

Understanding and Improving Human Data Relations

Alex Bowyer

- [7 Discussion II: Designing and Pursuing Better Human Data Relations](#)
 - [7.1 Introduction & Background](#)
 - [7.1.1 External Research Settings](#)
 - [7.1.2 Attribution of Insights](#)
 - [7.1.3 Theories of Change](#)
 - [7.1.4 Better Human Data Relations as a Recursive Public](#)
 - [7.2 Expanding the Concept of Human Data Relations](#)
 - [7.2.1 The Role of Personal Data](#)
 - [7.2.2 Human Data Interaction or Human Information Interaction?](#)
 - [7.2.3 The Two Distinct Motivations for Human Data Relations](#)
 - [7.2.3.1 Life Information Utilisation](#)
 - [7.2.3.2 Personal Data Ecosystem Control](#)
 - [7.3 Obstacles to Better Human Data Relations and Resulting Insights](#)
 - [7.3.1 Data Awareness & Understanding](#)
 - [7.3.1.1 Obstacle 1: Data Legibility](#)
 - [7.3.1.2 Insight 1: Life Information Makes Data Relatable](#)
 - [7.3.1.3 Obstacle 2: The Personal Data Diaspora](#)
 - [7.3.1.4 Insight 2: Ecosystem Information is an antidote to Digital Life Complexity](#)
 - [7.3.2 Data Useability](#)
 - [7.3.2.1 Obstacle 3: Data isn't free](#)
 - [7.3.2.2 Obstacle 4: Data that is Unmalleable and Non-Interrogable](#)
 - [7.3.2.3 Insight 3: Life Information & Ecosystem Information as a Material](#)
 - [7.3.3 Ecosystem transparency](#)
 - [7.3.3.1 Obstacle 5: A Complex and Invisible Data Ecosystem](#)
 - [7.3.3.2 Obstacle 6: A Lack of Metadata](#)
 - [7.3.3.3 Insight 4: Data with Provenance](#)
 - [7.3.4 Ecosystem negotiability](#)
 - [7.3.4.1 Obstacle 7: Provider Hegemony](#)
 - [7.3.4.2 Insight 5: The Four Levers of Power](#)
 - [7.3.4.3 Obstacle 8: Closed, Insular and Introspective Practices](#)
 - [7.3.4.4 Obstacle 9: A Trend of Actively Diminishing Individuals' Agency](#)

- [7.3.4.5 Insight 6: The Inaccessible Data Self](#)
- [7.3.5 General Human Challenges](#)
 - [7.3.5.1 Obstacle 10: A lack of demand and HDR motivation](#)
 - [7.3.5.2 Insight 7: New Life Capabilities](#)
 - [7.3.5.3 Obstacle 11: Digital Life Management is hard work](#)
 - [7.3.5.4 Insight 8: Always Serve a Need](#)
- [7.3.6 General Technical challenges](#)
 - [7.3.6.1 Obstacle 12: A lack of Interoperability](#)
 - [7.3.6.2 Obstacle 13: Insufficient machine understanding of human data](#)
 - [7.3.6.3 Insight 9: The Power of Semantic Analysis and Information Standards](#)
- [7.4 Working in the present to build the future: Current Work & Future Opportunities](#)
 - [7.4.1 Learn & Discover](#)
 - [7.4.1.1 Opportunity 1: Auditing Data Holders](#)
 - [7.4.1.2 Opportunity 2: Collective Investigation](#)
 - [7.4.1.3 Opportunity 3: A 'Data Understanding' Industry](#)
 - [7.4.2 Defend the Status Quo and Push Back](#)
 - [7.4.2.1 Opportunity 4: Tapping and Monitoring The Seams](#)
 - [7.4.3 Create New Structures and Systems](#)
 - [7.4.3.1 Opportunity 5: A central home for your personal data](#)
 - [7.4.3.2 Opportunity 6: Algorithmic Meaning Extraction](#)
 - [7.4.4 Create New Information Interfaces](#)
 - [7.4.4.1 Opportunity 7: Life Information interfaces](#)
 - [7.4.4.2 Opportunity 8: Ecosystem Detection & Visualisation](#)
 - [7.4.4.3 Opportunity 9: Information Flows](#)
 - [7.4.5 Create New Capabilities](#)
 - [7.4.5.1 Opportunity 10: Exploratory Actions & Asking Tools](#)
 - [7.4.5.2 Opportunity 11: Self Profiling & Curating your Digital Self](#)
 - [7.4.6 Influence & Motivate](#)
 - [7.4.6.1 Opportunity 12: Regulating the Information Landscape](#)
 - [7.4.6.2 Opportunity 13: Information Unification and Schematisation](#)
 - [7.4.6.3 Opportunity 14: Life/Ecosystem Information as Boundary Objects \(Multi-stakeholder Design\)](#)
 - [7.4.6.4 Opportunity 15: The Business Value of Transparency and Human Centricity](#)
 - [7.4.7 Educate & Empower](#)
 - [7.4.7.1 Opportunity 16: Life Information literacy](#)
 - [7.4.7.2 Opportunity 17: Personal Data Ecosystem Literacy](#)
 - [7.4.7.3 Opportunity 18: Individual Discovery: Mapping Your Personal Data Ecosystem](#)
- [7.5 Thesis Conclusion](#)

- [Bibliography](#)

7 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

7.1 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. 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.

In section 7.1.1 these external activities 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.

7.1.1 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).

7.1.2 Attribution of Insights

While this thesis is my own original work, 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, Alex Ballantyne and others at BBC R&D;
- Paul-Olivier Dehay and Jessica Pidoux at Hestia.ai;
- Stuart Wheeler 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.

7.1.3 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 is 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.

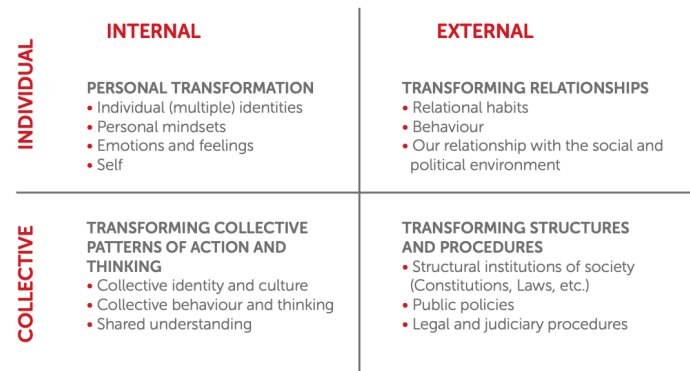


Figure 29: The Four Dimensions of Change¹

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 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.

7.1.4 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) 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?”

7.2 Expanding the Concept of Human Data Relations

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.

7.2.1 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 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.

Way of thinking about data	Explanation & Implications
Data as part of oneself	A photo or recording of you, or a typed note or search that popped into your head could be deeply personal. This lens on data highlights issues around emotional attachment/impact, privacy, and ethics.
Data as memory	Data can be considered as an augmentation to one's memory, a digital record of your life. This lens facilitates design thinking around search and recall, browsing, summarising, cognitive offloading, significance/relevance, and the personal value of data.
Data as creative work	Some of the data we produce (e.g. writing, videos, images) can be considered as an artistic creation. This lens enables thinking about attribution, derivation, copying, legacy and cultural value to others.
Data as new information about the world	Data created by others can inform us about previously unknown occurrences in our immediate digital life or the wider world. This lens is useful for thinking about discovery, recommendations, bias, censorship, filter bubbles, and who controls the information sources we use, as well as who will see and interpret data that we generate and what effects our data has on others.
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.

7.2.2 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 useability, 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.

7.2.3 The Two Distinct Motivations for Human Data Relations

By making this distinction between the two types of information which people might interact with, and considering the six wants in Chapter 6, it becomes clear that there

are two very different reasons why people might want better HDR:

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

The two distinct processes that individuals might go through in pursuit of these motives are exemplified in Figure 30. PDEC is a process of holding organisations to account over and managing *what happens to personal data*, often regardless of what it means, whereas LIU is more concerned with *what the data means* and its inherent value as encoded life information, regardless of where it is stored and how it is used³. This novel way of modelling the motivations for data interaction were first proposed in my 2021 workshop paper (Bowyer, 2021).

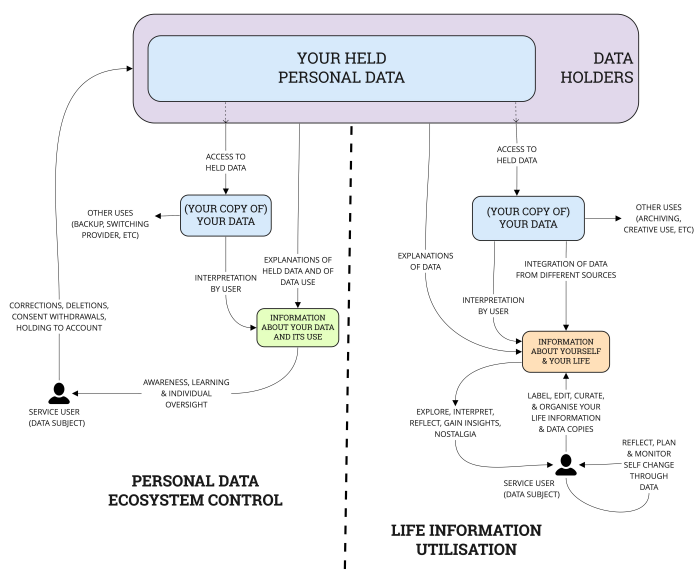


Figure 30: The Two Motivations for HDR: Controlling your personal data ecosystem and utilising your information about your life, with 'idealised'⁴ processes illustrated

7.2.3.1 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, journaling, scrapbooking, arts and crafts.

In the LIU context, the most important wants to focus on improving are *data understandability* (6.1.2) and *data useability*¹³ (6.1.3), which relate closely to the HDI concepts of *legibility* and *agency* respectively.

7.2.3.2 Personal Data Ecosystem Control

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

7.3 Obstacles to Better Human Data Relations and Resulting Insights

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: General Human Challenges
- 7.3.6: General Technical Challenges

[Add diagram here]

7.3.1 Data Awareness & Understanding

7.3.1.1 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.

7.3.1.2 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.

7.3.1.3 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.

7.3.1.4 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.

7.3.2 Data Useability

7.3.2.1 Obstacle 3: Data isn't free

Almost all data is constrained in some way, limiting its useability. 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 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.

7.3.2.2 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 to 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.

7.3.2.3 Insight 3: Life Information & Ecosystem Information 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.

7.3.3 Ecosystem transparency

7.3.3.1 Obstacle 5: A Complex and Invisible Data Ecosystem

7.3.3.2 Obstacle 6: A Lack of Metadata

7.3.3.3 Insight 4: Data with Provenance

7.3.4 Ecosystem negotiability

7.3.4.1 Obstacle 7: Provider Hegemony

7.3.4.2 Insight 5: The Four Levers of Power

7.3.4.3 Obstacle 8: Closed, Insular and Introspective Practices

7.3.4.4 Obstacle 9: A Trend of Actively Diminishing Individuals' Agency

7.3.4.5 Insight 6: The Inaccessible Data Self

7.3.5 General Human Challenges

7.3.5.1 Obstacle 10: A lack of demand and HDR motivation

7.3.5.2 Insight 7: New Life Capabilities

7.3.5.3 Obstacle 11: Digital Life Management is hard work

7.3.5.4 Insight 8: Always Serve a Need

7.3.6 General Technical challenges

7.3.6.1 Obstacle 12: A lack of Interoperability

7.3.6.2 Obstacle 13: Insufficient machine understanding of human data

7.3.6.3 Insight 9: The Power of Semantic Analysis and Information Standards

7.4 Working in the present to build the future:

Current Work & Future Opportunities

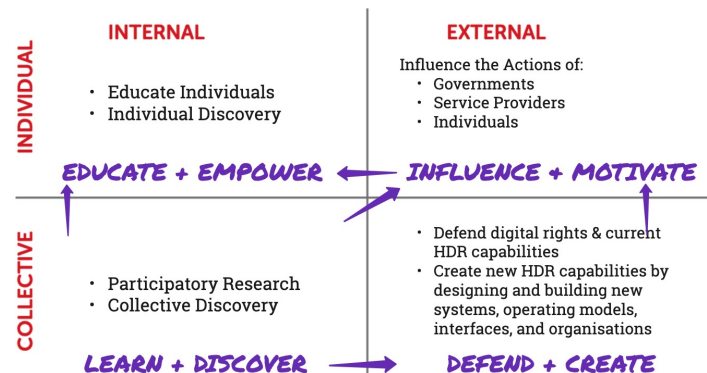


Figure 31: ToC chart showing Trajectories of Change for improving Human Data Relations

In Figure 31, 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, 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 Back
 - 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 be 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?]

7.4.1 Learn & Discover

7.4.1.1 Opportunity 1: Auditing Data Holders

7.4.1.2 Opportunity 2: Collective Investigation

7.4.1.3 Opportunity 3: A 'Data Understanding' Industry

7.4.2 Defend the Status Quo and Push Back

7.4.2.1 Opportunity 4: Tapping and Monitoring The Seams

7.4.3 Create New Structures and Systems

7.4.3.1 Opportunity 5: A central home for your personal data

7.4.3.2 Opportunity 6: Algorithmic Meaning Extraction

7.4.4 Create New Information Interfaces

7.4.4.1 Opportunity 7: Life Information interfaces

7.4.4.2 Opportunity 8: Ecosystem Detection & Visualisation

7.4.4.3 Opportunity 9: Information Flows

7.4.5 Create New Capabilities

7.4.5.1 Opportunity 10: Exploratory Actions & Asking Tools

7.4.5.2 Opportunity 11: Self Profiling & Curating your Digital Self

7.4.6 Influence & Motivate

7.4.6.1 Opportunity 12: Regulating the Information Landscape

7.4.6.2 Opportunity 13: Information Unification and Schematisation

7.4.6.3 Opportunity 14: Life/Ecosystem Information as Boundary Objects (Multi-stakeholder Design)

7.4.6.4 Opportunity 15: The Business Value of Transparency and Human Centricity

7.4.7 Educate & Empower

7.4.7.1 Opportunity 16: Life Information literacy

7.4.7.2 Opportunity 17: Personal Data Ecosystem Literacy

7.4.7.3 Opportunity 18: Individual Discovery: Mapping Your Personal Data Ecosystem

7.5 Thesis Conclusion

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

[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

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](https://doi.org/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](https://doi.org/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).

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

Harris, T. (2013) 'A Call to Minimize Distraction Respect Users' Attention'. Available at: <http://www.minimizedistraction.com/>.

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.

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](https://doi.org/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.

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](https://doi.org/10.1016/j.lisr.2008.07.001).

Mortier, R. *et al.* (2013) 'Challenges & opportunities in human-data interaction', *University of Cambridge, Computer Laboratory*. Citeseer. doi: [10.5210/fm.v17i5.4013](https://doi.org/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](https://doi.org/10.2139/ssrn.2508051).

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

'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).

Taplin, D. H. and Clark, H. (2012) 'Theory of change basics: A primer on theory of change', *ActKnowledge*, p. 9. Available at: http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.

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](https://doi.org/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](https://doi.org/10.1145/2851581.2886436).

1. Diagram 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).↩
2. The 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'.↩
3. Of 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.↩
4. The 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]↩