

# 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 Process Transparency](#)
    - [6.2.2 Individual 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](#); Strohmayr *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 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.

**People need to be able to access the valuable knowledge locked within their data.** People need to see the complete picture of their data in order that they might access at least glean as much value and insight from it as data holders do (5.4.3.1). For self improvement and improving one's situation (a key goal of Early Help), access to metrics visible in data are extremely important, so that one might measure progress (4.3.2.4). While support workers do endeavour to provide this and sometime show data extracts, this can never be as empowering as having full capabilities to explore that data *"in your own time, at your own pace"* [Bowyer *et al.* (2018); 4.3.2.4]. In commercial service provider relationships, data interfaces



present data in ways that are configured to reflect the profit motives of the organisation (explored further in my work with Goffe *et al.* (Goffe *et al.*, 2021)), and so accessing a relevant view of one's own data and having the means to access the knowledge within it is similarly difficult. Participants in Case Study Two found returned data unusable (5.4.3.2), and wanted visualisations that would allow them to discover patterns and insights, and tools to explore their data (5.4.3.1, 5.5.3). If data holders can make data not just visible but useable, individuals can learn about the valuable knowledge that data holders are currently extracting from their data, and hopefully how to access that knowledge for themselves.

**If holders do not make data usable, 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 upon one's data (Mortier *et al.*, 2014). Personal data contains valuable and actionable information about individuals and their lives (5.5.3). The SI field has identified that there are already many practical barriers to working with one's personal data effectively, including not just access but challenges in integration, sensemaking and goal-tracking – the 'barriers cascade' (Li, Dey and Forlizzi, 2010) – which hinder the ability to use one's data for personal benefit. As observed in Case Study Two, an inability to access the value in your own data can lead to feelings of resignation, concern, suspicion or distrust (5.6). Even if one can see and understand one's data, an inability to act upon it can reinforce feelings of being passive and uninvolved; without this, any opportunity to feel engaged and motivated is lost (Bowyer *et al.*, 2018, p. p8). Being able to use data for one's own purposes is a critical ingredient of empowerment and rebalancing power (Hoffman, 2014a). Without data usability, individuals are in effect digitally impaired, leading a less functional society where innovation and growth is limited (Abiteboul, André and Kaplan, 2015, p. 2.1.4).

## 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. *Process 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. *Individual 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 in Decisionmaking*: 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 Process Transparency

**People need a window into how their data is used; this means transparency of processes not just of data.** It is well established that there is currently extensive use of personal data by service providers and other parties that is beyond an individual's view (Hoffman, 2011), forming an ecosystem of data use based upon one's data, which is currently not centred on or visible to the individual concerned (2.3.4). Decisions made based upon personal data directly affect people's lives through policy decisions (in the care context) or business/functionality decisions (in the commercial context). People need to understand the value created by the use of their data and how (if at all) they are compensated for this (Hoffman, 2011). Even with full data access, understanding and useability, individuals cannot see into this opaque world of data use; the data is just an artifact produced and shaped by unseen processes. Like an archaeologist trying to infer the customs of lost civilisations through ancient relics, observation of the data can only reveal so much. Andrew Cormack, writing before GDPR, observed that "*it is more important to know how information is processed than the actual values involved*" (Cormack, 2016). The SILVER project found that families had

very little awareness or understanding how their data was used, and that consent was therefore not meaningful because consent had been given without processual understanding [4.2.2]. In Case Study One participants agreed that people need rights to see how their data is used (4.2.6). Case Study Two revealed a clear desire for awareness of how data is used, how decisions are made, and how this might affect them (5.4.2.1), with over 74% of goals in pursuing GDPR requests relating to wanting greater insight into personal data use practices (5.3.2). 70% of participants wanted to understand what providers infer from their data and this was unmet in 73% of cases and fully met in only 7% of cases (5.3.2).

**Process transparency is required to enable accountability.** In Case Study Two, participants recognised that organisations had collected data about them which could be exploited, and wanted to understand the extent of that capability (5.4.3.3). Data access can provide a window into collection capability, but only process transparency can reveal the extent of data use capability. Many participants expressed a desire to assess the trustworthiness of their service providers; they had curiosity, suspicion and unanswered questions that only transparency could address (5.3.2) and sought to judge whether data use practices were “appropriate” (5.2.4.1). In the Early Help context, all data processing is hidden from individual view and no access or questioning capability except through their support worker (functioning as a selective gatekeeper) (4.1.1, 4.2.1, 4.4.1). Returning to Case Study Two, there was evidently some transparency available in the form of the ability to make a GDPR request, but many participants found GDPR responses inadequate for holding providers accountable (5.2.4.3). Nonetheless, data access request handling is itself a data process, and so in this sense, the GDPR process did offer some ability to judge the trustworthiness and integrity of providers in data handling, in part informed by the breadth and quality of data returns (Table 11) but perhaps moreso by the experience of the GDPR process. Many participants formed or revise their perceptions of companies, with perceptions of providers having a lack of care and making access difficult, or of providers being helpful and open having a strong impact on participants’ attitudes toward them (5.4.4.3).

**There is no accountability, processes are not transparent, and thus power remains imbalanced.**

Across both case studies, the lack of process transparency is clear. Early Help services have no obligation to describe or share their data use practices with supported families, and apparently only even attempt to do so at the point of initial onboarding and consent collection (Bowyer *et al.* (2018), 4.2). This equates in practice to a complete lack of accountability over data practice (4.4.1, 4.5) Meanwhile in the commercial context, some companies failed to respond at all to GDPR, which is a barrier to accountability service providers. Many routinely failed to adequately meet the transparency rights stated by GDPR, without repercussion or consequence (5.4.2.2). In both sectors, data holders’ freedom to collect and use data without adequate transparency or ability to be held to account can be seen as an exertion of power over individuals. The power imbalance (Hoffman (2011), Hoffman (2014a)) and the dominance of data holders over the individuals about whom data is held, is reinforced by a lack of transparency.

**People face an incomplete picture of their data ecosystem, even after using all available means to achieve transparency.** In SILVER and my prior work in the Early Help context, supported families expressed concern that data could not adequately represent the complexity of their lives (Bowyer *et al.* (2018)), a view supported by literature (Cornford, Baines and Wilson (2013); Gitelman (2013)) and sustained by support workers & staff in Case Study One (4.2.6, 4.3.3.1). Therefore, transparency is vital in order to ensure data is fair and accurate to them, yet it is not available (4.3.4.1) so they have no means to ensure this. Similarly in Case Study Two, the most popular GDPR goals around understanding inferences made from data about people (5.3.3) remained unmet. Data was incomplete, delayed, or inaccessible (5.3.2) and the potentially most informative type of data when it comes to understanding processing – metadata, derived and acquired data – were typically absent. Apparently broad responses were discovered to be very limited when viewed through the lens of privacy policy commitments and GDPR transparency rights (5.3.4).

**Trust in data holders is needed, and gaps in transparency create distrust and a risk of broken expectations, harming relations.** Individuals need a functional understanding of their data and its handling, and this is crucial to trust. Good explanations (as were often found lacking in Case Study Two

(5.4.3.2)) can deliver some of this needed understanding and subsequently increase trust (Glavic *et al.*, 2021), as observed in a minority of cases (14%) where a good GDPR response led to the participant's trust in their provider increasing. Conversely we found that incomplete data (or a general lack of transparency/difficulties of access (5.4.4, 5.6)) can harm trust, as in the majority of cases (52%) can harm trust (5.3.4), leading to thoughts such as 'what are they hiding?' (5.4.4.3). Privacy policies that contradicted expectations or lacked sufficient explanations also led to distrust (5.3.4). Trust in the independence and integrity of data holders is essential (Dijck, 2014), and this was often a concern in the Early Help context, where trust between support worker and supported family is especially critical in order for the support relationship to be effective. Earlier work found that families wanted to be confident that their data would be handled sensitively and fairly only by those with a need to know, and believed that greater visibility of data processing would allow them to trust that that was the case (4.2.2 and (Bowyer *et al.*, 2018)). When families felt alienated from their data, trust was absent (4.4.1), and a lack of transparency and accountability makes it hard for families to maintain trust in the system (4.5). These issues of transparency and trust are inherent in a data-centric operating model, and the World Economic Forum have summarised this problem thus:

“*A crisis of trust is developing, stemming from the use of personal data in ways that are inconsistent with individuals' preferences or expectations.*” (Hoffman, 2014b)

**Information facilitates trust; transparency therefore offers an opportunity to earn trust & improve relations.** In both studies, the findings led us to conclude that increased transparency from data holders / service providers would improve trust; in Case Study One we concluded that support workers and organisations should be as open as possible about data handling and sharing (4.3.4.3), while in Case Study Two we highlighted the potential benefits of increased consumer loyalty that greater transparency might bring (5.5.2), as well as the need for policymakers to legislate in favour of increasing individuals' understanding of data practices (5.5.1). In doing so, we are recommending a level of transparency that goes beyond current GDPR practice, and even beyond current GDPR policy; in order to redistribute power, GDPR needs to deliver meaningful transparency, not just the “*box ticking*” delivery of unhelpful files that our participants sometimes observed (5.5.1). Crabtree describes meaningful transparency by saying that it cannot be a “*one-way street*” that reduces individuals to “*being spectators*” on how their data is used; he says that it involves “*making the whole ecosystem transparent, not just the front end*” (Crabtree and Mortier, 2016). Access to good information about practices is the most effective way to earn trust (5.4.4.2), and both studies' findings suggest that a proactive attitude can do just that. (4.4.1, 4.3.4.2, 5.5.2, 5.6).

**Initially, transparency may cause distrust, but only where practice is problematic; this is accountability becoming real and catalysing better data practices.** It is important to note that in Case Study Two we saw the transparency of GDPR cause elevated distrust (5.3.4), however this does not mean it should be avoided. The reasons cited for distrust arising were invariably due to the discovery of practices that participants did not approve of. This is a clear illustration of the link between transparency and accountability; the transparency reveals the non-consensual or unsatisfactory practices that providers must change if they wish to maintain trust and loyalty, such as unclear data practices, data over-use or data sharing that the individual would not have consented to had they been asked. This shows that in some cases trust is fragile, where unfavourable practices are hidden and only the individual's unawareness is keeping the relationship intact (5.4.4). Data holders should not only be transparent, but should follow this up by acting upon subsequent feedback, improving practices that individuals discover and challenge (5.4.4.3). By shining a light, accountability becomes real and change for the better can occur. Ultimately, increasing transparency can help providers uncover exactly what they need to do to earn greater trust (5.6).

**Without transparency of data and processes, individual action is blocked and power remains imbalanced.** Across the two studies, a clear pattern emerges: transparency can increase trust, enable accountability, empower individuals, and (provided organisations respond favourably) actually tilt the power balance back toward an equitable and fair relationship where data is collected and used in clear sight of the

individuals it concerns, where they might hold those organisations to account and immediately challenge any unsatisfactory practice, unauthorised processing/sharing or inaccurate data. Thanks to GDPR, individuals are now able to take direct action to educate themselves and pursue greater transparency, and utilise their rights to motivate incremental changes from data holders (5.5.3). Without transparency, data holders will continue to hold the balance of power, and individuals will lack agency and accountability.

## 6.2.2 Individual Oversight

**Data visibility and 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. And therefore, people want something more than data and process transparency, the natural next step is the ability to make decisions about what happens to their data. Participants’ goals in Case Study Two included curiosity, suspicion and a desire to shed light on specific incidents (5.3.3), mirroring the desires families in (Bowyer *et al.*, 2018) exhibited to be able to know and see what data was held and used about them. In both cases, individuals wanted to have a say over what happens. Current models of informed consent have been found to be inadequate, with the initial handover of data acting as a ‘*point of severance*’ (Luger and Rodden, 2013). This was echoed in the experiences of families in the Early Help context, who gave consent at the point of initial onboarding, but lost all ability to influence what happens to their data thereafter [Bowyer *et al.* (2018); 4.2.2]. GDPR aims to adhere more to a *dynamic consent* model (Kaye *et al.*, 2015; Williams *et al.*, 2015) by giving people an ongoing set of rights, including the right to be informed about the use of your data, the right to object to certain data uses, and the right to get your data corrected or deleted (Information Commissioner’s Office, 2018). In line with the *accountability principle* [ADD REF TO EU ACCOUNTABILITY PRINCIPLE Article 29 Data Protection Working Party (2010) Opinion 3/2010 on the Principle of Accountability, 00062/10/EN WP 173. AND Article 29 Data Protection Working Party (2014) Opinion 8/2014 on Recent Developments on the Internet of Things, 14/EN WP233], this in effect would allow people to act as overseers or regulators over their own data: watching how it is used, and demanding action or change to practice when they see data use that goes against their wishes.

**People need agency and negotiability over held data about them, in order to ensure fairness and accuracy and reduce risk.** As my earlier work in the Early Help context (Bowyer *et al.*, 2018) showed, there is a strong desire to ensure data is *fair* and *accurate*, because that data is used to inform judgements and make decisions that can directly affect the individuals concerned. Data in Case Study Two showed a clear problem with the accuracy of unseen data: while in 92% of cases volunteered data (which by definition, has been seen by the individual) was found accurate, derived and acquired data (previously unseen by the individual) was found inaccurate in 50% and 80% of cases respectively (5.3.2). Being able to ensure fair and accurate data goes beyond being able to see and understand the data, but requires also *agency* (the ability to act within a data system, such as to delete or correct data, or withdraw consent) and *negotiability* (the ability to continue to have a voice and make changes as circumstances change) (Mortier *et al.*, 2014). People need a relationship with their data (2.1.5). In both Case Studies and in earlier work, individuals perceived tangible risks both of data being held beyond their reach, but also of potentially inaccurate data being used to make decisions. Risk factors identified in the Early Help context included facilitating or encouraging crime, causing social and psychological harm, and enabling medical mismanagement or welfare support failures (Bowyer *et al.*, 2018). In Case Study Two, participants felt that held data about them that is not visible or controllable was a liability that might lead to privacy violations, commercial exploitation, and an increased risk of data leaks (5.4.3.3). Clearly people feel that for their data to be safe, they must be able to see and verify its storage and use for themselves and enforce action when something is not right. As early as 1980, when the world was less data-centric, it was already recognised that individuals would need the ability to challenge data use, as the OECD observed in their guidelines:

“The right of individuals to access and challenge personal data is generally regarded as perhaps the most important privacy protection safeguard.” (Organisation for Economic Co-operation and Development, 1980)

**Individual oversight capabilities must be supported by governance, so that they can effect desired changes.** As Gurstein notes, a key element of effective data access is governance, that is, mandating data holders to support individuals in accessing their data and respecting their wishes over what should happen to that data (Gurstein, [2011](#)). Individuals need to be able to give instructions, make changes and express permissions that have weight; they need to be listened to, so that they can meaningfully effect change (4.4.2). Bakardjieva, examining the use of data about others in a different context (research), identified that individuals whose data is used need the ability to influence not only the data about them, but the actual decisionmaking that occurs based on that data: both the data *and* the decisionmaking should become objects that the individual subject can manipulate (Bakardjieva and Feenberg, [2001](#)). At the time of writing, much of the focus on GDPR has been about access to data, perhaps because this is more tangible, and very little about GDPR's other rights that can influence decision-making (5.1.2). This was backed up by participant experiences in Case Study Two, where desires to influence or change practices or delete data were either not actionable or ineffective (5.4.3.3). Governance over individual data rights has two elements. First, to support individuals in complaints or challenges, which are currently unevenly enforced (5.5.1). But more importantly than this, given the extensive use of data by organisations and the great potential for misuse or harm, individuals need to be able to trust that systems are in place that mandate the behaviour of data holders to be trustworthy; to compel organisations to maintain good data practices such as data security and dynamic consent in the first place (5.4.4.1). In the GDPR context the bodies that can do this already exist - the Data Protection Authorities. In the public sector/care context, the picture is less clear. Participants identified a need for oversight bodies to compel good practice, identify appropriate access rules, and to provide independent oversight in contentious cases (4.3.4.3); this is particularly difficult given that no organisation can see the full picture of an individual's civic data.

**Individual oversight would bring individuals back to the centre of their personal data ecosystem as an active participant.** As outlined in 2.3.4, the ideal has been established that individuals need to be at the centre of their own personal data ecosystem, overseeing and controlling their data selves as easily as their physical selves. Currently, as seen in both contexts, data functions as a proxy for their direct involvement [Bowyer *et al.* ([2018](#)); 5.4.4.1]. Decisionmakers consult data first, as the primary source of truth (4.1.2, 2.1.2), and the individual second (if at all). For transparency to be meaningful, data flows need to open up to include individuals as part of the loop, changing them from passive spectators to active participants (Hoffman, [2014a](#); Crabtree and Mortier, [2016](#)) in the processing of their data. Examples of specific oversight abilities desired in Case Study One were the ability to explain or annotate datapoints (4.3.3.4), to be able to check data together with support workers, with the record of that check becoming part of the data (4.3.3.2), or to have granular access controls over precisely which data could be seen by whom (4.3.3.5). In Case Study Two, a clear picture emerged that what participants want is the ability to make choices. They want control over the data they are forced to sacrifice to companies (5.4.4.1); to avoid the '*point of severance*' Luger describes, data sacrifice should be a loan or sublicense, not a taking-possession-of.

**Given the changing and complex nature of human life, data is inadequate and consent is never complete, so longitudinal participation and oversight is needed.** Too often, data is treated as a static source of truth (see above). Attempting to represent people as data in order to require less human contact is a reasonable goal from an organisational efficiency or cost-saving perspective, but any representation will never be complete or adequate (Cornford, Baines and Wilson, [2013](#); Bowyer *et al.*, [2018](#)). In Case Study One, the findings showed the need for numerous efforts to augment data in order to combat its inherent inadequacy, such as support workers seeking to understand the people behind the records (4.3.3.1) and maintaining a constant attitude of seeking to understand more deeply than the data record can allow (4.3.4.2). Even if a data record can be corrected or completed, it will still be inadequate, because human lives change continuously: people move, start and end relationships and jobs, marry, divorce, have children, pursue new interests, become incapacitated, or die. The passage of time can radically change the context or relevance of data (Bowyer *et al.*, [2018](#)). A one-time informed consent upon data collection is inadequate in this ever-changing context (4.2.2, 4.2.6, 4.3.3.3). And of course, if consent needs to be ongoing (and in order for it to be meaningful) this means that engagement with the individual concerned, and that individual having a view of their data and its use, need to be ongoing too. In order to avoid storing

or using data beyond its need, ongoing data access is needed, in order to enable ongoing individual oversight (4.2.2, 4.3.2.4). Systems and processes must treat data as dynamic (4.3.3.3), as something that will become inaccurate without sustained engagement. In Case Study Two, we note that GDPR data access is currently based around viewing a one-time snapshot of your data, and does not take this need for negotiability (Mortier *et al.*, 2014) into account at all (5.5.1) (though some companies now offer download dashboards that come closer to providing ongoing access). Ongoing access, consent and participation do carry cost implications for providers, and effort implications for individuals - but these can be improved over time: the former through automation, standards and education, and the latter through holistic approaches to personal data ecosystems; these mitigations will be explored further in Chapter 7.

**There is scant individual oversight available today. Governance is lacking. If people cannot make choices about their data, they will remain powerless.** Participants in both contexts faced an inability to see the full picture of how their data is processed and used. Despite families in Case Study One workshops spending time designing interfaces for seeing and correcting their data and changing permissions (3.5.3, 4.3.2.1, 4.3.2.3), no such interfaces exist. The entirety of their data access and influence is limited to what can be achieved verbally with their support workers (4.2.4, 4.4.3). Without transparency and dynamic consent mechanisms, those families lack accountability. They are excluded with no ability to oversee or participate in the life of their data. In Case Study Two, of the 41% of participant goals that concerned gaining insight into and control over the use of their data, 66% were unmet. Participants reported seeing no clear pathway on how to access rights to control their data and only 1 of the 10 cases where a participant wanted to delete their data was successful (5.3.3). At the time of writing only one company, Apple, has a privacy hub that offers clear routes to access data rights other than access. Participants also reported in some cases being unable to check the accuracy of their data, or to investigate specific incidents where they had concerns (5.3.3). The general view was one of widespread disappointment, that despite the promise of GDPR it did not confer any power to the individual to influence data use. (5.2.4.3), leading in some cases to a reluctance to submit GDPR requests in future. Access requests were also rarely seen as useful in the care context, and our understanding is they typically only occur in the case of complaints. In the GDPR context, the inability to restrict data use or delete data was seen as a lack of control, and the retaining of data against their wishes as a liability (5.4.3.3). Ultimately oversight means having choices, which is essential in the data-centric world. The case studies' findings show that, in general, participants felt they had been forced to sacrifice data to access services, and offered *no* practical choices or control over that data. Without individual oversight, there is no choice and people remain powerless.

### 6.2.3 Involvement

**Data represents people. But people are not records. There is a need to engage the human behind the data, as people can never be fully represented in data.** Intrinsic in the move towards data-centricity has been a move away from human involvement. In the commercial sector this is due to cost-saving (call centres and web portals being cheaper than individual customer interactions) (2.1.2). In the care context it is similar but there is also a desire to create a society that functions at large without individuals requiring special handling and support (3.4.1). Both case studies' findings, consistent with literature (Abiteboul, André and Kaplan, 2015; Crabtree and Mortier, 2016), reveal myriad problems created by the exclusion (Bowyer *et al.*, 2018) of people from matters that affect them - from feelings of alienation or disengagement (5.4.4.1) to actual harms caused by erroneous or unfair judgements (4.2.2). Service providers holding data need to contextualise data as an incomplete view into the complex human world, and seek greater understanding (4.3.4.2) while looking for positives in data (4.3.4.1). It is interesting to note that the Troubled Families programme was created help find the human situations of people slipping through the cracks of the system, which highlights the inadequacy of purely data-based decision-making (3.4.1).

**Consent to access and use data needs to be dynamic and meaningful, which can only happen through ongoing involvement.** As established in 4.4.1 and above in 6.2.2, ongoing data consent is essential, and this is especially important where that data is used to make decisions (4.4.3). One-time consent is ineffective and meaningless (4.5). Asking individuals for consent and subsequently less involved



in decision-making reinforces a hierarchical, rather than an equitable, power relationship, as Bakardjieva and Feenberg found in their work looking at how to involve virtual subjects in research (Bakardjieva and Feenberg, 2001). Without ongoing consent, the power imbalance is amplified (4.5). In the commercial context, companies view data as their asset to exploit (Hoffman, 2011; Toonders, 2014), and the simple fact of having the ability to collect or access to data about people has proved in practice, sufficient to enable a variety of practices which would be likely to be refused consent if made visible to users (Melendez and Pasternack, 2019; Evans, 2021; Claburn, 2021). Individuals feel forced into a one-sided arrangement of sacrificing data in exchange for service benefits; with no choice upfront on signup, and minimal practical choices afterwards, their only choice is Hobson's choice ('Hobson's Choice', [no date](#)). Consent has become commoditised, and from a corporate perspective the focus has become constructing a legal justification for using an individual's data rather than practically engaging with them and verifying if they approve (Woods and Böhme, 2022). This can only happen when the individual about which data is held is excluded from data handling processes.

**A human channel for conversation is wanted, to enable explanations, questions, and consultation.**

In my earlier work (Bowyer *et al.*, 2018) and in Case Study One, all participants viewed that individuals should be able to talk to someone about their data (4.2.6, 4.3.2.4), in order to ask questions or explain datapoints. In Case Study Two, participants had questions about their data that they wanted to answer (5.3.3), yet these questions remained unanswered (5.4.2.3). Participants regularly experienced painful and ineffective processes when trying to answer their questions, found that GDPR responses, often unhelpful, provided no backchannel for followup questions or further communication (5.2.4.3). They were left '*in the dark*'. This highlights the need for a human support channel, which is not mandated by GDPR, not just to better understand the data itself (6.1.2), but to enable ongoing consent negotiation within the relationship with data holders.

**Individuals should be consulted in decision-making. This improves accuracy, perspective, and fairness and reduces consent liability.** A common theme in the findings of both my earlier work with families and Case Study One was the idea that data is more likely to be fair and accurate if in the individual has the opportunity to express their perspective on it. Families advocated checking data together to identify gaps (4.3.3.2), valued the prospect of making their own contributions to the data to '*tell their own story*' (4.3.3.4) and wanted a '*right to explain*' or annotate their data (Bowyer *et al.*, 2018). Support workers recognised the need to work with families to understand their situation better (4.3.2.2). These findings show it is important to give the human data subject a role in data creation or checking, and that if they are involved involved, to ensure a fairer and more complete view can be obtained than the limited view presented by the data record (4.3.3.1). Without involvement, agency (Mortier *et al.*, 2014) will always be limited. Data is not neutral (Gitelman, 2013; Neff, 2013), and this means all stakeholders should be given a role (Bowker, 2005) in order to avoid errors, harm or disempowerment (4.4.1). While involvement is negligible in the care context (being limited to the TAF and the support worker relationship, in the commercial context, there is no such involvement, and the results of this can be seen in the low accuracy of previously unseen data (5.3.3) and in the low trust ratings (5.3.4) given to many providers after seeing data returns. It seems that data and knowledge would become more accurate when it is closer the data is to the individuals concerned (4.4.3, 4.5); data created and handled far from the family would intuitively be less accurate and less likely to have been considered from their perspective (4.2.2). Organisations on both sides value data accuracy (4.2.3, 2.1.2) so greater involvement could help achieve this. Data holders would also benefit from involving individuals because responsibility for consent would become shared, resulting in a higher 'buy-in' from individuals and a reduced liability; provided communication is effective and without barriers, involved individuals would be inclined to speak up if they see something they do not agree with (4.4.1). Of course, this call for involvement in decision-making is not absolute. Businesses need to be free to exercise their expertise. Patients are not best placed to decide what medicines they need, and over-involvement of users in product design can result in mediocre products that suffer from '*design by committee*'. In fact, the attitude conveyed toward the individual is critical (as borne out in our Case Study Two findings (5.4.4.2, 5.4.4.3); Edwards and Elwyn, in their paper on shared decisionmaking, argue that *feeling* involved is actually more important than actually contributing to decisions (Edwards and Elwyn, 2006).

**Effective collaboration can be achieved by coming together around the data, using it as evidence (of facts or of opinions) and as a boundary object.** Workshop C in Case Study One specifically explored the prospects of shared data interaction, which had emerged from SILVER and phase 1 (Workshop A/B, see 4.2.5) findings as a possible way to meet the needs of both supported families and support workers. Shared values were identified (4.2.6) and a model for shared data interaction builds upon the findings (4.4.3). Central to this is the idea that evidence-based decision making can be more effective (both from a rapport-building perspective but also in terms of the quality of decision made) as guidance has advised [OFSTED (2015); Department for Education (2018); 4.1.2] and as seen in the findings (4.3.2.2, 4.3.3, 4.4.1). Families and staff saw potential benefits from checking data together (4.3.3.2), using specific datapoints in discussions as evidence (4.3.2.2), or to help families open up (4.4.2), on top of the simple benefits of making all evidence equally visible (4.3.2.2, 6.1.1). Throughout such shared data interactions, representations of data perform an important function as a boundary object (Star, 2010; Bowker *et al.*, 2015); it provides a common focus to discussions that is relatable to both parties, “*things to think with*” (Papert, 1980; Brandt and Messeter, 2004). This helps improve legibility by surfacing the (perfectly valid) differences in different parties’ perspectives (6.1.2, Mortier *et al.*, 2014). Working with families in 2017 I observed that gathering around data representations facilitated a less confrontational interaction than an across-the-table interview would (Bowyer *et al.*, 2018).

**Being involved means being able to learn and take action at any time, including on one’s own and away from official contact or interactions with service representatives.** Both Case Studies reveal that people want an ongoing window into their data and its use. Limiting data access and process transparency to specific times spent with a gatekeeper, as in the Early Help case (4.1.1), or to a process that only provides a one-time snapshot of data (5.5.1) reinforces the hierarchical power imbalance. Giving access to data and abilities to explore and ask questions “*in your own time*” (4.3.2.4) can unlock new individual benefits from data (5.4.3.2, 6.1.3) and provide more immediate feedback that can help people measure their progress towards improvement goals (2.2.3, 4.4.2, 4.4.3), as well as enabling ongoing individual oversight (4.4.2) and dynamic consent (4.3.4.3). In the face of providers making decisions based on unseen data using processes that cannot be observed, people feel excluded and powerless (5.4.4.1, 6.1.1). To ensure effective access (Gurstein, 2011), accountability and trust, there must be ongoing involvement, data access and transparency (5.5.1, 6.2.2).

**Indirect data use enforces an uneasy trust; services that use data need a human face or point of contact, in order to grow understanding, earn trust and improve relations.** There is a coldness to data. Facts, judgments and mistakes appear in print with equal weight, and without explanation or context. In both studies, the value of human contact accompanying data access was evident; in Case Study Two participants valued the GDPR responses that felt most human (5.4.4.3) and lamented the inability to discuss data or resolve questions (5.3.3, 5.4.2.3). In (Bowyer *et al.*, 2018) participants wanted ‘*to have a conversation*’ about their data and in Case Study One participant data interface designs included buttons to chat to their support worker or ask questions (4.3.2.3). People do not want to be severed or alienated from their data, yet they feel they have no choice but to relinquish access and involvement (4.4.2, 5.4.4.1). Ideally, all data-using organisations would have a human face or point of contact that individuals can address questions to and in whom their trust of the organisation can be embodied (4.4.2).

**Without involvement, people can never take a full and equitable role in processes that affect their life.** When data is used by organisations, this inherently serves as a proxy for their involvement [Bowyer *et al.* (2018); 5.4.4.1]. People have consequently lost control and agency (Crabtree and Mortier, 2016), creating a crisis of trust and a power imbalance (Hoffman, 2014a, 2014b). Exclusion from data handling and decision-making inevitably reinforces a hierarchy, with the individual destined to have less say and influence over services that affect their daily lives. Taken together, an ongoing involvement with the data held by service providers and the processes that use that data would change indirect data interaction (6.2) into direct data interaction (6.1) and transform the individual experience of service use. Of course such a shift is difficult and costly to offer, for both small and large companies – the data-centric world emerged in part as a means to reduce costly human interactions and facilitate large-scale scaling up to serve larger

user bases – but what we see suggests that the balance has tipped too far towards exclusion of individuals, and that giving them even a limited role in reviewing data, consenting and contributing to decisions, and just being informed, could carry significant benefits for both individual empowerment and organisational reputation, and trust.

### 6.3 Achieving Individual Empowerment

1. Through these summarised insights, I have shown a multifaceted 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

Abiteboul, S., André, B. and Kaplan, D. (2015) *Managing your digital life with a Personal information management system*. 5. ACM, pp. 32–35. doi: [10.1145/2670528](https://doi.org/10.1145/2670528).

‘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).

Bakardjieva, M. and Feenberg, A. (2001) ‘Involving the Virtual Subjects’, *Ethics and Information Technology*, 2, pp. 233–240. doi: [10.1023/A:1011454606534](https://doi.org/10.1023/A:1011454606534).

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

Bowker, G. C. (2005) *Memory practices in the sciences*. MIT Press, p. 261.

Bowker, G. C. *et al.* (2015) *Boundary objects and beyond : working with Leigh Star*. MIT Press, p. 548. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=nmSkCwAAQBAJ&oi=fnd&pg=PR5&dq=Boundary+Objects+and+Beyond:+Working+with+Leigh+Star+Objects+and+Beyond%3A+Working+with+Leigh+Star&f=false>.

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>.

Claburn, T. (2021) 'Android's Messages, Dialer apps quietly sent text, call info to Google'. Available at: [https://www.theregister.com/2022/03/21/google\\_messages\\_gdpr/](https://www.theregister.com/2022/03/21/google_messages_gdpr/).

Cormack, A. (2016) 'Is the Subject Access Right Now Too Great a Threat to Privacy?', *European Data Protection Law Review*, 2(1), pp. 15–27. doi: [10.21552/edpl/2016/1/5](https://doi.org/10.21552/edpl/2016/1/5).

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).

Department for Education (2018) *Working Together to Safeguard Children*. March, p. 393. doi: [10.1080/13561820020003919](https://doi.org/10.1080/13561820020003919).

Dijck, J. van (2014) 'Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology', *Surveillance and Society*. Surveillance Studies Network, 12(2), pp. 197–208. doi: [10.24908/ss.v12i2.4776](https://doi.org/10.24908/ss.v12i2.4776).

Edwards, A. and Elwyn, G. (2006) 'Inside the black box of shared decision making: distinguishing between the process of involvement and who makes the decision', *Health Expectations*. Wiley/Blackwell (10.1111), 9(4), pp. 307–320. doi: [10.1111/j.1369-7625.2006.00401.x](https://doi.org/10.1111/j.1369-7625.2006.00401.x).

Evans, W. (2021) 'Amazon's dark secret: It has failed to protect your data'.

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>.

Glavic, B. *et al.* (2021) 'Trends in Explanations: Understanding and Debugging Data-driven Systems', *Foundations and Trends in Databases*. Now Publishers, Inc., 11(3), pp. 226–318. doi: [10.1561/XXXXXXX.Boris](https://doi.org/10.1561/XXXXXXX.Boris).

Goffe, L. *et al.* (2021) 'Appetite for Disruption: Designing Human-Centred Augmentations to an Online Food Ordering Platform', *34th British Human Computer Interaction Conference Interaction Conference, BCS HCI 2021*, pp. 155–167. doi: [10.14236/ewic/HCI2021.16](https://doi.org/10.14236/ewic/HCI2021.16).

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).

'Hobson's Choice' (no date). Available at: <https://www.britannica.com/dictionary/Hobson%27s-choice>.

Hoffman, W. (2011) *Personal data : The emergence of a new asset class*. World Economic Forum, pp. 1–40. Available at: <http://www.weforum.org/reports/personal-data-emergence-new-asset-class>.

Hoffman, W. (2014a) *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](http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf).

Hoffman, W. (2014b) *Rethinking personal data: Trust and context in user-centred data ecosystems*. May. World Economic Forum, p. 35. Available at: [http://www3.weforum.org/docs/WEF\\_RethinkingPersonalData\\_TrustandContext\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf).

Information Commissioner's Office (2018) 'Your data matters - Your rights'. Available at:

<https://ico.org.uk/your-data-matters/>.

Kaye, J. *et al.* (2015) 'Dynamic consent: a patient interface for twenty-first century research networks', *European Journal of Human Genetics*. Nature Publishing Group, 23(2), pp. 141–146. doi: [10.1038/ejhg.2014.71](https://doi.org/10.1038/ejhg.2014.71).

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.

Melendez, S. and Pasternack, A. (2019) 'The data brokers quietly buying and selling your personal information'. Available at: <https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information>.

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).

OFSTED (2015) *Early help: whose responsibility?*, p. 32. Available at: [www.ofsted.gov.uk](http://www.ofsted.gov.uk)  
[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/410378/Early\\_help\\_whose\\_respons](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/410378/Early_help_whose_respons)

Organisation for Economic Co-operation and Development (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Available at: <https://www.oecd.org/digital/economy/oecdguidelinesontheprivacyandtransborderflowsofpersonaldata>

Papert, S. (1980) 'Mindstorms: children, computers, and powerful ideas'. Basic Books, Inc.

Star, S. L. (2010) 'This is not a boundary object: Reflections on the origin of a concept', *Science Technology and Human Values*, 35(5), pp. 601–617. doi: [10.1177/0162243910377624](https://doi.org/10.1177/0162243910377624).

Strohmayr, 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).

Toonders, J. (2014) 'Data Is the New Oil of the Digital Economy'. Available at: <https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/>.

Williams, H. *et al.* (2015) 'Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research.', *JMIR medical informatics*. JMIR Publications Inc., 3(1), p. e3. doi: [10.2196/medinform.3525](https://doi.org/10.2196/medinform.3525).

Woods, D. W. and Böhme, R. (2022) 'The commodification of consent', *Computers & Security*. Elsevier Ltd, 115, p. 102605. doi: [10.1016/j.cose.2022.102605](https://doi.org/10.1016/j.cose.2022.102605).

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).