

Understanding and Improving Human Data Relations

Alex Bowyer

- [6 Discussion Part 1: An Understanding of Human Data Relations](#)
 - [6.1 Answering RQ1: What do people want in *direct* data relations?](#)
 - [6.1.1 Visible](#)
 - [6.1.2 Understandable](#)
 - [6.1.3 Useable](#)
 - [6.2 Answering RQ2: What do people want in *indirect* data relations?](#)
 - [6.2.1 Transparency](#)
 - [6.2.2 Oversight](#)
 - [6.2.3 Involvement](#)
 - [6.3 Achieving Individual Empowerment](#)
- [Bibliography](#)

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 first 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), supplying your 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 data 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 required for 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 capabilities made available in terms of software, hardware and Internet access (sufficiently powerful, sufficiently available and affordable), and skills (ensuring that individuals are able to interpret 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 and access to 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, systemic support is needed, including governance, advocacy and assistance.

Offering access to data is not solvable at a purely technical level. Even a well-built data interface with 24/7 access would not provide the depth and breadth of visibility people want. As observed in Case Study Two, even those companies that provide instant data access portals such as Google and Facebook did not provide participants with all the data they desired, nor all the answers they sought (5.4.2.3), and most companies offered negligible follow-up support after data had been delivered (5.3.3). Further investigations into data access conducted as part of the #digipower investigation (3.4.3.4)

confirmed that SAR requests and data portals rarely provide insight into some of the most desired types of data including derived and acquired data and data transfers. Effective access and visibility also requires advocacy (Gurstein, [2011](#)): people require support and training to make use of their data.

Furthermore, given the insufficient breadth of returned data from companies (5.3.3, 5.4.2.2) and near-total lack of access to data on the public sector side (4.3.2.2, 4.3.2.3), it is clear that external governance (Gurstein, [2011](#)) to ensure effective access is needed. Without the sort of pressure on data-holders that only policymakers can exert, organisations will not be compelled to provide richer responses or better information-access support (5.4.2.2, 5.5.1), and while small improvements can be achieved through individual action, people generally lack the means to effectively demand the increased visibility required (5.5.3). The impact of this lack of governance is most keenly felt in the PDE/MyData space (2.3.4), where emergent actors seek to encourage data-holding organisations to enable greater information access so that they might build better data access tools for individuals, but are hampered by a lack of top-down governance supporting their requests as well as a lack of funding and investment by data-holders in data advocacy.

If data is not visible, this can lead to subjection, alienation and exclusion. Throughout both Case Studies, we have seen the negative psychological effects of people not being able to see their data. Families in both my earlier work with families (Bowyer *et al.*, [2018](#)) and Case Study One were caused significant worries by not being able to see their data. People do not want to be treated like *subjects* (in either sense of on being subjugated (Bowyer *et al.*, [2018](#)) , or a topic being discussed) and reducing people to a set of assertions in data causes them to become, in effect ‘objects to be administered’,

which is harmful and disempowering (4.2.3, 4.3.4.2, (Cornford, Baines and Wilson, [2013](#))). Supported families felt helpless and resigned to being judged through data and sometimes suspicious of those holding or using that data (4.3.4.1). This led in some cases to withholding of information or distrust of support workers, harming the effectiveness of a relationship that is designed to empower (4.4.1). In that same section (and in (Bowyer *et al.*, [2018](#)) Page 7) I outline how ongoing individual access to data has the potential to transform attitudes, remove dependence and a feeling of being a subject, and could empower families to help themselves. Such fears and worries about unseen data were echoed in Case Study Two, with participants exhibiting great unawareness of held data (Table 12, 5.3.3), and concerns over data being held out of their sight for long periods of time (5.4.3.3) as well as similar feelings of resignation or lack of choice (5.4.4.1, 5.6). Denying access to held data was seen as a key source of holding power over individuals (5.4.4.1), and visibility of data is a key part of assessing 'to what extent the bargain' (of data sacrifice for value as described above) 'is fair' (2.1.4, 5.5.3 and (Larsson, [2018](#))). It is an inherent consequence of representing people through data and then using that data to make decisions (2.1.2, (Cornford, Baines and Wilson, [2013](#); Bowyer *et al.*, [2018](#))) that individuals become sidelined and excluded (2.3.3 and (Crabtree and Mortier, [2016](#))). Without visibility of data, consent is not meaningful, and individual needs are more easily ignored or overlooked.

6.1.2 Understandable

Visibility and access to see data is not enough, people need to be able to interpret it. Data is only valuable in so much as it enables us to access the information which it

encodes (2.1.1). People need to be able to make sense of it. When humans look at data, we inevitably attempt to interpret it to see what it can tell them; in Early Help, support workers try to learn more about people's lives by examining data about them (4.2.3, 4.3.3.1). In doing so they apply their own knowledge and expectations in an attempt to extract facts. Similarly in the context of everyday digital life data, individuals search for value and meaning in that data, they reflect upon it and try to relate it to their own lives (5.4.3.1). While Early Help staff receive training on how to understand families' data, individuals struggle to understand their data without sufficient support, as discussed above. Returned data from GDPR requests is often dry and technical. It may contain codes, internal notations or abbreviations that a layperson cannot understand (5.4.3.1). Raw data is rarely sufficient to provide clear, unambiguous and unbiased information to the reader (Gitelman, [2013](#); Neff, [2013](#)). In line with one of the three core principles of HDI, *legibility*, data should be understandable by those it concerns (Mortier *et al.*, [2014](#)). In both Case Studies (and my prior work), individuals shared a desire to not just be aware of, but to **understand** what data was held about them and how it was used ((Bowyer *et al.*, [2018](#)), 4.3.2.4, 5.4.2.1). People are only just beginning to understand the significance of a data-centric world that uses data to make decisions that affect their lives (2.1.2, 2.1.4, 5.5.3).

People need understandable summaries of information content and context. It was very clear from the findings of both Case Studies that all humans looking at data need *summaries* to help them digest and locate key information. In Case Study Two participants were often overwhelmed or "drowning" (P1) at the volume or technical complexity of the data returned from access requests, "so much of it that's

impossible to know what it all means" (P4) (5.4.3.1). These feelings were mirrored in Case Study One, by support workers who feared the liability of having to "*trawl through*" large volumes of data and know all the relevant and important facts about a family so that they do not make mistakes (4.3.2.1). Participants on both sides talked of needing help to see the whole picture, something that is hard to achieve from individual datapoints or sets of files. In both cases, summaries of data would help comprehension. However, the task of creating a summary is not straightforward and places power in the hands of the summary-maker, who can decide what is relevant, how the data is framed and what is omitted from the summary. People look at information for different reasons, to answer different questions, so the question of who decides what is relevant or most important within a body of data is a critical one. Different summaries would be needed for different audiences. As Mortier reminds us, effective legibility requires a recognition that individuals' viewpoints of data can and should differ (Mortier *et al.*, [2014](#)). There is a question about who decides what the viewer of a summary 'needs to know' (4.3.3.1, 4.3.4.3). This is further complicated by the fact that the data itself is not neutral (Gitelman, [2013](#)); in the Early Help context it was clear that opinions as well as facts are recorded ((Bowyer *et al.*, [2018](#), p. p6) and from SILVER project), and a focus on the recording of data most helpful to the support worker. Commercial data holders record data in ways that are optimised for their existing systems and processes, as seen through the presence of internal codes, system screenshots and filenames in returned data (5.4.3.2). In Case Study Two most participants' comments on returned data indicated that it had not been not presented in a way optimised for understanding (5.4.3.1), failing to support *sensemaking*;"Information presentation should be as clear as

possible so that people can interpret their data and extract meaningful information from it." (Gurstein, [2011](#))

Rather than raw data, people need information and visualisations, arranged and optimised for understanding.

Data, by itself, is not meaningful. In order to be able to answer questions and acquire knowledge, people need information (2.1.1). Access to raw data files or database records or spreadsheets does not satisfy this, and prior research in the civic data context states that this would be inadequate and limiting (Cornford, Baines and Wilson, [2013](#)). To comprehend the meaning of data, visualisations and explanations can help (4.4.2); as one support worker in Case Study One observed, some families might find data tables too technical, *"I think sometimes it's easier to do it in pictures"* (4.3.2.1). Participant-designed interfaces in Case Study One included pie charts, graphs, spider diagrams and timelines, all designed to convey information more intuitively (4.3.2.1). In Case Study Two and prior GDPR requests, it was often the case that companies often returned data not in understandable forms that were less useful than the apps or websites those service providers offer. For example, run tracking apps such as Nike+ and Strava return route log information in XML-based TCX files which are meaningless without some analysis tool or visualisation. JSON files, a commonly returned data format, often use a timestamp format that is just a long number, not understandable by humans without extra work. Data was often returned in formats that were more a reflection of internal systems (e.g. screenshots, table dumps or exports) than being optimised for understanding (5.4.3.1), which some participants found useless. As P5 observed, *"It's like being given the bricks to a house... It doesn't really mean anything when it's just bricks, if you don't know how to put it together"* (5.4.3.2). It is

clear that visualisations are key to accelerating understanding (and are also subject to the same challenges of selection and bias as summaries). Furthermore, visualisations of data can function as powerful boundary objects and “things to think with” ((Brandt and Messeter, [2004](#); Bowyer *et al.*, [2018](#)) and 3.5.2, 4.4.2) and in the care context the use of data representations as a focal point or evidence for was thought to be more productive and empowering (4.3.2.2, 4.4.1).

Information becomes most meaningful when it is recognisable and relatable and can be mapped back to life experiences. Across both case studies, there was a clear

search for meaning in data, which manifested as a desire to use that data to build a fuller picture of the individual (or family)’s life. In Case Study One and in SILVER, it was evident that support workers seek a broader view into supported families’ lives by reviewing data about them (4.3.2, 4.3.3.1).

Both supported families and staff saw value in seeing all the data about each individual in a common place - in other words, structuring the information around the individual’s life, rather than the information silos of different agencies (4.3.2.3). In

Case Study Two, where participants were asked about the value they saw in the data companies held about them (5.3.3), it was clear that data was most valuable to participants when it was recognisable and relatable to events in their lives. This is particularly important for data that participants have never seen before such as derived, acquired or metadata - without a way to connect it to one’s own life, it is impossible to relate to

(5.3.3). Echoing goals surfaces in personal informatics literature (Li, Dey and Forlizzi, [2010](#)), participants sought insights about themselves in data, and so valued data that spanned a longer time period such that they might use it to spot patterns in relation to events in their lives (5.4.3.1). They

valued the opportunity to use data as “*a window into the past*” (P11, 5.4.3.1). Long term data was seen as a liability (5.4.3.3), and when it came to considering attitudes to providers, participants were most concerned about data collection by the larger data companies like Google and Facebook who, through their myriad apps, websites, devices and other means, had many touchpoints into their lives and thus had a broader picture of their life activity (5.4.3.3. 5.4.4.2). It is clear that to transform dry, technical data into meaningful information, it should be contextualised in relation to events in the life of the individual it describes.

In practice, ongoing human support is needed to facilitate the understanding of data. In both case studies, participants felt that they had questions to ask of data holders, especially when the data was difficult to understand but also more generally when the answer to a given question could not be found in the data. In my earlier work (Bowyer *et al.*, [2018](#)) and in Case Study One, participants emphasised the ability to talk to someone about their data (4.3.2.4), and included such features in their designs - both to understand and receive help in understanding, but also to feedback and provide additional information, explanation or context so that they might be better understood (4.3.3.4). In some cases families needed additional support for reasons of accessibility or technical literacy. Human communication channels for data support need to be available at convenient times too; in Case Study One individuals felt constrained by needing to wait until their next meeting with a support worker, and wanted a communication channel they could use in their own time. In Case Study Two, participants were regularly frustrated by data they could not fully understand and could not ask about: sometimes literally, when internal codes were not explained or technical formats

were used, and sometimes when they could not understand how the information had been derived, such as insurer's driving scores (5.4.3.1) or Instagram's inferred interests. The general pattern of GDPR handling by organisations was to deliver data, often handled by a back-end team with no customer face or means to ask follow-up questions. Where questions were asked they were typically hindered by delays and middlemen preventing an effective conversation (5.3.3). Considering the systemic changes toward data-centricity that the world has undergone, as described in 2.1.2 and 2.2.4, it is no surprise that human support has reduced. Across both studies we see the cost of that shift toward dealing with data instead of dealing with people; individuals get left behind, without the means to understand or ask about their data.

If data is not understandable, distrust can arise. In both Case Studies, the costs of individuals not being able to understand their data and being left in an unsatisfactory position of being unable to resolve concerns or ask questions is evident: Without understanding, comes distrust. In Case Study One, participants were concerned when they saw assertions on their records that they disputed or could not identify the source of. In Case Study Two, privacy policies that were too vague in their explanations of data (such as Google's) caused participant distrust because they seemed broad and lacked tangible examples, and large volumes of technical data caused suspicion. It is clear that understandable information about what is represented in data, and about the context and use of that data, can help individuals to trust the data holder. In the minority of cases in Case Study Two where the GDPR had a positive impact, the reasons given included understandable data and helpful human responses, for example P10 cited Niantic's detailed data and positive attitude,

while P6 described Sunderland AFC as upfront, and said that their data *'just made sense'*. P7 specifically distrusted LinkedIn because she felt they had not bothered to adequately engage with her (5.4.4.3). The importance of trust will be discussed in more detail in 6.2.1 below.

6.1.3 Useable

People need to be able to explore and interrogate data to ask questions of it. In both case studies, many participants showed they are aware that their personal data contains insights and value they cannot access. In Case Study One and (Bowyer *et al.*, [2018](#)) this manifested as concern over what unseen incriminating judgements might be storied in their records, and an awareness that the only data supported families can actually see is whatever subset of data their support worker chooses to share with them (4.3.4.1). In Case Study Two, this manifested as feelings of not having the technical skills necessary to explore their returned data (5.6) and that without better tools individuals were not getting the same view of data that service provider staff have (P2, 5.4.3.1). What we can see from these observations is that even visible, understandable data is not enough to meet people's needs. People need to be able to interact with their data, in order to explore it, gain insights or answer questions. There is a need for tools not only to access data, but to help people make sense of (often technically formatted) data (5.4.3.1), and explore it in different ways to answer questions - which implies the need for operations such as filtering, searching, comparing and orienteering (see 2.1.4 for others), in order to understand history, context and patterns in the data, as suggested in self informatics (SI) literature (2.2.3). Several participants in Case Study Two hoped to receive data in

formats they could visualise, “mashup” and play with, but did not get this - not only was data not optimally formatted for such use, it lacked sufficient explanation to enable individuals to pursue such goals for themselves (5.4.3.2).

Data needs to be useable - correctly formatted and explained in a portable and standardised form. As

discussed in 5.6 and 5.4.3.2, people have multiple needs around data: For understanding, people need *understandable information*, both the facts and assertions encoded within the data, but also information about the data itself - its context, history, use, and significance. But there is a distinct need for *usable data*, which is somewhat orthogonal to this. A PDF containing a screenshot from an internal system might potentially be sufficient for understanding, but is useless for exploration or visualisation-building. Similarly a technical log file might contain rich data that can be queried and visualised given the right tools, but without those tools or an informational summary, is of no immediate value (5.4.3.2). This dilemma was alluded to by P4 in Case Study Two thus: *“If you want to view the data they have about you, it’s quite usable. If you want to do something [analytical], then it’s not”* (5.4.3.2). This dichotomy of needs is discussed in 5.5.1, where the introduction of standard formats is proposed as a means to catalyse the building of data insight interfaces. As P1 stated, *“it would be nice if these companies had a standardised model of how this information is presented to people”* (5.4.3.1).

People need to be able to interact with data, which means interfaces are needed By themselves, even standardised files as described above are not practically usable. As discussed in 4.4.2 and consistent with effective access (Gurstein, [2011](#)) participant data designs in Case Study Two remind us that

simply providing data is not sufficient: to be meaningfully able to act upon data requires some form of interface not only for visualisation and interrogation as mentioned above, but also so that any physical, cognitive or accessibility needs can be met. In Case Study Two, some participants wanted not just for data access, but for tools to help them find insights from their data (5.5.3, 5.6). While several companies are starting to create interfaces for data access, most of these are still focussed upon file delivery, with the notable exception of Google Timeline and Google My Activity, which provide a glimpse how an interface to explore data could be more useful than providing a bundle of files. Both of these examples also re-iterate the value of unifying data around an individual's life, as discussed in 4.3.2.3 and above in 6.1.2 above.

Data needs to be explorable from a temporal perspective

Another aspect of usable data that goes beyond what a data file can offer is the ability to view it over time. The importance of this temporal capability, as identified in literature cited in 2.1.4 (practical information access) and 2.2.2 (temporal PIM systems), and my prior writing (Bowyer, [2011](#)), was evident in both Case Studies. In SILVER and in Case Study One, being able to access historical data for a full picture regularly surfaced as a desire in discussions - even though the exact bounds and mechanisms for achieving this were contentious (4.2.6, 4.3.2.1, 4.3.3.1). In Case Study Two, as Table 12 shows, 26% of participant goals related to SI-type reflection (2.2.3) on one's past to enable self-insights, nostalgia and creative uses of data. No participant was able to achieve this, and this was in large part due to the lack of temporal data exploration capability, meaning that significant potential value (the value of a long-term dataset as described above) remained locked away and inaccessible (5.3, 5.6). Case Study

Two also suggested a lack of thought to this temporal perspective from data holders, who delivered data as a one-off snapshot that was already out-of-date when delivered. Indeed the GDPR explicitly discourages excessive data requests, rendering an ongoing view of data as proposed in 5.5.1 near impossible. Having better temporal data exploration capabilities would enable people to understand themselves and their data ecosystems better, informing both personal self-improvement goals and better decisions about personal data practices and provider choices (5.6). Other exploration perspectives that could be powerful include location-based views or person/company/relationship-based views.

If holders make data unusable, this is a barrier to individual agency and power. In Human-Data Interaction terms, people need not just legibility, but *agency* - the ability to act. [PARA UNFINISHED]

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 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.

[reference the fact that individuals' viewpoints of data may differ but be valid (6.1.2)]

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](https://doi.org/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. (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](https://doi.org/10.1145/3173574.3173710).

Brandt, E. and Messeter, J. (2004) ‘Facilitating collaboration through design games’, in *Proceedings of the eighth conference on participatory design artful integration: Interweaving media, materials and practices - pdc 04*. New York, New York, USA: ACM Press, p. 121. doi:

[10.1145/1011870.1011885](https://doi.org/10.1145/1011870.1011885).

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](https://doi.org/10.1332/030557312X645838).

Crabtree, A. and Mortier, R. (2016) 'Personal Data, Privacy and the Internet of Things: The Shifting Locus of Agency and Control', *SSRN Electronic Journal*, pp. 1–20. doi: [10.2139/ssrn.2874312](https://doi.org/10.2139/ssrn.2874312).

'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](https://doi.org/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](https://doi.org/10.2139/ssrn.3887097).

Larsson, S. (2018) 'Algorithmic governance and the need for consumer empowerment in data-driven markets', *Internet Policy Review*, 7(2). doi: [10.14763/2018.2.791](https://doi.org/10.14763/2018.2.791).

Li, I., Dey, A. and Forlizzi, J. (2010) 'A stage-based model of personal informatics systems', *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*. New York, New York, USA: ACM Press, p. 557. doi: [10.1145/1753326.1753409](https://doi.org/10.1145/1753326.1753409).

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](https://doi.org/10.1145/2493432.2493446).

McNamee, R. (2019) *Zucked: Waking up to the Facebook Catastrophe*, 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](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).

Neff, G. (2013) 'Why Big Data Won't Cure Us', *Big Data*, 1(3), pp. 117–123. doi: [10.1089/big.2013.0029](https://doi.org/10.1089/big.2013.0029).

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](https://doi.org/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](https://doi.org/10.1017/ipo.2021.30).