Understanding and Improving Human Data Relations

Alex Bowyer

7 Discussion II: Designing and Pursuing Better Human Data Relations

7.1 Chapter Overview

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 answer RQ3, which asks the practical question *'What challenges and opportunities are relevant when pursuing better Human Data Relations?'*. Throughout Context Three (3.4.3), which consists of my embedded work in four industrial and academic research projects during the course of my PhD, I have been able to explore how better Human Data Relations (HDR) can be pursued in practice. This chapter aims to provide a roadmap, illustrated with real world insights and activities, and building upon the theoretical insights from the Case Studies, to inform the design of future research, innovation and policy.

In section 7.2, the concept of HDR is expanded to identify its two key purposes, which are used in section 7.3 as a framing to present identified obstacles and opportunities. Section 7.4 summarises the roadmap in a Theory of Change context, and section 7.5 concludes the thesis, summarising its contributions and answering the overall research question.

7.1.1 Practical Research Settings

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

1. **BBC R&D's Cornmarket Project**, which explored through UX design 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 data rights and 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 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 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 through close collaboration,

discussion and ideation with other researchers, including:

- Jasmine Cox, Suzanne Clarke, Tim Broom and Alex Ballantyne at BBC R&D;
- Paul-Olivier Dehaye and Jessica Pidoux 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

Due to these collaborations and the ongoing and parallel nature of all 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 each project's trajectory 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.

7.2 Expanding the Concept of Human Data Relations

Chapter 6 established six 'wants' in HDR: visible, understandable and usable data; process transparency and individual oversight and 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 purposes, to which those six wants apply differently, therefore it is worth spending the time to develop the concept of HDR further.

7.2.1 Human Data Interaction or Human Information Interaction?

Human Data Interaction (HDI) concerns the way people interact with data. Mortier *et. al* (Mortier *et al.*, 2013, 2014) define 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. Originating in library sciences, the parallel field of Human Information Interaction (HII) is well established, considering the way humans relate to information without regard to the technologies involved (Marchionini, 2008). As William Jones observes, these two fields should be considered together (Jones *et al.*, 2006). As DIKW theory highlights (see 2.1), 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 need to relate both to data, and to 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.

In considering Human Data Relations, there are in fact three distinct aspects to

consider:

1. *data* - the stored digital artifacts held by organisations for algorithmic processing, 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 interpretation (or in some organisations' case, through analytical inference).
3. *information about data* (also identified in Table 9 / 5.3.1 as *metadata*) - facts about the data, such as where it has been stored, who has accessed it, how it was collected, or when it has been shared externally.

7.2.2 The Two Distinct Purposes of 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 - to use :

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 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 purposes are exemplified in Figure 29. 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*, regardless of where it is stored and how it is used[1]. The two purposes of HDR are further detailed in (Bowyer, 2021).

[INSERT DIAGRAM - Figure 29]

7.2.2.1 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 (Bowyer *et al.*, 2018). While the existence of digitally-encoded information clearly unlocks new possibilities, LIU has existed in some form throughout modern 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 Understandability of data (6.1.2) and Useability of data (6.1.3), which relate closely to the HDI concepts of *legibility* and *agency* respectively. Obstacles and opportunities for Understandability and Useability will be explored in 7.3.1.1 and 7.3.1.2 respectively.

7.2.2.2 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). It was only when organisations began to collect and store facts about people as a substitute for direct communication and involvement that it became 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 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 will be handled collectively as *"Ecosystem Transparency"*, and the latter two as *"Ecosystem Negotiability"* (drawing on the HDI concept of *negotiability*). Obstacles and opportunities for improving HDR in these two grouped areas will be explored in 7.3.2.1 and 7.3.2.2 respectively.

7.2.3 The Role of Personal Data

In the modern world, where almost anything can be encoded as data, and many previously analogue objects and activities now have digital equivalents, the concept of data has become broad and hard to pin down. It is worthwhile in considering HDR (and indeed part of this thesis's main RQ) to expand upon what role data plays in people's lives. Through the Case Studies and Research Settings and my prior learning, I have identified 8 distinct ways in which people might relate to their data, which are presented in Table X

Table 15. **Eight lenses for thinking about 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.

Way of thinking about data	Explanation & Implications
Data as a source of knowledge 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 use it yourself for reflection, self-improvement and planning.
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, 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 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 our information sources
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 their data into curated collections, or use it to convey facts, ideas or to evoke an emotional impact. This lens is useful to consider data uses such as lists, annotation, curation, editing, remixing, 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 in any or all of these ways at any given time, and any data interaction design should take these into account.

Looking behind these lenses on what data *is* to people, 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 actions of others that may affect you;
3. Data has a role as **a material with which to make changes** in your life;
4. Data has a role as **a means to monitor** data holders' behaviours, digital influences upon you and changes within your life.

7.3 Answering RQ3: What are the challenges and opportunities?

[What challenges and opportunities are relevant when attempting to meet the six wants of human data relations? - How the six wants fit into these two purposes]

7.3.1 Challenges and Opportunities in Life Information Utilisation

7.3.1.1 Understandable Data

7.3.1.1.1 Obstacles to Data Understandability

[Meaningfulness / relatability -> relate it to people/places/events]

[Context - Life - > need life interfaces]

[Information within Data -> Lack of Visualisations and Tools]

[Scatteredness -> holistic/unification, place to centralise]

[Complexity -> common formats/abstractions/summarisations]

7.3.1.1.2 Improving Data Understandability

[Personal data Stores as place to put stuff]

[Build systems to extract meaning - interpreting and combining signals]

[Use standards & semantics to convert data to life information]

[presenting and visualising life information]

7.3.1.2 Useable Data[2]

7.3.1.2.1 Obstacles to Data Useability

[Trapped Data -> Force unlocking of data through technical means or regulatory influence]

[Integration challenges -> Need to be able to bring data together and connect and

combine]

[Lack of malleability -> need to be able to slice/group/view from different perspectives]

[inability to investigate -> enable questions, comparisons, investigations etc]

7.3.1.2.2 Improving Data Useability

[supporting useful actions on data - filtering, referencing, cross referencing, conjecturing/whatiffing - data action verbs]

[data as material, interface features as tools to use that material]

[supporting appropriation, annotation, organisation, curation, use & re-use]

[temporal, entity-based/relational and geographical exploration]

[support goal setting, tracking and reflection]

[an information operating system]

[asking tools rather than answers or insights]

7.3.1.3 Other Factors in Life Information Utilisation

[Motivation -> Showing the potential]

[Effort -> doing as much as possible automatically, conjecture and assertion over blank pages. training rather than meticulous instrcution.]

[how the other wants fit in, visibility as it pertains to Life info, transparency/oversight/involvement etc]

[agency over trapped data (by tech or by companies (lead into next)]

7.3.2 Challenges and Opportunities in Personal Data Ecosystem Control

7.3.2.1 Ecosystem Transparency

7.3.2.1.1 Obstacles to Ecosystem Transparency

[hidden data and closed processes -> closed by default thinking -> encourage or legislate for openness.. e.g. data portability/access rights, rights to explanatione etc, but more needed]

[silos and motives towards closed proprietary systems -> highlight the pains]

[lack of information *about* our data -> awareness and accountability even where access is difficult ->]

[lack of standards, motivations against interoperability -> motivate standards and

unconver opportunities for interoperability]

7.3.2.1.2 Improving Ecosystem Transparency

[ecosystem visualisation and overviews]

[exploiting the seams - the battle for the seams]

[standards creation and the benefits of enabling a 'data understanding' industry]

[regulation - forcing openness transparency and interop. DSA ?]

[collectives - as a means to exert individual power]

7.3.2.2 Ecosystem Negotiability

7.3.2.2.1 Obstacles to Ecosystem Negotiability

[structural power, resource control, centralisation etc -> uneven landscape -> awareness as first step and systemic change needed to change.]

[the four levers of infrastructural power. accumulation of info/surveillance as power. changing available information/actions as power]

[data self affects you but cannot see (proxy for involvement, unseen inferences etc)- > find a way to produce better digital selves]

[Controlling the landscape of what is knowable, and what is do-able -> recognise the importance of free information landscapes, and make them happen through tech or through regulation]

7.3.2.2.2 Improving Ecosystem Negotiability

[better digital selves -> people as source of data. profiles and curated as better representation of self, ref past calls in C4&5 for stewardship, user-contributed data etc]

[collectives, supported by policy [uber, ref GDPR guidelines?]]

[the battle for landscape control - RSS, API, 3P interfaces, etc, Defending The Seams And Protecting Interface Freedom]

[-> exploiting the seams in order to produce new information presentations… ref JE paper (+colin?) -> web aug, firefox containers. Taking Back Power In The Browser, resist moves to apps]

[-> better policies to protect the information landscape? DSA?]

7.3.2.3 Other Factors in Personal Data Ecosystem Control

[complex data ecoystems]

[Inconsistent and difficult data rights offerings]

[Lack of up to date insights / delay]

[data literacy and rights awareness - you should teach this in schools]

[Feeds and flows that loop in the data subject (default not opt in)]

7.4 A Theory Of Change Perspective on Better Human Data Relations

[the four change quadrants for each of the two purposes. diagrams to work out].

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#.

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.

Collins English Dictionary (no date a) 'Useability'. Available at: https://www.thefreedictionary.com/useability.

Collins English Dictionary (no date b) 'Useable'. Available at: https://www.thefreedictionary.com/useable.

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.

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.

Merriam-Webster Dictionary (no date a) 'Usability'. Available at: https://www.merriam-webster.com/dictionary/usability.

Merriam-Webster Dictionary (no date b) 'Usable'. Available at: https://www.merriam-webster.com/dictionary/usable.

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.

Nielsen, J. (2012) 'Usability 101: Introduction to Usability'. Available at: https://www.nngroup.com/articles/usability-101-introduction-to-usability/.

---

1. 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 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 and so is part of the reason why one might want to exert control over data use rather than being part of exploiting data to gain self-insights and personal benefits.↵

2. The words *'usability'* and *'usable'* (spelt without an 'e') most commonly refer to a judgement of the degree to which a website or user interface is easy to use (Nielsen, 2012). Throughout this thesis, I deliberately use the alternative word spellings of *'useability'* and *'useable'* (Collins English Dictionary, no date a, no date b) respectively, to clearly distinguish from this ease-of-use concept and to denote that I am referring a different meaning: the more literal definition, i.e. *"the quality or state of being convenient and practicable for use"* (Merriam-Webster Dictionary, no date a, no date b). Any usages without an 'e' can be taken to refer to the interface ease-of-use concept.↵