that humans have with organisations that hold data about them, in service of the **indirect data wants** [6.2] of transparency, individual oversight and involvement.
 3. **Human/Data Relations:** the ways that organisations **manage their customers with respect to personal data**. Similar to ‘*public relations*’ or ‘*customer relations*’, this concerns the ways that organisations present their data practices (so as to build trust), and the ways in which they could involve users with data and provide support to understand data to their users (in order to empower individuals and build more effective customer relationships) [4.4.1; 5.5.2; 6.1.2].
-

[TODO Format this as an inset box not a table]

Having defined the scope of HDR, we can say that ‘better’ HDR can be achieved by working to improve upon the identified six aspects of human data relations. However, as this section will explain, HDR is motivated in two distinct ways, to which those six wants apply differently. As background understanding for

this duality of motivation, it is first necessary to examine more closely what role data plays in people's lives.

1.2.1 The Role of Personal Data

In the modern world, where almost anything can be encoded as data, and given many previously analogue objects and activities now have digital equivalents, the concept of data has become broad and hard to pin down. Underlying Human Data Relations is to explain what roles data can play in people's lives – what it is to people. Through the Case Studies, external work and my prior learning, I have so far identified 8 distinct lenses to consider how people might relate to it. These are modelled in Table 15.

Table 2: Table 15. **Eight lenses on data.**{#table-15}

Way of thinking about data	Explanation & Implications
Data as property	Data can be considered as a possession. This highlights issues of ownership, responsibility, liability and theft.
Data as a source of information about you	Knowing that data contains encoded assertions about you and can be used to derive further conjectures enables thinking about how it might be exploited by others, but also how you can explore and use it yourself for reflection, asking questions, self-improvement and planning. It invites consideration of the right to access, data protection, and issues around accuracy, fairness and misinterpretation / misuse.
Data as part of oneself	A photo or recording of you, or a typed note or search that popped into your head could be deeply personal. This lens on data highlights issues around emotional attachment/impact, privacy, and ethics.
Data as memory	Data can be considered as an augmentation to one's memory, a digital record of your life. This lens facilitates design thinking around search and recall, browsing, summarising, cognitive offloading, significance/relevance, and the personal value of data.
Data as creative work	Some of the data we produce (e.g. writing, videos, images) can be considered as an artistic creation. This lens enables thinking about attribution, derivation, copying, legacy and cultural value to others.
Data as new information about the world	Data created by others can inform us about previously unknown occurrences in our immediate digital life or the wider world. This lens is useful for thinking about discovery, recommendations, bias, censorship, filter bubbles, and who controls the information sources we use, as well as who will see and interpret data that we generate and what effects our data has on others.

Way of thinking about data	Explanation & Implications
Data as currency	Many data-centric services require data to be sacrificed in exchange for access to functionality, and some businesses now explicitly enable you to sell your own data. This lens highlights that data can be thought of as a tradable asset, and invites consideration of issues of data's worth, individual privacy, exploitation and loss of control.
Data as a medium for thinking, communicating and expression	Some people collect and organise data into curated collections, or use it to convey facts and ideas, to persuade or to evoke an emotional impact. This lens is useful to consider data uses such as lists, annotation, curation, editing, remixing, visualisation and producing different views of data for different audiences.

When considering HDR, it is important to recognise that people may think of their personal data through any or all of these '*lenses*' [Karger *et al.* (2005);2.2.2] at any given time, and any process or system design involving data interaction should take these into account.

Looking across this set of lenses, it is possible to identify four specific roles that data can serve:

1. Data has a role as an **artifact of value** to your life;
2. Data has a role in **informing** you about yourself, the world, and the prior or recent actions of others that may affect you;
3. Data has a role as a **usable material with which to effect change** in your life;
4. Data has a role as a **means to monitor changes** in data holders' behaviours, digital influences upon you or changes within your life.

1.2.2 Human Data Interaction or Human Information Interaction?

To unpack HDR further, it is important to highlight the difference between humans relating to data, and humans relating to information. Human Data Interaction (HDI) concerns the way people interact with data. Mortier *et al.* (Mortier *et al.*, 2013, 2014) defined the field of HDI without distinguishing data (the digital artifact stored on computer) from information (the facts or assertions that said data can provide when interpreted). This is an important distinction. The parallel field of Human Information Interaction (HII) originated in library sciences, and considers the way humans relate to information without regard to the technologies involved (Marchionini, 2008). William Jones *et al.* called for a new sub-field of HII in an HCI context², observing that it is important to include

²The group of HCI researchers involved in this panel were (with the exception of Raya Fidel) seemingly unaware of the existing HII field in library sciences as they positioned the

a focus on information interaction because HCI can “*unduly focus attention on the computer when, for most people, the computer is a means to an end – the effective use of information*” (Jones *et al.*, 2006). DIKW theory [2.1] highlights that **interpretation of data to obtain information** is a discrete activity. This was borne out in the findings of Case Study Two, where it became clear that participants have distinct needs from data, and from information (5.4.3.2). Access to data **and** information is critical to both understanding and usability, as detailed in section 6.1.2 and 6.1.3.

Drawing on this theory, we can see then that in considering Human Data Relations, there are in fact three distinct artifacts to consider:

1. ***data*** - the stored digital artifacts pertaining to users held by organisations for algorithmic processing and human reference, copies of which can be obtained using individual data rights.
2. ***information about individuals*** - the collection of facts and assertions about the individual and their life, which are obtained through human or algorithmic interpretation of stored data (or in some organisations' case, through analytical inference).
3. ***information about data*** (also categorised in Table 9 / 5.3.1 as ***metadata***)
- stored facts about the data, such as where it has been stored, who has accessed it, how it was collected, what it means, or when it has been shared externally.

1.2.3 The Two Distinct Motivations for Human Data Relations

By making this distinction between the two types of information which people might interact with, and considering the six wants in Chapter 6, it becomes clear that there are two very different reasons why people might want better HDR:

- (i) to acquire *information about one's data*, so that one might exert control over and make informed choices about where *the data* is held and how it is used, in order to be treated fairly and gain more control over the use of one's personal data. This is **Personal Data Ecosystem Control (PDEC)**.
- (ii) to acquire *information about oneself*, so that one might gain insights into one's own behaviour and gain personal benefits from those insights or them to make changes in one's life. This is **Life Information Utilisation (LIU)**.

The two distinct processes that individuals might go through in pursuit of these motives are exemplified in Figure 30. PDEC is a process of holding organisations to account over and managing *what happens to personal data*, often regardless of what it means, whereas LIU is more concerned with *what the data means* and its inherent value as encoded life information, regardless of where it is stored and

publication as a call for a ‘new field’.

how it is used³. This novel way of modelling the motivations for data interaction were first proposed in my 2021 workshop paper (Bowyer, 2021).

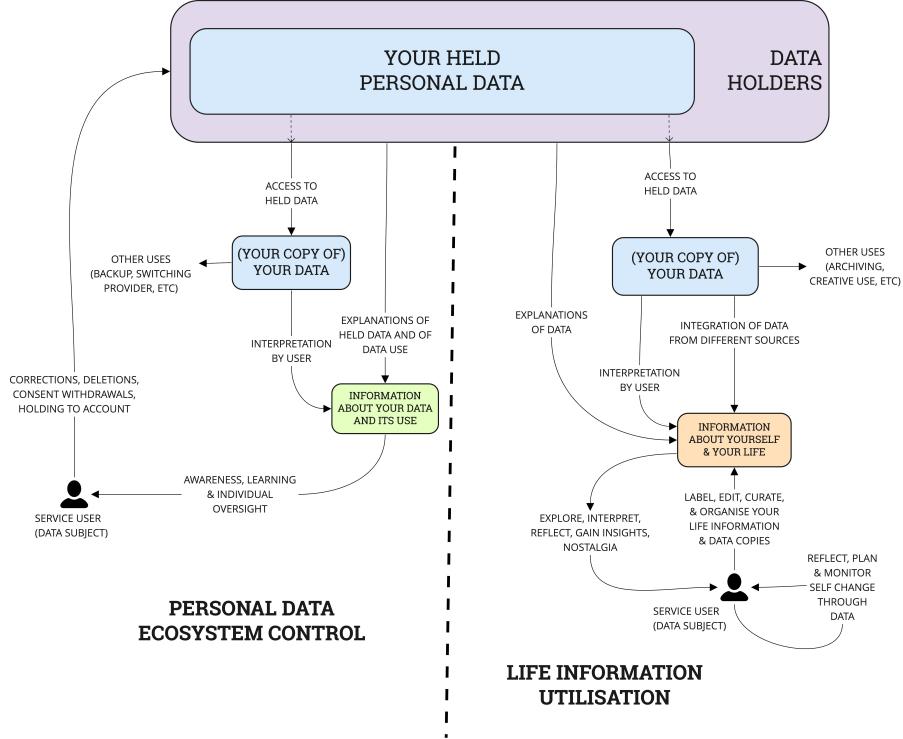


Figure 1: Figure 30: The Two Motivations for HDR: Controlling your personal data ecosystem and utilising your information about your life, with ‘idealised’⁵ processes illustrated

1.2.3.1 Life Information Utilisation *Life Information Utilisation*

is a superset of *Self Informatics (SI)* 2.2.3. It includes all purposes relating

³Of course, there is some overlap; the reason that organisations hold data is so that they can interpret it (usually algorithmically) to inform decision-making. In this way, organisations could be seen to be doing LIU of service users’ lives for their own benefit. From a human-centric perspective, this grey area is situated as part of PDEC, as from the individual perspective, how organisations understand you through information will inform decisions that affect your life. Thus, this can be considered part of the reason why one might want to exert control over use of your data, rather than being part of exploiting data to gain self-insights and personal benefits.

⁵The illustrated processes assume reliance on existing data access processes such as GDPR, where the only access is through provision of a copy of one’s data. This is in fact, not ideal, as it creates divergent versions and will quickly become out-of-sync, however for the sake of simplicity this inefficiency is ignored here. Improvements upon this approach are explored in [INSERT REF]

to self-monitoring and self-improvement through data, but also includes all other uses of personal data including creative expression, evidence gathering, nostalgia, keeping, and sharing. Many of these desires were expressed in Case Study Two (see Table 12 in 5.3.3), and also hinted at in the Early Help context [4.4.1]. While the existence of digitally-encoded information clearly unlocks new possibilities, LIU has existed in some form throughout human civilisation, as seen through analogue processes such as storytelling, journalling, scrapbooking, arts and crafts.

In the LIU context, the most important wants to focus on improving are *data understandability* (6.1.2) and *data useability*¹⁵ (6.1.3), which relate closely to the HDI concepts of *legibility* and *agency* respectively.

1.2.3.2 Personal Data Ecosystem Control Unlike LIU, **Personal Data Ecosystem Control** is an individual need that is new; arising as a result of the emergence of the data-centric world (2.1, 2.2.4). Only when organisations began to collect and store facts about people as a substitute for direct communication and involvement did it become necessary. The more data is collected about individuals, and the more parties collect and share that data, the greater the need for individuals to learn about that data so that they might influence its use (or risk their lives being affected in unexpected or potentially unfair ways). PDEC is a direct response to the power imbalance between data holders and individuals that the World Economic Forum described in 2014 [Hoffman (2014);2.1.2].

In the PDEC context, multiple data wants are important: visible data and transparent processes, as well as individual oversight and involvement. For simplicity, the former two wants can be referred to collectively as “*ecosystem transparency*”, and the latter two as “*ecosystem negotiability*” (drawing on the HDI concept of *negotiability*), and these terms will be used below.

1.2.4 Four Objectives for Human Data Relations

In order to provide value to future researchers, activists and innovators, this chapter contributes a map of the HDR opportunity landscape. This map is expressed in two parts across this section and 7.4. As a first step, we can take the “six wants from data relations” Chapter 6, and map reduce those to four simple ‘*landscape objectives*’ which shape our ultimate goals for effective HDR in this landscape of opportunity:

1. Data Awareness & Understanding;
2. Data Useability¹⁵;
3. Ecosystem Awareness & Understanding and
4. Ecosystem Negotiability.

As Figure X shows, the need for data to be understandable, visible and useable, applies to all types of data, whether that data is interpretable as **life information** (information within the data, that says something about the individual) or **ecosystem information** (information *about* the data, where it is held and how

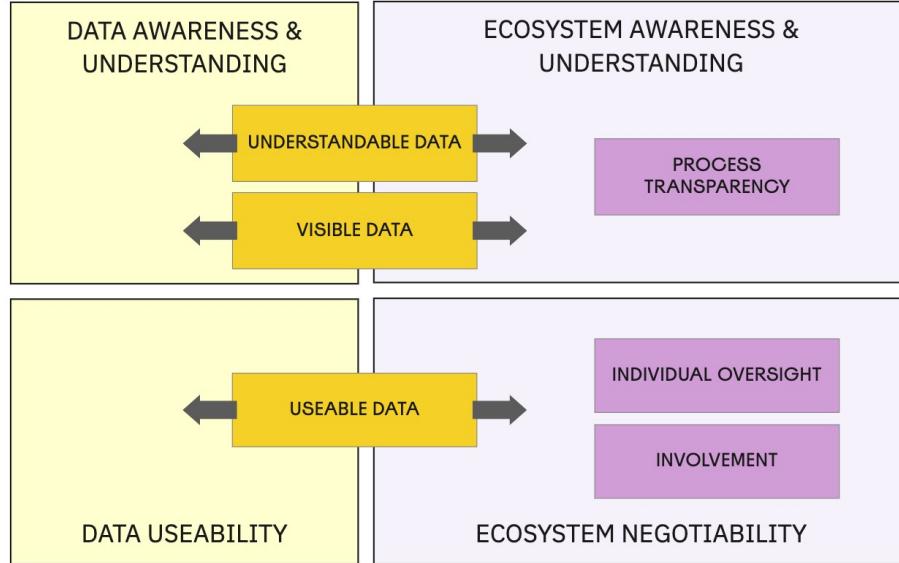


Figure 2: Figure X: Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

it is used). These two types of information will collectively be referred to as **human information**, and will be used in describing the HDR landscape in subsequent sections.

1.2.5 Better Human Data Relations as a Recursive Public

Before engaging with the practicalities of pursuing these HDR objectives, it is valuable to revisit the stance from which we approach this change. As outlined in 3.2, the research of this PhD has been grounded in participatory action research and experience-centred design; by using a Digital Civics (Vlachokyriakos *et al.*, 2016) frame to gain deep understanding of people’s needs and the ways those needs are not fully met, we can interpret and model how the world needs to change. Section 3.2 already outlined that we can consider such research as political, seeking to correct an imbalance in the world. In this chapter, we look beyond identifying what change is needed, and step into the role of activist, exploring how individuals and groups can actually change the world they inhabit.

In doing so, we can consider ourselves (those who pursue better Human Data Relations, or *HDR reformers* as a shorthand) as a *recursive public* (Kelty, 2008; ‘Recursive Public (Discussion Page)’, no date), albeit a nascent one. This is a term originating in the free software movement to describe a “*collective, independent of other forms of constituted power, capable of speaking to existing forms of power through the production of actually existing alternatives*”. This term captures the idea that through various means at our disposal: participatory research,

experience-centred design, engineering software prototypes, exertion of legal rights, and efforts to raise public awareness, we seek to modify the systems and practices we live within in pursuit of our goals. This idea of reconfiguring society in this way has been conceived as *civic hacking* (Crabtree, 2007; Levitas, 2013; Tauberer, 2014). The collective around better Human Data Relations does not yet exist as a named and identifiable *public* (Le Dantec, 2016) but its members congregate around emergent collectives in interconnected and overlapping spaces, most notably the MyData community (MyData, 2017) and its members, but also research and activism agendas including but not limited to: personal data lockers (CitizenMe, 2021; ‘Digi.me’, no date); digital rights ('Open rights group: Who we are', no date), gig economy worker rights (Kirven, 2018; ‘Worker info exchange’, 2022), privacy by design (Cavoukian, 2010), privacy activism (Davies, 1990; ‘Bits of freedom: Our focus’, 2000), data justice (Taylor, 2017; Crivellaro *et al.*, 2019), critical algorithm studies (Gillespie and Seaver, 2016), humane technology (Harris, 2013) and explainable AI ('Explainable AI: Making machines understandable for humans', no date).

Whether these disparate groups coalesce into a single identifiable public remains to be seen, and so too whether the term this thesis offers of *Human Data Relations* is sufficient to capture that public. At the least, it provides a descriptive umbrella term. Nonetheless, the breadth of research and innovation and activism happening in this space validates both the need and the desire for such a recursive public around better HDR to exist – in fact, it already does, whether we name it such or not. Therefore, this chapter takes an unashamedly critical view of the status quo, favouring disruptive societal changes that would further the objectives of better Human Data Relations and providing actionable approaches that will be of use to the members of this public. The chapter asks, “*How can we change the world into the one we want?*”

1.3 The landscape of opportunity: Obstacles to better Human Data Relations, and how we might overcome them

Using the four objectives established in 7.2.4 as our goals, and considering how they might be tackled, specific *obstacles* have been identified. These are analogous to Li’s ‘*barriers cascade*’ [2.2.3; Li, Forlizzi and Dey (2010)] and represent the obstacles that individuals or system designers must be empowered to overcome if the objectives are to be met. These obstacles are followed by useful *insights* I have identified that might help overcome those obstacles. This is summarised in Figure X, which shows an HDR-specific barriers cascade: a route of overcoming obstacles through which individuals might be empowered and by which organisations might become more HDR-friendly.

The obstacles and insights in the figure are explained in the following subsections. The last of these (corresponding to the ‘solution space’ box) covers some of the more pervasive obstacles that apply to all of the previous four HDR objectives.

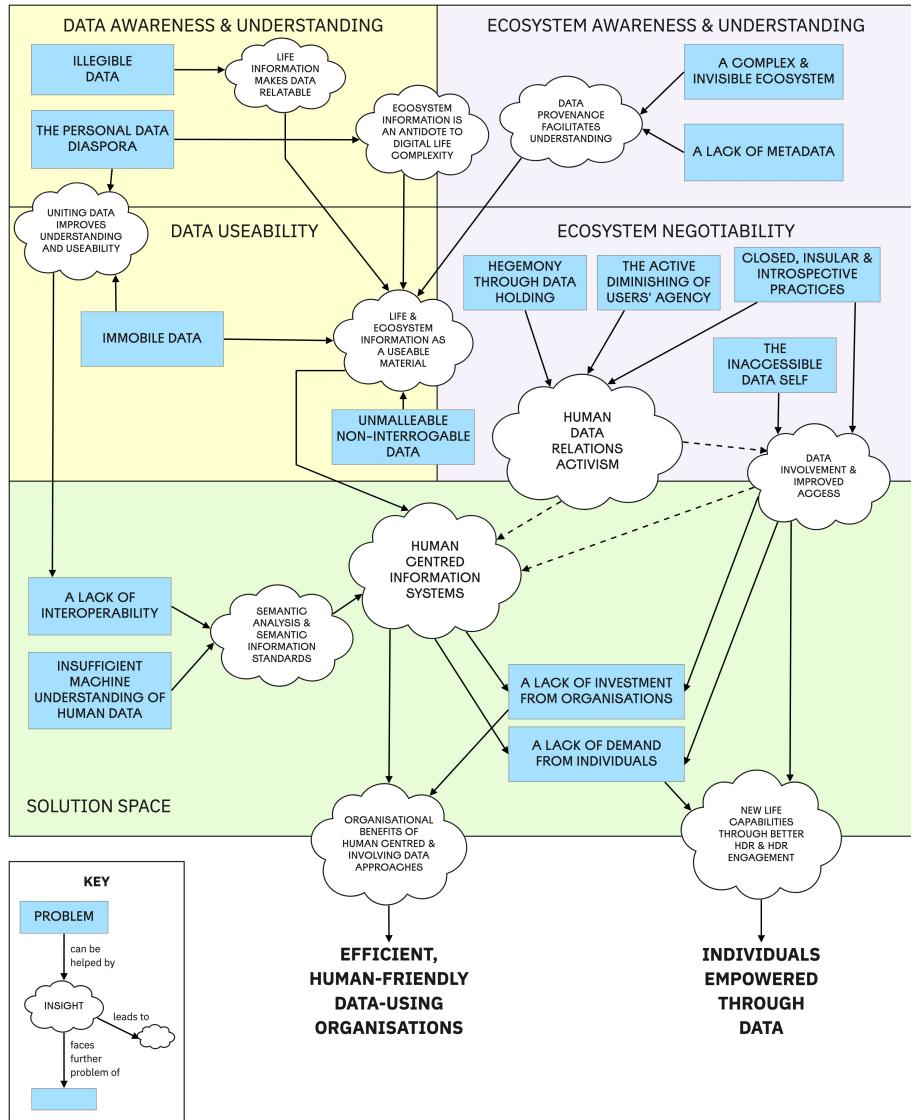


Figure 3: Figure X: Obstacles and Resulting Insights in the HDR Opportunity Landscape

1.3.1 Obstacles to the HDR Objective of Data Awareness & Understanding

In pursuit of visible, understandable data [6.1.1; 6.1.2], the first obstacle encountered is that in today's complex digital landscape, **most personal data is invisible, inaccessible or unrelated**. It is trapped in service providers' databases, or on different devices or hard drives, or inaccessible due to software limitations or proprietary file formats (Abiteboul, André and Kaplan, 2015; Bowyer, 2018), or in formats. Participants of both Case Studies talked of 'not knowing' what data exists and of being 'in the dark'. As Case Study Two showed, even where data is accessible, it is not relatable ('legible' [Mortier *et al.* (2014); 2.3.2]). Thus the objective here is to tackle this obstacle and ensure that people not only have awareness of their data, but can understand ('make sense' [Gurstein (2011); 2.1.4]) of what it means.

INSIGHT 1: Life Information makes Data Relatable

In the pilot study and Case Study One, 'data cards' were used to represent common types of civic data [Figure 8?X]. In Case Study Two [ADD REF to Types diagram in 3.X], and in Hestia.ai's digipower investigation (7.1.1), categories of provider-held data were illustrated with examples. In my research report for BBC Cornmarket [ADD REF], the use of **relatable examples** was identified as an important way to help people understand what a piece of data represents. Recalling that to make data meaningful, we must be able to interpret it as information [2.1.1, this can be refined further: **To make data meaningful, it needs to be expressed as life information**. Tables, spreadsheets and 'big data' sound dry and (to some) dauntingly technical, but once those same datapoints are expressed as '*facts about your life*', the hurdle of relatability is overcome. The application (and effectiveness) of this principle is evident in successful online services like Netflix, Spotify and Strava, and in social media platforms like Facebook: these interfaces show understandable everyday concepts like Friends, Events, Movies, Playlists, not files, records, folders or database rows. These examples have successfully '*pushed the technology into the background*', in line with Weiser's vision (Weiser, 1991), Rogers' calm computing, and the quote that opens this chapter. While exploring this idea of mapping *life concepts* further at BBC R&D, I produced Figure X, which shows a near-exhaustive overview of the many different pieces of information in an individual's life that might be held as data by service providers:

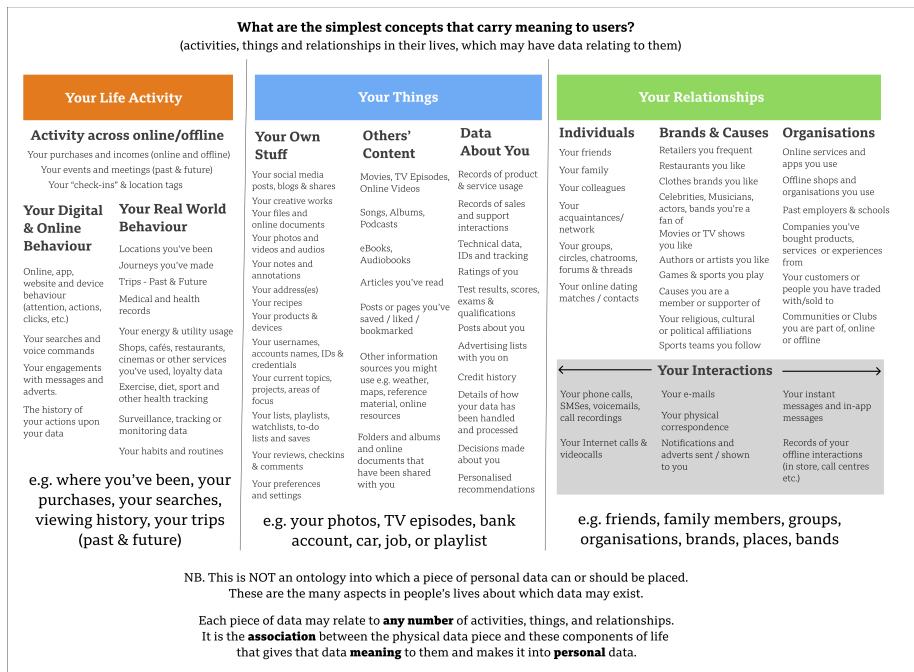


Figure 4: Figure X: Life Concept Modelling

(continues...)

This diagram shows how most common personal data types handled by providers today can be mapped to more relatable life information concepts. These life concepts (illustrated with examples where possible) are the best way to make data meaningful and relatable to individuals, and to begin to help people in their search for value in their data [5.4.3.1].

[TODO make this an inset box not a table] [TODO Fix non showing caption]

Another important obstacle to consider here is what I call the *Personal Data Diaspora*⁶. As illustrated by Imogen Heap's quote opening Chapter 1, an individual's personal data is typically very widely dispersed. For example, if I consider just my movement tracking data, I have over time accumulated activity logs from walking, running, cycling, and driving which are stored by Nike+, MyFitnessPal, Strava, Google Fit, Fitbit, Apple Health and Google Maps, not to mention the records remaining on my different smart watches, smartphones and hard drives. This is the problem of Integration (Li, Forlizzi and Dey, 2010) that SI enthusiasts face [2.2.3]. Even aside from the issues this creates in terms of managing one's data ecosystem [2.2.4], it means that it is impossible to view the history of my physical activity side by side, to spot patterns over time or make comparisons. To overcome this obstacle, approaches to data interfaces and life information modelling must be identified that recognise the **scattered, complex reality of each individual's personal data ecosystem** and begin to make it visible and understandable. This is explored further in 7.3.3 and 7.3.4 below.

The takeaway for this HDR objective is that data awareness and understanding is a problem of representation. Invisible data should be visibly represented, and all data should be represented in the context of its interpretation as life information.

1.3.2 Obstacles to the HDR Objective of Data Useability¹⁵

To consider how to improve the useability of data, we must first consider what properties of data, as it typically exists today, make it hard to use. The primary obstacles are that **most personal data is immobile, inaccessible, unmaliable and not interrogable**.

It is **immobile**, in that it is very difficult to move a dataset out of the environment where it exists: most data exists in organisations' internal databases, where it is tightly coupled to technology stacks, interfaces and business processes that use it. Separating one's data from the service that holds the data is difficult and often impossible.

⁶The word '*diaspora*' is typically used with reference to populations, but is an apt term, derived from the Greek '*diaspeirein*' meaning '*scattered about*' or '*dispersed*'.

[TODO possibly move this paragraph elsewhere to avoid repetition with previous section] This setting of personal data also explains why it is **inaccessible** to individuals (in the sense of ‘effective access’ (Gurstein, 2011)). Data access requests such as GDPR are typically satisfied by creating a copy of the data, which creates problems of delay, divergence and understanding. Even then, as Case Study Two showed, this is incomplete [5.4.2.2] and much of the data is never made available. Its accessibility is also hindered by the technical nature of data. For organisational efficiency, data will often be stored in complex proprietary structures which are designed for the algorithmic efficiency of the specific operations the service provider wants to perform, rather than for general-purpose re-use.

Evident from individuals’ goals for use of their data [Table 12] is that people need to be able to ask questions of their data. This highlights the problem that data is not **interrogable**. It must stand for itself, and there is no obvious way to ask a question about the meaning of the data or about what the data says about a particular question, without either the co-operation of the data holder, or advanced technical skills in querying and data analysis. To be able to ask questions of data, it needs to be malleable - one needs the ability to break it down, look at it from different perspectives, and reconstitute it in different ways. This requires more than just an ability to produce visual representations of the data, but an ability to **interact with the data** and produce new interpretations and insights that can help to answer specific questions.

To overcome these obstacles, we need to find ways to extract data from its current constraints into environments where it can move freely and be examined and reconstituted without restriction.

To address these obstacles, the following insights could help:

INSIGHT 2: Data Needs to be United and Unified

It is clear that better HDR involves recognising this scattered, splintered reality (Lemley, 2021) and moving beyond it. To make data useable for individuals, the diaspora must be united. This means that data from different sources must first be **united** – brought together – and then **unified**, which means making it into a collection of data about the individual and their life, rather than scattered slices of that person's life held separately in ways that are optimised for specific services. This is a multi-faceted sociotechnical problem of access, interpretation and **integration** (as recognised in self-informatics [Li, Forlizzi and Dey (2010); [2.2.3]]). The negotiability aspects are important (we can only unite data that we can access, and only those that stored information can fully explain it) but these aspects will be explored in 7.3.3 and 7.3.4 below. Setting that aspect aside and focusing on the practical, the way ahead begins with creating a space where data can be held, combined, controlled and **owned** by the individual - '*place for your personal data*' (Jones, 2011, p. [2.2.4]), forming the seed of their new human-centric personal data ecosystem. This is in line with Bergman's '*subjective classification principle*': that '*all related items should be classified together regardless of technological format*' (Bergman, Beyth-Marom and Nachmias, 2003) (We could add: '_‘regardless of where they are held’'). This vision is embodied in the concept of **Personal Data Stores** (PDSs) [2.3.4]. The BBC R&D Cornmarket project [7.1.1] is one project which is examining how to build PDSs, and in section 7.4 I explore possible design approaches. At this stage, we must recognise the importance of the *concept*, though. Once data is united and unified, this enables the creation of new views of data that were not previously possible. For example, today each separate TV app, device or streaming service maintains separate records of what you have watched. Once unified in a PDS, it would be possible to present you with a unified view of all the past content you had viewed, across all channels, as this mockup I made at the BBC shows:

[TODO make this an inset box not a table, image appears in box]

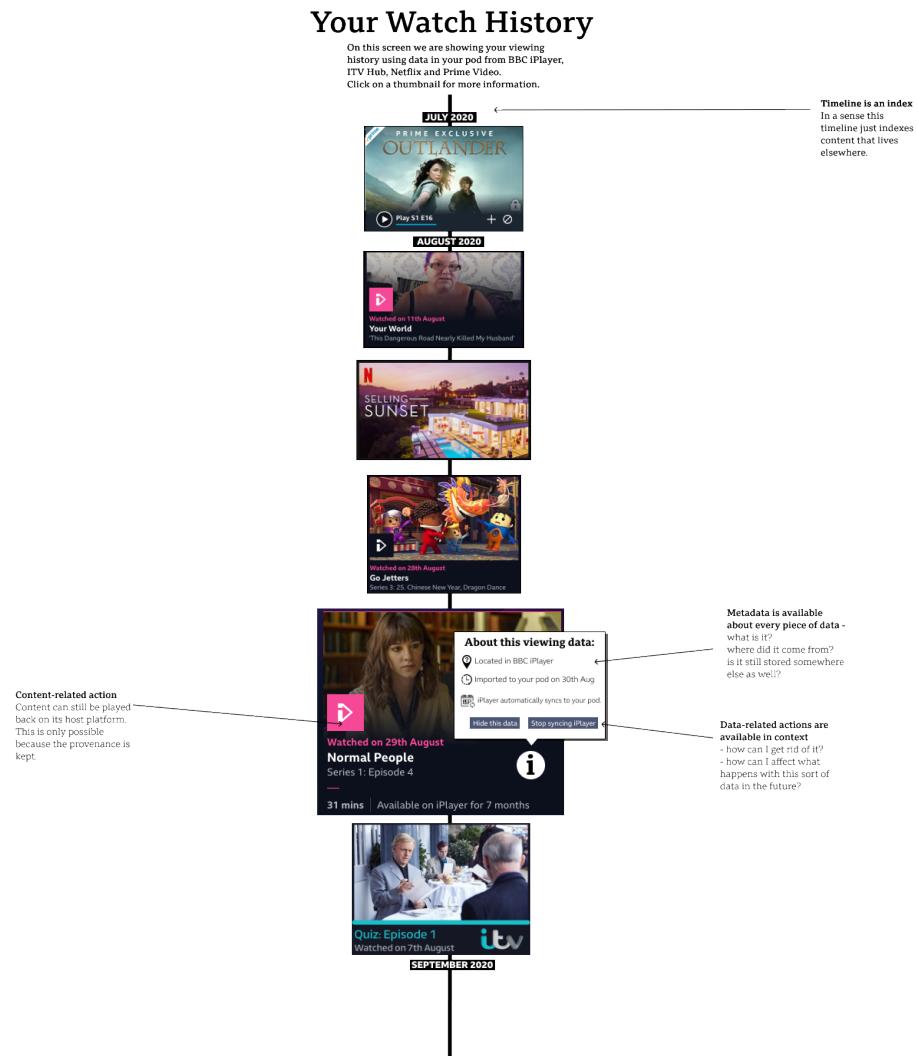


Figure 5: Figure X: A mockup of a unified TV viewing history which I created for the BBC R&D Cornmarket project

INSIGHT 3: Data Must Be Transformed into a Versatile Material

Looking at the specific individual goals Case Study Two participants had with personal data [Table 12] (e.g. reflection, pattern-finding, goal-tracking, and creative use) - and also at the many mechanisms that innovators in the PIM space have identified [2.2.2] (e.g. associative exploration, spatial arrangement, embodied interaction for different contexts), what we can infer is that somehow, unified data must be **transformed into a versatile material**. To truly empower users to make use of their data, we need to move to a model where data - represented as facts (or assertions) about their life – can be created, deleted, moved, grouped, annotated, copied, shared, modified, labelled, organised, separated or otherwise manipulated. This idea of data being a material is new for everyone but data scientists: it is new not just to end users but for designers too. Eva Deckers, in her work on *data-enabled design*, an approach to design which also calls for data to become a material, notes that designers (and we could expand this to laypeople too) “*are in general not trained and prepared to work with data. They’re not equipped with the right tools, data manipulation is not part of the schools’ curriculum and designers [people] are rarely interested in understanding data*” (Deckers, 2018). Her work with colleagues on the ‘connected baby bottle’ illustrated how such an approach can create a space for the iterative user-centred development of new capabilities (Bogers *et al.*, 2016). Based on this thesis’s theorisation of human data relations, the best candidate for what this material should *be* is the two information concepts we have identified - life information and ecosystem information. So the goal of data usability calls for the creation of **systems that enable human information to be treated as a material**.

So, for data to be useable, we must change its nature. We have been trained by the computers that have existed up to now that the basic units for interacting with computer systems are files - these are the material of today’s personal computers. Where we do interact with data as information instead of files, that information is typically presented in limited contexts within certain products or apps. In line with the goal to move up the DIKW pyramid [2.1], we need smarter computer systems, that move beyond files (Bowyer, 2011) - systems whose basic units of interaction are pieces of human information. **We need a human information operating system.**

1.3.3 Obstacles to the HDR Objective of Ecosystem Awareness & Understanding

As I have established in 2.2.5, 2.3, 6.2 and 7.2, human data relations cannot be made effective without a sea change in the way that individuals are able to interact with the complex ecosystem of personal data that we each inhabit. **Our personal data ecosystems are incredibly complex and largely invisible.** For example, it is very easy to allow a handful of communication and social media apps access to your address book or contact list, and before you know it you

have created a complex and unmanageable network of connections that silently sync and propagate your addresses and phone numbers across the Internet. And there are deeper layers which are not even slightly visible to users: networks of data brokers, advertisers and digital cookie companies exchange user identifiers, activity data and personal information about you while you browse or use apps (Pidoux *et al.*, 2022). As the Case Studies showed, the ability to build up a meaningful picture of your personal data ecosystem is completely absent [4.3.4.1] or severely limited, causing people to remain '*in the dark*' and leads to feelings of fear (Bowyer *et al.*, 2018), overload [2.2.4] and resignation [5.4.4.1]. Managing one's personal data ecosystem is an **overwhelming, unmanageable task** that even personal data experts are not fully able to get a handle on. We do not feel '*in control*' [Teevan (2001); 2.2.2]. The ability to provide a user with ecosystem transparency is hindered by the complexity and multiplicity of the data relationships they have been encouraged to set up, and by a lack of tools to provide a meaningful, or indeed any, view of those relationships. A further aspect to this obstacle is that in both Case Study contexts, no one individual or organisation has the ability to see the whole of a user's data ecosystem [4.3.4.3; Cornford, Baines and Wilson (2013)], and there is little commercial motive to try and solve this problem, as every provider focuses just on their own apps, websites and services. Making one's ecosystem visible, transparent and understandable is therefore an essential objective for better HDR.

INSIGHT 4: Ecosystem Information is an Antidote to Digital Life Complexity

Having identified that acquiring ecosystem information and understanding is a key motivator for many people (constituting 74% of participant goals in Case Study Two [Table 12]) and is an essential objective for better HDR, we can view the building of systems for **ecosystem detection** and **ecosystem information display** as ingredients to help overcome the obstacle. As a representative example to help show what this could look like, we can look to a new app called SubsCrab, pictured in Figure X.

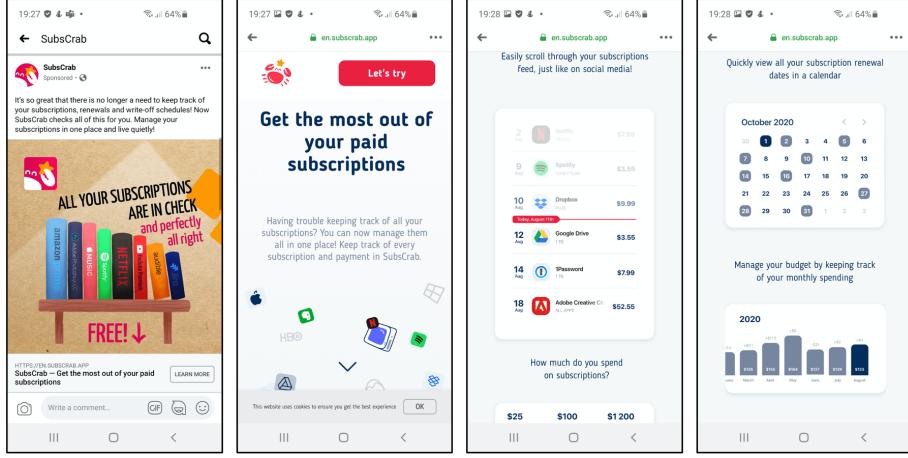


Figure 6: Figure X: SubsCrab: An example application for ecosystem detection and visualisation

(continues...)

This app connects to the user's e-mail account, and searches it and monitors it for e-mails from service providers such as Netflix, Spotify, Dropbox, or Google with which the user has monthly subscriptions. In doing so, it is **detecting part of the user's ecosystem** - identifying which companies they have a payment relationship with, and parsing the e-mails to identify billing dates and payment amounts. It then **provides additional representations of that ecosystem information to the user**, so that they might get on top of their subscriptions, see what they need to pay (or cancel), and feel more 'in control' [Teevan (2001); 2.2.2] of this aspect of their digital life. Thanks to this illustration it is easy to imagine other types of ecosystem detectors, for example detecting relationships with free services and websites, identifying account numbers and e-mail addresses, password resets, addressbook syncs, OAuth logins, family identities and more. Each of these could then power new interfaces, contributing to the **simplification of the user's digital life** and **giving people more visibility and control** over their previously unmanageable data ecosystem.

(continues...)

A further interpretation from this insight is that a key element of the required ‘sea change’ in approaches to human information relations mentioned above, is to challenge the current life-information-centric model that pervades in PDS and SI approaches, which all **assume that the only way to unite data is to collect it**. The difficulty in such an approach is that you can only collect that which you can extract. To address this, I drawing inspiration from a computer programming concept known as ‘*pass by reference*’ (as opposed to ‘*pass by value*’) (Ananya, 2020) where data is ‘pointed to’ rather than moved and also from productivity guru David Allen who recommends the use of ‘*placeholders*’ (Allen, 2015) to keep track of tasks you cannot otherwise bring into your planning. To be able to build a complete map of a user’s ecosystem we must **be able to keep track of accounts and data that are remote**, much like a search engine points to information on different pages around the web. We can create **proxy representations of service-provider-held or otherwise inaccessible data** (e.g. offline or restricted). These representations can become part of the manipulable material in the user interface, and could be augmented with links to visit those remote services.

Once we begin to think about storing and representing human information in ways that go beyond simply representing the information that is encoded within the data, and into the realm of *what the data is about*, new possibilities are unlocked. We can envisage building a PDS-type system that is not only a repository of personal data, but (thanks to proxy representations), a collection of ecosystem information *and contextually-situated* life information too, including information about relationships with data holders or other entities. This, however, exposes a secondary problem that any builder of such a system would face: **a lack of metadata** (as discussed in 2.2.2). Typically, much of the information stored on our hard drives lacks context about where it has come from, and how it relates to the individual in a holistic life/ecosystem sense. Where data access rights are executed (or data is shared via human means such as in 4.3.2.2), the attention is on the data itself: what it says. Case Study Two showed that some of the most desired information was not the data itself, but how it is used and shared and what is inferred from it (i.e. **metadata** [Table 9]), yet this was rarely forthcoming [Table 10]. There are many facets that can be quantified and recorded about a datapoint or dataset, as illustrated in Figure X, which I created at BBC R&D:

These facets can be mapped back to the 5 W’s that collectively make up the user’s *context* [Abowd and Mynatt (2000); #2.2.2-contextual]. It is notable that many of these facets are not explicitly recorded today, or would take significant work to capture; nonetheless, this exploration can serve as a useful reference for how information can be better contextualised (supporting the context-based and associative information management described in 2.2.2). Taking a step back to view this lack of metadata at a more conceptual level, leads us to the next

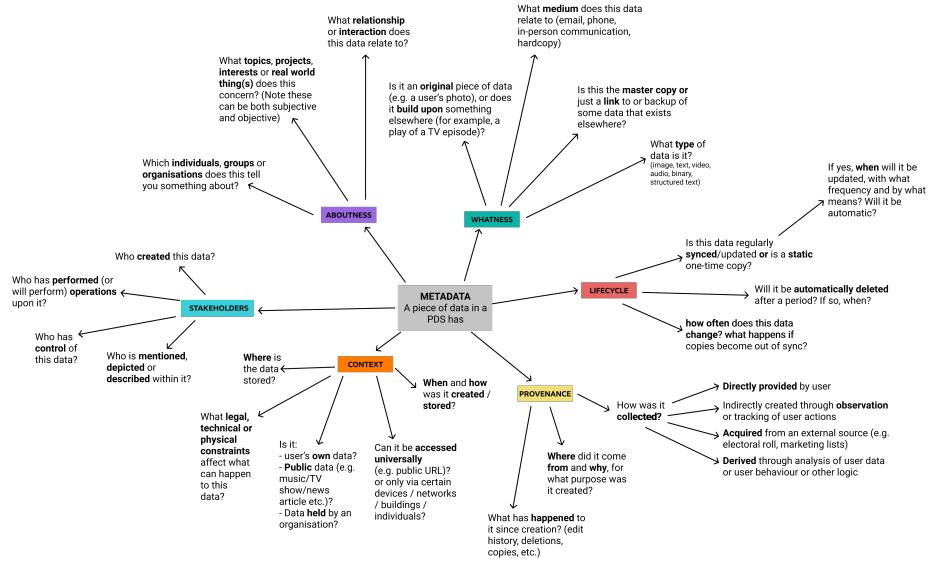


Figure 7: Figure X: Some of the many aspects of metadata that might exist about a datapoint or dataset

insight:

INSIGHT 5: We Must Know Data's Provenance

Metadata is what gives information *context*, which is critical to sensemaking [2.2.3] and enables good experience-centred design [2.3.2, 2.3.3]. Without context, data loses meaning (as observed by a participant in Case Study Two [5.4.3.1]). Collecting historical data about the individual is important from an SI reflection perspective [2.2.3]), but knowing the **history of a piece of data** is vital to understanding its nature and context. Data is not neutral and in fact is inherently biased, since it was created for a specific purpose with a specific agenda in mind (Gitelman, 2013; Neff, 2013). To address this, more context is clearly needed, Significant research in this space has been undertaken by Professors Mike Martin and Rob Wilson at Northumbria University, formerly Newcastle University, who promote the idea of ***data with provenance***; in other words that **data must carry with it the details of why it exists, how it came to be, and what has happened to it since its inception**, and that provenance must be communicated alongside any visualisation of the data, in order for it to be fairly assessed through full understanding of its context. Provenance is essential for data to be trusted, argues Martin, and should be quite granular: a piece of data should be attributed not just to an individual or organisation, but to the relationship between role-holding individuals in a specific context, and greater insights can be gained when **considering all actions upon data as motivated communications from one party to another**; only by capturing this information in-situ can the data be fully appreciated (Martin, 2022). This framing essentially advances the concept of history tracking into the sociotechnical, ecosystem-aware problem space this section is addressing. While everyday system designs have not approached this level of granularity, the importance of data provenance has been recognised in the PIM space: As described in 2.2.2, temporal PIM systems, from Lifestreams (Freeman and Gelernter, 1996) to activity streams (Hart-Davidson, Zachry and Spinuzzi, 2012) rely upon data provenance in some form. A study by Jensen *et al.* concluded that provenance tracking can be valuable for identifying related documents, a critical part of knowledge work today (Jensen *et al.*, 2010). Odom, Lindley and colleagues proposed the idea of file biographies, which view the lifetime of a file as something that should remain connected, and can be traversed in order to understand the context of the file at its different interaction points (Lindley *et al.*, 2018). This comes close to Martin's vision but does not capture the motivation for each interaction. While provenance capture is not a solution in its own right to the understanding of data and of ecosystems, it is clear that data with provenance is very likely to be a valuable part of any design that aims to help individuals with managing to get an overview of their complex and invisible personal data ecosystems.

What we can see in this section is that by paying attention to ecosystem information, metadata and provenance, we can open up a new space that, at the time of writing in 2022, almost no-one is building for. For people to manage

their digital world, they need a map. This is the first step on the road to giving individuals the ability to have oversight of their personal data ecosystem and take action within it.

1.3.4 Obstacles to the HDR Objective of Ecosystem Negotiability

This section explains three distinct obstacles to ecosystem negotiability: the intrinsic structures that give data holders power, the trend of actively diminishing user agency, and the intractable data self.

1.3.4.1 Hegemony through Data Holding It is in the pursuit of individual oversight [6.2.2] and decision-making involvement [6.2.3] that the impact of the power imbalance between data holders and individuals [2.1.2] becomes most clear; unlike the other HDR objectives, **individuals cannot act to claim ecosystem negotiability for themselves**. Negotiability means having the power to act, and in the context of systems and interfaces owned and designed by service providers that **power can only be given**. The hegemony of data holders is therefore the greatest obstacle to this objective, so it is vital to examine the nature of that power - where does it come from?

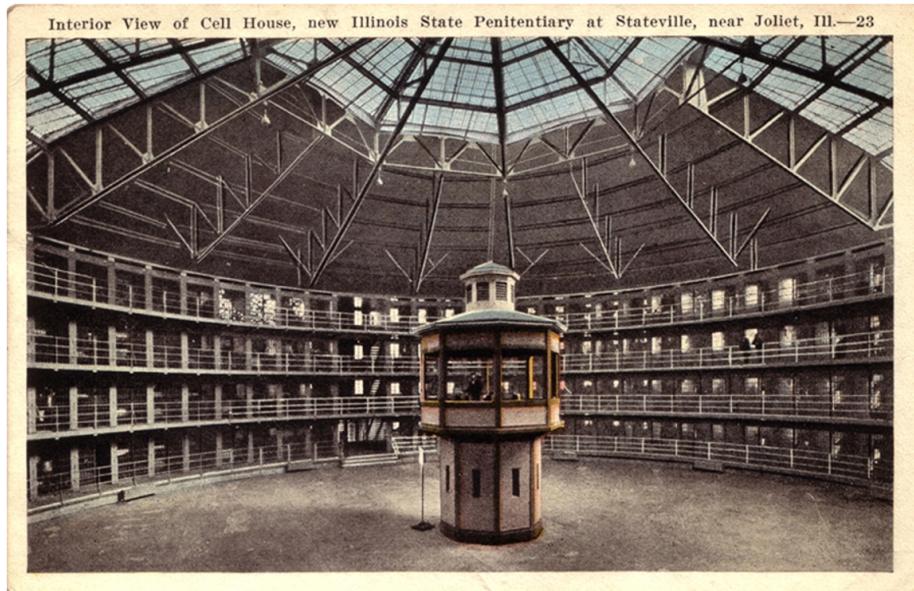


Figure 8: Figure X: The Panopticon Structure of the Illinois State Penitentiary

A helpful analogy for the relationship between provider and user can be seen in the design of Jeremy Bentham's *Panopticon* (Bentham and Bozovic, 2011), a real-world version of which is pictured in Figure X: an 18th century prison architecture design that would elevate the power of the (hidden) prison guards

to observe all the prisoners easily at any time while removing prisoners' privacy and providing no ability to observe those in power. As in Orwell's *Nineteen Eighty-Four*, individuals are unable to know when they are being watched (in this case because the guards are hidden from view by one-way screens), thus are forced into compliance. Structuralist philosopher Foucault interpreted the Panopticon as a political design, recognising that human environments can be configured to influence or regulate behaviour, in order to defend the power of the ruling class (Foucault, 1975). Such designs embody his four principles:

- **Pervasive Power:** the guards see everything all the prisoners do, all the time
- **Obscure Power:** the guards can see into any cell at any time, but the prisoners can't know when, how or why they are being observed
- **Direct Violence Made Structural:** the structure motivates the prisoners to self-regulate their behaviour without being coerced (through beating or punishment)
- **Structural Violence Made Profitable:** having been made compliant by the structure, the prisoners can be put to work for the benefit of those in power, as it is the only option available to them.

We can see at least three of these traits in modern Internet platforms such as Facebook today: these platforms monitor user behaviour (pervasive power) without their knowledge and without accountability (obscure power). Interfaces are designed to offer only those actions that benefit the platforms (for example, clicking ads, sharing content or spending more time on site (structural violence made profitable). This has happened through the processes of *platformisation* and *infrastructurisation* (Helmond, 2015; Plantin *et al.*, 2018), which have supplanted the Web 2.0-era promise of a free, open Internet that could be a great leveller and empower individuals.

Through the control of the data and of the design of the interfaces through which the data is made available – the only channel through which they can be observed – **service providers and platforms assert a structural power over the digital landscape**. Just as the design of the panopticon regulates the behaviour of the prisoners, so the configuration of the platforms, apps and service interfaces we use regulate and limit the behaviour of users. As Lessig wrote, '**Code is law.**' (Lessig, 2000). This infrastructural power is explained further in the insight below.

Looking deeper into theories of power reveals that structural power is not the only form of power which modern-day data-centric service providers hold. Jasperson *et al.*'s extensive review of types of power in the context of technology organisations (Jasperson *et al.*, 2002) identifies 23 different power paradigms, of which at least 13 can be, and are, asserted by data-centric organisations today:

- **authority:** ownership of technology or infrastructure (for example of websites, servers and code)
- **resource control:** controlling the flow of resources (in this case of infor-

mation/data)

- ***systems/structural power***: structural manipulation of others (as detailed above)
- ***rational power***: controlling decision-making processes
- ***disciplinary power***: using an influential position to affect others' mental models (for example, positioning location tracking as theft resilience)
- ***zero sum power***: winning a battle for ownership/resource control at the other party's expense (e.g. losing control of your sacrificed data)
- ***behavioural influence***: persuading others to carry out the desired behaviour (e.g. restricting features to motivate subscription payments, or promoting certain content or actions)
- ***interpretative influence***: determining how reality is externally represented (e.g. Facebook determining the way in which your social network is represented to you)
- ***network centrality***: becoming an indispensable hub of a wider ecosystem (for example, Facebook/Google dominance in online ad-brokering)
- ***processual power***: changing processes for competitive advantage (for example, platforms offering preferential APIs or rates to compliant partners)
- ***socially shaped power***: influencing a wide audience to settle upon a preferred interpretation (e.g. using dominant market position to dominate debates e.g. about privacy norms)
- ***interpretive power***: creating the internal representations of reality within an organisation (for example, presenting unpopular attitudes to data privacy to staff as normal/acceptable/beneficial for business)

INSIGHT 6: Data Holders use Four Levers of Infrastructural Power

Hestia.ai [7.1.1] have produced a model to explain the mechanisms by which powerful technology companies gain power and use it to shape today's digital landscape. In this model, ***infrastructural power*** comes from three things: ***technical ability***, ***organisational ability***, and ***the acquisition of data about individuals and populations***. Thus, as organisations (especially platforms) collect more data, and grow in market influence or technical capability, they gain power over individuals and over other organisations. They exert power in four quadrants, using four 'levers'. Simplified and expressed in the terms of this thesis, these are:

1. **Collect & Interpret Data to Acquire Knowledge**: Data and signals are collected from individuals and interpreted in order to infer their intents and interests. For example, Google collects raw GPS and wi-fi hotspot data from mobile phones, which it then statistically analyses to infer which shops or venues you visited and what forms of transport you used, increasing Google's knowledge about individuals and populations.

INSIGHT 6: Data Holders use Four Levers of Infrastructural Power

2. Present Content and Configure Structures to Influence Individual Behaviour:

Knowledge of individual intents and interests is exploited within user interfaces to influence desired individual actions. For example, Facebook or presents a user with a product relevant to their interests, which they are motivated to click upon, generating ad revenue. Another example would be Twitter manipulating the content of the user's feed to show more tweets from conversation topics where they can show promoted tweets, increasing ad revenue.

3. Configure Structures to Improve Knowledge Acquisition: A provider uses its dominant position to force other organisations to improve the provider's ability to acquire knowledge. For example, Google provides free analytics tools to web developers, but requires the end users of those client websites to supply visitor data back to Google, increasing their ability to acquire knowledge about individuals and populations.

4. Configure Structures to Disadvantage Others: Certain providers (typically of operating systems or popular devices) can configure the structural relationships between other parties. For example, a smartphone manufacturer could limit data exchange between other apps, while still extensively collecting data signals themselves, such as when Google was found to be collecting call history from Android's dialer app.

The precise mechanisms and techniques employed by platforms and providers when exerting their infrastructural powers as well as the social and market consequences of these practices are explored in detail in Hestia.ai's digipower technical reports, of which I was a co-author (Bowyer, Pidoux, *et al.*, 2022; Pidoux *et al.*, 2022).

An important aspect highlighted in the research is that providers' power is far greater than many realise: Unlike in the physical realm, providers of popular online platforms can **reconfigure the landscape to change the way that individuals perceive reality**, in line with the powers of interpretative influence, behavioural influence and socially shaped power described above (Bowyer, Pidoux, *et al.*, 2022). Providers control the extent to which (if at all) the data stored behind the scenes, and the internal processes that use that data, are visible, and how such data and processes are represented.

The above model shows that the accumulation of data (and hence, information) is implicitly and objectively a form of power. This theory is consistent with participants' observations in 5.4.4.1 that data holding and limiting access to it is a source of power. We can predict that as long as current platforms and service providers are free to collect so much personal information, the information landscape will remain imbalanced and individuals will not be able to acquire ecosystem negotiability.

INSIGHT 6: Data Holders use Four Levers of Infrastructural Power

Through this insight it is clear that the most powerful data holders exert huge influence over the digital landscape, in terms of what is knowable and what is doable. Individuals or activists' abilities to balance the landscape are hindered by the fact that they are operating in a landscape that the incumbent platform and service providers effectively control.

[TODO make this an inset box not a table]

1.3.4.2 The Active Diminishing of User Agency The second major obstacle to the objective of ecosystem negotiability we must recognise, is that the above processes of platformisation and power exertion are not a one-off transition, but rather an ongoing process which has not ended. Today's digital platforms exhibit **a continuing trend of actively diminishing individuals' agency**, especially evident in the last decade. When software was sold in a box, manufacturers competed based upon which product would let the user take home the greatest range of features and capabilities. New releases with new features drove new product sales. But in the cloud computing era, a smaller set of core features done well is sufficient to guarantee an ongoing subscription revenue from a user. Cost savings in development and support costs can be made by reducing feature sets. Constrained, compliant users are easier to manage. The relentless pursuit of increased profits and further cost saving sees products lose, not gain, features. Interfaces are reshaped to serve businesses' interests first and foremost. As described in 2.3.5, the primary concern is about making user behaviours constrained, predictable and profitable, moreso than meeting their needs or providing maximal value. Plantin *et al.* describe the particular harmful influence on the ecosystem of Facebook's power exertions: "*Facebook is a formidable force in a profit-motivated platformisation which is beginning to eat away at the Open Web. This entails moving away from published URIs and open HTTP transactions in favor of closed apps that undertake hidden transactions with Facebook through a Facebook-controlled API.*" (Plantin *et al.*, 2018)

Here are just a few examples of the ways in which users' agency has been, and continues to be, diminished:

- Facebook closed their RSS feeds, and later parts of their APIs, meaning that users could no longer consume their friends' posts in any other environment than the ad-filled and manipulated Facebook main feed. Later they removed features such as friend-list feeds and favorite-page feeds, removing users' ability to compartmentalise their content viewing or focus on certain friends. The 'Friends' page on Facebook currently shows a list of recommended new friends; to access your current friend list requires an extra click. Encouraging users to grow their networks is prioritised over user convenience.

- Twitter closed the parts of its APIs that allowed real-time notifications and access to one's home feed, killing off primary functionality for a healthy ecosystem of third party Twitter clients that gave users choice (Newton, 2018). TweetDeck, a major third party Twitter client was acquired, and later shut down, as was Twitter's own desktop client. Eventually, the only option left to users was to use the web interface. (Gayomali, 2015; Hatmaker, 2018; Siegal, 2022)
- Apple has been diminishing users' agency for a long time. Users cannot open up iPhones even to change the battery without invalidating their warranty. Apple have removed disk drives, headphone ports, SD card slots and other ports. Certain parts of the hard drive on macOS devices are now read-only and unwritable by users.
- Facebook recently announced that they will no longer store users' historical location data (though they will still use location unformation) (Pegoraro, 2022). This means that users will lose the capability to access historical location records, but also this makes it harder for users to see how their location data will be used in future, as there will be no historical log to examine. This shows that data-centric companies can change their practices to limit agency and reduce accountability too.
- Online news and discussion site Reddit has removed access to content for non-logged in users, and uses deceptive techniques to present advertisements that look like posts from users and to discourage users from appearing offline. These patterns have been described as '*disrespectful design*' (Regoje, 2021).
- In an example from the public sector, through my work on the SILVER project [3.4.1.1] just prior to the introduction of the GDPR in 2018, I heard whispers in at least one local authority of plans to 'shift from getting data collection consent from supported families towards simply informing them of our practices' (in other words, removing their choice). This shows that the instinct to further organisational interests over those of the individual is not limited to commercial data holders.
- In a similar vein, in 2022 TikTok announced that it would rely on legitimate interest rather than consent when it comes to using users' activity data to personalise the app experience, removing users' ability to withdraw consent to such use. This plan has subsequently been paused after warnings that this might breach GDPR (Lomas, 2022).

Unchecked, it is clear that trends to reduce users' agency and further providers' interests will continue. Therefore this trend to diminish users' agency is a particular obstacle that would need to explicitly targeted if data interfaces are to become more free-flowing (Bowyer, 2018), and if the objective of ecosystem negotiability is to be realised. Somehow, the trend needs to be halted, before it can be reversed. Judging by the TikTok example, perhaps only regulatory changes can force such a change.

The Intractable Data Self The third and final obstacle I have identified to the objective of ecosystem negotiability is **the intractable data self**. As identified in the pilot study (Bowyer *et al.*, 2018), and in Case Study Two [5.4.4.1] data about individuals serves as their proxy. It serves as their data self, and if it is incomplete, inaccurate or unfair, which is highly likely given the difficulties of representing people in data (Martin, 2007; Cornford, Baines and Wilson, 2013), this can cause harm (Bowyer *et al.*, 2018) or undermine attempts to help individuals (Cornford, Baines and Wilson, 2013). Yet currently, although some legal rights to data correction exist (Information Commissioner's Office, 2018), people lack practical abilities to modify or assert control over the most important version of themselves as far as providers are concerned: the version of them that exists in data. Even when data can be seen (such as via a support worker or GDPR data access requests) **people lack the ability to exert influence over their data self** [5.5.2; Cornford, Baines and Wilson (2013)]. To address this obstacle, the most likely direction would be to explore possibilities by which people could take a role in the curation of their data self, as both Case Studies have proposed [4.4.3; 5.5.2] and 6.3 have proposed.

To conclude consideration of this objective, it is noteworthy that to date, research and innovation on personal data ecosystem negotiability has been very limited. It is much easier to find business models and research funding for specific, well-defined contexts. Due to the lack of business incentive, only non-profit socially-focussed research organisations such as BBC R&D and Sitra have found themselves well-equipped to explore this problem space. Nonetheless, despite these challenges, there is an urgent societal need for researchers, designers, policymakers and innovators to explore how trends of diminishing agency can be reversed to involve people with their data. People need to be reconnected with their data selves, and given control over their digital lives, at the broadest level, rather than being excluded.

1.3.5 Obstacles to the HDR Objective of Effective, Commercially Viable and Desirable Systems

In the previous four subsections, the obstacles to the four identified HDR objectives [7.2.4] were considered. However, during attempts to tackle these objectives, and through observation of how the public and businesses were engaging with the growing Personal Data Economy, it became clear that there are certain obstacles that are specifically faced in this sector that affect *all* efforts to make progress towards improving HDR. The main challenge is around building such disruptive systems that are so different from the status quo: **businesses and individuals will not readily invest time and money in HDR, because it is unfamiliar**. Customers are not demanding HDR capabilities in their lives, and, all but the most socially responsible businesses do not immediately see the value in something that runs so contrary to their current business models which are based on the accumulation of data and the control of customer experiences.

Today, data is overwhelming, complex, and ‘sounds boring’. There is no denying that currently, engaging with one’s personal data economy to any degree more than that of passive consumer, is hard work. People routinely accept data sacrifice, click through T&Cs and cookie banners and are unwilling (or in some cases lack sufficient technical literacy, comprehension or skill) to do the work of asserting control over their digital lives. There is not a clear demand for holistic and novel ways of managing your digital life and exerting agency and negotiability over it. Across both Case Studies and the PDS work at BBC R&D, it was clear that even if new **human-centric information systems** and more inclusive service interaction practices could be created, we cannot assume that people will be inclined to use them in great numbers, especially not if it seems like hard work or not worthwhile. This can be seen as an obstacle that affects all HDR improvement approaches we see, and indeed is why many companies in the emergent PDE economy [2.3.4] struggle to find a business model - while there are clear benefits, better HDR does not appear to something that a mainstream audience would be directly be willing to pay for. But this should not deter disruptive innovation nor does it indicate that such offerings would not be useful. As automobile pioneer Henry Ford famously said, *“If I had asked people what they wanted, they would have said faster horses.”* Nonetheless, it is a clear overarching obstacle to overcome.

INSIGHT 7: Human-centred Information Systems must serve Human Values, Relieve Pain and Deliver New Life Capabilities

Through work at BBC R&D exploring how to better connect people with their data, it became clear that there is a way to combat such indifference and apathy of users. It emerges from the realisation that the way people find value in data is to connect it to their lives. The more that people see relatable life information and can imagine ways to harness that information in their everyday life, the more motivated they will be. BBC R&D conducted some research (Forrester, 2021) that identified fourteen specific Human Values that people seek to satisfy in their lives, which are shown in Figure X. These are, at the most abstract, goals that people care about in their daily existence.



Figure 9: Figure X: Human Values, as identified in BBC R&D research funded by Nesta

(continues...)

Given these and the earlier observation that life information is what makes data relatable, the insight I offer here is that **the way to make people care about their data is to use it to help them in their life**. By starting with a focus on a user's world, one can then focus in on their life, and then the data that represents elements of that life. Then, the individual has a vested interest.

Systems and features should be designed from this life-centric perspective. This is known as **value-centred design** (Reber and Duffy, 2005) and it has been argued that this should become the guiding design philosophy in HCI (Cockton, 2004). And to offer true individual value, all human-centric system designs must also take into account *context* [2.3.2], *environment* (Abowd, 2012) and *experience* [3.2]. In business modelling, there is a tool called the *value proposition canvas*, which identifies three ways of conceptualising value: **gain creators, pain relievers and jobs-to-be-done**. If we can use those concepts to inform our designs, we can produce better human-centric functionality - relieve an individual's pain points, help them complete their tasks, or offer them some gain over the status quo. In the HDR space, given the lack of existing tools for digital life management, we have the opportunity to create quite a unique type of gain: **new capabilities over your digital life** that you have never had before. This ability to do new things has been identified as key ingredient of user empowerment (Meschtscherjakov, Wilfinger and Tscheligi, 2014; Schneider *et al.*, 2018).

(continues...)

Here is an example of what this value-centric approach might look like in the HDR space: Myself and BBC R&D colleague Jasmine Cox imagined focusing on address books and contact lists as a strong relatable starting point to generate demand for a human-centric interface. This could provide people with new life capabilities while also relieving pains. Many people have address and contact information scattered far and wide, and face a complexity they cannot easily manage when it comes to the automated syncing and sharing of potentially sensitive contact information between devices, apps and providers. Developing human-centric personal information management capabilities to bring that messy situation under control would offer a clear and tangible benefit to users. In Figure X, we show how there could be a strategic path, beginning with detecting ecosystem and life information from the individual's calendar and e-mail inbox, through to building up to more holistic life-level PDS capabilities.

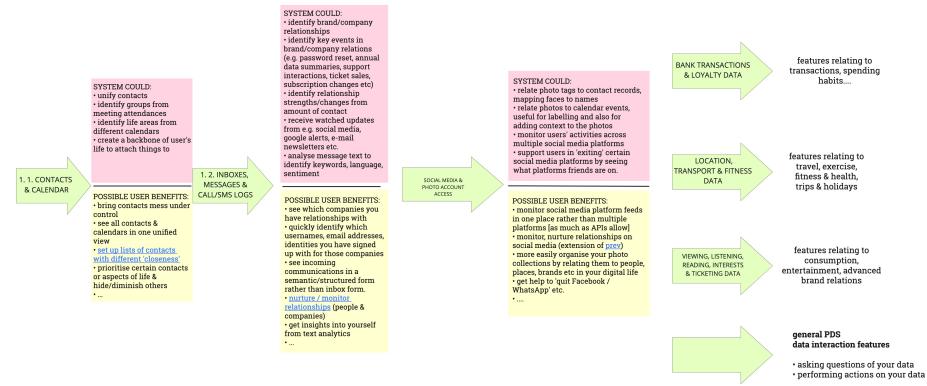
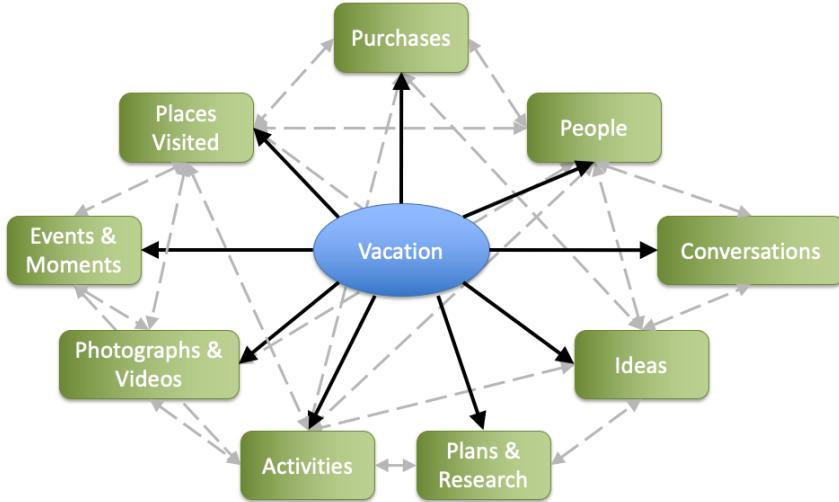


Figure 10: Figure X: A contact-and-calendar centric PDS approach

(continues...)

Another example that is helpful to consider is my the example from my 2011 article: that of a vacation, as shown in Figure X (Bowyer, 2011). Today, all the information around such a holiday is scattered into multiple systems - emails, online provider bookings, chat logs, cloud synced photos, web browser bookmarks, smartphone location logs, etc. It is not hard to imagine that a system that was able to bring all related information about that vacation together in one central interface (mockup in Figure X) could deliver huge value to users and be very compelling. Such context-targeted human-centric offerings can have a much greater chance of generating interest and impact than offerings that merely allow you to 'organise your data' or some other abstract phrasing.



[TODO make this inset box not a table, images appear in box]

The kind of life-spanning, unifying interfaces described in the insight above are nothing like the interfaces that are built today, as they span across different providers' data and services. This highlights the secondary obstacle that all HDR system builders will face, whichever objective they wish to target: **closed, self-interested organisations with a lack of interoperability**. Building an HDR system will necessarily involve connecting to systems of different providers that have different touchpoints into an individual's life and world. Yet most companies act in closed, introspective and non-cooperative ways to further their own interest. Companies like Apple, Amazon, Microsoft, Facebook and Google (the so-called '*big five*') build **proprietary, incompatible silos** or '*walled*

gardens' – sub-Internets that pretend that the alternatives do not even exist, in order to encourage a flow of money and attention to their own products and services. Commercial motives encourage them to get users to spend time in their own proprietary spaces (so that resultant ad revenue can be captured) and in order to maintain subscription revenues it is in providers' interests to make it hard for individuals to leave or switch providers. In effect, providers build for a world that does not exist, where **every individual is imagined to only interact with that single company's interfaces**. I would argue, for example, that Google's venture into social networking with Google+ did not succeed because it failed to build for a reality where most people and their friends were already on Facebook. But one can understand their motives; there is little incentive to open up the ecosystem when the free flow of information and of users might result in loss of income for the company in question. Users with negotiability would be more able to leave. And this also encourages keeping users in the dark [5.4.2]. The less agency and negotiability that users have, the more freedom the provider has to do exactly what they want with their data. In this context, users are '*pathetic dots*' (Lessig, 2000) or '*docile bodies*' (Foucault, 1975).

The tendency of organisations to work in closed, introspective practices and to be resistant to opening up data or services is not solely motivated by commercial reasons: the public sector has a vastly complex, closed and fragmented ecosystem [Pollock (2011); Copeland (2015); 4.1.2]. Efforts to build a system to share health data with support workers for the SILVER project [7.1.1] proved hugely challenging. Sometimes the challenge was a more technical one - incompatible data formats that are hard to reconcile, or data being stored in legacy systems with no public API that would allow programmatic access to that data, or issues around licensing. Data sharing agreements must be established, especially in the public sector which is by its nature more liable to scrutiny and accountability. But more than these technical or procedural issues there was resistance to change in data processes and an unwillingness to share data between agencies, often motivated by a fear of legal repercussions. **Data-centricism encourages insular thinking:** it encourages organisations to codify the world into their own systems, processes and formats for their own use.

And yet, for effective HDR, **data needs to be separable from services**. The more users data is tightly coupled to specific services, the less agency users have and **the harder it is to build life-centric systems**. On BBC R&D's Cornmarket project, attempts to build an interface for users to import data from multiple popular Internet services proved to be a hugely complicated endeavour, requiring access to many different APIs or manual exports and imports of data by users. There needs to be greater interoperability and **greater establishment and adoption of standard formats for exchanging human information** (as distinct from establishing standards for data or service-specific APIs). As mentioned above, platformisation breaks the Open Web (Plantin *et al.*, 2018). To overcome this, companies must be persuaded that human-centric thinking, interoperability and transparency has not just social benefits, but business

benefits too.

But at an abstract level the technical obstacle, the problem is one that has always faced the tech industry, which is that there often is no universally agreed way to represent important concepts - in this case human-centric information concepts such as events, social media posts, website visits, location history information, app activity, etc. And any entity that does create a standard then faces the challenge of trying to persuade others that their standard is the best one to use. In general, standards work best when established by non-commercial industrial standards bodies (for example the World Wide Web Consortium (W3C) or International Organisation for Standardization (ISO) and then mandated through policy such as European Union law. Such standards must be established with input from industry experts.

INSIGHT 8: We Need to Teach Computers To Understand Human Information

In order to move towards standardised ways to store and unify personal data from multiple sources, computer systems must be taught to understand the information within the data, and how it relates to an individual and the world. This moves beyond just capturing data provenance: put simply, **computers need to understand human information**. They need to move beyond files (Bowyer, 2011) and databases, and begin to perform operations on human informational concepts, and to associate those concepts according to what they mean - i.e. ***semantically***. This is a preliminary step that will enable the building of systems and interfaces that are able to deal in human concepts and represent the elements of everyday life.

We need to store **semantic context and semantic associations**, i.e the meaning of things, not just raw bundles of data. This is advocated by the Web's inventor Tim Berners-Lee in his vision of a Semantic Web (Berners-Lee, Hendler and Lassila, 2001) and by proponents of *networked* and *semantic* PIM systems, as detailed in 2.2.2. There is a need to develop standard ways to digitally model facts and assertions about users' lives, so that those disparate pieces of data can be unified, connected, correlated and compared. Some standards are already developing, such as *data shapes* ('ShapeRepo: Make your apps interoperable', 2022). And the extraction of meaning from data is a problem domain all of its own. Sizable industries have built up around Content Analytics and Enterprise Content Management. But to consider the problem at its simplest level, I offer this insight: Through **the capture of metadata** at the point of data recording, and through **subsequent programmatic analysis** of stored data, as illustrated in Figure X, we can begin to teach computers what the data we store represent.

[TODO inset box not table, including image in box]

Machine learning technologies and Artificial Intelligence have pushed machine understanding of human words, images and content to impressive levels in recent

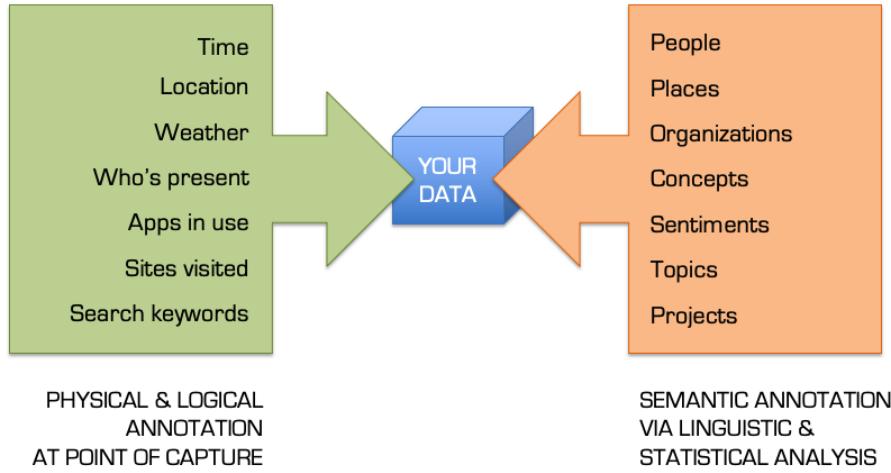


Figure 11: Figure X: Annotating Data with Semantic Context

years and such technologies can certainly be helpful, but in fact at the core what we are talking about here is something much simpler than AI; It is simply about labelling datapoints in as many different ways as possible so that those datapoints can be associatively retrieved from many different angles, and providing humans with ways to amend incorrect labels and to reclassify data or apply new semantic associations. Issues of interoperability for PDS systems are being actively explored and developed in the ‘Solid’ community (Bansal, 2018; Berners-Lee, 2022) in pursuit of a decentralised web (Verborgh, 2017).

Such approaches are in their infancy, and have not yet been adopted extensively in commercial settings. Even after addressing the obstacles of end-user buy-in and the technical complexities of building human-centric systems, data-driven corporations, motivated as they are by profit and business success (and smaller online organisations too) need to be persuaded of the business value of transparency, interoperability and human-centrality. This is explored further in 7.4.5.

In summary, whichever of the above four HDR objectives are targeted, all HDR reformers involved in building HDR systems must:

1. create, adopt and co-ordinate around **new standards** for human information storage and management
2. invest in systems that elevate computers from data-processing machines to **human-information-processing machines**, and
3. make a persuasive case to both businesses and individuals that the new approach offers **tangible, previously unavailable value**.

1.4 The landscape of opportunity: Four approaches to improving Human Data Relations

In this section, I will present four different ‘flavours’ of activity, that we can pursue as HDR reformers. These are represented diagrammatically as trajectories of change using a model known as *Theories of Change (ToC)* which is explained in 7.4.1.

These trajectories for change could be supported or pursued in many different ways, which may appeal to different readers:

- prototyping and creating proofs of concepts;
- fundraising or investment; design activities;
- market research, participatory research or usability research;
- ‘early adopter’ testing and quality assurance of PDE / PDS offerings;
- promotion, advocacy and journalism of HDR issues;
- critical audits of provider practices;
- policy design;
- political pressure on governments and regulators;
- participation in open data, PDE, civic hacking or MyData communities; or even
- individual self-experimentation with HDR tools, rights and capabilities.

1.4.1 A Primer on Theories of Change (ToC)

To provide a structure for cataloguing the insights conveyed by this section, I use a *Theory of Change (ToC)* framing. ToC is a set of methodologies commonly used by philanthropists, educators and those trying to improve the lives of disadvantaged populations (Brest, 2010); the theories can be used in different ways including planning, participatory design and field evaluation of the effectiveness of new initiatives. There are many different implementations, but common to most of them is a focus on explicitly mapping out desired outcomes (Taplin and Clark, 2012) with a clear focus on who is acting and whether the change being brought about is a change in action, or a change in thinking (Es, Guijt and Vogel, 2015). In this section, ToC theory will be used in a very limited way, not as a methodology but simply to provide a structural frame for proposed changes, as described below. Using ToC to perform evaluation of the effectiveness of proposed change approaches in action in society would be well beyond the scope of this thesis. Nonetheless, this frame is a useful way to map out the different approaches to changing the world in pursuit of the ideal of better HDR.

Figure 29 illustrates the aspects of ToC thinking that are important when using this frame. Specifically, desired changes can be broken down into:

⁷Diagram used here unchanged from *Hivos ToC Guidelines* (Es, Guijt and Vogel, 2015, p. 90) under a CC-BY-NC-SA 3.0 license, whose authors state that this diagram was adapted from earlier work by Wilber (1996), Keystone (2008) and Retolaza (2010, 2012). [TODO footnotes may need renumbering now]

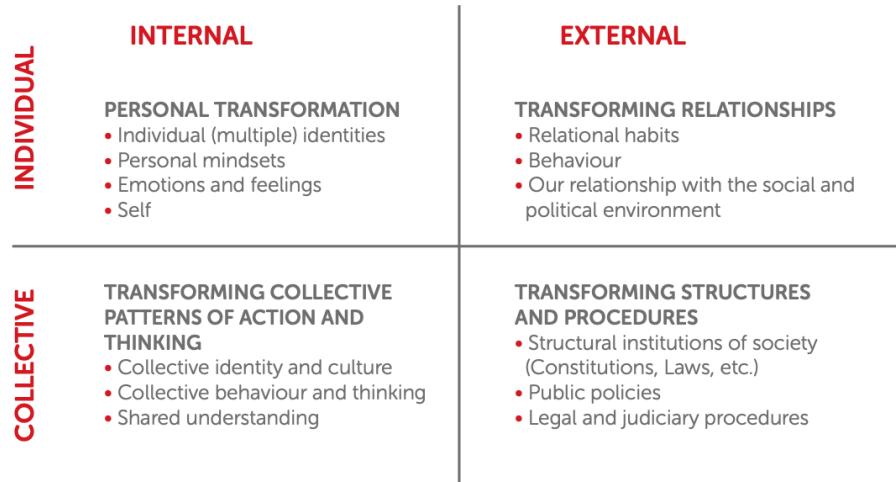


Figure 12: Figure 29: The Four Dimensions of Change⁷

- **Internal changes:** changes in thinking, feeling, reasoning, understanding, attitudes or identity.
- **External changes:** changes in actions, behaviour, interactions, structure, policy, technological capability, processes and the external environment.

At the same time, desired changes can be broken down into:

- **Individual changes:** changes to individual thought or actions
- **Collective changes:** changes to the thoughts or actions of groups of people together, or to the systems, practices and norms of society at large.

These two splits produce four dimensions of change, and form four quadrants representing different types of change, which are shown in Figure 29 and described here:

- **Individual/Internal (II):** This top-left quadrant represents changes to what individuals know and understand, and to how they think, feel and plan to take action.
- **Individual/External (IE):** This top-right quadrant represents changes to how individuals' relationships with others; acting (or being enabled to act) differently in their daily lives and when interacting within society.
- **Collective/Internal (CI):** This bottom-left quadrant represents changes in the shared knowledge of groups of people or to the collective identity or values of social groups.
- **Collective/External (CE):** This bottom-right quadrant represents changes to the structures and procedures within which people operate, including technology, law, societal norms and communications.

Key to ToC thinking is the idea that making changes in one quadrant can

stimulate change in others; for example, collective learning about data attitudes and practices, such as the research conducted in this PhD, (lower left quadrant) could inform the design of new technologies, interfaces or processes (lower right quadrant), which if built could make new structures available to have an impact on improving individual-provider relationships (upper-right quadrant). The changes to those relationships could then in turn lead to individuals thinking and feeling differently (upper left quadrant), for example feeling more empowered or having greater awareness of data practices. The diagram at the start of each of the following four subsections illustrate that approach's trajectory of change through the ToC quadrants.

1.4.2 Approach 1 to Improving HDR: Discovery-Driven Activism

[DIAGRAM PLAN - TODO ADD DIAGRAM] [individuals informing/powering collectives Collectives helping individuals Using Data to Demand Change in Practice => which in turn enables individuals with stronger capabilities and better transparency & insights also data access & understanding services helping collectives and individuals info pulling from the bottom right and top left into bottom left, then pushing up into top two.]

The approach to HDR reform presented in this section embraces the *activist* aspect of being a recursive public [7.2.5], as well as the idea of reconfiguring one's world. This approach focuses on the realities of the current data-centric provider ecosystem, and is focused upon deeply understanding it so that it can be challenged from a grounded position of strength. The approach, which would be applied typically to a single service provider, app or platform (which could also be a public sector service) is fourfold:

1. **Gather Evidence:** Using all available means, gather personal data and information from the target organisation.
2. **Determine Current Practice:** Analyse the gathered data and information, using targeted approaches to dig deeper where needed, to establish facts about how the target organisation stores and handles personal data.
3. **Identify Desired Change:** By consideration of the specific practices observed from the analysis, specific problematic practices, gaps or areas where improvement or change is needed are identified.
4. **Appeal for, or Compel, Others to Act:** Using all available means, directly or indirectly persuade or force the target organisation to execute the desired change.

I describe this approach as **Discovery-Driven Activism**. The discovery phase, which aims simply to establish facts about the past or current data practices of the target organisation, can be broad ('*let's see what we find*', as in the digipower investigation [7.1.1]), or highly targeted, such as when The Citizens (a non-profit pressure group in the UK who '*use impact journalism to hold government and big tech to account*' ('The citizens - about us', 2020)) used Subject Access Requests to investigate a breach of personal data by the Labour Party in the UK (Colbert,

2022).

Subject Access Requests and **Data Portability Requests** (Information Commissioner's Office, 2018) are two of the most powerful tools for this kind of investigation. Similarly to Freedom of Information Requests which have used to obtain otherwise hidden data and information from governments and public sector organisations ('More than 400,000 freedom of information requests made since 2005', 2014), so these new data access rights are beginning to be used to force commercial organisations to release personal data or information about data processing. There are challenges in non-compliance, as discussed in 5.4.2.2 and 5.5.1, but the ability for the individual to ask very broad or very precisely targeted questions and to be able to threaten a complaint to a Data Protection Authority if that question is not answered, is a significant new power that HDR reformers can exploit.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

In this inset box, I will show, using a mini case-study from my own direct personal experience, how one person can apply the discovery-driven activist approach to compel a multi-billion dollar international data-centric organisation to improve their HDR.

As an avid user for several years of the music streaming service Spotify who has built up a large library of playlists, I have made a number of GDPR requests to get copies of my personal data. When I was first given a copy of my personal data, I was returned a basic ZIP file including 12 JSON files containing playlists, search queries, account information, my last 12 months of track play history, and inferences about my musical tastes. Spotify also make an extended data download available, including technical log data, and extended play history (which covers the lifetime of my account). I requested this extended download and received a much larger dataset with 175 JSON files, including granular details of when I had used different interface features and the precise details of every song I had ever played. Thinking that I would like to use this data to build a view of my listening history that was not tied to the Spotify platform (in line with the idea of increasing agency by separating one's data from the service that holds it [7.3]), I examined the streaming history and playlist data with this purpose in mind. What I found was that individual songs were identified only by textual strings of the title, artist and album name. This information is insufficient for a programmer's use - there is no unique identifier or Uniform Resource Indicator (URI) to uniquely identify the specific version and release of a track played. Also without such an identifier, it would not be possible to generate a thumbnail image of the track, or build functionality such as a clickable link to 'play this track in Spotify'.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

This highlights a common issue that occurs with data access requests, as highlighted in 5.4.3.2 - there is ambiguity over whether providers should identify data in a machine-readable way (useful for programming), or in a human-readable way (to optimise understanding). In my case, I needed both. I e-mailed Spotify back and was provided with an alternative fileset which contained only Spotify Track URIs, such as `spotify:track:5CKqyYTZqp6Nb4b3kJjUL5`. These met the programmer need to uniquely identify the track, but not the human need - I had no idea which artist or track each of these URIs corresponded to, as there was no human-readable text accompanying each entry. So, I e-mailed Spotify back, making the case that my GDPR rights had not been fully satisfied, because I needed *for each play history entry*, both machine-readable ID and human-readable track title and artist name. I sent Spotify over 30 e-mails on this matter between October 2020 and May 2021. There is little continuity of conversation between support agents, and hard to be escalated to the correct staff with the technical or legal expertise to assist with such nuanced questions. However, by persistently and politely repeating my questions and not accepting No for an answer, I was able to achieve a notable outcome, Spotify **changed the format of their data returns**, not just for me but **for all future customers**. Now, each item in the playback history data you get back from Spotify, every item includes textual track and artist details AND a Spotify track URI. The data can now be understood by both human and machine. The likely interpretation here is that I successfully able to persuade their Data Protection Officers (who handle GDPR requests) the importance of returning data that is both machine-readable and human-understandable. Perhaps they also recognised the amount of work they had invested in supporting my query, and wanted to **avoid having to do such work ever again** should I or any other customer make the same request in future. This was a tiny impact, but a lasting one, and it shows that the discovery-driven activism / civic hacking approach can have an effect in improving HDR with a target organisation.

INSIGHT 9: Individual GDPR requests can compel companies to change data practices.

A larger scale example of individuals forcing giant corporations to change is seen in the case of Facebook. In the early 2010s, Austrian lawyer Max Schrems began to pressure Facebook to disclose more personal data to their users. He created a tool to enable people to make their own data access requests, which over 40,000 people used. Faced with an overwhelming volume of work and massive liability of future data access requests, Facebook was forced to launch the self-service ‘Download Your Information’ (DYI) download tool, increasing transparency for all Facebook users worldwide (Solon, 2012). Facebook was forced to increase its transparency further when Paul-Olivier Dehaye (now CEO of Hestia.ai [7.1.1]) made a GDPR request (later backed by legal action) to force Facebook to disclose more information about which advertisers Facebook had enabled to target him using the Facebook Custom Audiences feature. Apparently in order to avoid being embarrassed in court, Facebook updated DYI so that your downloaded information includes a list of advertisers who have added you to a Custom Audience (Dehaye, 2017). Dehaye and Schrems both continue to act as HDR reformers and civic hackers following the discovery-driven activism approach, through their organisations Hestia.ai [7.1.1] and privacy rights organisation noyb.eu ('none of your business') (Schrems, 2017) respectively.

Facebook’s DYI tool, mentioned in the insight above, represents a useful class of tool in the arsenal of the activist HDR reformer. Along with Google Takeout, it is one of number of '**data download portals**' that allow users to download their own data. Since GDPR’s introduction in 2018, an increasing number of large online platforms including Facebook, Google, Apple, Netflix, Twitter, Spotify, Uber, Instagram and Strava, faced with the need to reduce the cost impact of GDPR request handling for their large userbases, have developed and augmented online self-service portals available where users can download a copy of their personal data. This has some advantages over Subject Access Requests in that data can usually be obtained within minutes or hours rather than taking up to 30 days, but has some disadvantages in that the data returned is a voluntary offering by the company, that may not cover the data that the individual is seeking and does not provide any ability to ask follow-up questions. This technique was sometimes used as a fallback means to obtain data in Case Study Two, and was used more strategically in the digipower project, where its merits and limitations are discussed [7.1.1; Bowyer, Pidoux, *et al.* (2022)]

Both access requests and download portals rely on the organisation in question to be transparent, accurate and thorough in their provision of information, but an alternative technique of **data flow auditing** allows individuals to investigate and collect data on the *actual* behaviour of a target organisation. This was used effectively in the digipower project [7.1.1]. Using an Android app called TrackerControl (Kollnig, 2021), a service provider’s app can be monitored while the user is using it normally, to see which servers or domains that app

is contacting (and one can imply, exchanging data with). Apple has recently introduced an equivalent function on iOS known as App Activity Reports (Apple, 2022), providing iPhone users with the same ability as part of the phone's operating system. This has limitations in that the content of the data exchanges is not known, but can serve as a valuable tool to verify claims made in privacy policies or GDPR responses, and also as a means to *generate questions* for further investigation, for example by identifying third parties such as data brokers which the target organisation may be sharing personal data with. This technique is described further in (Bowyer, Pidoux, *et al.*, 2022), along with a comparison of the different techniques of data flow auditing, data download portals and data access requests.

In general, what the discovery-driven activism approach highlights is that there is a role for pro-active citizens to play in challenging the power of data-holding organisations by treating those organisations as a subject of investigation, both in research (Walby and Larsen, 2012) and in the pursuit of improving civic society (Schrock, 2016).

Once information has been obtained, the HDR reformer activist can use a variety of means to try to bring about the desired change. If a target organisation fails to comply with a data access request, or a demand to erase or correct data, they can be *reported to the Data Protection Authority*. In some cases even the threat of this (which can carry a large fine) can be enough to compel the organisation to change. If a breach of law is found, the target organisation could be *taken to court*, as seen in the Schrems case above, which resulted in new legislation that Facebook had to comply with (Kuchler, 2018). As well as individual cases, this also often happens in the form of class action lawsuits, as with Facebook and Cambridge Analytica (Bowcott and Hern, 2018). Increasingly, unethical or illegal data practices are being challenged. In some cases, such extreme measures are not needed. Simply *making data available to the public* can be empowering to society at large. This approach has been demonstrated by the UK website TheyWorkForYou, which increases democratic accountability of MPs by making MP's votes and public statements more readily accessible (mySociety, 2004). As well as structured '*impact journalism*' such as that conducted by The Citizens as mentioned above, another technique available to individual activists is *public shaming of misbehaving organisations*, especially on Twitter. While the ethics of this are complex and it does not always succeed, the technique has been used effectively to force organisations to change, in order that they might avoid further bad publicity (Silver, 2014; Braw, 2022).

INSIGHT 10: Collectives can compare and unify their data and use their pooled knowledge to demand change.

Increasingly, the Internet experience that individuals experience is not the same as anyone else's. Thanks to recommendations, targeted ads and social media feeds personalised to your interests, no two people will see the same digital reality. This means it is very difficult for regulators or individuals to hold digital service providers to account. In recent years, many activists have embraced **the power of collectives**, and realised that together, they can discover far more than they can alone.

An example of this is the WhoTargetsMe project, launched in 2017 (Jeffers and Webb, 2017). The objective of this project was to monitor political advertising in the UK. Recognising (as larger studies have shown (Bakshy, Messing and Adamic, 2015)) that everyone was seeing different advertisements, the goal was to have each individual report what adverts they see on Facebook, so that these can be pooled and compared with others. Over 50,000 people participated, building up an otherwise unavailable clear picture of the ways in which different political demographics were being targeted. This is a powerful mechanism available to collectives in this space: the ability to **have individuals obtain their own datapoints and then compare them**.

Another example is seen in the Worker Info Exchange ('Worker info exchange', 2022), a collective that helps gig economy workers such as Uber drivers and Deliveroo riders to make data requests. Using the pooled data, they conduct investigations to understand algorithmic inequalities and identify unfair treatment of worker by employers. They then help those workers to fight for better working conditions, much **like a traditional trade union, but powered by collectively-sourced data**. This resulted in Uber being taken to court, and some gains being made for drivers (Lomas, 2021; Foucault-Dumas, 2021).

As the aforementioned case with Max Schrems showed Insight 9, collectives can be particularly powerful when **exerting their data access rights *en masse***, and this can improve HDR and force greater transparency. René Mahieu and Jef Ausloos have published an exhaustive list of collective actions taken using GDPR rights, addressing issues such as discrimination by US colleges, corporate surveillance of climate activists, identifying gaps in data disclosures, and manipulation of users on dating apps (R. Mahieu and Ausloos, 2020). The authors identify that the GDPR provides an '**architecture of empowerment**' and have called for better enforcement and for European authorities to provide better support for the ability for collectives to make data access requests together (R. L. P. Mahieu and Ausloos, 2020). Hestia.ai's digipower investigation concluded that data-discovery driven collectives are a vital step on the road to a more digitally empowered society (Pidoux *et al.*, 2022, p. 70). It is clear that organised collectives exploiting data access rights represents a powerful vector for impactful discovery-driven activism.

Having identified that there is a clear trajectory where individuals and collectives can obtain data to empower them, it is clear that this complex work can be supported. We see the emergence of '***data access & understanding services***', with entrepreneurs and activist enthusiasts:

- creating **tools to help people get their data back** from organisations, such as Tap My Data ('Tap my data', 2022), AccessMyInfo ('Access my info', 2022) and My Data Done Right ('Privacy salon & my data done right', 2022);
- creating **tools to help people understand their personal data**, such as Ethi (Jelly, 2021) and digipower.academy ('About digipower.academy', 2022); and
- helping collectives and journalists with **training, workshops and skills development** to take advantage of their data rights and **understand data ecosystems**, like Hestia.ai (Dehaye, 2019).

If such emergent endeavours can be supported and enabled to flourish, that could make HDR reformers using the discovery-driven activism approach more successful by ensuring that a lack of legal, technical or investigative skill does not become a barrier to any HDR reformer wanting to use this approach.

This approach shows that there is a role for independent actors and organisations to carry out discovery-driven activism - access requests, complaints, legal challenges, public campaigns and more. Discovery-driven activism can empower individuals and collectives to incrementally work towards building the world of better HDR that this thesis outlines.

1.4.3 Approach 2 to Improving HDR: Building the Human-centric Future

[TODO diagram FRAME AS DIAGRAM Building new designs (reaching into understanding, LITREF data enabled design and Human values) Delivering new structural capabilities. Enabling new individual and collective perspectives.]

The approach to HDR reform presented in this section focuses on the gaps in individual data interaction capability that exist today. The objective here is to design and build proofs of concept for **novel human-centric information systems that can deliver people new capabilities over their data**. In this approach, the focus is more introspective than Approach 1: it is about how the individual can improve their relationship with data in the context of their own digital life. The bulk of this section describes **specific design ideas** developed by myself and colleagues at BBC R&D during my 2020-2021 research internship on the Cornmarket project. As established in Insight 2, one of the most promising models for giving people a new and improved relationship with their data is to create a place where one's scattered (Abiteboul, André and Kaplan, 2015) personal data can be stored and aggregated in one place (Jones, 2011), and based on Insight 3, that place should offer individuals the means to use their life and ecosystem information as a material they can explore, examine

or repurpose as they see fit. This leads to the vision of a Personal Data Store [2.3.4] or into which one can unify the data from the different parts of your digital life (as depicted in the AllOfMe vision video ('AllofMe.com Teaser Clip', 2008)), and meet public demand for "control over your data" (Teevan, 2001; Hartman *et al.*, 2020). The Cornmarket R&D project therefore sought to develop a human-centred (ie. non-commercial) PDS proof of concept, as conceptualised in Figure X.

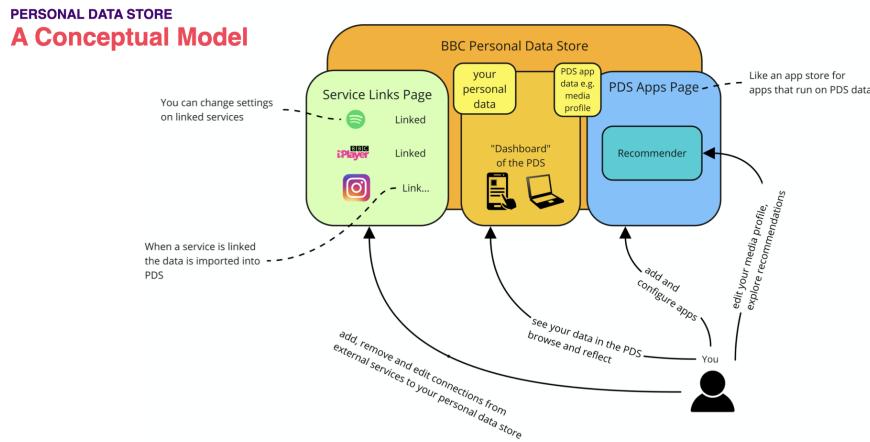


Figure 13: Figure X: A conceptual model for a Personal Data Store system that I developed for the BBC Cornmarket project

1.4.3.1 Designing a Life Interface: Transforming Data into Life Information The first challenge in designing such a system is to consider what data will be stored, and how it might be represented. At a high level, I identified a number of different types of data that a user might wish to store in a PDS:

Then, considering the earlier observation by myself (Bowyer, 2011) and proponents of temporal PIM systems [2.2.2] that time can be used as a unifying concept for personal information, I considered how we might represent occurrences within personal data as *happenings*, anchored against a particular point in time. In effect, this would focus on different properties of the data, much like Karger's *lenses* [Karger *et al.* (2005); #2.2.2-contextual].

In modelling data as life information using the concepts shown in Figure X [TODO link to Life Concepts diagram] and Figure X [TODO link to Happenings above], it is important to come up with as simple a model as possible, so that the *life interface* is not overwhelmingly complex and unmanageable. Over several iterations, I was able to reduce the modelling of life information in a PDS down

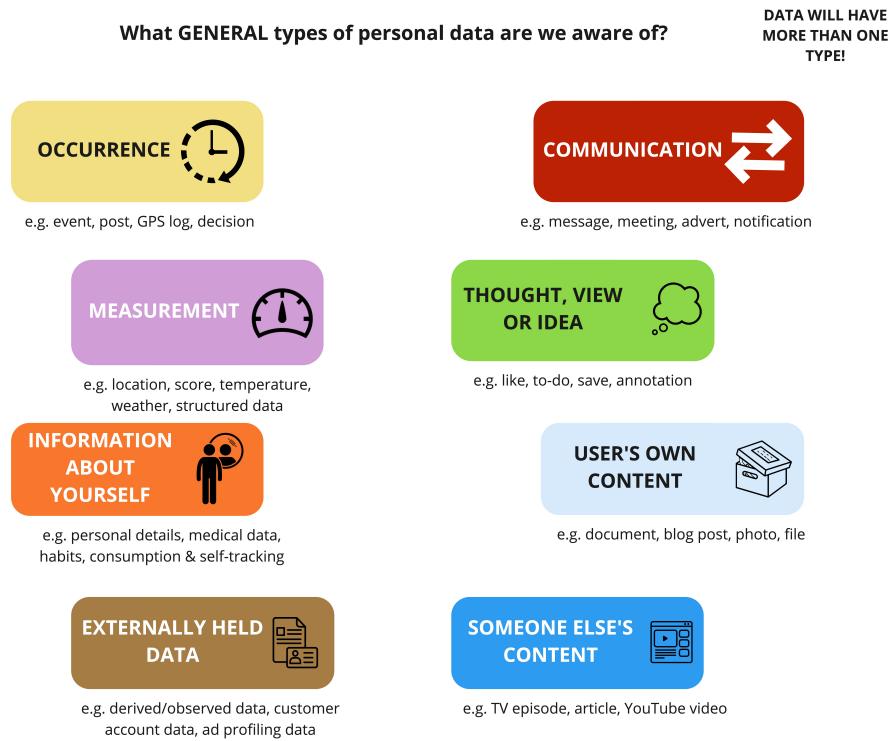


Figure 14: Figure X: High Level Data Types

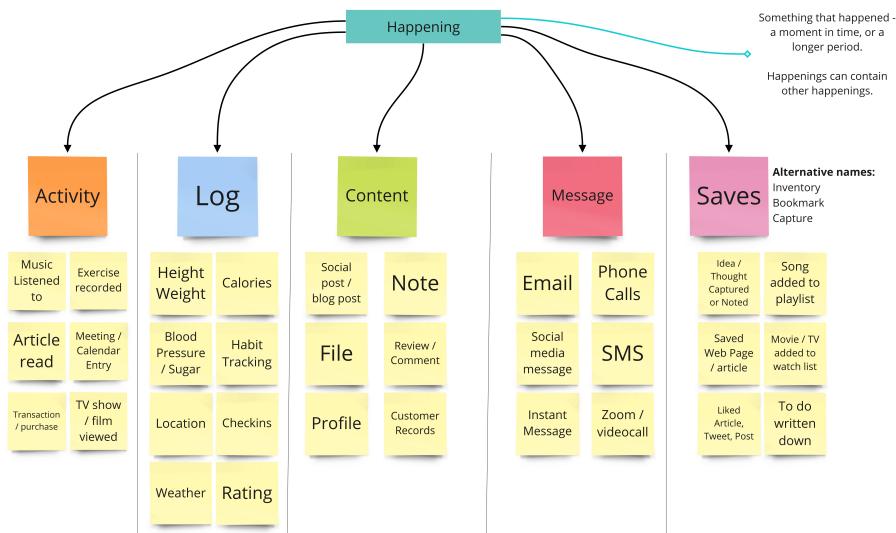


Figure 15: Figure X: Happenings

to the following four types - *activities*, *interactions*, *transactions* and *datapoints*, which could correspond to four views or lenses within the user interface:

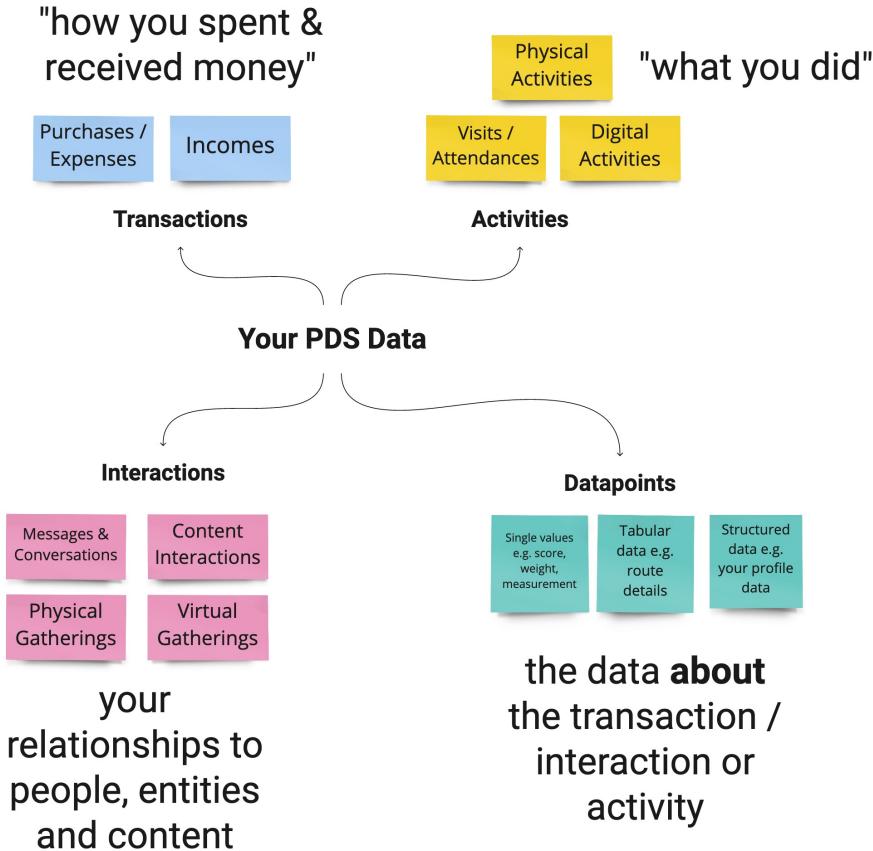


Figure 16: Figure X: A Simple Life Information Presentation Model for use in a PDS

Having decided upon models for the information within a PDS, it becomes much easier to design visual presentations of that information. Figure X shows a design mockup of some of the views of life information that one could offer within a PDS:

1.4.3.2 Life Partitioning What has been shown in Figure X is the beginnings of imagining what a *life interface* might look like, a single holistic interface covering all aspects of one's digital life.

Previous HCI design work [2.3.1] has focused on, in the first wave, the functionality of the machine; in the second wave, on the common tasks of a work

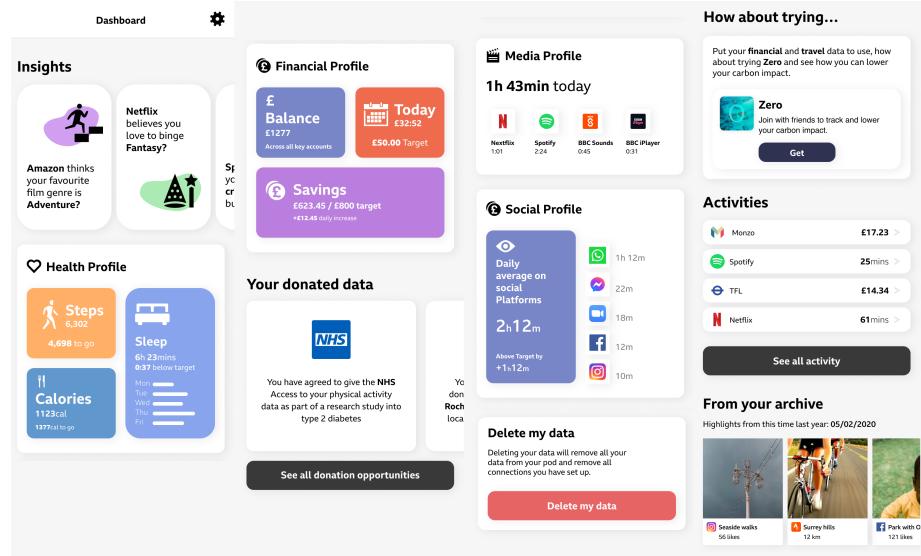


Figure 17: Figure X: Mockup of Life Information presented in a PDS interface, designed by Alex Ballantyne of BBC R&D

environment; and in the third wave, on classes of users and the commonalities of experiences in everyday life. But the challenge designers of life interfaces face is even more difficult. As Lindley noted, no single arrangement of information will suffice because in the same context two people may have different needs [Lindley *et al.* (2018); #2.2.2-contextual]. Because of the *subjectivity principles* [Bergman, Beyth-Marom and Nachmias (2003); #2.2.2-subjective], we are now faced with the need to build an interface that is suited to the individual, even though **every individual has different needs**. The design requirements for the interface are likely unique to the individual. Therefore, any life interface design must be able **support different mental models** - and as the life sketching exercise [Figure 10] in Case Study Two showed, people have many different ways of compartmentalising their life.

During a life interface design hack week at BBC R&D, colleague Jasmine Cox and I explored this problem. We considered that a key element of the interface design would need to be the versatility to *partition one's life in different ways*, and then to be able to assign different collections of life information to the different partitions. As a visual illustration of one possible mental model for life partitioning, consider Figure X, where, based on an idea I had had years earlier, I modified a Cluedo™ board to represent the ‘rooms of your digital life’:

We also imagined other mental models, such as partitioning your life according to parts of the body (heart for relationships, body for fitness, brain for current projects, etc.) or a landscape with forest, lake and buildings representing different



Figure 18: Figure X: Life Partitioning Analogy using a Cluedo™ board. Cluedo board design is a copyright of Hasbro, Inc., fair use applies.

things.

Whatever the visual metaphor engaged, the important thing is the functionality - being able to use these partitions to *filter* [2.1.4] your life information and focus on a particular *perspective* [Lansdale and Edmonds (1992); Krishnan and Jones (2005); 2.2.2]. Deciding to focus on a particular aspect of one's life is analogous to fixing a *conceptual anchor* [Teevan (2001); 2.2.2].

We then produced a mock app workflow for assigning different elements of your life (such as people, places or topics) to different partitions of your life, and viewing a visualisation of those different partitions in some subjectively meaningful form before filtering on a particular life partition and then being presented with a *timeline* [2.2.2] of life information pieces associated with that part of life. This is illustrated in Figure X, with artwork by Jasmine Cox:

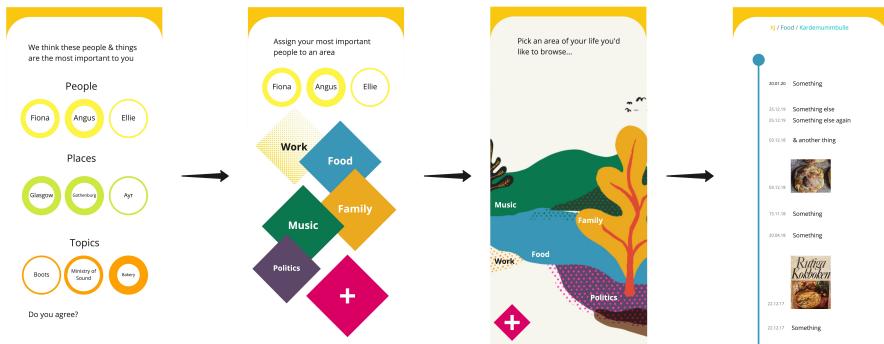


Figure 19: Figure X: Browsing By Areas of Life

1.4.3.3 Entity Extraction from Pooled Life Data

INSIGHT 11: Automating the Identification of Entities can enhance Machine Understanding and Unburden Life Interface Users.

Having identified the need to assign every piece of a user's life information to a particular partition (or multiple partitions) of their life, it quickly becomes apparent that this would be **too much work for the user** to do alone. Systems that use manual categorisation and tagging to classify information work best with a large userbase to contribute effort to the classification operation (Golder and Huberman, 2006). As part of the explorations of PDS approaches at BBC R&D I therefore also examined how this challenge might be addressed (considering also that having a lot of effort could be a deterrent to adoption [Objective 5]). I identified an approach that could help with this problem: If the *entities* (for example, a person, a place, an event or a topic) associated with a piece of data can be identified, then a lot of the **assignment of data to life partitions can be handled automatically**. For example, association with your office location would indicate that any data associated to that location is likely to relate to the 'work' part of your life, and this could be done automatically, reducing the effort for the PDS user. The process of **identifying entities within data**, known as *entity extraction* or *named entity recognition (NER)* is a well-established technique, which relies on the trained recognition of proper nouns and keywords combined with the statistical analysis of sentence grammar (Marshall, 2019). This technique is used extensively in text-mining products within the Content Analytics industry such as those produced by my former employer, OpenText ('What is text mining and content analytics?', 2022). However, in the context of a PDS, I propose that new techniques can be applied, making use of the data touchpoints into different parts of an individual's life to identify entities relevant to them personally (including, for example, names of friends or private projects that a standard NER solution would not detect). Data is full of references to entities that have personal relevance in your life. Finding these allows meaningful metadata to be attached to each datapoint. Figure X shows how a large number of entities could be detected from different parts of an individual's data once it has been imported into a PDS environment:

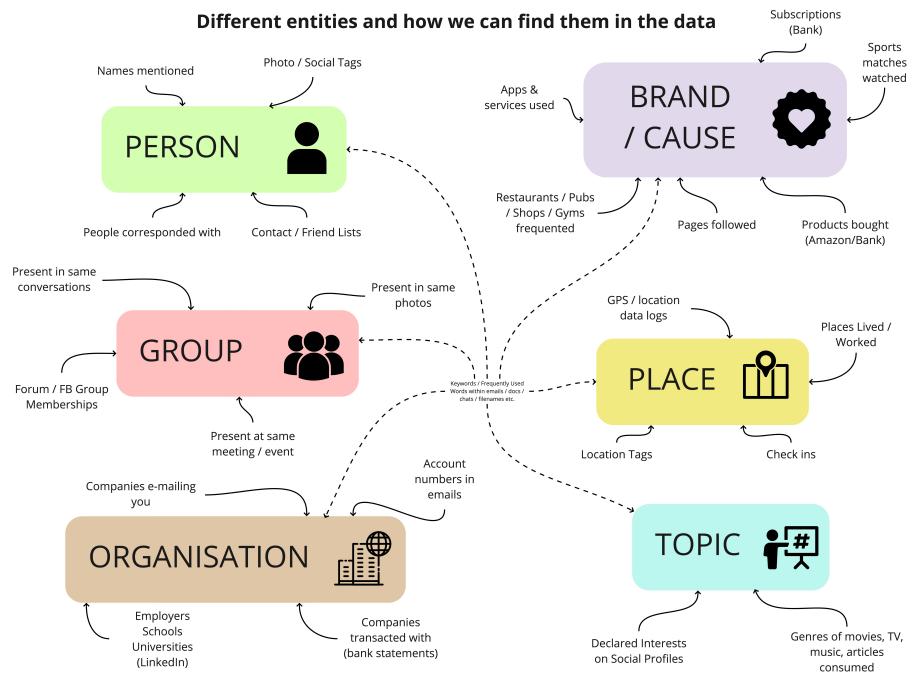


Figure 20: Figure X: Identifying entities in data

(continues...)

This sort of approach could be quite powerful in reducing the effort for life interface users. By scanning the data, the most prevalent entities could be identified, and the user need only assign the entities to different parts of their life, as illustrated in the first two frames of Figure X [REF to mockup]. This would then allow hundreds of associated data points which had been programmatically associated to that entity, to be assigned to the correct ‘bucket’ or life partition. While such an approach would not be perfect, and there would need to be some corrections made by the user, this is far preferable to them having to provide all the classifications and is likely to motivate greater engagement. I have observed in user experience design and consideration of productivity systems that users are more motivated to correct errors, than to fill in a blank page. Philosophically, we are moving here towards a *learning system*, a system that can be told when it is right and when it is wrong, and get better at classifying things correctly, analogous to the way an executive might train an assistant to anticipate his/her needs better, a sort of *digital life assistant* [ADD REF]. Bayesian classification techniques could also be used to help with the learning here (Authors, 2022b). . This approach is also useful for *ecosystem detection*, as outlined in Insight 4 - as identification of relationships with external entities is a key first step to mapping a user’s ecosystem.

Insight 11 offers a practical, theory-informed approach as to how we might start to build systems that are more able to understand the meaning of human information [Insight 8]. This is a technique that is already being used by corporations. For example, Facebook has developed an internal system called *word2vec*, shown in Figure X, through which it mines information about the world from the data that people generate through social media posts, messages and interactions. This information is exploited for commercial benefit. It is about time similar techniques were used for individual benefit and personal data empowerment.

Automatic ‘best guess’ attempts to arrange personal data in a PDS can go further than just association of related entities. In many cases, it is possible to programmatically determine the nature of a piece of data. In Figure X[examine a piece] is shown a detailed approach for identifying and classifying data automatically in a PDS, drawing on the previously identified data types in Figure X[types], the data attributes in Figure X[attributes], and the entities identified in Figure X[entities]. In Figure X[determining nature], a more abstracted approach is shown looking at how data could be classified according to the four different life information concepts in Figure X[4 ways PDS].

1.4.3.4 Designing Capabilities for the Manipulation of Life Information Moving beyond the initial challenge of classifying and arranging human information within a *life interface*, a core consideration in designing such an interface is to consider what capabilities the user might be given over the pieces

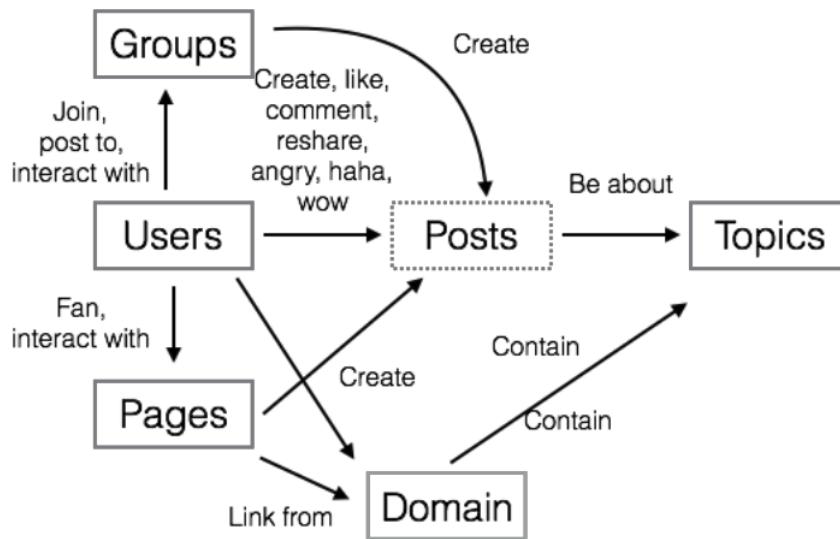


Figure 21: Figure X: A design for Facebook’s world2vec model, semantically modelling data from social media posts on Facebook (Peysakhovic and Wu, 2020)

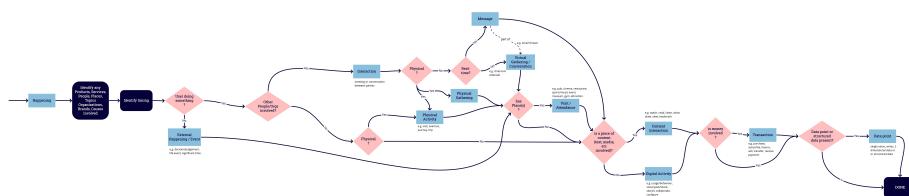


Figure 22: Figure X: Determining the nature of a piece of data

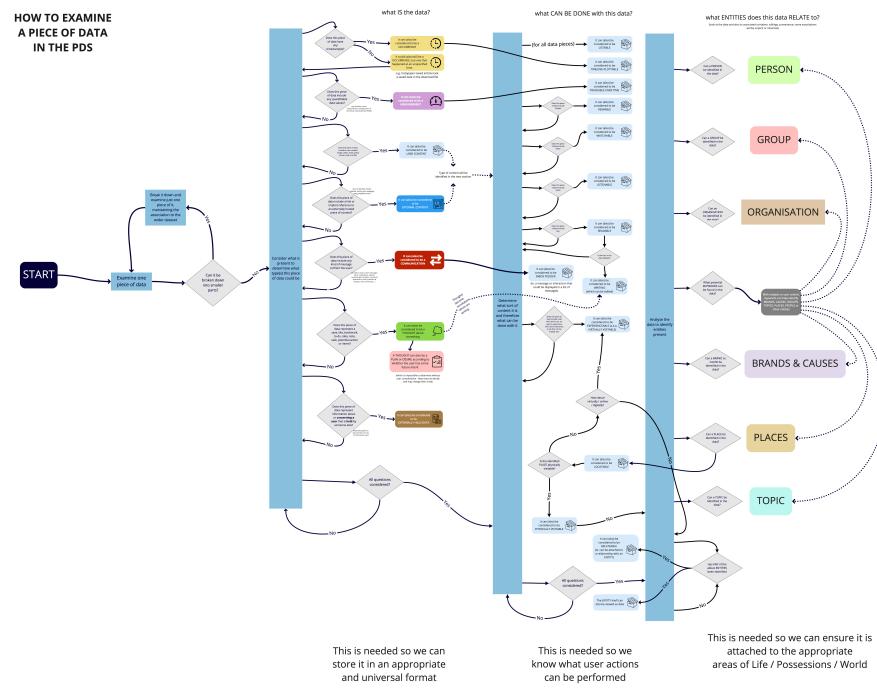


Figure 23: Figure X: Identifying the attributes of data

of life information represented in the interface. It is through the provision of a wide range of useful operations upon information items that the information will start to feel like *a material* as described in Insight 3 and deliver the *new capabilities* needed for success Insight 7.

There are two ways to consider this problem - one starting with the data, and another starting with the user.

Starting with the data, and considering the different types of data that a user might be able to gather from across their digital life, I identified that each datapoint or dataset is likely to have properties which determine *what can be done with it*. A number of possibilities are shown in Figure X:

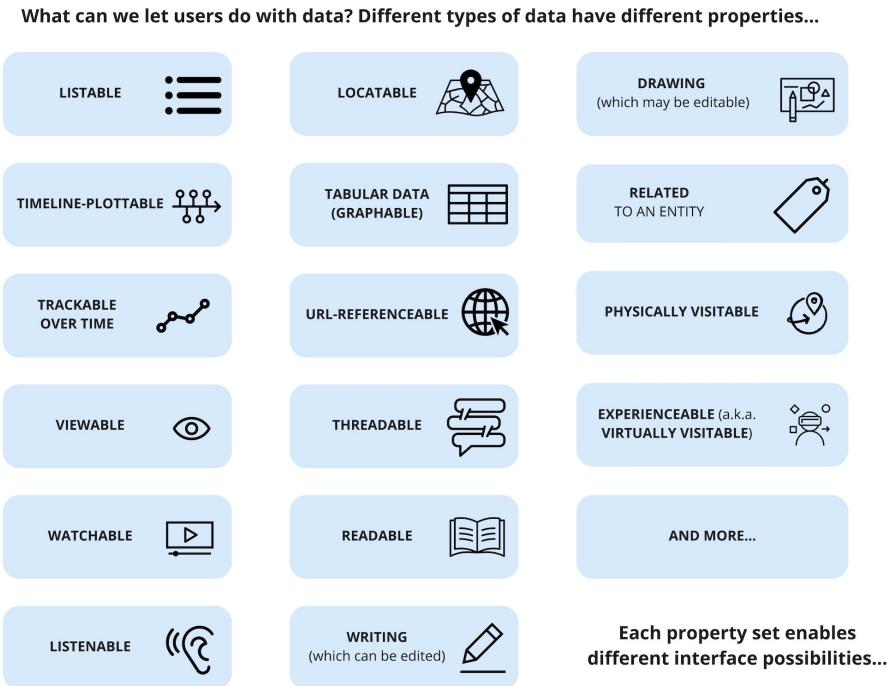


Figure 24: Figure X: Attributes of Data

These attributes could then be used to determine which datapoints would be eligible for inclusion in different views (or *lenses* (Karger *et al.*, 2005)) of the data, such as tables, graphs, maps, lists, media thumbnail views, etc.

If we consider capabilities from the user's perspective, it is useful to recognise that visualisations alone (such as those in [REF FigureX-dashboard]) are not sufficient to meet the need for either interrogability or malleability [7.3.2]. To address the need for malleability, we need to determine what actions an individual might want to perform on a piece of life information. To explore this design question,

we can draw on multiple sources of inspiration:

- actions identified in prior PIM research & design work [2.2.2] such as arranging, browsing, keeping, and associating;
- actions needed for SI [2.2.3] such as combining, history viewing, interpreting patterns;
- actions needed for effective access [2.1.4] such as filtering, zooming, getting detail on demand; and
- actions inspired by the eight lenses of how people think of data [Table 15] such as summarising, deriving or remixing.

Having developed a large set of initial design ideas for possible actions, I worked with my BBC colleague Chris Gameson to distil this information design thinking into a 12 *data card* designs that could encapsulate some of the most common *actions a user might want to perform* upon personal data in a PDS. These are shown in Figure X. These had a similar purpose to the cards in my pilot study [1.3.1] and Case Study One [4.2.1], acting as ‘*things to think with*’ (Bowyer *et al.*, 2018).

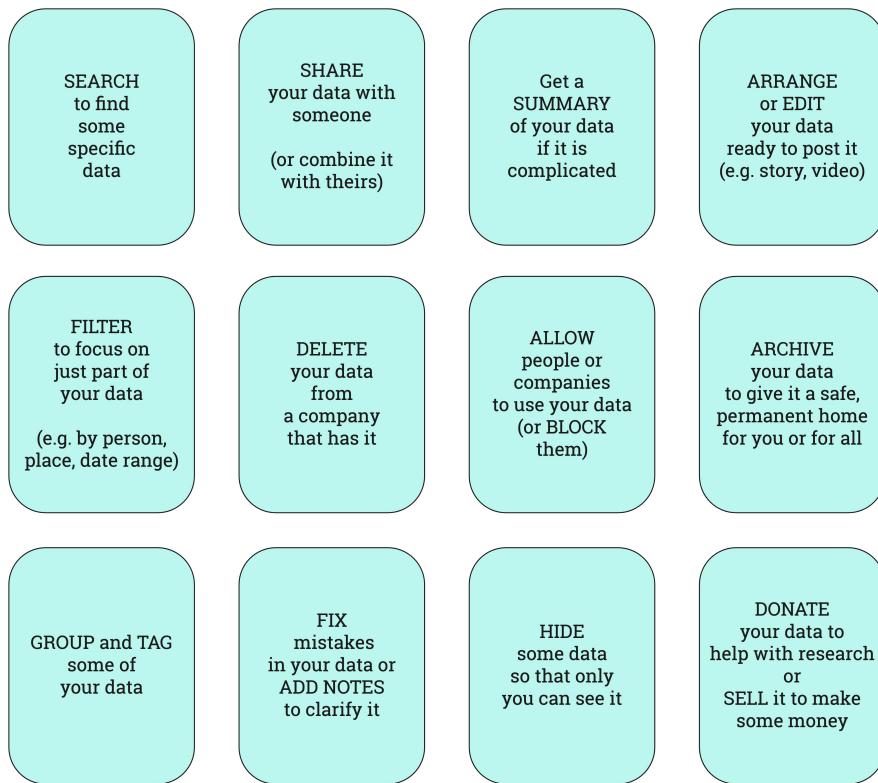


Figure 25: Figure X: Data actions

As identified in chapter 5 [5.3.3; 5.4.2.3] and explored further in 6.1.3, being able to ask questions of data is important to individuals. Addressing this need for interrogability in a similar approach, we produced 10 data card designs to encapsulate some of the different common questions that users might want to ask of their data, as shown in Figure X.

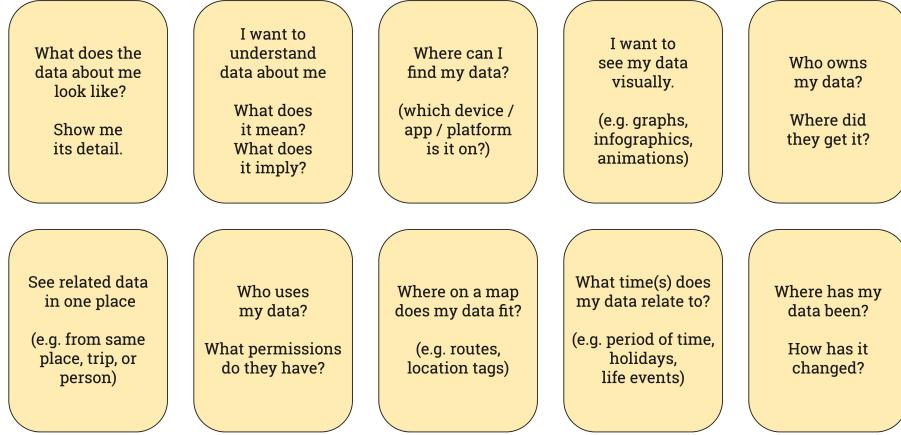


Figure 26: Figure X: Data questions

These actions and questions act as a design building blocks, a precursor that can inform the design of features for a life interface that could provide the user with new capabilities to interrogate and manipulate the data in a much more interactive way.

Once information has been correctly identified semantically (using approaches such as those in Figure X[naturedetermining] and Figure X[identifingattributes]), and associated to entities, it then becomes possible to enrich users' search and browse capabilities by enabling the use of *facets* {Televiciute (2020)} to allow users to precisely target a set of data points or a particular data item. An example of how this could be done is shown in Figure X, another design artefact from my work at BBC R&D. This is not a user interface mockup but rather a three-stage strategy, with examples, for the different types of facets that one could offer a user to help them explore their data in powerful ways:

1.4.3.5 Digital Self Curation & Inclusive Data Flows A final element of Approach 2, which, unlike the previous subsections, has a broader focus than just the creation of an effective life interface because it has sociotechnical/business process design implications too, aims to address the problem identified in 7.3.4 - the intractable data self. Drawing upon the philosophy behind VRM [2.3.4], we can imagine (as implied by both Case Studies [4.3.3.4; 5.5.2]) that were individuals able to **create or contribute to the digital representation of**

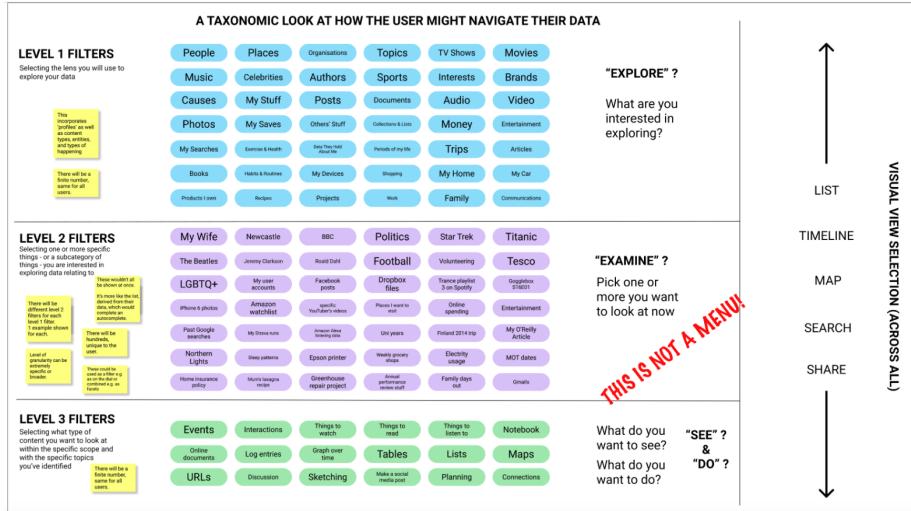


Figure 27: Figure X: Taxonomies For Navigation

themselves in data, that this would be more accurate. Furthermore, this would provide an opportunity to address the identified HDR want of *involvement* [6.2.3]. In the BBC Cornmarket project, this was explored through the concept of recommendations. Functional interfaces were built which allowed users to construct and edit a representation of their own musical/media tastes known as a '*media profile*', using personal data such as viewing/listening history imported from multiple sources as inputs. This profile would then be used to recommend new content, which would be likely to be better suited to the individual's tastes since it would be based on a more accurate representation of self. This aspect of the Cornmarket project (Orphanides, 2021) is not only a practical manifestation of the ideas of pull-centric/VRM approaches, but is also illustrative of a new way of thinking about personal data, where it can be **directly generated (or edited/fixed) by the individual**, rather than just asserted on inferred by the service provider without verification. Such thinking could be expanded further to accommodate the idea that data is not static. Once we make the mindset shift from data as a static bundle of facts to **data as an ever-changing flow or stream** [2.2.5; 4.3.3.3; 6.2.2], it becomes much easier to think about how processes and interfaces might be updated to give users a direct role in the evolution and shaping of that data. This also makes it easier to maintain knowledge of data's provenance [Insight 5]. Such '**inclusive data flows**' and processes could yield benefits for users in terms of increased ecosystem negotiability Objective 4, but also for businesses, as will be explored in Approach 4.

The future-centric designs and insights⁸ I have presented as part of Approach 2

⁸The research and design work and insights from my time with BBC R&D's Cornmarket project are documented in more detail in the external publications detailed in 1.3.4).

show that while developing interfaces and processes for a better HDR future is a challenging design problem (due to the variety in individuals' mental models and the breadth of data across an individual's digital life), it is certainly possible to tackle these challenges. It is likely that the reasons that few have ventured into this space are largely due to the lack of an obvious business model or incentive rather than any question of the merits of these approaches. It is possible to **empower individuals with new capabilities** over the life and ecosystem information encoded within their data, and even to design **new data interaction approaches** that give users an ongoing role in the curation of that data, and HDR reformers can help make a world of better HDR a reality by **building, or investing, in the building of these more life-centric technologies and processes**.

1.4.4 Approach 3 to Improving HDR: Defending Autonomy and Nurturing the Information Landscape

[TODO diagram: taking external *protective* action as collectives, surfacing, challenging, pushing for better enforcement of existing regulation]

While Approach 2 focuses on designing a better future, and Approach 1 focuses on identifying and campaigning for positive changes in provider practices, there is a need for a more direct approach to tackle an urgent and evident problem head on: the active diminishing of users' agency [7.3.4.2]. In pursuit of mass market appeal and cost reduction, there is a trend towards the *dumbing down* of technology, which invariably leads to fewer features and less agency. Increasingly, companies encourage users to think of technology as a black box, which you cannot and should not look inside, as illustrated in Figure X:



Figure 28: Figure X: The 'Black Box' view that many modern technology providers present

This disempowering trend is seen across hardware, software and service industries, where providers attempt to **restrict the ways in which technology can be used** even after it has been purchased - cars that can only be serviced by official dealers, printers that can only use officially branded ink cartridges, smartphones

that cannot be repaired without special tools and warranty invalidation, media files that cannot be copied or shared, operating systems that block or discourage certain software from running, applications that forbid screenshots. Apple, for example, encourages users to consider technology as ‘magical’, rather than as tools to be harnessed and understood. This approach is highly problematic for user agency:

“Magical design prioritises pleasing and surprising a passive user who can only use the solution as authorised.” – Cristiano Storni (Storni, 2014)

In order to challenge this trend towards disempowered, compliant users and **protect individual autonomy** and our ability to make sound judgements and hold organisations to account, direct activism and grassroots civic action is called for. This forms the core objective of this approach: to **identify ways in which agency is being diminished**, and to find ways to **circumvent the new barriers** in order to maintain existing levels of capability.

It is bad enough when a particular technology blocks you from doing what you want to do, but the trend is even worse when viewed at a broader level: As the digipower investigation highlighted [Insight 6], the information landscape that we inhabit as digital citizens is being manipulated to **change what we see, what we understand of the world, and what we can do, in order to manipulate our behaviour**, which can have harmful effects on democracy, agency and individual autonomy. Therefore a further goal here is to protect our ability to **experience unbiased and unmanipulated information landscapes**. A recent US Congress investigation shows how the big five use their dominant positions as device manufacturers, search engines or major Internet platforms to promote their own interests (Sisco, 2022). In a paper looking at the growing manipulation of search results, Shah and Bender consider how this landscape could be protected and accountability preserved. They describe their approach as establishing **‘guardrails for the status quo’**, and I find this phrase to be an excellent summary of what Approach 3 is trying to do (Shah and Bender, 2022). We can also take this idea further through the idea of **adversarial design** (DiSalvo, 2012) which advocated **using design processes to challenge the status quo**. To understand how to gain traction here and carry out **‘design after design’** (Storni, 2014), it is important to understand the concept of **seams**, how they can be exploited, and the risks of losing them. This concept is explained in Insight 12.

INSIGHT 12: The ‘Seams’ of Digital Services need to be Identified, Exploited and Protected.

As we have already identified [7.3.4.1] product design (be it hardware or software) is political. Designers pass some power to the user through their design, but also, users should be able to take some power on their own terms. This is the case made by Cristiano Storni in his 2014 paper on ‘*the politics of seams*’, Cristiano Storni identified the idea of ‘**empowerment-in-use**’ which advocates the idea that people need to appropriate their technologies to different users than the designers may not have foreseen (Storni, 2014), something that current black box, limit-what-the-user-can-do thinking blocks. Central to this capability is the concept of **seams** - those **exposed areas which the user is free to change**. This concept was proposed by Mark Weiser and developed by Chalmers and others (Weiser, 1994; Weiser and Brown, 1997; Chalmers, MacColl and Bell, 2003). The changes such as closures of APIs or removal of ports as described in 7.3.4.2 can be seen as the removal of seams. As Storni highlights, the availability of design seams is a **critical determiner of user power**, and companies gain power and reduce agency when they remove those seams. It follows therefore that by identifying, exploiting and protecting the seams of digital services and devices, user autonomy and the viability of data-unification efforts can be protected.

An unseen battle is for the free flow of information is underway at the seams of today’s digital products. Hackers, civic activists and makers seek to repurpose and exploit the edges of products for their own means, while digital service providers and platforms try to block such activity. For example:

- A successful tool called Findings allowed people to clip and share their favourite quotes from Kindle books. Amazon blocked and banned this tool, and the company shut down (Owen, 2012; Maldre, 2012).
- Louis Barclay created a tool called Unfollow Everything, which allowed Facebook users to automatically unfollow all friends and pages, in order to give them greater control of their News Feed reading experience and avoid being manipulated into reading more than they want to. He was banned for life from Facebook and threatened with legal action should he ever build any tools that manipulated the Facebook experience (Barclay, 2021).
- Various activist groups have for several years been fighting to give individuals the legal **right to repair** their own products (Miller, 2021), which has often been blocked through planned obsolescence, inaccessible seams or restricting access to parts. The problem has been described as **device tenancy**, the idea that our relationship with our technology products is more like a tenant, where a landlord retains overall control and permits us to perform certain activities (Tufekci, 2019). New laws have been introduced in the EU (Tett, 2022), forcing companies to support individuals to repair their devices. Apple has subsequently released self-service repair kits, though these themselves force parts to be paired with particular phones, limiting the utility of self-repair (Moore-Colyer, 2022).

INSIGHT 12: The ‘Seams’ of Digital Services need to be Identified, Exploited and Protected.

- As detailed in my co-authored paper with Louis Goffe and colleagues (Goffe *et al.*, 2021, 2022), ***web extensions*** and ***web augmentation*** offer a powerful technique for modifying web experiences and repurposing user interfaces. This is because once a website is loaded into your browser, it is no longer under the control of the remote site, and by creating a *web extension* to run code within your local web browser that loaded website can be edited, *scraped* (P., 2021), or otherwise repurposed. This has been successfully used to stop clickbait, dispute fake news, combat addiction, filter explicit words and more. However, in order to re-assert control over these customisations, Google has announced changes to the way Chrome extensions will work, which will ‘*stifle innovation*’ and limit what developers can do within the web extension (Miagkov, Gillula and Cyphers, 2019).
- An example from 2016 shows how seams can be exploited to obtain information and increase transparency. By brute force querying of a Facebook API, researchers were able to identify a complete list of 282,000 interests on Facebook and identify the relative popularity of each interest. (Havlak and Abelson, 2016).
- A number of HDR reformers, myself included, had identified a new seam for subverting some of Facebook’s control over how its content is consumed [7.3.4.2]: *accessibility tags* or *ARIA tags* (Authors, 2022a). These are specially marked-up tags which are used by screenreaders to display or read content in a more accessible way for partially-sighted or blind people. Because these show page content in a standard format (whereas the HTML of most web pages varies widely and often changes), they present a reliable way to more easily scrape content from the loaded web page within a web extension. In experiments at Open Lab we successfully scraped posts from friends’ feeds (which Facebook do not make available anywhere except the News Feed) so that they could be consumed separately in a more human-focused user experience. This technique has been used successfully to monitor Facebook ads by NYU’s Ad Observatory (Watzman, 2021). In 2021, Facebook was found to have deliberately obfuscated content within ARIA tags to prevent such investigations, resulting in **visually impaired users being unable to differentiate ads from posts and hearing junk characters read aloud**. This can be taken as an adversarial stance against researchers, activists and HDR reformers, and shows that companies like Facebook will go to extraordinary lengths to assert their dominance and reduce user agency (Faife, 2021).

INSIGHT 12: The ‘Seams’ of Digital Services need to be Identified, Exploited and Protected.

- One reason why many companies and services have produced apps is because these are much more locked down and controllable than the web browser environment; there are fewer seams. However, adopting the same philosophy as using web extensions to modify web-based experiences, and drawing on data flow auditing technologies like TrackerControl [7.4.3] researchers at Oxford University have now developed techniques by which mobile apps can be reverse engineered and modified to change user experiences to better meet users’ needs, offering the promise of a ‘**right to fair programs**’ (Kollnig, Datta and Kleek, 2021).

These examples make it quite clear that Storni was right: product seams are the place where control can be asserted or regained. They are the setting for the battle for the freedom and integrity of today’s information landscape, and it is important for HDR reform that this space is specifically targeted. The role of the HDR reformer here is twofold:

- 1) To **surface information injustices** and especially the closures of seams.
- 2) To **push the seams to gain transparency and re-assert control**, using both and campaigns for policy change.

In this context, the work of whistleblowers such as Frances Augen (Horwitz *et al.*, 2021) and Edward Snowden (Macaskill *et al.*, 2013) is particularly validated and important. Whistleblowers can expose internal procedure changes that affect the information landscape’s integrity that are not otherwise visible. In order to hold online platforms to account, the public must be **aware** and **able to attribute** any restriction in freedom or information access to the correct source, in other words to know that the information or functionality is being modified or restricted. Seams should be much more in the public consciousness than they are.

[TODO make this an inset box not a table]

The reason that seams are so important, is that because they a point of friction, a point of resistance to the idea that *data should be separable from services* [Objective 5]. One of the goals of HDR, and particularly this approach, is to wrest control of data away from service providers, so that it can be truly free-flowing, as I envision in (Bowyer, 2018). In this approach, I have identified some of the key avenues for progressing this goal - through web extensions, accessibility tags, API exploitation and reverse engineering. But in the face of such extreme power, it is clear that such grassroots advocacy will not be enough. For example, web extensions are a powerful approach but are limited in that only those ‘power users’ who install the extension will gain the benefits. They do not directly help the layperson with their diminishing agency. Therefore, as outlined in [5.5.1], HDR reformers must also work to educate and persuade policymakers of the need for change. As an example of this, in my Case Study Two paper (Bowyer, Holt, *et al.*, 2022) I called for better guidelines from the

EU towards data holders. The European Data Protection Board launched a public consultation on a new set of guidelines they had constructed around GDPR (European Data Protection Board, 2022) and I contributed a detailed set of recommendations to that consultation, emphasising some of the important insights I have identified in this thesis, including a focus on human information, delivering ongoing understanding, establishing standards, and viewing data as an ongoing flow over time (Bowyer, 2022). This sort of engagement by HDR reformers is critical to help shape future policies and guidance, contributing expertise to help the civil servants making those rules. As HDR reformers we must continue to push for better regulation - as it really is the only force that can significantly change data-holding organisations' practices. Better regulation is needed not only to help individual user autonomy, but to combat online extremism (Arthur, 2017). In this context it is important to note that the European Union is, at the time of writing in July 2022, developing a wide range of new laws that could help to improve the agency of individuals and the integrity of the information landscape:

- The Data Act (Board, 2022)
- The Data Governance Act (Commission, 2022a)
- The Digital Markets Act (Commission, 2022b)
- The Digital Services Act (Commission, 2022c)
- The Artificial Intelligence Act (Brakel, Uuk and Boine, 2022)

It is too early to evaluate exactly what these laws will change or what their impact will be, but HDR reformers should observe carefully: they will almost certainly play as pivotal a role in protecting the information landscape as the GDPR has in opening up data access to individuals.

In this approach, I have highlighted how important it is for HDR reformers to **seize and harness the powers we are given, and fight to hold onto them**. Groups of HDR reformers can combine development skills, innovation and disruptive design approaches to find and publicise new ways to circumvent providers' efforts to control and limit users' agency, and policymakers can prevent further erosion of individual agency through legislation and enforcement over the information landscape. As the popularised adage based on the words of John Philpot Curran and Thomas Jefferson goes, "*The price of freedom is eternal vigilance*".

1.4.5 Approach 4 to Improving HDR: Winning Hearts and Minds: Teaching, Championing and Selling the Vision

[TODO diagram: Structural work in upper right - standards. Selling work in top level - show value to individuals. Selling work in top level - show value to organisations. Structural work in bottom right - systems. Individual work in top left - empower and educate individuals. all leading to new action of individuals in top right]

- MAIN POINT: That the nature of pursuing Human Data Relations causes

for a radical reconfiguration of today's data world. We need new systems (which means not only there need to be business drivers for those systems but also that existing organisations must choose or be compelled to invest in them), and people need to understand, use and see value in those systems. Therefore, there needs to be specific investment:

- SUBPOINT in Education, and Data Literacy
- SUBPOINT in Systems Building (just see above)
- SUBPOINT in standards, information uniting the diaspora
- SUBPOINT in Researching New Business Models and Demonstrating Value of transparency and human centricity

INSIGHT 13: It is possible to demonstrate business benefits of Transparency and Human-centricity

As outlined in 7.3.5 and in this section, it is essential that work is done to persuade data-holding organisations of the benefits of moving towards the new paradigms outlined in this thesis. The following avenues for possible future research and advocacy toward data holding organisations have been identified:

- **Trust & Reputation:** In line with the third aspect of HDR [7.2] as well as the recommendations in [4.3.4], [4.4.1], [5.5.2] and [6.2.1], displaying a more inclusive, open and supportive attitude to data handling could strengthen the service relationship and increase customer loyalty and trust. Organisations that are seen to have good human data relations are preferred.
- **Consent:** In the wake of the GDPR, ensuring consent is becoming an increasing concern to organisations, and the risks of legal consequences for mistakes are high. It makes sense that a more dynamic [Bowyer *et al.* (2018); 4.4.1; 5.5.2; 6.2.2] consent approach that involves individuals [6.2.3] and keeps them in the loop, will enable individuals to speak up much earlier and express consent wishes that might otherwise go undetected.
- **Accuracy:** The best placed person to spot errors in data's accuracy or fairness is the individual about whom the data is concerned. Therefore, increasing their involvement is likely to improve the quality of the data, especially if additional data is contributed or curated by the service user [4.3.3.4, 6.2.3]
- **Liability:** In an increasingly litigious society, storage of personal data, especially health or financial data, is a significant liability for businesses, especially if something goes wrong. Investment in human-centered personal ecosystems would outsource the storage of sensitive data to data trusts or PDS providers, reducing liability for the service business. By ensuring that data is accessed only in ways that are centralised outside of the business and remaining in the user's control – such as PDS company digi.me's Private Sharing model (digi.me, 2019) – organisations can ensure that have negligible risk of mishandling customer data.

INSIGHT 13: It is possible to demonstrate business benefits of Transparency and Human-centricity

- **Better Customer Targeting** The most radical, but perhaps the most persuasive, business model relating to better HDR, is the Vendor Relationship Management approach [2.3.4], where individuals express their own service or product desires explicitly, which vendors then respond to. This turns traditional models inside out, and would empower users more, but due to the inherently improved accuracy of a self-declared interest, might also give businesses a greater confidence that their investment in converting those customers to a sale would be worthwhile. It is important to remember that the current drive towards collecting more data that drives the platformisation trend is in order to improve ad targeting, so that businesses can get a better return on their investment. A VRM approach, or any other approach where the individual contributes improved data to their data self, is in line with that current business objective.

[TODO make this an inset box, not a table]

- SUBPOINT in supporting Data Understanding Industry. empowering individuals as investigators. Tools to map their own ecosystems and unite their own personal data diaspora.
- ENDING: that this is not just a technical problem, and not just a case of building new things. It's about beginning and catalysing a cycle of constant feedback, of data enabled design and action research / iterative software and business model development - finding what works, championing it, selling it.

1.5 Thesis Conclusion

- frame this chapter as a call to arms
- emphasise the practical nature of this chapter
- recap HDR field as contribution
- recap the findings of the thesis and the map of the landscape
- recap the obstacles, insights and opportunities
- reiterate the answer to the main RQ
- highlight future value/societal implications of the work
- we have defined HDR. who the recursive public are (roles). what they might do... what the field is... etc
- [TODO add stuff from here]

Bibliography

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 digipower.academy’ (2022). Available at: <https://digipower.academy/about>.
- 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. and Mynatt, E. D. (2000) *Charting Past, Present, and Future Research in Ubiquitous Computing*. 1, pp. 29–58. Available at: <https://www.cc.gatech.edu/fce/pubs/abowd-mynatt-tochi-millenium.pdf>.
- ‘Access my info’ (2022). Available at: <https://accessmyinfo.ca/>.
- Allen, D. (2015) *Getting things done: The art of stress-free productivity*. Penguin.
- ‘AllofMe.com Teaser Clip’ (2008). YouTube. Available at: <https://www.youtube.com/watch?v=JWyqt4WL6xE> (Accessed: 21 March 2021).
- Ananya (2020) ‘Java: Pass by value or pass by reference’, *Medium*. Available at: <https://medium.com/swlh/java-passing-by-value-or-passing-by-reference-c75e312069ed>.
- Apple (2022) ‘About app activity report’. Available at: <https://support.apple.com/en-us/HT212958>.
- Arthur, C. (2017) ‘Theresa may wants to tackle online extremism. Here’s how to do it’, *The Guardian*. Available at: <https://www.theguardian.com/commentisfree/2017/may/26/theresa-may-online-extremism-google-facebook-twitter-manchester>.
- Authors, V. (2022a) ‘ARIA accessibility tags’, *Mozilla Developer Network*. Available at: <https://developer.mozilla.org/en-US/docs/Web/Accessibility/ARIA>.
- Authors, V. (2022b) ‘Naive bayes classifiers’, *GeeksForGeeks*. Available at: <https://www.geeksforgeeks.org/naive-bayes-classifiers/>.
- Bakshy, E., Messing, S. and Adamic, L. A. (2015) ‘Exposure to ideologically diverse news and opinion on facebook’, *Science*. American Association for the Advancement of Science, 348, pp. 1130–1132. doi: 10.1126/SCIENCE.AAA1160/SUPPL_FILE/PAPV2.PDF.
- Bansal, A. (2018) ‘An introduction to SOLID, tim bernes-lee’s new, re-decentralized web’, *FreeCodeCamp*. Available at: <https://www.freecodecamp.org/news/an-introduction-to-solid-tim-berners-lees-new-re-decentralized-web-25d6b78c523b/>.
- Barclay, L. (2021) ‘Facebook banned me for life because i help people use it less’, *Slate*.
- Bentham, J. and Bozovic, M. (2011) *The panopticon writings*. Verso Books (Radical thinkers). Available at: <https://books.google.co.uk/books?id=VbpvDwAAQBAJ>.
- 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.
- Berners-Lee, T. (2022) ‘Solid: Sir tim bernes-lee’s vision of a vibrant web for all’. Inrupt. Available at: <https://inrupt.com/solid/>.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) ‘The Semantic Web’, *Scientific*

- American*, 284(5), pp. 34–43. Available at: <https://jstor.org/stable/10.2307/26059207>.
- ‘Bits of freedom: Our focus’ (2000). Available at: <https://www.bitsoffreedom.nl/english/>.
- Board, E. D. P. (2022) ‘The EU’s data act: Data protection must prevail to empower data subjects’. Available at: https://edpb.europa.eu/news/news/2022/eus-data-act-data-protection-must-prevail-empower-data-subjects_en.
- Bogers, S. et al. (2016) ‘Connected baby bottle’, pp. 301–311. doi: 10.1145/2901790.2901855.
- Bowcott, O. and Hern, A. (2018) ‘Facebook and cambridge analytica face class action lawsuit’, *The Guardian*. Available at: <https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit>.
- 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>.
- Bowyer, A. et al. (2018) ‘Understanding the Family Perspective on the Storage, Sharing and Handling of Family Civic Data’, in *Conference on human factors in computing systems - proceedings*. New York, New York, USA: ACM Press, pp. 1–13. doi: 10.1145/3173574.3173710.
- Bowyer, A. (2021) ‘Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration’. Available at: <https://eprints.ncl.ac.uk/273832#>.
- Bowyer, A., Pidoux, J., et al. (2022) *Digipower technical reports: Auditing the data economy through personal data access*. doi: 10.5281/zenodo.6554177.
- Bowyer, A., Holt, J., et al. (2022) ‘Human-GDPR interaction : Practical experiences of accessing personal data’, *CHI ’22*.
- Bowyer, A. (2022) ‘Response to guidelines 01/2022 on data subject rights - right of access version 1.0’. Available at: https://edpb.europa.eu/sites/default/files/webform/public_consultation_reply/Response%20to%20Guidelinesv1.0-AlexBowyer-March2022.pdf.
- Brakel, M., Uuk, R. and Boine, C. (2022) ‘The artificial intelligence act’. Future of Life Institute. Available at: <https://artificialintelligenceact.eu/>.
- Braw, E. (2022) ‘How corporate boycotts could backfire’, *Foreign Policy*. Available at: <https://foreignpolicy.com/2022/03/28/russia-sanctions-ukraine-corporate-boycotts-could-backfire/>.
- Brest, P. (2010) ‘The Power of Theories of Change’, *Stanford Social Innovation Review*, 8(2), pp. 47–51.
- 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.
- Chalmers, M., MacColl, I. and Bell, M. (2003) ‘Seamful design: Showing the seams in wearable computing’, *IEE Eurowearable*.
- CitizenMe (2021) ‘Become a Citizen and unlock the value of your data’. Available at: <https://www.citizenme.com/for-citizens/> (Accessed: 23 August 2021).

- Cockton, G. (2004) ‘Value-centred HCI’, *NordiCHI*.
- Colbert, M. (2022) ‘Labour members left in the dark over data breach’, *Byline Times*. Available at: <https://bylinetimes.com/2022/06/06/labour-members-left-in-dark-over-data-breach/>.
- Commission, E. (2022a) ‘Data governance act explained | shaping europe’s digital future’. Available at: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>.
- Commission, E. (2022b) ‘The digital markets act: Ensuring fair and open digital markets’. Available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.
- Commission, E. (2022c) ‘The digital services act: Ensuring a safe and accountable online environment’. Available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.
- Copeland, E. (2015) *Small Pieces Loosely Joined: How smarter use of technology and data can deliver real reform of local government*. Available at: www.policyexchange.org.uk https://www.policyexchange.org.uk/publication/small-pieces-loosely-joined-how-smarter-use-of-technology-and-data-can-deliver-real-reform-of-local-government/.
- Cornford, J., Baines, S. and Wilson, R. (2013) ‘Representing the family: how does the state ‘think family’?’, *Policy & Politics*, 41(1), pp. 1–19. doi: 10.1332/030557312X645838.
- Crabtree, J. (2007) ‘Civic hacking: A new agenda for e-democracy’, *openDemocracy*. Available at: https://www.opendemocracy.net/en/civic_hacking_a_new_agenda_for_e_democracy/.
- 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.
- Davies, S. (1990) ‘Privacy international: About us’. Available at: <https://privacyinternational.org/about>.
- Deckers, E. (2018) ‘Data-enabled design’, *UXDX*. Available at: <https://uxdx.com/blog/data-enabled-design/>.
- Dehaye, P.-O. (2017) ‘Facebook forced to disclose more information about its ad targeting’, *Medium*. Available at: <https://medium.com/personaldata-io/facebook-forced-to-disclose-more-information-about-its-ad-targeting-7e6c0127722>.
- Dehaye, P.-O. (2019) ‘Hestia.ai: About us’. Available at: <https://hestia.ai/en/about/>.
- digi.me (2019) ‘Digi.me private sharing: See how you can do more with your personal data’. YouTube video; YouTube. Available at: https://www.youtube.com/watch?v=pGcnK_KraXs.
- ‘Digi.me’ (no date). Available at: <https://digi.me/> (Accessed: 23 August 2021).
- DiSalvo, C. (2012) *Adversarial Design*. MIT Press (Design thinking, design theory). doi: 10.7551/mitpress/8732.003.0007.
- Es, M. van, Guijt, I. and Vogel, I. (2015) ‘Hivos ToC Guidelines: Theory of Change Thinking in Practice’. The Hague, The Netherlands: Hivos.

- European Data Protection Board (2022) ‘Guidelines 01/2022 on data subject rights - right of access version 1.0 - public consultation’. Available at: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-012022-data-subject-rights-right_en.
- ‘Explainable AI: Making machines understandable for humans’ (no date). Available at: <https://explainableai.com/> (Accessed: 16 June 2022).
- Faife, C. (2021) ‘Facebook rolls out news feed change that blocks watchdogs from gathering data’, *The Markup*. Available at: <https://themarkup.org/citizen-browser/2021/09/21/facebook-rolls-out-news-feed-change-that-blocks-watchdogs-from-gathering-data>.
- Forrester, I. (2021) ‘Talking about human values and design’, *BBC Research & Development*. Available at: <https://www.bbc.co.uk/rd/blog/2021-07-talking-about-human-values-and-design>.
- Foucault, M. (1975) ‘Discipline and punish: The birth of the prison’, *New York*. Pantheon Books.
- Foucault-Dumas, C. (2021) ““This case forces uber to do a lot more than ever before”, says data protection expert”, *Medium*. Available at: <https://medium.com/personaldata-io/uber-vs-drivers-trial-interview-data-protection-expert-rene-mahieu-55359f8cdd9d>.
- 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.
- Gayomali, C. (2015) ‘Why twitter is killing TweetDeck’. The Week. Available at: <https://theweek.com/articles/467040/why-twitter-killing-tweetdeck>.
- Gillespie, T. and Seaver, N. (2016) ‘Critical Algorithm Studies - A Reading List’. Available at: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>.
- Gitelman, L. (2013) *Raw data is an oxymoron*. Edited by Lisa Gitelman. MIT Press, p. 182. Available at: <https://mitpress.mit.edu/books/raw-data-oxymoron>.
- Goffe, L. et al. (2021) ‘Appetite for disruption: Designing human-centred augmentations to an online food ordering platform’, in *34th british HCI conference 34*, pp. 155–167.
- Goffe, L. et al. (2022) ‘Web augmentation for well-being: The human-centred design of a takeaway food ordering digital platform’, *Interacting with Computers*.
- Golder, S. A. and Huberman, B. A. (2006) ‘Usage patterns of collaborative tagging systems’, *Journal of Information Science*, 32, pp. 198–208. doi: 10.1177/0165551506062337.
- 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. (2013) ‘A Call to Minimize Distraction Respect Users’ Attention’. Available at: <http://www.minimizedistraction.com/>.
- Hart-Davidson, W., Zachry, M. and Spinuzzi, C. (2012) ‘Activity streams: Building context to coordinate writing activity in collaborative teams’, in *SIG-*

- DOC'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.
- Hartman, T. *et al.* (2020) ‘Public perceptions of good data management: Findings from a UK-based survey:’, <https://doi.org/10.1177/2053951720935616>. SAGE PublicationsSage UK: London, England, 7. doi: 10.1177/2053951720935616.
- Hatmaker, T. (2018) ‘Twitter is killing its twitter for mac desktop client’. TechCrunch.
- Havlak, H. and Abelson, B. (2016) ‘The definitive list of what everyone likes on facebook’, *The Verge*. Available at: <https://www.theverge.com/2016/2/1/10872792/facebook-interests-ranked-preferred-audience-size>.
- Helmond, A. (2015) ‘The platformization of the web: Making web data platform ready’, *Social media + society*. Sage Publications Sage UK: London, England, 1(2), p. 2056305115603080.
- Hoffman, W. (2014) *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.
- Hogan, T. (2012) *Toward a phenomenology of human-data relations*. Available at: http://www.manovich.net/DOCS/data_art.doc.
- Horwitz, J. *et al.* (2021) ‘The facebook files’, *Wall Street Journal*. Available at: <https://www.wsj.com/articles/the-facebook-files-11631713039>.
- Ihde, D. (1990) *Technology and the lifeworld: From garden to earth*. Indiana University Press.
- Information Commissioner’s Office (2018) ‘Your data matters - Your rights’. Available at: <https://ico.org.uk/your-data-matters/>.
- Jasperson, J. (Sean). *et al.* (2002) ‘Review: Power and Information Technology Research: A Metatriangulation Review’. Society for Information Management; The Management Information Systems Research Center. doi: 10.2307/4132315.
- Jeffers, S. and Webb, L. K. (2017) ‘About who targets me’.
- Jelly, M. (2021) ‘The Mission’. ethi.me. Available at: <https://web.archive.org/web/20220312232859/https://www.ethi.me/the-mission> (Accessed: 12 March 2021).
- Jensen, C. *et al.* (2010) ‘The life and times of files and information: A study of desktop provenance’.
- Jones, W. *et al.* (2006) “It’s about the information stupid!”: Why we need a separate field of human-information interaction’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 65–68. doi: 10.1145/1125451.1125469.
- Jones, W. (2011) ‘The Future of Personal Information Management Part I: Our Information, Always and Forever’.
- 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>.
- Kelty, C. M. (2008) *Geeks and Recursive Publics*. Duke University Press, pp. 27–63.
- Kirven, A. (2018) ‘Whose gig is it anyway: Technological change, workplace

- control and supervision, and workers' rights in the gig economy', *U. Colo. L. Rev.* HeinOnline, 89, p. 249.
- Kollnig, K. (2021) 'TrackerControl'. Available at: <https://www.trackercontrol.org/>.
- Kollnig, K., Datta, S. and Kleek, M. V. (2021) 'I want my app that way: Reclaiming sovereignty over personal devices; i want my app that way: Reclaiming sovereignty over personal devices', *CHI Extended Abstracts*. ACM. doi: 10.1145/3411763.3451632.
- 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.
- Kuchler, H. (2018) 'Max schrems: The man who took on facebook — and won', *The Financial Times*. Available at: <https://www.ft.com/content/86d1ce50-3799-11e8-8eee-e06bde01c544>.
- 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.
- Le Dantec, C. A. (2016) *Designing publics*. MIT Press.
- Lemley, M. A. (2021) 'The splinternet', *Duke Law Journal*, pp. 1397–1428. Available at: <https://perma.cc/92LZ-B8DN%5D>.
- Lessig, L. (2000) 'Code is law: On liberty in cyberspace', *Harvard Magazine*. Available at: <https://www.harvardmagazine.com/2000/01/code-is-law-html>.
- Levitas, J. (2013) 'Defining civic hacking', *Medium*. Available at: <https://medium.com/civic-innovation/defining-civic-hacking-16844fc161cd>.
- Li, I., Forlizzi, J. and Dey, A. (2010) 'Know thyself: Monitoring and reflecting on facets of one's life', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 4489–4492. doi: 10.1145/1753846.1754181.
- 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.
- Lomas, N. (2021) 'Dutch court rejects uber drivers' "robo-firing" charge but tells ola to explain algo-deductions', *TechCrunch*. Available at: <https://techcrunch.com/2021/03/12/dutch-court-rejects-uber-drivers-robo-firing-charge-but-tells-ola-to-explain-algo-deductions/>.
- Lomas, N. (2022) 'TikTok 'pauses' privacy policy switch in europe after regulatory scrutiny'. TechCrunch. Available at: <https://techcrunch.com/2022/07/12/tiktok-pauses-privacy-policy-switch/>.
- Macaskill, E. et al. (2013) 'NSA files decoded: Edward snowden's surveillance revelations explained', *The Guardian*. Available at: <https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded>.
- Mahieu, R. L. P. and Ausloos, J. (2020) 'Harnessing the collective potential of GDPR access rights : towards an ecology of transparency', *Internet Policy Review*. Available at: <https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487>.
- Mahieu, R. and Ausloos, J. (2020) 'Recognising and Enabling the Collective

- Dimension of the GDPR and the Right of Access’.
- Maldre, M. (2012) ‘Amazon makes kindle less social -’. Available at: <https://www.mattmaldre.com/2012/10/01/amazon-makes-kindle-less-social/>.
- Marchionini, G. (2008) ‘Human-information interaction research and development’, *Library and Information Science Research*, 30(3), pp. 165–174. doi: 10.1016/j.lisr.2008.07.001.
- Marshall, C. (2019) ‘What is named entity recognition (NER) and how can i use it?’, *Medium*. Available at: <https://medium.com/mysuperai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d>.
- Martin, M. (2007) *Research note: Representing identity and relationships in information systems*, *International Journal of Business Science & Applied Management (IJBSAM)* Suggested Citation: Martin, pp. 47–51. Available at: <http://hdl.handle.net/10419/190583><https://creativecommons.org/licenses/by/2.0/uk/>.
- Martin, M. (2022) ‘The trustworthy, governable platform: A concept of safe social spaces in the public network’. in prep.
- Meschtscherjakov, A., Wilfinger, D. and Tscheligi, M. (2014) ‘Mobile attachment-Causes and consequences for emotional bonding with mobile phones’, *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2317–2326. doi: 10.1145/2556288.2557295.
- Miagkov, A., Gillula, J. and Cyphers, B. (2019) ‘Google’s plans for chrome extensions won’t really help security’, *Electronic Frontier Foundation*. Available at: <https://www.eff.org/deeplinks/2019/07/googles-plans-chrome-extensions-wont-really-help-security>.
- Miller, A. (2021) ‘The right to repair movement, explained – BMC software | blogs’, *The Business IT Blog*. Available at: <https://www.bmc.com/blogs/right-to-repair/>.
- Moore-Colyer, R. (2022) ‘Apple launches iPhone self service repair kits — but there’s a big catch | tom’s guide’, *Tom’s Guide*. Available at: <https://www.tomsguide.com/news/apple-launches-iphone-self-service-repair-kits-but-theres-a-big-catch>.
- ‘More than 400,000 freedom of information requests made since 2005’ (2014) *BBC News*. Available at: <https://www.bbc.co.uk/news/uk-politics-30632221>.
- 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.
- MyData (2017) ‘Declaration - MyData.org’. Available at: <https://mydata.org/declaration/> (Accessed: 8 November 2019).
- mySociety (2004) ‘About TheyWorkForYou’. Available at: <https://www.theyworkforyou.com/about/>.
- Neff, G. (2013) ‘Why Big Data Won’t Cure Us’, *Big Data*, 1(3), pp. 117–123. doi: 10.1089/big.2013.0029.
- Newton, C. (2018) ‘Twitter officially kills off key features in third-party apps’. The Verge. Available at: <https://www.theverge.com/2018/8/16/17699626/twitter-third-party-apps-streaming-api-deprecation>.

- ‘Open rights group: Who we are’ (no date). Available at: <https://www.openrightsgroup.org/who-we-are/> (Accessed: 16 June 2022).
- Orphanides, K. G. (2021) ‘The BBC’s radical new data plan takes aim at netflix’, *Wired UK*. Available at: <https://www.wired.co.uk/article/bbc-data-personalisation>.
- Owen, L. H. (2012) ‘Betaworks’ findings shifts to web clipping, as amazon bans kindle clips’, *GigaOm*. Available at: <https://web.archive.org/web/20120922222936/http://gigaom.com/2012/09/18/betaworks-findings-pivots-as-amazon-bans-kindle-clips/>.
- P., R. (2021) ‘Web scraping extensions: The easy gateway to web data’, *Medium*. Available at: <https://raluca-p.medium.com/web-scraping-extensions-the-easy-gateway-to-web-data-40e8592e13bf>.
- Pegoraro, R. (2022) ‘Facebook will soon stop tracking your location and delete your location history’, *Fast Company*. Available at: <https://www.fastcompany.com/90750241/facebook-will-soon-stop-tracking-your-location-and-delete-your-location-history>.
- Peysakhovic, A. and Wu, L. (2020) ‘CS 4803 / 7643: Deep learning guest lecture: Embeddings and world2vec, a guest lecture by ledell wu’. Available at: https://web.archive.org/web/20211018015836/https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_EMBEDDING_world2vec_final_version.pdf.
- Pidoux, J. et al. (2022) *Digipower technical reports: Understanding influence and power in the data economy*. doi: 10.5281/zenodo.6554155.
- Plantin, J. C. et al. (2018) ‘Infrastructure studies meet platform studies in the age of google and facebook’, *New Media and Society*. SAGE Publications Ltd, 20, pp. 293–310. doi: 10.1177/1461444816661553.
- Pollock, R. (2011) ‘Building the (Open) Data Ecosystem – Open Knowledge Foundation Blog’. Available at: <https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/> (Accessed: 23 July 2019).
- ‘Privacy salon & my data done right’ (2022). Available at: <http://www.privacysalon.org/my-data-done-right-right>.
- Reber, M. and Duffy, A. (2005) ‘Value centred design: Understanding the nature of value’, *International Conference on Engineering Design*.
- ‘Recursive Public (Discussion Page)’ (no date). Available at: https://wiki.p2pfoundation.net/Recursive_Public (Accessed: 16 June 2022).
- Regoje, O. (2021) ‘Reddit’s disrespectful design’. Available at: <https://ognjen.io/reddits-disrespectful-design/>.
- Schneider, H. et al. (2018) ‘Empowerment in HCI - A survey and framework’, in *Conference on human factors in computing systems - proceedings*. Association for Computing Machinery. doi: 10.1145/3173574.3173818.
- Schrems, M. (2017) ‘Noyb.eu - our detailed concept’, *noyb.eu*. Available at: <https://noyb.eu/en/our-detailed-concept>.
- Schrock, A. R. (2016) ‘Civic hacking as data activism and advocacy: A history from publicity to open government data’, *New Media and Society*. SAGE Publications Ltd, 18, pp. 581–599. doi: 10.1177/1461444816629469.
- Shah, C. and Bender, E. M. (2022) ‘Situating search’, in *CHIIR 2022 - Proceedings*

- of the 2022 Conference on Human Information Interaction and Retrieval.* Association for Computing Machinery, Inc, pp. 221–232. doi: 10.1145/3498366.3505816.
- ‘ShapeRepo: Make your apps interoperable’ (2022). Available at: <https://shaperepo.com/>.
- Siegal, J. (2022) ‘Twitter is killing TweetDeck for mac on july 1st and everyone’s angry’. BGR. Available at: <https://bgr.com/tech/twitter-is-killing-tweetdeck-for-mac-on-july-1st-and-everyones-angry/>.
- Silver, J. (2014) ‘Shamed on twitter, corporations do an about-face’, *Ars Technica*. Available at: Shamed on Twitter, corporations do an about-face.
- Sisco, J. (2022) ‘New internal documents show tech giants pushing out competitors’, *Politico*. Available at: <https://www.politico.com/news/2022/07/19/documents-antitrust-case-google-amazon-00046522>.
- Solon, O. (2012) ‘How much data did facebook have on one man? 1,200 pages of data in 57 categories’, *Wired*. Available at: <https://www.wired.co.uk/article/privacy-versus-facebook>.
- Storni, C. (2014) ‘The problem of De-sign as conjuring: Empowerment-in-use and the politics of seams’, *Proceedings of the 13th Participatory Design Conference on Research Papers - PDC ’14*. New York, New York, USA: ACM Press, pp. 161–170. doi: 10.1145/2661435.2661436.
- ‘Tap my data’ (2022). Available at: <https://tapmydata.com/>.
- Taplin, D. H. and Clark, H. (2012) ‘Theory of change basics: A primer on theory of change’, *ActKnowledge*, p. 9. Available at: http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.
- Tauberer, J. (2014) ‘Civic hacking’, *Open Government Data: The Book*. Available at: <https://opengovdata.io/2014/civic-hacking/>.
- 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. B. (2001) ‘Displaying dynamic information’, in *Conference on human factors in computing systems - proceedings*, pp. 417–418. doi: 10.1145/634067.634311.
- Televiciute, J. (2020) ‘What is faceted search and how does it improve ecommerce UX?’ Searchanise. Available at: <https://searchanise.io/blog/what-is-faceted-search/>.
- Tett, G. (2022) “Right to repair” campaign forces rethink by big tech’, *Financial Times*. Available at: <https://www.ft.com/content/aabe2aee-cd2b-42e4-9a0b-51f838da89db>.
- ‘The citizens - about us’ (2020). Available at: <https://the-citizens.com/about-us/>.
- Tufecki, Z. (2019) ‘We are tenants on our own devices’, *Wired*. Available at: <https://www.wired.com/story/right-to-repair-tenants-on-our-own-devices/>.
- Verborgh, R. (2017) ‘Paradigm shifts for the decentralized web’. Available at: <https://ruben.verborgh.org/blog/2017/12/20/paradigm-shifts-for-the-decentralized-web/>.
- Vlachokyriakos, V. et al. (2016) ‘Digital civics: Citizen empowerment with and through technology’, *Conference on Human Factors in Computing Systems - Proceedings*, 07-12-May-, pp. 1096–1099. doi: 10.1145/2851581.2886436.

- Walby, K. and Larsen, M. (2012) ‘Access to information and freedom of information requests: Neglected means of data production in the social sciences’, *Qualitative Inquiry*, 18, pp. 31–42. doi: 10.1177/1077800411427844.
- Watzman, N. (2021) ‘The political ads facebook won’t show you’, *Medium*. Available at: <https://medium.com/cybersecurity-for-democracy/the-political-ads-facebook-wont-show-you-e0d6181bca25>.
- Weiser, M. (1991) ‘The computer for the 21st century’, *Scientific American*, 265(3), pp. 94–105. doi: 10.1145/329124.329126.
- Weiser, M. (1994) ‘Creating the invisible interface: (Invited talk)’, in *Proceedings of the 7th annual ACM symposium on user interface software and technology*, p. 1.
- Weiser, M. and Brown, J. S. (1997) ‘The coming age of calm technology’, in *Beyond calculation*. Springer, pp. 75–85.
- ‘What is text mining and content analytics?’ (2022). OpenText. Available at: <https://www.opentext.com/products-and-solutions/products/ai-and-analytics/opentext-magellan/magellan-text-mining>.
- Windeyer, R. C. (2021) *Black box exposures: Enriching public engagement with human-data relations through intermedial performance strategies*.
- ‘Worker info exchange’ (2022). Available at: <https://www.workerinfoexchange.org/>.