

Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Discussion II: Designing and Pursuing Better Human Data Relations	3
1.1 Introduction & Background	3
1.1.1 External Research Settings	5
1.1.2 Attribution of Insights	5
1.1.3 Theories of Change	6
1.1.4 Better Human Data Relations as a Recursive Public	8
1.2 Refining and Defining ‘Human Data Relations’ (HDR)	9
1.2.1 The Role of Personal Data	9
1.2.2 Human Data Interaction or Human Information Interaction?	10
1.2.3 The Two Distinct Motivations for Human Data Relations	11
1.2.3.1 Life Information Utilisation	12
1.2.3.2 Personal Data Ecosystem Control	12
1.3 The landscape of opportunity for better HDR, with its obstacles and resulting insights	14
1.3.1 Data Awareness & Understanding	15
1.3.1.1 Obstacle 1: Data Legibility	15
1.3.1.2 Insight 1: Life Information Makes Data Relatable	15
1.3.1.3 Obstacle 2: The Personal Data Diaspora	15
1.3.1.4 Insight 2: Ecosystem Information is an antidote to Digital Life Complexity	15
1.3.2 Data Usability	15
1.3.2.1 Obstacle 3: Data isn’t free	15
1.3.2.2 Obstacle 4: Data that is Unmalleable and Non-Interrogable	18
1.3.2.3 Insight 3: Life Information & Ecosystem Information should be Useable as a Material	18
1.3.3 Ecosystem transparency	18
1.3.3.1 Obstacle 5: A Complex and Invisible Data Ecosystem	18
1.3.3.2 Obstacle 6: A Lack of Metadata	19

1.3.3.3	Insight 4: Data Needs Provenance	20
1.3.4	Ecosystem negotiability	20
1.3.4.1	Obstacle 7: Provider Hegemony and the Nature of Digital Power	20
1.3.4.2	Insight 5: Data Holders Exploit Four Levers of Power to Manipulate the Digital Landscape	21
1.3.4.3	Obstacle 8: Closed, Insular and Introspective Practices	22
1.3.4.4	Obstacle 9: A Trend of Actively Diminishing Individuals' Agency	23
1.3.4.5	Obstacle 10: The Inaccessible Data Self	24
1.3.5	Obstacles in the Solution Space	24
1.3.5.1	Obstacle 11: A lack of demand and HDR motivation, and perceived hard work	24
1.3.5.2	Insight 6: New Life Capabilities and Pain Relievers Drive User Demand	25
1.3.5.3	Obstacle 12: A lack of Interoperability	26
1.3.5.4	Obstacle 13: Insufficient machine understanding of human data	27
1.3.5.5	Insight 7: Semantic Analysis and Information Standards can Transform Data Storage and facilitate human-centric interface building	27
1.3.5.6	Insight 8: Better HDR can deliver business value through increased accuracy and consent, and decreased liability.	28
1.4	Four approaches to improving HDR, and their specific opportunities	28
1.4.1	Learn & Discover	30
1.4.1.1	Opportunity 1: Auditing Data Holders	30
1.4.1.2	Opportunity 2: Collective Investigation	30
1.4.1.3	Opportunity 3: A 'Data Understanding' Industry	30
1.4.2	Defending the Status Quo and Pushing for Compliance .	31
1.4.2.1	Opportunity 4: Tapping the Seams in pursuit of a Free Information Landscape	31
1.4.2.2	Opportunity 5: Collective Activism	32
1.4.3	Create New Structures and Systems	32
1.4.3.1	Opportunity 6: A central home for your personal data	32
1.4.3.2	Opportunity 7: Modelling Data as Life Information	34
1.4.3.3	Opportunity 8: Algorithmic Meaning Extraction and Learning Systems	36
1.4.4	Create New Information Interfaces	42
1.4.4.1	Opportunity 8: Life Information interfaces	42
1.4.4.2	Opportunity 9: Ecosystem Detection & Visualisation	45
1.4.4.3	Opportunity 10: Inclusive Information Flows	45
1.4.5	Create New Capabilities	45

1.4.5.1	Opportunity 11: Exploratory Actions & Asking Tools	45
1.4.5.2	Opportunity 12: Self Profiling & Curating your Digital Self	45
1.4.6	Influence & Motivate	47
1.4.6.1	Opportunity 13: Regulating the Information Landscape	47
1.4.6.2	Opportunity 14: Information Unification and Schematisation	47
1.4.6.3	Opportunity 15: Life/Ecosystem Information as Boundary Objects (Multi-stakeholder Design) . .	47
1.4.6.4	Opportunity 16: The Business Value of Transparency and Human Centricity	47
1.4.7	Educate & Empower	47
1.4.7.1	Opportunity 17: Life Information & Personal Data Ecosystem Literacy	47
1.4.7.2	Opportunity 18: Individual Discovery: Mapping Your Personal Data Ecosystem	47
1.5	Thesis Conclusion	47
	Bibliography	47

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 “*What relationship do people need with their personal data, and how might that be achieved?*” and look specifically at its second clause. This chapter describes practical approaches for future research and innovation, in a way that is deliberately broad and shallow, from the perspective that it is more useful to introduce a wide range of applicable ideas than to go into great detail on just a few. This is not intended to form a complete or exhaustive roadmap; it is a snapshot of ongoing work, identified challenges and known opportunities, forming an *anthology of reference material* for designers and innovators in this space. These ideas are illustrated through real world insights and activities from the four industrial and academic research projects I was part of, and from the work of other innovators and activists. This

chapter also builds upon the theoretical insights from the Case Studies in order to inform the design of future research, innovation and policy as to how the better Human Data Relations conceived in this thesis thus far might be achieved.

The approach this chapter takes is to name and illustrate what challenges and opportunities are relevant when attempting to bring about changes in the world that would bring people closer to the six HDR wants that this thesis has uncovered. There are many aspects to such a wide-reaching objective: technical, design, commercial, legal, moral, social and political and this chapter does not pretend to cover them all nor to be formal empirical research. However I have been fortunate to have undertaken, during the same time period as I have been working on this PhD but outside of the research, direct embedded work in personal data interaction related projects (3.4.3) in academic and industrial research that directly contribute to the question of how to bring about better human data relations in practice. As a result, some of these 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.

In section 7.1.1 the external activities I undertook are described; they form a primary point of reference for insights and illustrations shared in this chapter, as they have allowed me to learn enough to provide a useful overview and highlight many important and evolving areas where different actors are trying to bring about changes that often align well to the six data wants uncovered in the previous chapters.

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 7.1.3, I introduce some additional background on *Theories of Change (ToC)*, which are used as a framing device for structuring the insights described in the main body of this chapter into a series of different possible trajectories for change.

In section 7.1.4, I consider the researcher-turned-activist stance that drives this chapter, framing the pursuit of better HDR as a *recursive public*.

In section 7.2, to provide deeper context for what follows, the concept of HDR is expanded to identify some additional insights into how people relate to data, and an important dichotomy of two distinct drivers that motivate people's needs for better relations with their data.

Section 7.3 and 7.4 form the main body of this chapter, with obstacles and insights being detailed in section 7.3 and specific opportunities into how better Human Data Relations can be pursued in practice described in 7.4. 7.4 is structured using the ToC framing described in 7.1.3, as a series of named opportunities fitting into each trajectory of change.

Section 7.5 concludes the thesis, summarising the change trajectories presented in 7.4, the thesis' contributions as a whole, and answering the overall research

question.

1.1.1 External Research 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 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)

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 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.

1.1.3 Theories of Change

To provide a structure for cataloguing the insights conveyed by this chapter, 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 chapter, 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 section 7.4 will use as its frame. Specifically, desired changes can be broken down into:

- **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

¹Diagram used here unchanged from *Hivos ToC Guidelines* (Es, Guijt and Vogel, 2015, p. p90) 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).

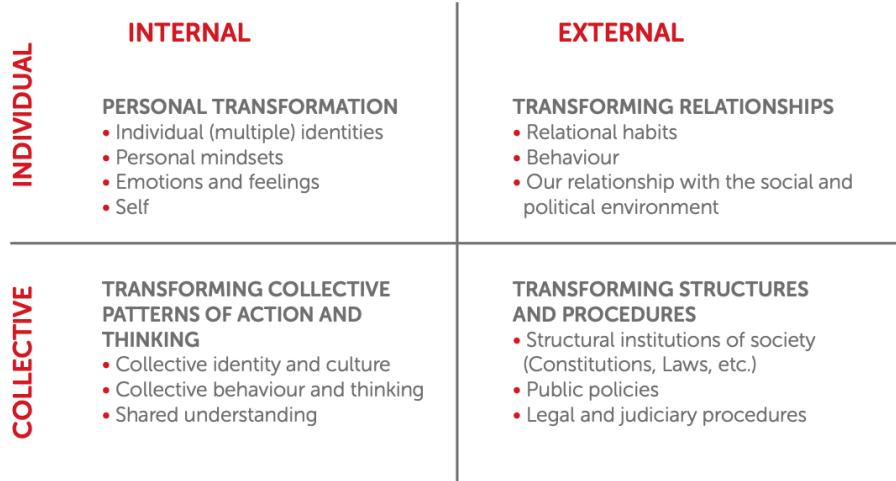


Figure 1: Figure 29: The Four Dimensions of Change¹

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.

1.1.4 Better Human Data Relations as a Recursive Public

Before engaging with the practicalities of pursuing change, 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 see 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 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: digital rights ('Open rights group: Who we are', no date), gig economy worker rights (Kirven, 2018), privacy by design (Cavoukian, 2010), 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 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. 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?"

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

Chapter 6 established six ‘wants’ in HDR: **visible**, **understandable** and **usable data**; **process transparency**, **individual oversight** and **decision-making involvement**. At a simplistic level therefore ‘better’ HDR can be achieved by working to improve upon those six aspects of data interaction. However, as this section will explain, HDR can be conceptually split into two distinct motives, to which those six wants apply differently, therefore it is useful to develop the concept of HDR further. 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 1: Table 15. **Eight lenses on data.**

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.

Way of thinking about data	Explanation & Implications
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.
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 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 (see 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

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 publication as a call for a ‘new field’.

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

1.2.3.1 Life Information Utilisation *Life Information Utilisation* is a superset of *Self Informatics (SI)*, as defined in 2.2.3. It includes all purposes relating 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 usability*¹³ (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).

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]

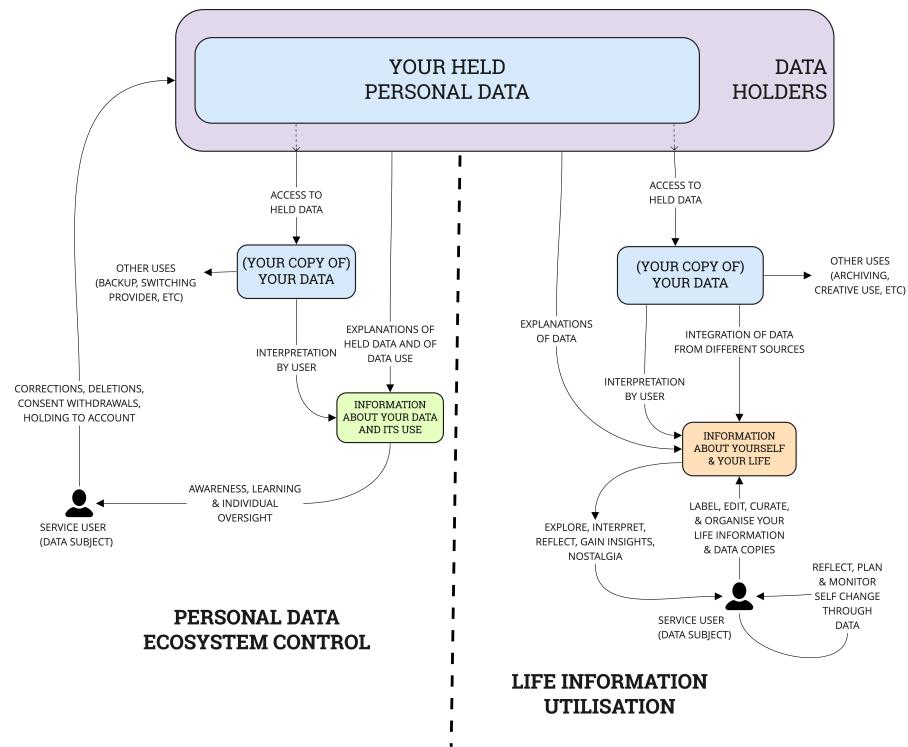


Figure 2: Figure 30: The Two Motivations for HDR: Controlling your personal data ecosystem and utilising your information about your life, with 'idealised'⁵ processes illustrated

PDEC is a direct response to the power imbalance between data holders and individuals that the World Economic Forum described in 2014 [2.1.2; Hoffman (2014)].

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.

1.3 The landscape of opportunity for better HDR, with its obstacles and resulting insights

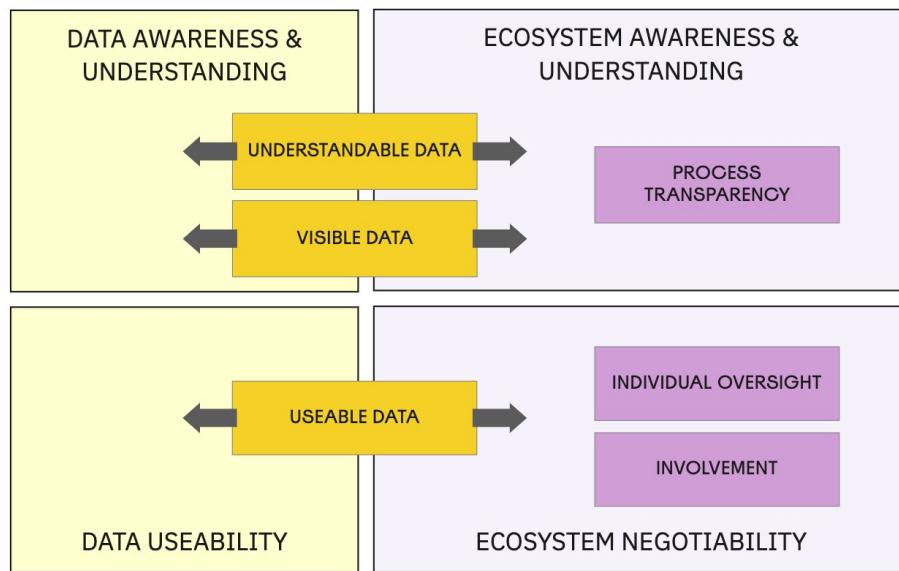


Figure 3: Figure X: Mapping the Six Wants into Objectives for the HDR Opportunity Landscape

In this section I will describe the high level obstacles to better HDR, in four sections. These are arranged into six groupings. The first four groupings correspond to the six wants identified in Chapter 6. Two additional groupings have been included to cover more general human and technical challenges that affect all endeavours in this space:

- 7.3.1: Obstacles to Data Awareness & Understanding
- 7.3.2: Obstacles to Data Useability
- 7.3.3: Obstacles to Ecosystem Transparency (including visible data & transparent processes)

- 7.3.4: Obstacles Ecosystem Negotiability (encompassing individual oversight and involvement) decision-making)
- 7.3.5: Obstacles in the Solution Space

1.3.1 Data Awareness & Understanding

1.3.1.1 Obstacle 1: Data Legibility People struggle to relate to data. It is not relatable because it is complex, not presented as meaningful information, and not easily interpretable as information. They lack tools to gain insights. To overcome this obstacle, more work is needed to make data relatable and to provide tools that can deliver valuable meaning and insights.

1.3.1.2 Insight 1: Life Information Makes Data Relatable When data is transformed into information that can be related back to moments, people, places or relationships in people's lives, it becomes instantly relatable. [from BBC: Data becomes meaningful when people are able to associate it with the real substance of their lives - people, places, organisations, causes or topics they care about. Therefore, the more associations you can find in data the more valuable it is.]

We can consider the different types of information in people's lives:

We need to model life information, not data.

1.3.1.3 Obstacle 2: The Personal Data Diaspora Every individual's personal data is scattered across multiple providers, devices, apps, held by hundreds of third parties. The complexity of a modern day digital life is unmanageable and overwhelming. People are inevitably ignorant of much of their data and its use. This can lead to resignation and apathy. To overcome this obstacle, approaches must be identified that recognise the scattered, complex reality of each individual's personal data ecosystem and begin to make it visible and understandable.

1.3.1.4 Insight 2: Ecosystem Information is an antidote to Digital Life Complexity No matter how understandable the data itself is, it is also critical that people can access information about their data ecosystem. Without this, there will always be aspects of their data that are beyond their awareness or beyond the reach of what they can access, control or manage. Many tools today do not recognise this, and build for a world that does not exist. It is important that people have tools that allow them to interact with multiple providers and data sources across their digital life.

1.3.2 Data Usability

1.3.2.1 Obstacle 3: Data isn't free Almost all data is constrained in some way, limiting its usability. It may be held by a particular provider and inaccessible. It may be stored in a format which is hard to use or change. It

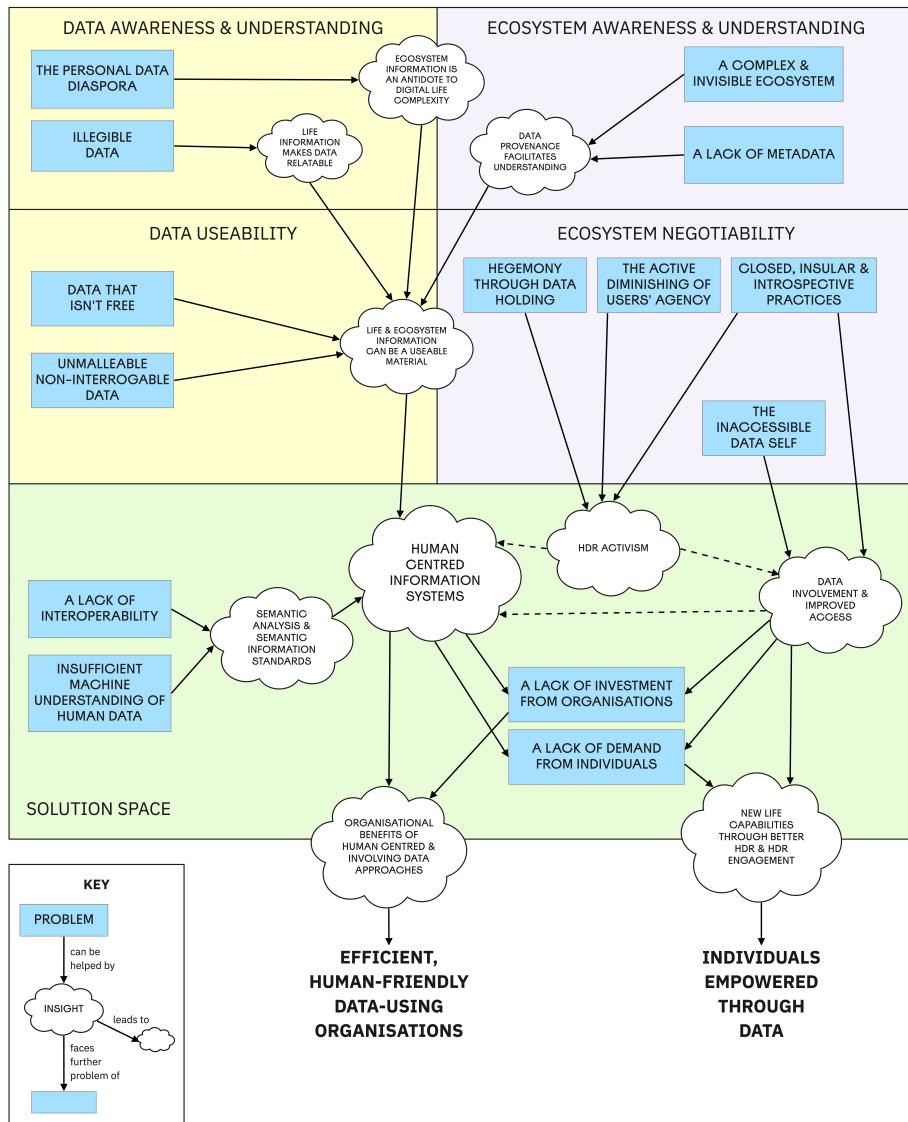


Figure 4: Figure X: Obstacles and Resulting Insights in the HDR Opportunity Landscape

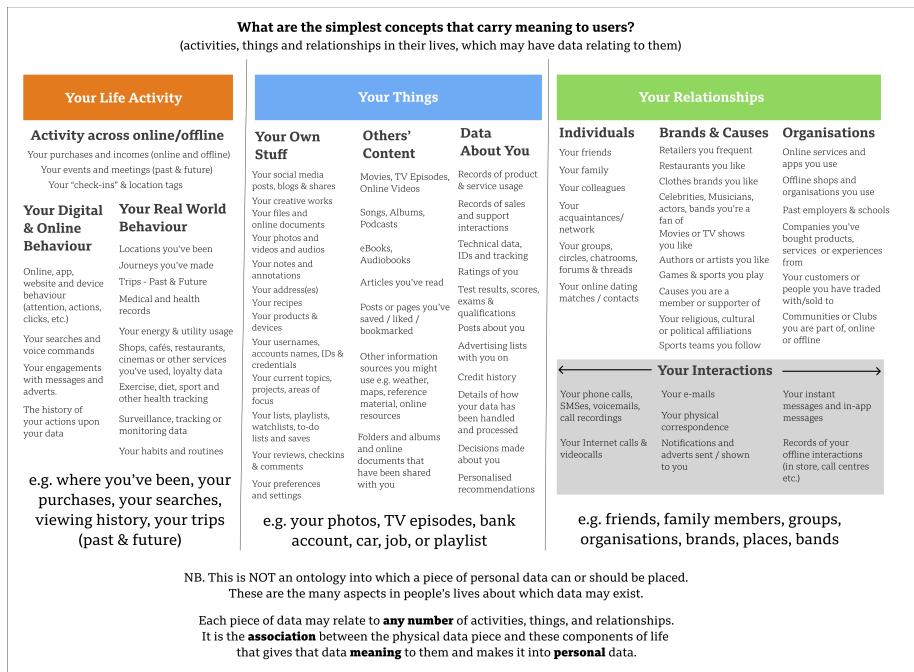


Figure 5: Figure X: Life Concept Modelling

may only be visible after a delay. It may be unchangeable. To overcome this obstacle, we need to find ways to extract data from its current constraints and to remove some of these technical or temporal limitations.

1.3.2.2 Obstacle 4: Data that is Unmalleable and Non-Interrogable

Even once an individual has gained possession or access to the relevant parts of their personal data, it can be extremely hard to use. This partly comes from a lack of malleability - the ability to break it down, look at it from different perspectives, reconstitute it in different ways. Put simply, people need to be able interrogate their data - ask questions of it. This requires more than just an ability to view visual representations of data, but an ability to interact with the data and produce new views and insights that can help to answer specific questions. Making some of the PIM and SI capabilities described in 2.2.2 and 2.2.3 can help to address this, but more capabilities can be made available and are needed to fully overcome this obstacle.

1.3.2.3 Insight 3: Life Information & Ecosystem Information should be Useable as a Material

Many computer operating systems and interfaces today treat files as the basic material that an individual can manipulate. To truly empower users to make use of their data, we need to move to a model where pieces of life information – facts (or assertions) – can be created, deleted, moved, grouped, annotated, copied, shared, modified, labelled, organised, separated or otherwise manipulated instead. So far, people access data within products. But what they need is a platform, not a product. We need an information operating system.

1.3.3 Ecosystem transparency

1.3.3.1 Obstacle 5: A Complex and Invisible Data Ecosystem The first and most obvious barrier that individuals face in managing a complex personal data ecosystem is that, to a great degree, they cannot see it. 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 exchanging user identifiers, activity data and personal information about you while you browse or use apps. As Chapter 5 showed, even though people have been granted new rights to access their data and information about provider data sharing activity, the ability to effectively execute those rights to build up a meaningful picture of your personal data ecosystem is severely limited by inconsistent, incomplete or unclear responses. The strong negative practical impacts of today's complex digital lives were already described in section 2.2.4; managing the complexity is an overwhelming, unmanageable task that even personal data experts are not fully able to get a handle on. 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 no individual or organisation has the ability to see the whole of a user's ecosystem, and there is little commercial motive to try and solve this problem, as every provider focuses just on their own apps, websites and services.

1.3.3.2 Obstacle 6: A Lack of Metadata From this complexity an additional obstacle becomes evident. There is scant attention to information **about** your data. Even where data access rights are executed (or data is shared via human means such as in Chapter 4), the attention is on the data itself: what it says. Chapter 5 shows 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, yet this was rarely forthcoming. There are many pieces of information that can be quantified about an individual's data, as illustrated in Figure X, which I created during my internship at BBC R&D:

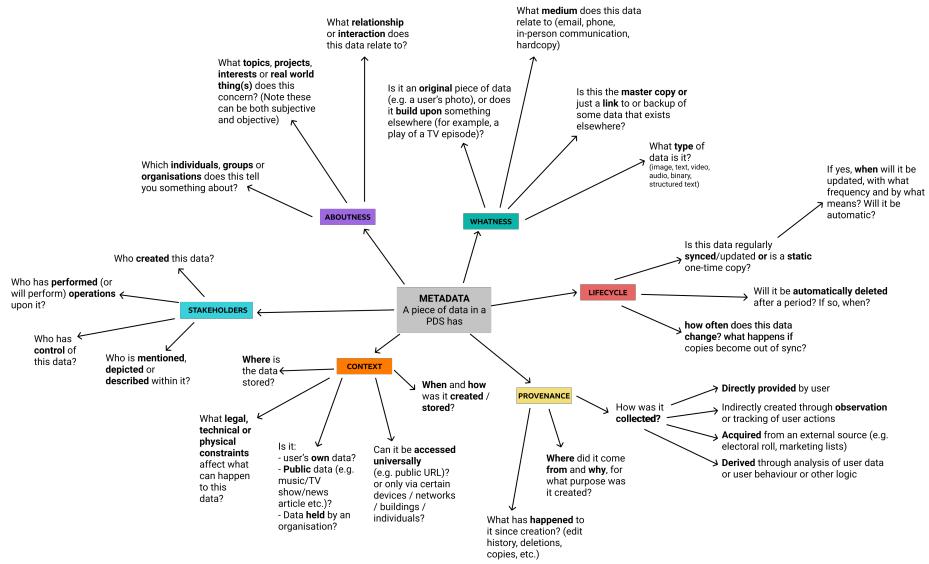


Figure 6: Figure X: Some of the many aspects of metadata that might exist about a datapoint or dataset

[EXPLAIN ASPECTS]

To provide users with meaningful transparency, many of these aspects will need to be tracked and visualised; not an easy task given the complexity and the potential to overwhelm a user, but nonetheless a vital first step on the road to giving individuals the ability to have oversight of their personal data ecosystem and take action within it.

[ADD REFERENCE BACK TO 2.2.2 METADATA]

1.3.3.3 Insight 4: Data Needs Provenance A number of researchers have independently identified the importance of keeping the history of a piece of data with it. Without context, data loses meaning (a phenomenon witnessed in Case Study Two – see 5.4.3.1). The idea that what has happened to not just an individual but to a piece of data over time is important is a key part of the thinking behind temporal PIM systems, from Lifestreams (Freeman and Gelernter, 1996) to activity streams (Hart-Davidson, Zachry and Spinuzzi, 2012) (see 2.2.2). William Odom, Siân Lindley and colleagues proposed the idea of file biographies, which view the lifetime of a file as something that should remain connected and traversal in order to understand the context of the file at its different interaction points. Significant research in this space has been undertaken by Professors Mike Martin and Rob Wilson at Northumbria University, formerly Newcastle University, who express 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, if it is to be fully understood (ie. for its context []). This plays into the ideas of Gitelman, Neff and others, that 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). [ADD MORE DETAIL FROM MIKE MARTIN PAPER AND EMAIL HERE]. While it is not a solution in its own right, it is clear that data with provenance is very likely to be a critical and valuable part of any design that aims to help individuals with managing to get an overview of their complex and invisible personal data ecosystems.

1.3.4 Ecosystem negotiability

1.3.4.1 Obstacle 7: Provider Hegemony and the Nature of Digital Power In the pursuit of individual oversight and greater involvement, the power imbalance between individuals and data holders (2.1.2) becomes most clear. While the Internet itself initially held the promise to be a great leveller and to empower individuals, this potential has largely been suppressed. Data is owned and controlled by service providers, who also design and control the interfaces, apps, websites and devices through which individuals access those services, controlling what (if any) of the data stored behind the scenes, and of the internal processes that use that data, is visible, and how such data and processes are represented. In Jasperson *et al.*'s detailed metatriangulation review of types of power that affect technology systems (Jasperson *et al.*, 2002) we can identify a number of specific types of power that clearly are in effect in today's digital data-centric service provider context:

[ADD TYPES OF POWER FROM JASPERSON WITH CONTEXTUAL EXPLANATIONS] [structural power, resource control, centralisation etc]

(Foucault, 1975)

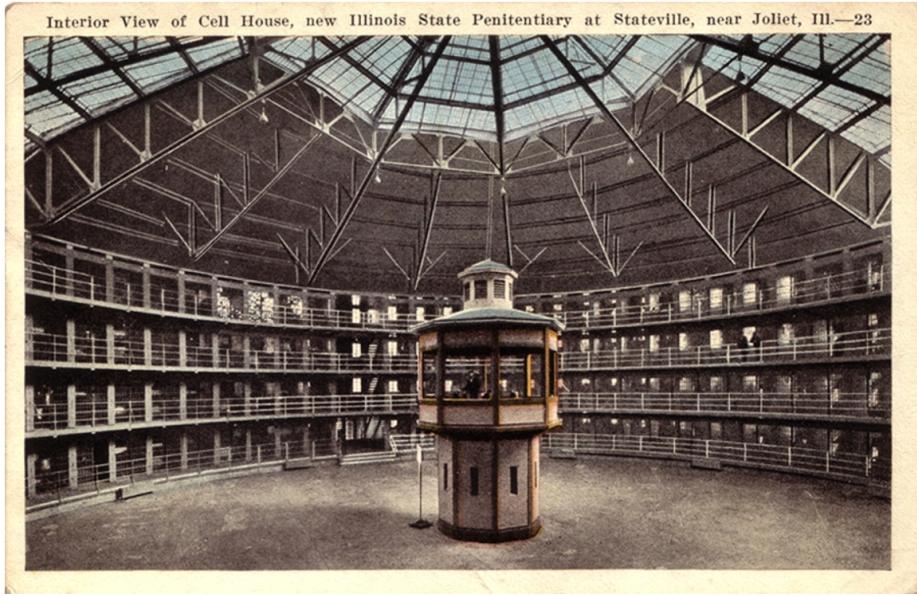
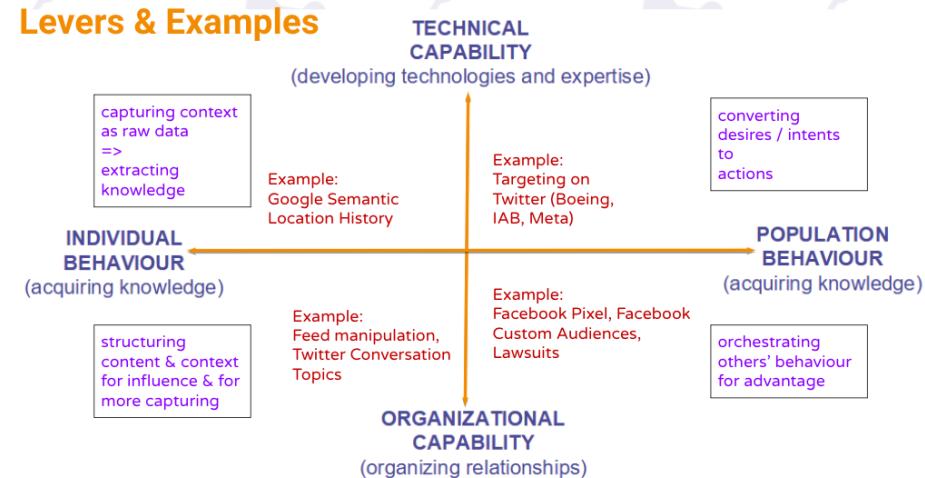


Figure 7: Figure X: The Panopticon Structure of the Illinois State Penitentiary

A helpful analogy for the relationship between provider and user can be seen in the design of Panopticon: A style of prison architecture designed to elevate the power of the prison guards to observe all the prisoners easily at any time and to diminish the ability of prisoners to operate in privacy or to see those in authority. Jeremy Bentham [REF], drawing on the philosophy of Foucault [REF], makes clear that such design is political, and shows that power can be enforced by the environment. This is a useful mental scaffold to keep in mind; as explained below [REF], we can think of today's digital landscape as similarly power-enforcing. Code is law [ADD REF Lessig], and interfaces limit what individuals can do. By holding data behind interfaces shaped to serve their own interests, the landscape is controlled by the data holders. [UPDATE THIS BASED ON OTHER WRITING ABOUT PANOPTICON]

1.3.4.2 Insight 5: Data Holders Exploit Four Levers of Power to Manipulate the Digital Landscape Sitra's #digipower investigation [REF], of which I was project leader for Hestia.ai, was a successor to my Case Study Two, but worked with high profile politicians and European influencers and added additional technical audit techniques. Its focus was not on the individual experience of data access, but on using those experiences and acquired datasets to better understand the data ecosystem. Through this research, a model was produced to understand the ways in which service providers (and in particular the larger ecosystem-level platform providers such as Google and Facebook) exert power over individuals and smaller organisations. This model is reproduced in

Figure X:



[TODO: redo this diagram]

[ADD EXPLANATION AND REFERENCE TO THE PIDOUX REPORT]

Through this landscape 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 do-able. 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.

A key mechanism to highlight here is that the accumulation of information is implicitly and objectively a form of power. This is consistent with participants' observations in 5.4.4.1 that data holding and limiting access to it is a source of power. In terms of this being an obstacle, we can therefore see 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.

1.3.4.3 Obstacle 8: Closed, Insular and Introspective Practices Today's digital landscape is fractured[REF Splinternet]; myriad providers vie to pull users into service relationships or connected ecosystems that will encourage a flow of money and attention to their own products and services, most evident from companies such as Apple, Amazon, Facebook, Google and Microsoft (the so-called 'big five') that have multiple touchpoints into people's lives through different devices, apps and services. We can think of these different providers' sub-Internets as walled gardens or silos [REF]. 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. 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, as Lawrence Lessig wrote, '*pathetic dots*' [ADD REF]. Thus service providers continue to build **proprietary, incompatible silos**.

But it is not only commercial motives that encourage insular attitudes to personal data and user service provision. In the SILVER project [ADD REF] meetings with local authorities and care providers revealed deep organisational and technical barriers within the public sector, with for example health organisations being typically unwilling to share health data with social care services, but also with different councils, community services and charities typically operating separate IT systems, each attempting to construct their own digital pictures within their own databases and very little operability. The problems of this technical reality are explored further in 4.1.2. From what we have observed, the introduction of GDPR and similar regulations has made this problem worse not better, as organisations and departments become increasingly paranoid about storing or sharing data they should not, or about the risks of acting upon data without sufficient consent. We learned of practices such as the sharing of information between care organisations verbally by telephone so that no digital trail was left.

It is clear that throughout society, there is a trend towards organisations being reluctant to work together around people's data, inclined towards collecting their own databases and not sharing them.

Also mention resistance to change

1.3.4.4 Obstacle 9: A Trend of Actively Diminishing Individuals' Agency As a result of the practices and motives described above, the last decade has seen much reduction in individuals' agency. 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. 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, rather than meeting their needs or providing maximal value. One of the most revealing examples is seen in the case of Facebook. Users used to be free to consume their friends' posts in other clients via RSS feeds. These were removed, forcing users to use only Facebook's interfaces, where their eyeballs

can be monetized (Twitter closed its APIs too to a great degree, killing off many third party readers). On Facebook users used to have the ability to view the latest updates from a particular list of friends or of news pages. These features too were removed, presumably to increase monetization through the main feed. 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.

Companies change their practices to limit users’ agency (and their own accountability to customers) too. For example, Facebook recently announced they will no longer collect historical location data from users (though they will still use location information). This makes it harder for users to see how their data has been used. Tiktok announced they will 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. Unchecked, it is clear that trends to reduce users’ agency and further providers’ interests will continue, creating another obstacle to be tackled.

1.3.4.5 Obstacle 10: The Inaccessible Data Self Earlier in this thesis the concept of a data self has been introduced (4.4.1, 4.4.3, 6.3). We know from both the preliminary study with families (Bowyer *et al.*, 2018) and Case Study Two that data serves as a proxy for direct human involvement of the served individual(s). Put simply, service providers try to minimise interaction with people, by maximising their usage of data to represent people. We are viewed through the distorted lens of our data selves. Despite the inherent challenge of representing people fairly and accurately in data [Bowyer *et al.* (2018); 4.4.1; 5.4.4.1], this is the default modus operandi for service provision today. This therefore represents a key obstacle to ecosystem negotiability today: how can individuals be given the ability to influence and shape the data self that providers will use to understand them and make decisions?

1.3.5 Obstacles in the Solution Space

While in the previous four subsections it was possible to identify obstacles relating to specific HDR wants, there are also some readily identifiable obstacles that will affect all our endeavours to improve HDR. Obstacles relating to human challenges are described in this section, and technical challenges are addressed in the following section, 7.3.6.

1.3.5.1 Obstacle 11: A lack of demand and HDR motivation, and perceived hard work In considering the recommendations of Case Study One (shared data interaction between the state and the individual) and of Case Study Two (new human-centric data practices by service providers), and in exploring possible new human-centric system and interface designs through my work with BBC R&D, it is evident that even if new human-centric types of computer system or service interaction practices can be created, we cannot assume that

people will be inclined to use them. 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. 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. But this should not deter disruptive innovation nor does it indicate that such offerings would not be useful. As 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.

1.3.5.2 Insight 6: New Life Capabilities and Pain Relievers Drive User Demand Through work at the 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. [include the three concentric circles diagram a bit like the one Rhianne used]

As an example, myself and BBC colleague Jasmine Cox imagined focusing on address books and contact lists as a strong relatable starting point that could easily generate a user demand. Many people 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, and developing human-centric personal information management capabilities to bring that messy situation under control would offer a clear and tangible benefit to users.

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 place 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.

As with any public offering of a product or service, it is important to start with identifying a problem or need, and to demonstrate a potential tool or solution that can help. In particular, there is a need to let people do **new** things that they could not do before. This has been identified as a key ingredient of user

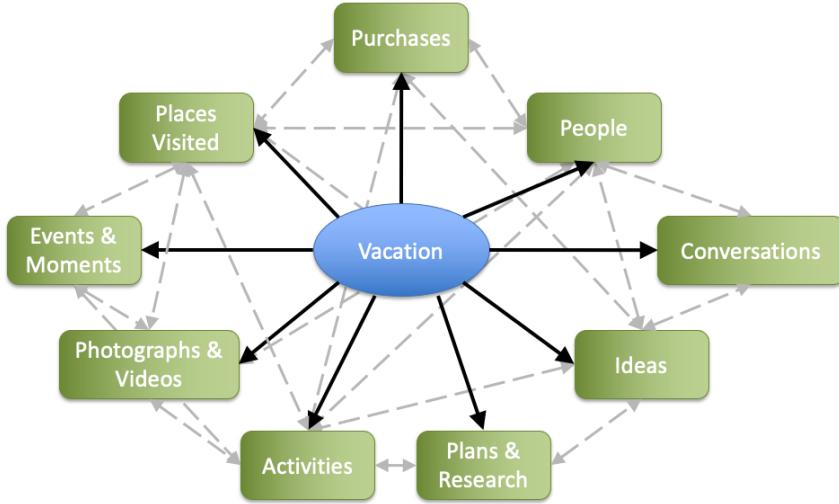


Figure 8: Figure X: Conceptual Semantic Grouping of Related Data for a Vacation

empowerment (Meschtscherjakov, Wilfinger and Tscheligi, 2014; Schneider *et al.*, 2018). This became a driving influence for design thinking on the BBC R&D Cornmarket project. It is not enough to believe that “If you build it, they will come.”

1.3.5.3 Obstacle 12: A lack of Interoperability Obstacle 8 (7.3.4.3) already touched on the issues around different companies developing different standalone walled garden or silo user experiences, from a sociotechnical or systemic standpoint. But there is a very specific technical problem that must be acknowledged across all HDR improvement approaches, and that is that it is very difficult to build technical systems that connect and exchange data with each other. This was witnessed first hand by our development team on the SILVER health data interface project [REF] which endeavoured to build a bridge making health data available to Early Help support workers. Not only are there a lack of standards, with each organisation using their own databases and formats for storing data, but often the programming interfaces (APIs) that would be needed to interface between different systems (sometimes legacy systems) do not exist, are insufficient. Furthermore, there can be issues around licensing and consent when data passes from one domain to another. Data sharing agreements must be established, especially in the public sector which is by its nature more liable to scrutiny and accountability. 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.

1.3.5.4 Obstacle 13: Insufficient machine understanding of human data Following on from the previous obstacle, but a subtly different point, is that it is technically difficult for machines to handle human information. Without deliberate coding, software can only understand streams of binary data as files or datasets, and does not understand what people, places, events or entities the facts within the data relate to. Therefore, it is necessary to consider how algorithms and systems can be designed to include an understanding of the semantics (meaning) of the information within the files and data records they handle. For example, the data record representing a post on Twitter looks entirely different to the data record representing a post on Facebook. No algorithm can recognise or unify these disparate pieces of data as two instances of the same semantic concept until its specifics of the data format can be mapped back to a common semantic abstraction of a “social media post”. [find meaning in user’s data]

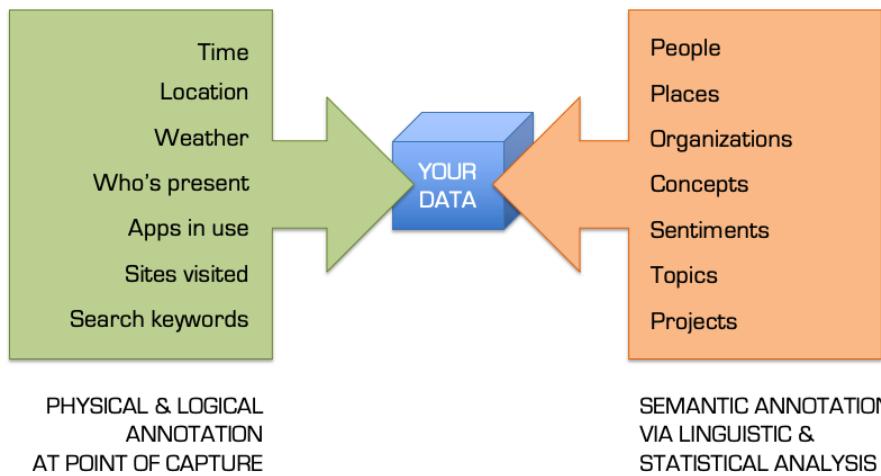


Figure 9: Figure X: Annotating Data with Semantic Context

1.3.5.5 Insight 7: Semantic Analysis and Information Standards can Transform Data Storage and facilitate human-centric interface building

This leads to the next insight: that to build systems and interfaces that are able to deal in human concepts and represent the elements of everyday life requires building systems that store semantic context and semantic associations, 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 PIM systems (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. Sizable industries have built up around Content Analytics and Enterprise Content Management. Through the capture of metadata at the point of data recording, and through subsequent programmatic analysis of stored data, as illustrated in Figure X (Bowyer, 2011), we can begin to teach computers what the data we store represent. Machine learning technologies and Artificial Intelligence have pushed machine understanding of human words, images and content to impressive levels in recent 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.

1.3.5.6 Insight 8: Better HDR can deliver business value through increased accuracy and consent, and decreased liability.

1.4 Four approaches to improving HDR, and their specific opportunities

Now, having established some of the key obstacles to improving HDR, we can move to considering what opportunities exist to pursue the HDR wants and to overcome those obstacles. This section will first introduce a framing for those opportunities, and then illustrate specific opportunities in detail.

In Figure X, the ToC frame introduced above in 7.1.3 / Figure 29 is used as a canvas upon which to position the different trajectories for changes that could improve HDR. By enumerating the possible types of activity that can bring about change, each of the four quadrants's core change trajectory can be named, as shown in purple, forming the backbone of the roadmap for improving HDR, which can be summarised thus:

- ***Learn & Discover*** (CI): In this quadrant, individuals, researchers, activists and other stakeholders work in groups to understand data attitudes and user needs, and to gain collective knowledge of data collection and usage practices which are sometimes hidden.
- ***Defend & Create*** (CE): In this quadrant, activists work to ensure current HDR capabilities are not eroded, while researchers, designers, technologists and social innovators design and create new technologies, operating models,

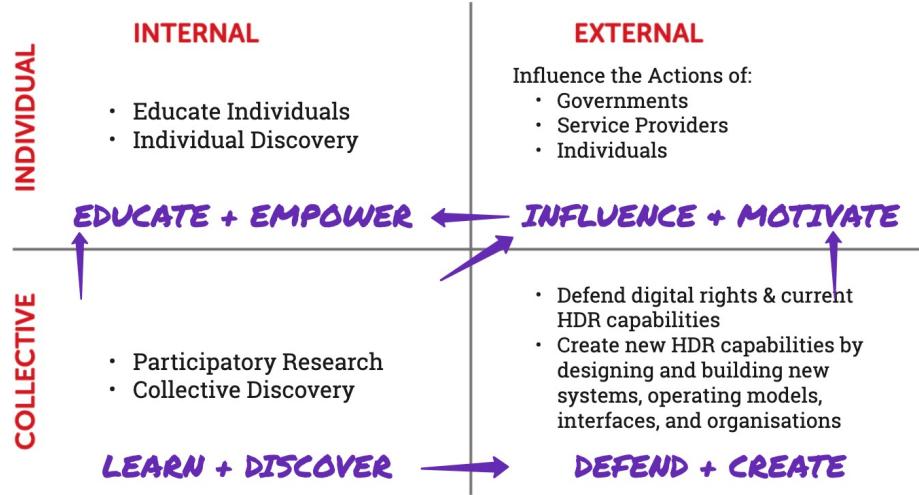


Figure 10: Figure X: ToC chart showing Trajectories of Change for improving Human Data Relations

organisations and interface designs, the structures enabling a world with better HDR. Given the majority of the opportunities for change identified in this chapter occur in this quadrant, this quadrant will be further subdivided into four different types of activity:

- Defend the Status Quo and Push for Compliance
- Create New Structures and Systems
- Create New Information Interfaces
- Create New Capabilities
- **Influence & Motivate (IE):** In this quadrant, individuals' relationships with data and with data holders, as well as data holders and policymakers' relationships, can improved. For us, as external actors seeking change, the task is to influence the many parties by showcasing and facilitating newly created structures and capabilities, and to harness our new collective knowledge to advocate the benefits of changing data-related behaviours.
- **Educate & Empower (II):** In this quadrant, individuals ways of thinking about data and data holders grow and evolve towards a state we could call 'feeling empowered'. This change is driven through education, improving Human Data Relations literacy, and through the experience of new capabilities and changed relationships with data and with data holders that can empower individuals to hold a more aware and equitable position in those relationships and in their digital life.

[TODO: do we need a summary diagram here?] [Figure X: SUMMARY OF OPPORTUNITIES]

1.4.1 Learn & Discover

Research such as that conducted in this PhD is an example of the collective, internal focused activity that can be done in this quadrant to further the goals of better HDR: Groups of people working together using a variety of techniques such as participatory co-design, interview-based quantitative studies, design prototype evaluation and other HCI techniques can gain new understandings of individual needs and experiences in HDR. However rather than mapping out such possibilities, this section will focus on more novel approaches that go beyond traditional HCI research towards activities that are potentially more socially impactful.

1.4.1.1 Opportunity 1: Auditing Data Holders Helps with: Ecosystem transparency

Through the emergence of new tracking tools such as TrackerControl and Apple's App Activity Reports, individuals can observe the actual behaviour of the apps they use, providing a new means to identify potential data sharing destinations, to assess whether providers are meeting their promises, and to uncover new questions that can be asked of providers using data access rights. By collectively examining and comparing such data, it is possible to begin to map out the data ecosystem, as was done in the digipower investigation [REF].

1.4.1.2 Opportunity 2: Collective Investigation Helps with: Ecosystem transparency

This sort of combination of individual observations is just one of many ways in which individuals can, through working together, discover more information about data usages and practices. Collectives offer a powerful means to examine how providers categorise users and process their data, for example by comparing field values to understand the range of possible values or inferences a data holder might have stored, or by comparing variations in information presentations, data rights handling or customer service experiences to reverse engineer provider practices.

1.4.1.3 Opportunity 3: A 'Data Understanding' Industry Helps with: Data Understanding, Ecosystem Transparency

Given the complexity of today's digital landscape and the forces that hinder better HDR, there is scope for an industry to develop around 'data understanding' services. This can encompass everything from self-service tools people can use to gain insights over their data (such as those provided by Ethi or Hestia.ai), to workshops helping consumer organisations, journalists, regulators, lawyers and other interested parties to collectively gain understanding and value from data so that they might better achieve their goals, as well as serving a general educational purpose for example in schools. This industry is beginning to emerge,

but faces challenges in funding, scalability, governance, and credibility and should be supported (REF Pidoux et al).

1.4.2 Defending the Status Quo and Pushing for Compliance

Given the shifting power balance of the information landscape outlined in 7.3.4.1, 7.3.4.2 and 7.3.4.4, it is clear that there is an opportunity, perhaps a need, for HDR reformers to carry out activities that monitor and publicise any changes that providers make that reduce individuals' HDR capabilities. Having identified such changes it is then easier for those HDR reformers, and indeed the wider public, to fight to protect and maintain current capabilities, as we see in the Right to Repair movement [REF] or the Net Neutrality movement [REF]. [mention also the idea of pushing to make sure what should be done, is done - e.g. in GDPR returns]

1.4.2.1 Opportunity 4: Tapping the Seams in pursuit of a Free Information Landscape Compounding the impacts of reducing agency described in 7.3.4.4 is the 'dumbing down' of technology. Apple, for example, encourages users to consider technology as 'magical', rather than as understandable tools to be harnessed and understood; Such thinking is manifested in their hardware design too: phones that cannot be opened up, expanded or repaired [REF]; the removal of accessory ports, disk drives, and headphone jacks [REF]; increased controls over what can be installed on users' hard drives and which areas of disk can be modified [REF]. These changes simplify the technology and bring it to a more mainstream audience, something that the iPhone and iPad must be given due credit for - but it happens at the cost of reducing user agency. Companies like Apple increasingly encourage users to think of technology as a black box, which you cannot and should not look inside.



Figure 11: Figure X: The 'Black Box' view that many modern technology providers present

[TODO rephrase this para as a better introduction to Seams] An important concept to understand in this space is that of *seams*. In 'The Politics of Seams' Storni outlines that current designs are incompatible with empowerment-in-use,

and highlights the role of design seams (and their removal) as being a key determiner of user power [REF]. He says that the designer passes some power to the user through their design, but also, that users should be able to take some power on their own terms (repurposing etc). He talks [says what] about the problems of technology as magic/design as conjuring:

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

Therefore part of what we need to be doing is (a) highlight and (b) removing seams/creating new seams between disconnected parts....

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 their users' agency, as illustrated by the use of web scrapers and web augmentation approaches to try and obtain information or functionality from providers that would otherwise be inaccessible. [also mention device tenancy (zeynep) and firefox containers/taking back power in the browser/browser as seam (reference Goffe et al)]

1.4.2.2 Opportunity 5: Collective Activism Collectives can also exert external influence in the adjacent ‘Defend & Create’ quadrant by using their learnings through data to demand change, as seen in the case of Uber drivers working together to obtain data on algorithmic judgements that affect their work and using that as evidence to help them demand fairer working conditions [REF]. [individuals collectively pressuring to improve GDPR responses, drive data portal improvements, etc] [traditional means e.g. press, public campaigns but also new ways e.g. mass GDPR or targeted GDPR] [mention dehaye’s pressure, leading to FB Off Site Activity, my success with Spotify] [noyb as example, also Privacy International, Bits of Freedom,] [cite examples from Mahieu papers][Mahieu, Asghari and Van Eeten (2018);mahieu2020a;Mahieu and Ausloos (2020)] [mention pooldata, data unions]

1.4.3 Create New Structures and Systems

1.4.3.1 Opportunity 6: A central home for your personal data Helps with: Data Understanding, Data Useability, Ecosystem Transparency, General Human Challenges Affected by: Lack of interoperability (7.3.6.1/12)

As others have identified, 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 personal data can be stored and aggregated in one place, a personal data locker (see 2.3.4). This prospect was explored through the BBC R&D Cornmarket project during my internship, as detailed in 3.4.3.3 [OR MOVE SOME OF THAT TEXT HERE]. As alluded to in the quote opening Chapter 1, people's data is scattered (see also 2.2.4 and (Abiteboul, André and Kaplan, 2015)), and simply providing the ability to bring data from sources together

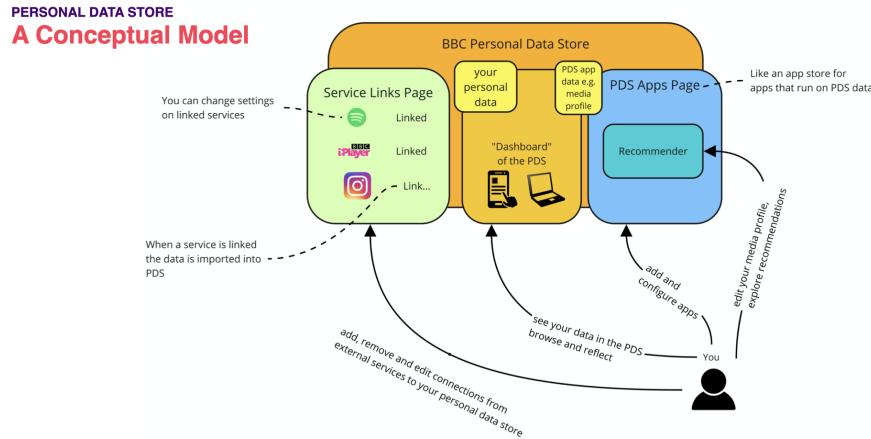
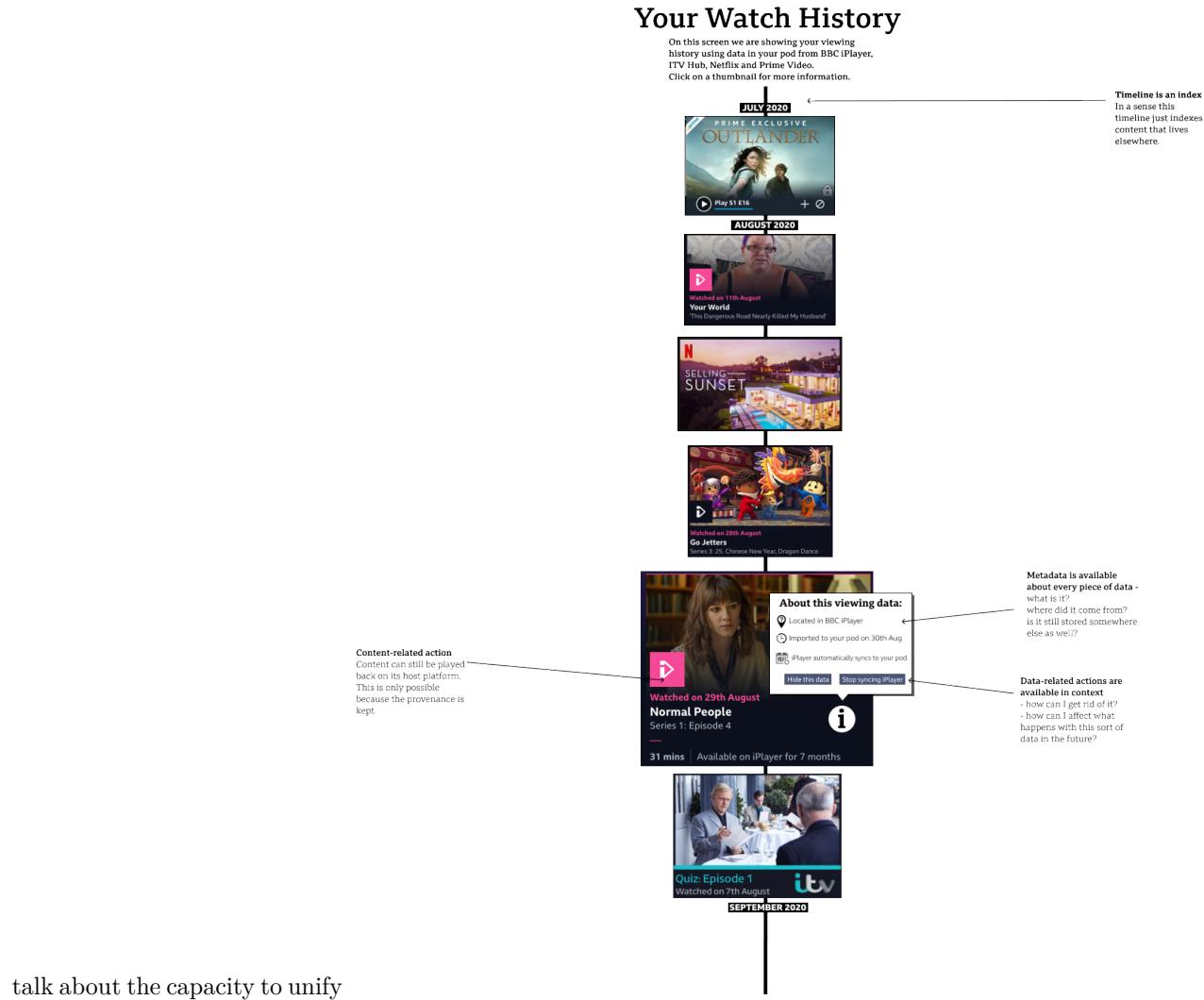


Figure 12: Figure X: A conceptual model for a Personal Data Store system that I developed for the BBC Cornmarket project

in one place can improve people's understanding of their data and its ecosystem. This integration requires technical standardisation but also [BRING IN SOME TEXT FROM BBC BLOG ARTICLE] [REF previous appetite for PDS <https://journals.sagepub.com/doi/full/10.1177/2053951720935616>]



talk about the capacity to unify

[add quote from BBC research where people liked the concept of a place for your data]

1.4.3.2 Opportunity 7: Modelling Data as Life Information

Helps with: Life Information as Material, Limited machine understanding of data

As part of the BBC R&D Cornmarket project, I carried extensive information modelling and design work with colleagues on how today's common types of data might be modelled as life information in order to help with the stated goals.

First we need to consider what a piece of data is, which is different from what data format it is or what semantic concept it represents.

What GENERAL types of personal data are we aware of?

DATA WILL HAVE
MORE THAN ONE
TYPE!

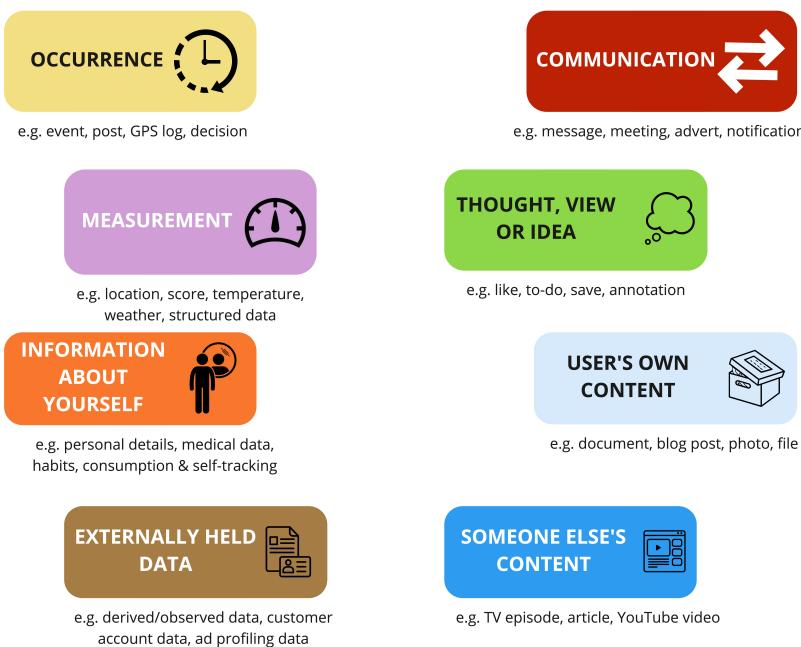


Figure 13: Figure X: High Level Data Types

It is possible to use some abstraction of commonalities to group together pieces of data that can perform a similar role:

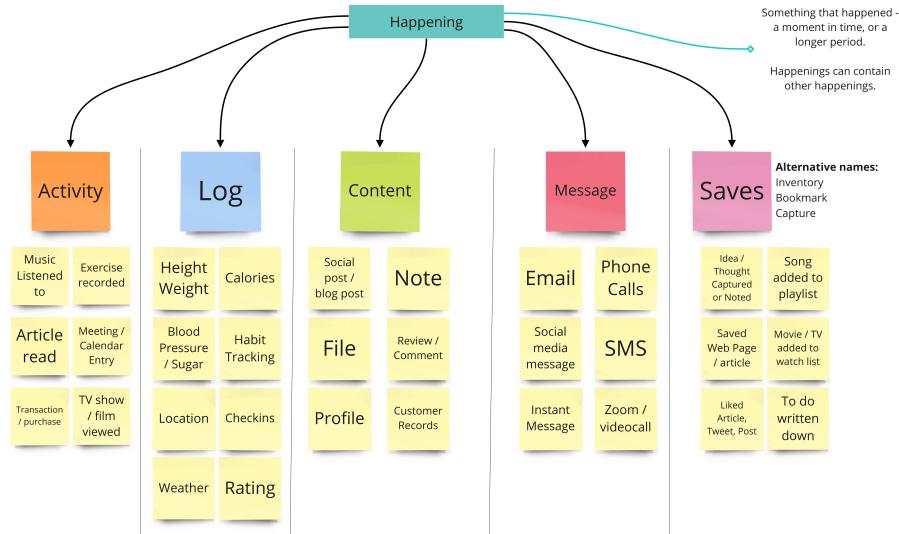


Figure 14: Figure X: Happenings

It can also be useful to model the different attributes of data in terms of what can be done to it.

We can imagine a simplified model of presenting information to users:

1.4.3.3 Opportunity 8: Algorithmic Meaning Extraction and Learning Systems

Helps with: engagement/efforts

Key idea to share: that the system should try and automatically associate data to entities. also: calendar/contact as start point. conjecture and assertion to reduce effort. learning, correcting, like an assistant [ref] word2vec as an example of the sophistication of what is being done and harnessed for provider purposes[REF]

[citation: diagram by Alex Peysakhovic, from CS 4803 / 7643: Deep Learning Guest Lecture: Embeddings and word2vec, a guest lecture at Georgia Tech by Facebook AI research engineer Ledell Wu, given Feb 18th 2020.

https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_EMBEDDING_WORLD2VEC_FINAL_VERSION.pdf, archived at https://web.archive.org/web/20211018015836/https://www.cc.gatech.edu/classes/AY2020/cs7643_spring/slides/L13_EMBEDDING_WORLD2VEC_FINAL_VERSION.pdf

one possible flow of how to identify data

show how attributes types etc (ref back to earlier diagrams) can be detected:

What can we let users do with data? Different types of data have different properties...

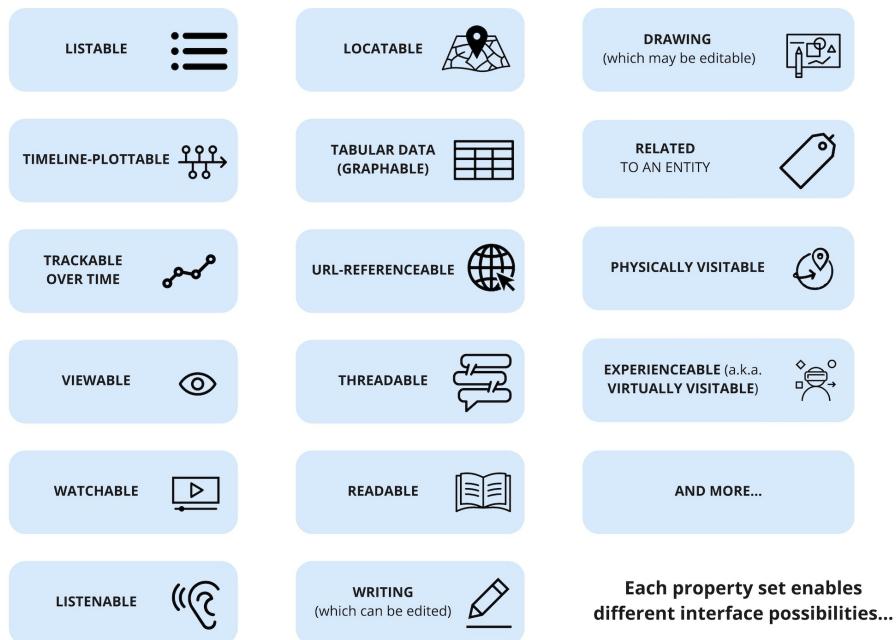


Figure 15: Figure X: Attributes of Data

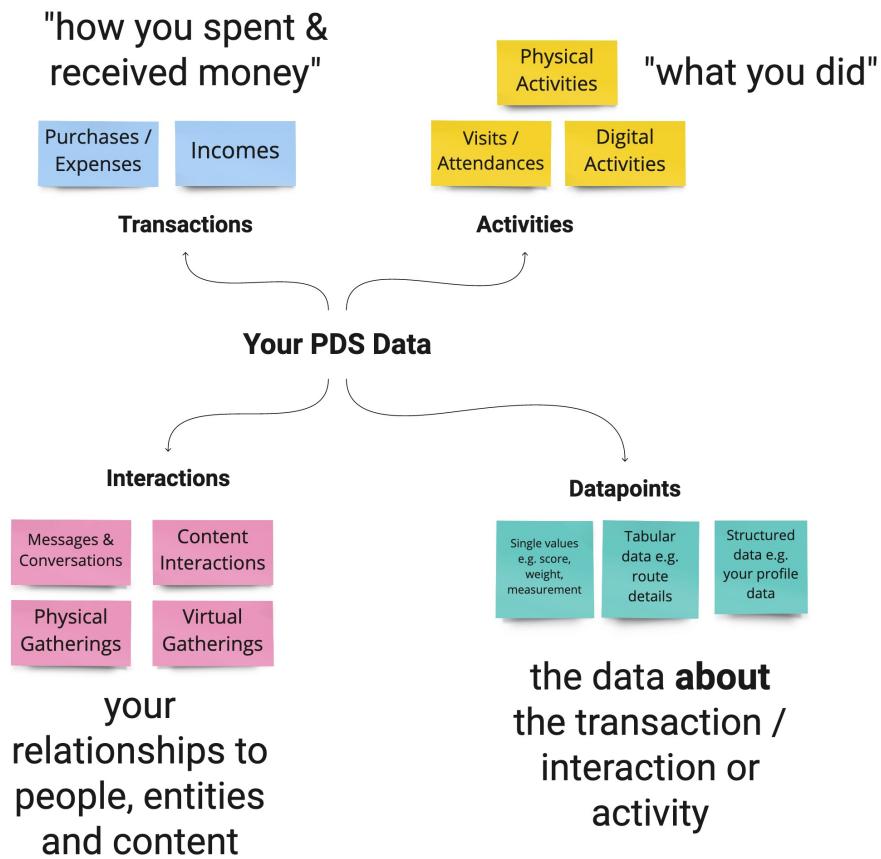


Figure 16: Figure X: Simple Life Information Presentation Model

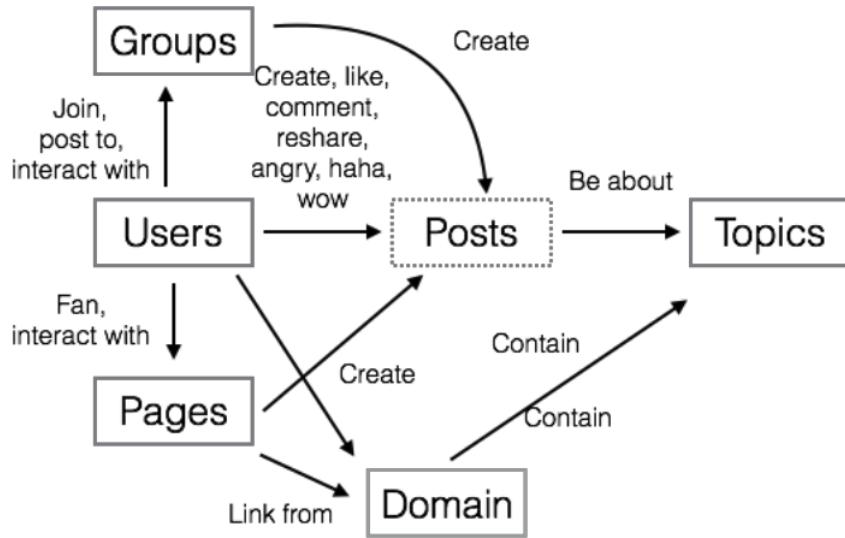


Figure 17: Figure X: A design for Facebook’s world2vec model, semantically modelling data from social media posts on Facebook

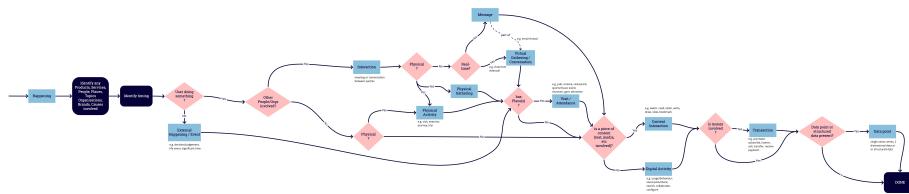


Figure 18: Figure X: Determining the nature of a piece of data

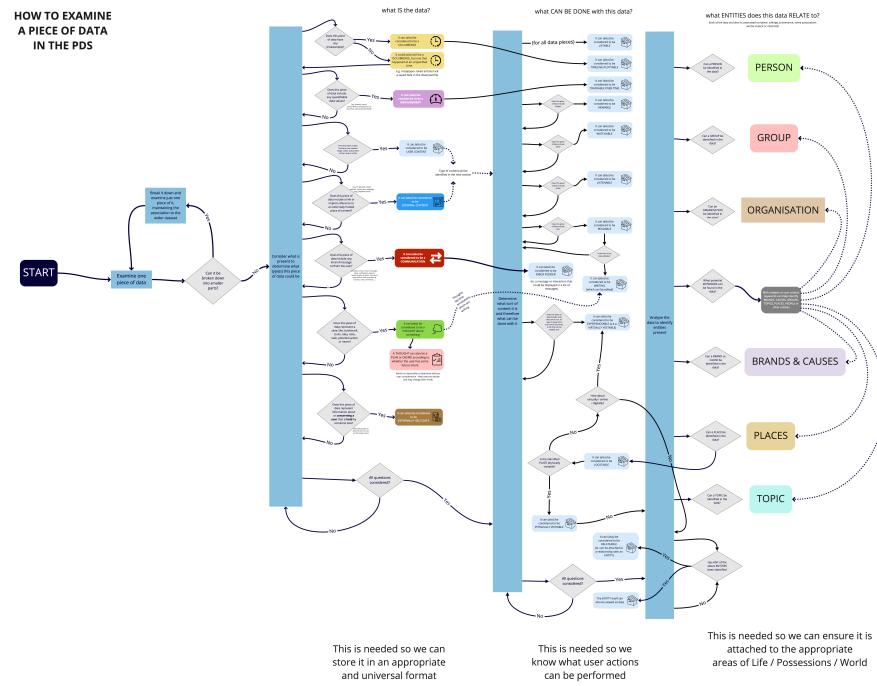


Figure 19: Figure X: Identifying the attributes of data

show how different types of entity can be identified. important to establish associations.

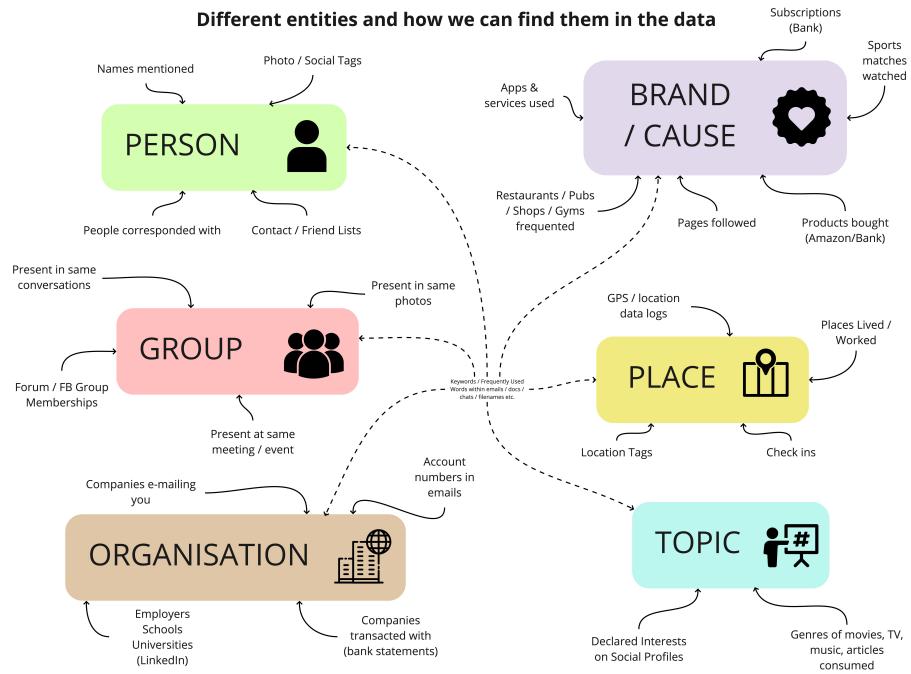
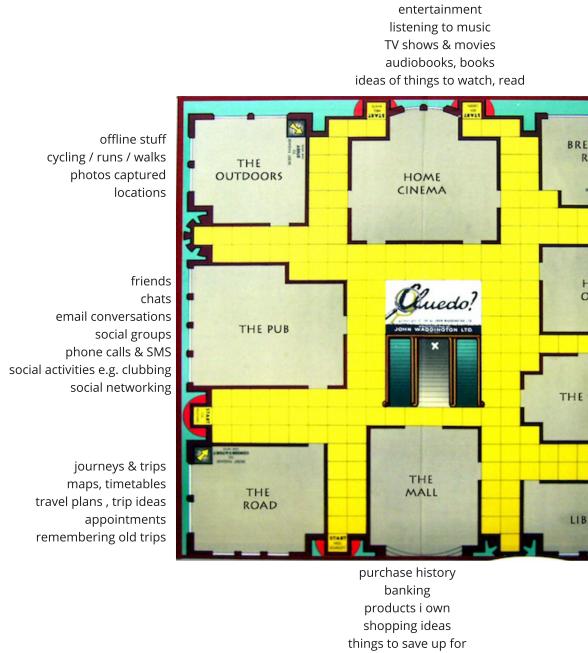


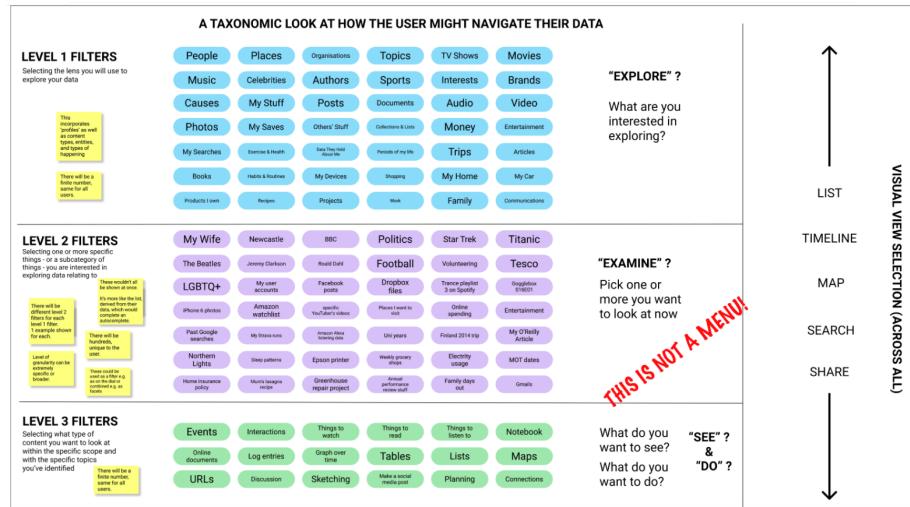
Figure 20: Figure X: Identifying entities in data

1.4.4 Create New Information Interfaces



1.4.4.1 Opportunity 8: Life Information interfaces (add credit for cluedo board)

Life partitioning would allow conceptually the user to navigate information according to what semantic concepts it is or relates to.



(add credit for cluedo board)

A mockup of how this might look in a user interface (ignore first frame for now)

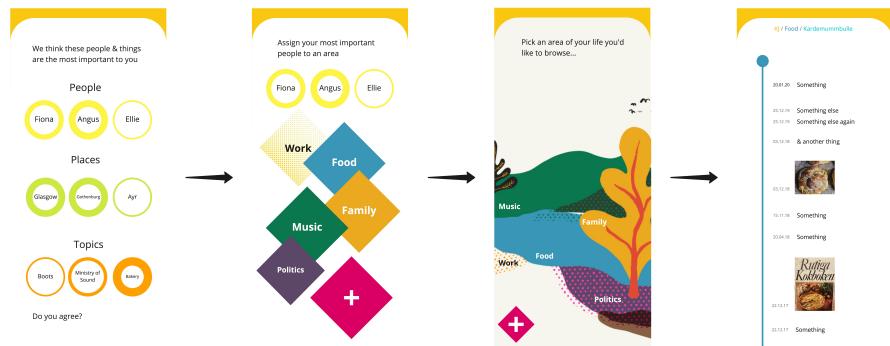


Figure 21: Figure X: Browsing By Areas of Life

A mockup of a life interface dashboard (by Alex Ballantyne)

The screenshot shows the Zero app's main dashboard with several sections:

- Insights:** Includes a purple icon with a person walking and the text "Amazon thinks your favourite film genre is Adventure?", followed by a green icon with a tree and the text "Netflix believes you love to binge Fantasy?".
- Health Profile:** Shows "Steps 8,500 to go" (4,600 taken), "Sleep 8h 22min (8h 27min active target)", and "Calories 1123.01 (1090.00 goal)".
- Financial Profile:** Displays "Balance £1277.00 Across all key accounts" and "Today £35.00 To get" (with a note: "£30.00 To get").
- Savings:** Shows "Savings £623.45 (£800 target) (£154.00 to reach)".
- Your donated data:** Features a blue "NHS" logo with a note: "You have agreed to give the NHS access to your physical activity data as part of a research study into type 2 diabetes". It also shows "96% done" and "Rochester location". A button "See all donation opportunities" is present.
- Media Profile:** Shows "1h 43min today" (with icons for Netflix, YouTube, BBC News, and BBC iPlayer) and a list of media consumption: "Netflix 1h 43min", "YouTube 22m", "BBC News 18m", "BBC iPlayer 12m", and "Facebook 10m".
- Social Profile:** Shows "Daily average on social media Platforms 2h 12m" (with icons for Twitter, LinkedIn, Facebook, and Instagram) and a list of social media activity: "Twitter 1h 12m", "LinkedIn 22m", "Facebook 18m", "Instagram 12m", and "Twitter 10m".
- How about trying...**: A section with a note: "Put your financial and travel data to use, how about trying Zero and see how you can lower your carbon impact." It features a "Zero" logo with the text "Join with friends to track and lower your carbon impact." and a "Get" button.
- Activities:** Shows activity data from "Mondays" and "Spotify" (25mins), "Tuesdays" and "Netflix" (61mins). A "See all activity" button is available.
- From your archive:** Highlights from this time last year: "05/02/2020". It shows three images: "Giant waves 56 miles", "Surfing 13hr", and "Surf wave 121 miles".
- Delete my data:** A note: "Deleting your data will remove all your data that you've put in, as well as all connections you've set up." and a red "Delete my data" button.

(to do: make this fit on a portrait page)

1.4.4.2 Opportunity 9: Ecosystem Detection & Visualisation Key idea: detecting the ecosystem. Example: the subscription detector.

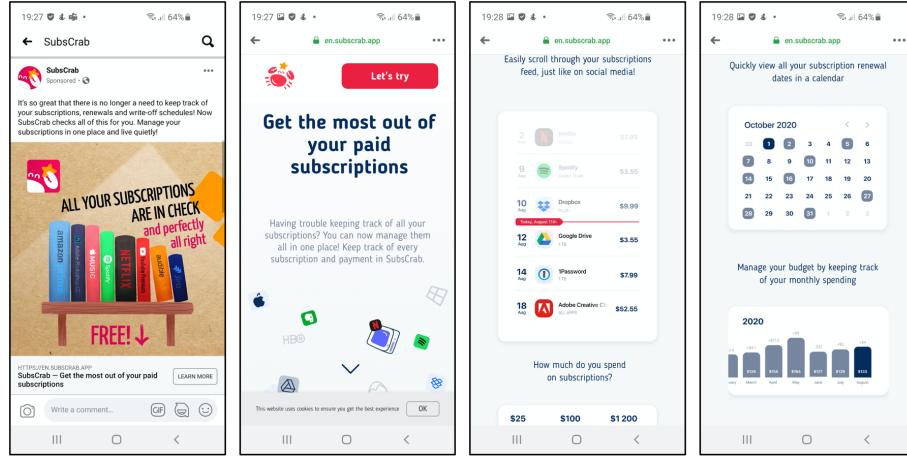


Figure 22: Figure X: Subscrab: An example application for ecosystem detection and visualisation

1.4.4.3 Opportunity 10: Inclusive Information Flows Key idea: Rivers of flowing info. Including people.

1.4.5 Create New Capabilities

1.4.5.1 Opportunity 11: Exploratory Actions & Asking Tools It is important to think about the capabilities people will have (expand on and map this back to all the PIM calls in 2.2.2)

refer back to dashboard mockup & data needs to be interrogable and malleable. What if there aren't visualisations for your questions. But also, visualisations raise questions

Here are a series of data cards we used in a user research activity at BBC. (initially conceived by me, condensed and reworked by myself in collaboration with Chris Gameson)

Key idea: verbs. types of question. asking tools not just predetermined insights.

1.4.5.2 Opportunity 12: Self Profiling & Curating your Digital Self Key idea: self profiling (a la BBC) but also more generally, the "I want a bicycle" VRM idea (unless already covered)

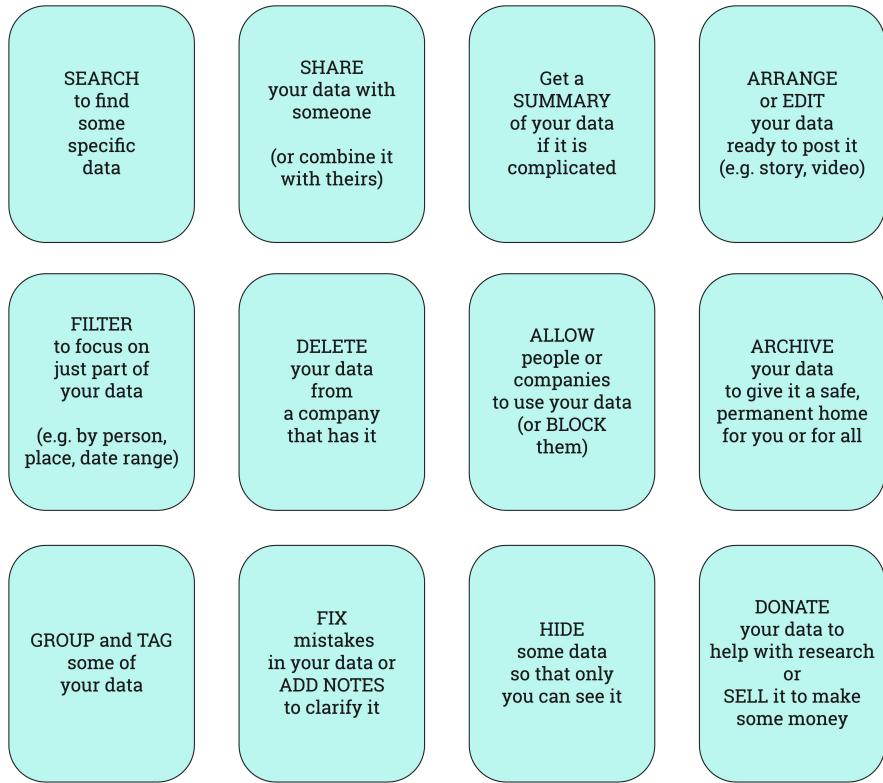


Figure 23: Figure X: Data actions

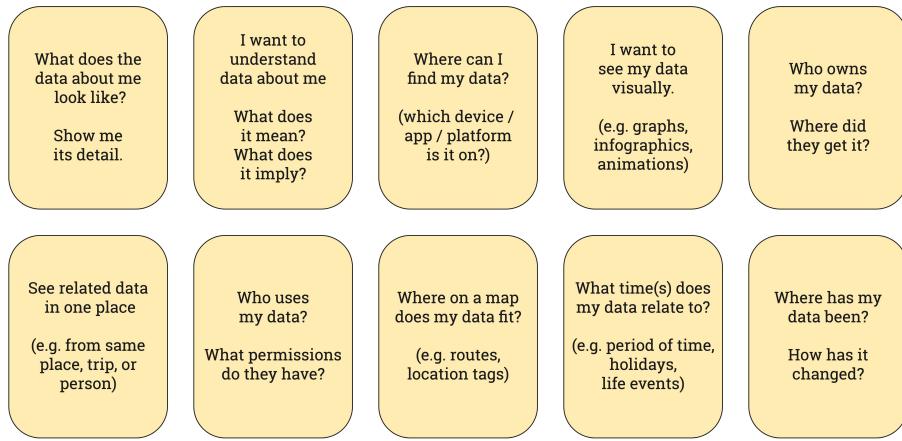


Figure 24: Figure X: Data questions

1.4.6 Influence & Motivate

1.4.6.1 Opportunity 13: Regulating the Information Landscape Key point: regulation of *landscape* -> new moves DSA maybe also some ref back to end C5 and to the GDPR Guidelines

1.4.6.2 Opportunity 14: Information Unification and Schematisation Key idea: generalised types of data (refer back to relevant model above) [reference to Solid Shapes etc]

1.4.6.3 Opportunity 15: Life/Ecosystem Information as Boundary Objects (Multi-stakeholder Design) [POSSIBLY CUT THIS ONE]

Key idea: Data as bringing different people together , ref living lab. also use of cards in my research and at BBC (and ref Urquhart?)

1.4.6.4 Opportunity 16: The Business Value of Transparency and Human Centricity key idea: tackling resistance reducing liability, improved consent . less waste cost on broadcast advertising (could express it as a development of all the ad personalisation today). selling the benefits.

1.4.7 Educate & Empower

1.4.7.1 Opportunity 17: Life Information & Personal Data Ecosystem Literacy Key point: define it, distinguish it from technical skills/literacy as well as from numbercrunching Literacy

1.4.7.2 Opportunity 18: Individual Discovery: Mapping Your Personal Data Ecosystem Key point: empowering individuals as investigators. Can help them with tools or learning programs.

1.5 Thesis Conclusion

[reiterate the answer to the question - the key 4 roles, 3 capabilities and N approaches needed for better human data relations] [how to group together the approaches]

[clarify the contribution of the thesis, with backreferences - 2 case studies, RQ answers, and the HDR roadmap]

[highlight future value/societal implications of the work]

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.

- 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>.
- Bowyer, A. (2011) 'Why files need to die'. Available at: <http://radar.oreilly.com/2011/07/why-files-need-to-die.html>.
- 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/#>.
- 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.
- 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.
- Es, M. van, Guijt, I. and Vogel, I. (2015) 'Hivos ToC Guidelines: Theory of Change Thinking in Practice'. The Hague, The Netherlands: Hivos.
- 'Explainable AI: Making machines understandable for humans' (no date). Available at: <https://explainableai.com/> (Accessed: 16 June 2022).
- Foucault, M. (1975) 'Discipline and punish: The birth of the prison', *New York*. Pantheon Books.
- 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.
- 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>.
- 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 *SIGDOC'12 - proceedings of the 30th ACM international conference on design of communication*. New York, New York, USA: ACM Press, pp. 279–287. doi: 10.1145/2379057.2379109.
- 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.
- 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.
- 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.
- 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.
- Le Dantec, C. A. (2016) *Designing publics*. MIT Press.
- Mahieu, R. L. P., Asghari, H. and Van Eeten, M. (2018) 'Collectively exercising the right of access: Individual effort, societal effect', *Internet Policy Review*, 7(3), pp. 1–23. doi: 10.14763/2018.3.927.
- 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>.
- 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.
- 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.
- 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).
- Neff, G. (2013) 'Why Big Data Won't Cure Us', *Big Data*, 1(3), pp. 117–123. doi: 10.1089/big.2013.0029.
- 'Open rights group: Who we are' (no date). Available at: <https://www.openrightsgroup.org/who-we-are/> (Accessed: 16 June 2022).
- 'Recursive Public (Discussion Page)' (no date). Available at: https://wiki.p2pfoundation.net/Recursive_Public (Accessed: 16 June 2022).
- 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.
- 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.
- 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/ToC Basics.pdf.
- Taylor, L. (2017) 'What is data justice? The case for connecting digital rights and freedoms globally', *Big Data and Society*, 4(2). doi: 10.1177/2053951717736335.
- 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.