Understanding and Improving Human Data Relations

Alex Bowyer

6 Discussion Part 1: An Understanding of Human Data Relations

It will be already evident to the reader that there are significant overlaps and parallels to be drawn across the findings and discursive insights in Case Study One and Two. In this mini-discussion chapter, I will draw on both Chapter 4 and Chapter 5 to produce a unified summary of findings and insights in terms of the first two research subquestions RQ1 and RQ2. To recap on the research objectives expressed in 3.3, these two RQs are:

- RQ1: "What is the human experience of personal data, and what do people want from their data?"
- RQ2: "What role does data play in people's service relationships and how could relationships involving data be improved?"

The answers to these research questions are best expressed as an understanding of individual *wants* relating to data. The word 'want' is used here in a broader sense than its everyday meaning, referring to the *lack* of something that would be beneficial (which may or may not be accompanied by conscious desire). By framing our accumulated understandings from the Case Studies in this way, we are exposing both the problem - the things that individuals do not have or cannot do, while also identifying the goals that any imagined solutions or improvements to the status quo would need to address. It logically follows that any solution that better delivers on individual *data wants* will lead to improved relations between individuals and their data. This is how we can conceptualise "Human Data Relations" as alluded to in the title of this thesis, and indeed this gives us a yardstick against which to understand what "better" means, which will be explored in Chapter 8. *"Human Data Relations"* is a term that I introduce here to expand upon the established theory of Human Data

Interaction (Mortier *et al.*, [2013](), [2014]()) in light of the Case Studies' findings from a more sociotechnical, interpersonal point of view. Humans have two kinds of relationships with data: *direct* interaction (such as through an interface in an app or website) and *indirect* interaction (through interacting with services, providers or individual representatives who themselves have access to personal data about the individual). Thus, Human Data Relations is a term that can encompass both the relationship humans have with their data, but also the relationships they have in which data plays an indirect role. In this context, RQ1 and RQ2 map quite cleanly onto these two types of Human Data Relations, and in answering RQ1 we can identify what people want from direct data relations, while RQ2 helps provide an answer as to what people want from indirect data relations.

6.1 Answering RQ1: What do people want in *direct* data relations?

By comparing and grouping elements of the findings from Case Study One (see 4.3) and from Case Study Two (see 5.4), three distinct data wants are evident when considering *direct* data relations. All data about them needs to be:

1. *Visible*: People need to have knowledge of data about them and an ability to see it and effectively access it;
2. *Understandable*: People need to be able to interpret this data to extract meaningful information from it (and about it), including through visualisations and summaries; and
3. *Useable*: People need to be able to take action upon this data, including exploring it, asking questions of it, using it to serve their own goals, and gaining personal value from it.

These wants are detailed in the following sections:

6.1.1 Visible

Data matters to every individual now, in a way that previously it did not. As described in 2.1.2, the role of data in our everyday lives has changed; data has become a material used by businesses to shape our world. In 2.2.2 and 2.2.4 I outlined how this change has practically manifested itself in people's lives; where once data was viewed as *ours* for our personal reference and use, the emergence of complex multi-party data ecosystems has meant that personal data management has become a sociotechnical & societal problem, not a practical individual one. In the past when businesses were more local, more personal and less data-centric, the data that businesses held about us was minimal and much less significant to our lives than the human relationships we had with those businesses. Businesses grew and data began to be considered as a resource to be processed at scale for customer insight and marketing exploitation, and though we didn't realise it, our need to understand those processes to protect one's own interests, began to grow. In the past, you didn't need to become aware of data storage and use, because it had little effect. Now, data has become a substitute for direct communication with the individual being served, as my research in both public sector ((Bowyer *et al.*, 2018) and 4.3.3) and the private sector (5.4.3.3) has shown. In both domains, people do not have awareness, let alone access, to the extent of data that exists about them. In 5.4.4.1 we saw individuals feeling that companies forced them to hand over data in exchange for service access, and then subsequently maintaining power over them through holding that data, using it to make decisions, and denying them access to that data or

even be clear about what data is held.

Data sacrifice is now required for many services, putting individuals at risk. Be it the personal financial, health and lifestyle data collected on an Early Help assessment form when a family signs up for Early Help support (4.1.2), or the contact details, payment information and preferences provided when individuals register with commercial service providers such as insurance providers or streaming media platforms (5.4.4.1), personal data is required to access services. Consent to hold and use this data is enforced upon signup, through waiver forms or Terms & Conditions agreements. Service providers in both sectors see the acquisition of more personal data as beneficial to their operations, be it support workers wanting to gain more data about families' lives (4.2.3, 4.2.6), or commercial providers using trackers to gain more insight about users that they can exploit for advertising (Binns, 2022). In both domains, this sacrifice is seen to have an emotional effect on people, ranging from curiosity to fear and distrust ((Bowyer *et al.*, 2018) and 5.4.4). Such fears are well-founded, with mistreatment through incorrect data known in both settings (4.2.2, 5.4.4.1). While data holders almost certainly do not *intend* to cause harm, data *can* be 'used against you' (P2's quote in 5.4.4.1) (Kröger, Miceli and Müller, 2021; Strohmayer *et al.*, 2021). In providers' eyes, people are now represented through data. Despite the fact that data is never truly objective (Gitelman, 2013; Taylor *et al.*, 2015) and a recognition (at least on the public sector side) that a data record can never tell the full story (4.2.6 and (Bowyer *et al.*, 2018)), the data record becomes the object to be administered, rather than the individual (Cornford, Baines and Wilson, 2013; Zuboff, 2019), and this in itself creates risk - through mishandling or inaccuracy. Given the data record is

seen as a source of truth (4.1.2, 5.5.3), it is vital that such information remains *fair*, and *accurate*; this is especially important in the commercial sector, where people are only just beginning to become aware of data misuses and data exploitations that are actually happening (Chang, 2018; McNamee, 2019; Zuckerman, 2021; 'Data Brokers: Everything You Need to Know', 2022). Ensuring fairness and accuracy of held data cannot be verified without individuals' awareness of data held about them.

Once data has been sacrificed, it enters a closed and opaque ecosystem, where the individual loses access and becomes unaware of that data's storage and use (Luger and Rodden's 'point of severance' (Luger and Rodden, 2013)). What was previously available for individuals to see becomes inaccessible and invisible. In the Early Help context, this manifested as families having a lack of awareness or direct access to data held about them and having to rely on support workers as gatekeepers to choose to inform or show them aspects of their data (4.1.1,4.5). In the commercial context, the situation is perhaps even worse, as not only is there rarely any kind of data viewing interface, there is not even a gatekeeper who might make people aware of their data and its use; and even if someone becomes motivated to gain awareness, the GDPR leaves them in the dark; in 62% of cases, the data that companies own privacy policies stated they collect, was not returned, and data that was returned was complete in only 22% of cases (5.3.2). In both context, no awareness is gained unless the information is actively sought. This means that the vast majority of people, busy and unaware, remain so. This is problematic because people cannot judge data accuracy or protect themselves from risk, because they may not even be aware of certain data's existence, or be able to access it even

when they are.

Across both contexts, we saw that **people want to see that which is hidden from them**. In the SILVER project (3.4.1.1, 4.2.2), and my prior work with families (Bowyer *et al.* (2018)), and in Case Study One (4.2.6, 4.3.2, 4.4.2), families wanted to see what data was held about them ("what they've got on me"). They wanted to be actively kept informed and to have the ability to see if data was fair and accurate. In the commercial context, the same feelings were found; participants expressed a great desire to see and know what companies are storing about them, especially data collected or inferred about them without their involvement. This is not just a desire, but a need, given that data can have impact on their daily lives as it is used to inform decisions on how content is presented and recommended to them, and what services they are advertised, offered or can access (5.5.3). We also see from Case Study Two that awareness is not just a binary; awareness includes having an appreciation of why the existence and use of certain data is significant and what its implications might be. Article 13.2.f of the GDPR ('Article 13: Information to be provided where personal data are collected from the data subject', 2020) states that, at least in the case of automated decision-making, people are entitled to meaningful information about the significance of the processing of their data, yet such explanations were not given to participants of Case Study Two.

Effective access to held data is a key element of visibility. Having gained awareness of data held and of the significance of it, people want that to be accompanied by meaningful access to the data itself. In Case Study One (4.3.2.1, 4.4.2), we saw families and support workers recognise the need to accommodate the differences in families digital literacy, mental

and physical handicaps, and technology skills while providing them access to the data held about them. This mirrors Gurstein's call for 'effective access for everyone' (Gurstein, 2011), which was detailed in 2.1.4. Four aspects in particular are relevant here: the content and formatting of the data (which should support different levels of linguistic and computer literacy), the capabililities made available in terms of software, hardware and Internet access (sufficiently powerful, sufficiently available and affordable), and skills (ensuring that individuals are able to intpret the data). In Case Study Two, we saw several participants feeling that data was delivered in too-technical formats (5.4.3.2), or that they lacked the skill to properly interpret the data (5.4.3.1). Effective access and interpreting data goes beyond visibility of data and includes understandability, which is explored in the next section.

Visibility of data must be timely, and ongoing. Given the ever-changing nature of data (and indeed of the lives of the people it represents), occasional or one-off access is not sufficient. In Human-Data Interaction theory, this concept is described as having *negotiability* (Mortier *et al.*, 2013, 2014): the ability to re-evaluate data and associated decisions as contexts change externally. It is also mentioned by Gurstein, who points out that time-limited access to data would not be effective (Gurstein, 2011). In the Early Help context, families wanted access to their data outside of support meetings; this implies some sort of self-service interfaces being available, that you can use in your own time rather than being reliant on the support worker as gatekeeper. People wanted to see all data about them directly, through a personal interface, as reflected in their workshop designs (4.3.2.3). This echoed findings of my earlier work with families, which had identified a need for continuing rights and visibility of data over time, in

order enable vigilance over keeping data accurate and meaningful as life changes (Bowyer *et al.* ([2018](#))). Timeliness also implies that access to an up-to-date view of the data does not require special and ongoing effort by the individual, it is always available. Both support workers and supported families saw value in notification feeds about changes to data records (4.3.3.3), so that changes are discovered and can be acted upon without having to wait for the next support meeting. In the commercial context we can see that GDPR provides a form of access that is not at all timely. The 30 day delay on request processing guarantees that data will be out of date by the time it is viewed, and individuals must repeatedly make GDPR requests to maintain an up-to-date view (and in doing so, they the imposition of charges as GDPR states that requests should not be excessive and that fees can be levied for additional copies). This lack of timeliness in the design of GDPR data access motivates my third suggestion to policymakers in 5.5.1, that they should offer ongoing access rather than the one-off delivery of data packages.

For held data to become visible,

7. Systemic support is needed, including governancy, advocacy and asssistance
8. If data is not visible, this leads to subjection and exclusion.

6.1.2 Understandable

1. Access to data is not enough, need to be able to interpret it.
2. Understandable summaries of information content and context are needed.
3. Information and visualisations should be arranged and optimised for understanding.

4. Information becomes most meaningful when it is recognisable and relatable and can be mapped back to life experiences.
5. In practice, ongoing human support is needed to facilitate the understanding of data.
6. If data is not understandable, distrust can arise.

6.1.3 Useable

1. People need to be able to explore and interrogate data and ask questions of it.
2. The data needs to be useable, which means correctly formatted and explained in a portable and standardised form.
3. The data needs to be explorable from a temporal perspective, which implies the ability to interact with the data is necessary.
4. If holders make data unusable, this is a barrier to individual agency and power.
5. Unusable data leads to disengagement and impairs individual independent action in society.

6.2 Answering RQ2: What do people want in *indirect* data relations?

By comparing and grouping elements of the findings from Case Study One (see 4.3) and from Case Study Two (see 5.4), especially in the context of individual relationships with care providers and digital service providers respectively, three distinct data wants are evident when considering *indirect* data relations:

1. *Transparency*: People need to know what data is being collected or held, and critically how it is being used, for

accountability and safety and in order to have trust in data holders;

2. *Oversight*: People need the ability to affect what data is held and how it is used, including reacting to changing circumstances, deleting data or withdrawing consent for certain uses; and

3. *Involvement*: People need to be invited and involved in decision-making based upon their data, so that they are not misrepresented and their needs are not overlooked. This can be aided by collaborative use of data, giving individuals a human point of contact, and consulting the person not just the record.

These wants are detailed in the following sections:

6.2.1 Transparency

1. People need a window into how their data is used; this means transparency not just of data but of processes.
2. Process transparency is required to enable accountability.
3. There is no accountability, processes are not transparent, and thus power remains imbalanced.
4. Even after using available transparency rights, people face an incomplete picture from and of their data.
5. Trust of data holders is needed, and gaps in transparency create distrust and a risk of broken expectations, harming relations.
6. Trust can be gained through transparency, which offers an opportunity to improve relations.
7. Transparency of data and processes enables individual action and facilitates the levelling of power balance.

6.2.2 Oversight

1. Data visibility + Process transparency naturally leads to a desire for individual oversight - if you see something that is 'not right', you are motivated to want to fix it.
2. We need to be able to oversee data use for fairness, accuracy and the reduction of risk.
3. The ability for individuals to meaningfully oversee the use of their data requires governance, so that they can effect the changes they desire.
4. Individual oversight of data use brings the individual back to the centre of their data ecosystem as as an active participant.
5. Given the changing nature of human life, all data is dynamic, consent is never complete, so longitudinal participation and oversight is needed.
6. There is very little oversight and governance available today.
7. Ultimately oversight means having choices, which is essential in the data-centric world. Without individual oversight, there is no choice and people remain powerless.

6.2.3 Involvement

1. There is a need to consider the human behind the data, people are not records and can never be fully reduced to data.
2. Consent to access and use data needs to be dynamic and meaningful, which can only happen through ongoing involvement.
3. A human channel for conversation is needed, to enable explanations, questions, and consultation.
4. Individuals should be consulted in decision-making, which improves accuracy, consideration, and fairness.

5. Effective collaboration can be achieved by bringing data subject and data holder together around the data, using it as evidence (of facts or opinions) and as a boundary object.

6. Being involved means being able to learn and act at any time, including on one's own and away from official contact or interactions with service representatives.

7. Data use enforces an uneasy trust; services need a human face or point of contact, in order to earn trust and improve relations.

8. Without involvement, people can never take a full and equitable role in processes that affect their life.

6.3 Achieving Individual Empowerment

1. Through these summarised insights, I have shown a multifacted set of needs and opportunities around data access and use. While all six of these data wants can produce improvements in their own right, the combination of all six is likely to produce more than the sum of its parts, an empowered form of digital citizen.

2. Giving people a role as co-stewards of their own data and involved in decision-making would be progressive and transformative, and this could be applied in different domains across society.

3. We can envision from this a new fully human-centred (or at least power-balanced) future - cooperative data stewardship and empowered, involved citizens.

Bibliography

'Article 13: Information to be provided where personal data are collected from the data subject' (2020). doi: 10.1093/oso/9780198826491.003.0044.

Binns, R. (2022) 'Tracking on the Web, Mobile and the Internet-of-Things'. Available at: http://arxiv.org/abs/2201.10831.

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.

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

Cornford, J., Baines, S. and Wilson, R. (2013) 'Representing the family: how does the state 'think family'?', *Policy & Politics*, 41(1), pp. 1–19. doi: 10.1332/030557312X645838.

'Data Brokers: Everything You Need to Know' (2022). Available at: https://www.avast.com/c-data-brokers (Accessed: 11 February 2022).

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.

Gurstein, M. B. (2011) 'Open data: Empowering the
empowered or effective data use for everyone?', *First Monday*.
First Monday, 16(2). doi: 10.5210/fm.v16i2.3316.

Kröger, J. L., Miceli, M. and Müller, F. (2021) 'How Data Can
Be Used Against People: A Classification of Personal Data
Misuses', *SSRN Electronic Journal*, (December). doi:
10.2139/ssrn.3887097.

Luger, E. and Rodden, T. (2013) 'An informed view on consent
for ubicomp', in *UbiComp 2013 - proceedings of the 2013 acm
international joint conference on pervasive and ubiquitous
computing*. New York, New York, USA: ACM Press, pp. 529–
538. doi: 10.1145/2493432.2493446.

McNamee, R. (2019) *Zucked: Waking up to the Facebook
Catasrophe*, p. 336.

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.

Strohmayer, A. *et al.* (2021) *Trust and Abusability Toolkit:
Centering Safety in Human-Data Interactions*.

Taylor, A. S. *et al.* (2015) 'Data-in-place: Thinking through the
relations between data and community', in *Conference on
human factors in computing systems - proceedings*, pp. 2863–
2872. doi: 10.1145/2702123.2702558.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile. Available at: https://books.google.co.uk/books?id=W7ZEDgAAQBAJ.

Zuckerman, E. (2021) *Mistrust: Why Losing Faith In Institutions Provides The Tools To Transform Them*. New York, NY, USA: W. W. Norton & Company, pp. 1–3. doi: 10.1017/ipo.2021.30.