Understanding and Improving Human Data Relations

Alex Bowyer

Contents

1 Case Study Two: The Human Experience of GDPR

> *"The Crystal Wind is the storm, and the storm is data, and the data is life. You have been slaves, denied the storm, denied the freedom of your data. That is now ended; the whirlwind is upon you . . . . . Whether you like it or not."* — from *'The Long Run: A Tale of the Continuing Time'* by Daniel Keys Moran (computer programmer and science fiction writer)

In this chapter, I will describe the second major case study of this PhD, in which I took 11 participants[1] through an longitudinal in-depth one-on-one process of three interviews with coaching and support in between, with the total engagement per participant lasting approximately 4 hours over a three month period. The purpose of the research was gain a deeper understanding of people's attitudes to the kinds of personal data held by companies in people's everyday lives and what they want from that data (in pursuit of RQ1) and specifically to examine the human experience of existing in a data-centric world [2.1], with each individual having a number of relationships with service providers that involve the use and holding of personal data; in line with RQ2 the goal is to better understand the role of that data in those relationships. In particular, having gained an initial understand of attitudes, hopes and expectations, a further objective was to examine how those expectations might change during the journey of digital life mapping, data request making, receiving and examining of data, and scrutiny of responses, collectively forming a holistic understanding of "the human experience of accessing your data with GDPR."

In section 5.1, I will expand on chapter 2 to explain the context of using GDPR in research as a means to retrieve personal data. In 5.2, I will explain the stages of the interview process (including details of how participants were sensitised) as well as the preparatory and intermediate steps I undertook as researcher. In section 5.3, I will explain the model of personal data types developed for this study, and will present quantitative and summary data from the interviews, explaining how participants' GDPR access requests progressed, highlighting participants' shared hopes and goals, and examining in particular how their perceptions of power and trust were affected by the experience. In section 5.4, I will describe the three themes uncovered through thematic analysis: that organisations provided participants with insufficient transparency to meet participants' hopes and their legal obligations [5.4.1], that people struggle to find meaning and value in their data when they do manage to access it [5.4.2], and that providers' data practices (in particular their GDPR request handling) can be harmful to their users' trust, but that greater openness can have an opposite, positive impact [5.4.3]. I will discuss the implications of these findings with reference to prior literature,

[1] 11 participants started the study but one dropped out after the first interview due to COVID-19, so only 10 participants conducted GDPR requests. 31 interviews were conducted in all.

from the perspective of policymakers [5.5.1], data-holding companies [5.5.2], and individuals [5.5.3]. Finally in 5.6, I will summarise these insights in terms of how they can advance our understanding of the research questions and their wider significance.

1.1 Context: Accessing Your Personal Data Using Your GDPR Rights

1.1.1 The Current Need for Data Access

As established in 2.1.2 and 2.2.4, people live digital lives, inevitably involving the use of myriad digital services that collect personal data, which is subsequently mined for value and exploited at scale, creating an imbalance of power between data holders and data subjects, and a exclusionary landscape around data use which is difficult for individuals to navigate: having acquired data about individuals, this becomes a focus for service providers' decision making and customer relations become less important. This everyday context is the chosen research setting for this case study.

Section 2.1.4 established how unaware many people are of this imbalance around data, that there is a want[2] for effective access to data to restore individual agency. As described in section 2.1.3, policymakers have been attempting since the 1970s to introduce legislation to tilt the balance of power back towards individuals, most recently and most notably the European Union's General Data Protection Regulation, which legally endows at least 513 million individuals[3] with new rights to timely data access, explanation, erasure and correction (Information Commissioner's Office, 2018).

Data protection and misuse issues have grown in the public awareness since the Snowden revelations in 2013 (Gellman, 2013), and have become even more important following the Cambridge Analytica scandal in 2018 ('Facebook–Cambridge Analytica Data Scandal', 2014; Chang, 2018), which may have resulted in manipulation of voting outcomes through personal data use, and the COVID-19 pandemic in 2021 (O'Donnell, 2020; Hamon *et al.*, 2021). Since the GDPR's launch in May 2018, it has undoubtedly resulted in new data access offerings; many large consumer companies have developed 'privacy hubs' or improved privacy policies where individuals14 can learn how their personal data is handled or access data download portals to easily download copies of it ('Privacy - Apple (UK)', no date; 'Privacy & Terms – Google', no date; 'Privacy', no date; 'Face-

[2]In this study and throughout this thesis, my usage of the word 'want' in the context of data capabilities deliberately includes both meanings of the word: the need or desire of the individual, but also that which they **lack** (see opening of Chapter 6).

[3]At the time of writing (summer 2022) the GPDR legally applies in both the European Union and the United Kingdom (which adopted a copy of EU legislation when it left the EU in 2020), which have a total population of 513 million individuals [TODO Replace with reference 37]. GDPR rights are also conferred to any individual who is a customer of businesses with registered offices in EU or UK countries, meaning that these rights are in effect globally available for non-EU, non-UK users of many multi-national digital service providers.

book - Data Policy', no date). Almost all data controllers and processors have now updated their privacy policies to include clear processes for data subjects to request copies of their personal data per their GDPR access rights.

However, it is not known how effective these offerings and processes are for service users, and how individuals feel about them in light of this backdrop of public concern. No service providers make data access statistics publicly available, but anecdotal reports from industry insiders suggest GDPR access rights and data download dashboards are not well-known and hardly used. This presents an opportunity to take individuals who have not previously used these capabilities on a journey of discovery that might enable us to assess the impact of these processes over time and whether—by compelling data holders to create such offerings and respond to access requests—GDPR succeeds in its goals to 'enhance the data protection rights of individuals' (Council of the European Union, 2015) and to give people 'control over their personal data' (The European Parliament and the Council of the European Union, 2016).

1.1.2 Current GDPR Research and its Limitations

Since it came into effect in May 2018, the GDPR has opened up new possibilities for research (Comandè and Schneider, 2021); the ability to obtain one's data records from organisations provides the general public with a potential deeper view inside those organisations, much like the UK's Freedom of Information Act has provided a view into governmental and public sector organisations, enabling research and improving accountability (Savage and Hyde, 2014). Such legally-enforced transparency can also provide researchers with a window into organisations and their processes that was previously only available based on goodwill. Ausloos and Veale (Ausloos, 2019; Ausloos and Veale, 2020) provide an outline approach for using the GDPR in research as well as describing the many ethical and methodological considerations that should be made. GDPR research can however be as simple as inviting participants to exercise their rights of access and talking to them about the experience and any changes in their perspective, which is the approach this study uses, as detailed below.

The GDPR process itself has also been examined from many perspectives by researchers: to understand data holder's compliance with legislation (Ausloos and Dewitte, 2018; Arfelt, Basin and Debois, 2019); to evaluate data portability (Wong and Henderson, 2018) and 'privacy by design' (Waldman, 2020); to compare its effectiveness in public/private sector contexts (Quinn, 2021) or in improving explainability (Hamon *et al.*, 2021), fairness (Kasirzadeh and Clifford, 2021), consent (Human and Cech, 2021), transparency (Spagnuelo, Ferreira and Lenzini, 2019) and the reduction of data breach risks (Gonscherowski and Bieker, 2018). Potential negative impacts have also been considered; the GDPR could be seen as a threat to privacy (Bufalieri *et al.*, 2020) or as an impediment to health research (Clarke *et al.*, 2019).

Clearly the GDPR has spurred a broad variety of research, spanning legal, social

and technology domains. Yet, there is scant research into the individual human experience of the GDPR. Alizadeh et al. conducted a study with 13 users of a German loyalty programme and interviewed them before, during and after they made GDPR data requests (Alizadeh *et al.*, 2019), finding better responses and GDPR education were needed. This is a good example of the sort of work that is needed to explore the human perspective on the GDPR journey, though this particular study was limited in breadth (only one service provider was targeted) and in depth (the data returned from companies was discussed largely at a high level of 'were your expectations met?' and potential to use the data for one's own benefits was not examined). The implications of the experience upon the participants' relationship with the provider were also not explored; it seems that impacts of data handling practice upon relationships is an under-researched area in general. Recent work (Bufalieri *et al.*, 2020; Glavic *et al.*, 2021; Zuckerman, 2021) has established that openness and transparency around data handling are key to services establishing individuals' trust; indeed an echo of this was seen in a public sector context in Case Study One (see Chapter 4). In a commercial context, such changes in trust can impact customer satisfaction and business success.

At a more fundamental level, there is a need to understand how people *experience* the GDPR; companies' GDPR processes have been designed to comply with litigation but often with insufficient design thinking(Cormack, 2021). GDPR-handling processes and data access systems have been motivated by a need to comply rather than by focusing on individual needs or desires [Abowd and Mynatt (2000); McCarthy and Wright (2004); Wright and McCarthy (2008); 3.2]. It is highly likely that many user needs or desires have been overlooked. Such experiential understanding could inform the design of improvements to companies' GDPR mechanisms (be they interface interactions or response-handling procedures), as well as identifying specific needs that might be best met through improvements to policy, including to the GDPR itself.

1.1.3 Human-Data Interaction: Towards a Human-centric Personal Data Ecosystem

Given the fact that data-centric services now span all aspects of our lives, and the amount of personal data about individuals has grown, it has become critical to think about the way people interact with data as a 'whole life' problem. This is one of the reasons this study focuses on the layman rather than a particular demographic, and 'everyday services' rather than a particular domain. Data has transcended the machine and now encodes facts about our lives, it exists across devices and across providers (Weiser, 1991; Mydex CIC, 2010; Abowd, 2012). This means that personal information management has become a sociotechnical problem [2.3.3], that can no longer be solved as a filing-and-retrieval problem as per traditional PIM approaches [2.2.2], but only when considered as multi-party negotiation over representation, ownership, access and consent. It is important to evaluate the GDPR in this context. Up to now, individuals have not had the

means to participate in or initiate such negotiations. On paper, it would seem that GDPR rights do convey this capability, but it is not known whether in practice, service providers' responses to GDPR can actually deliver data subjects the ability to take part in negotiations around data in a fully-informed way. While some research on relationships around data and data as a shared resource is now emerging [2.2.5], the relationship with data-holding service providers has not been examined in this way.

A roadmap for best practice in this space can be found in the emergence of the 'personal data ecosystem' concept [2.3.4]. Researchers have identified that a human-centric approach to personal data is needed, placing individuals at the centre, as controllers and overseers of their own personal data (Mydex CIC, 2010; Symons *et al.*, 2017). This is an emergent space of much activity and research ('Human Data Interaction Project at the Data to AI Lab, MIT', 2015; BBC R&D, 2017; MyData, 2017; Symons *et al.*, 2017; MyData.org, 2018; 'HDI Network Plus, University of Glasgow', 2018; 'HDI Lab, Heerlen', 2020) and provides a strong framing for us to evaluate the human experience of—and interaction with—the GDPR; given people's diminished agency and control over their data (Woolgar, 2014; Crabtree and Mortier, 2016), do the GDPR's access rights, as implemented by service providers, provide the effective access (Gurstein, 2011) people need? Does the GDPR help people to achieve legibility, agency and negotiability, the three tenets of Human-Data Interaction [2.3.2; Mortier *et al.* (2014)]?

This case study aims to explore the research gap identified in 5.1.2 above, from this perspective of greater human-centric need in a sociotechnical multi-party data use context. It will do so by scrutinizing the experience of using one's GDPR rights to discover how well the process meets individuals' needs and expectations; in the process the objective is to uncover problems in order to identify possible solutions that could address them.

1.2 Study Design and Configuration

To address these research objectives, 31 qualitative interviews were conducted, with a convenience sample of 11 individuals from a population of researchers and students at (or connected with) Newcastle University, aged 20-40 years; self-identifying as 5 females and 6 males. Participants were not data experts (only 1 had previously made a GDPR request), but were computer-literate, educated to degree level, and used to reflecting critically on their own behaviours and opinions. Participants were compensated for their time with Amazon vouchers worth £20.

Each participant's journey progressed at its own pace [Figure 5.1] with participants invited to three separate 1-on-1 interviews between December 2019 and April 2020. The scope and purpose of each interview was as follows:

1. **Interview 1: Sensitisation, Life Exploration and Company Selection** [1 hour, in person]. Participants were sensitised to the research context using an interactive tour of a poster display on the topics of GDPR

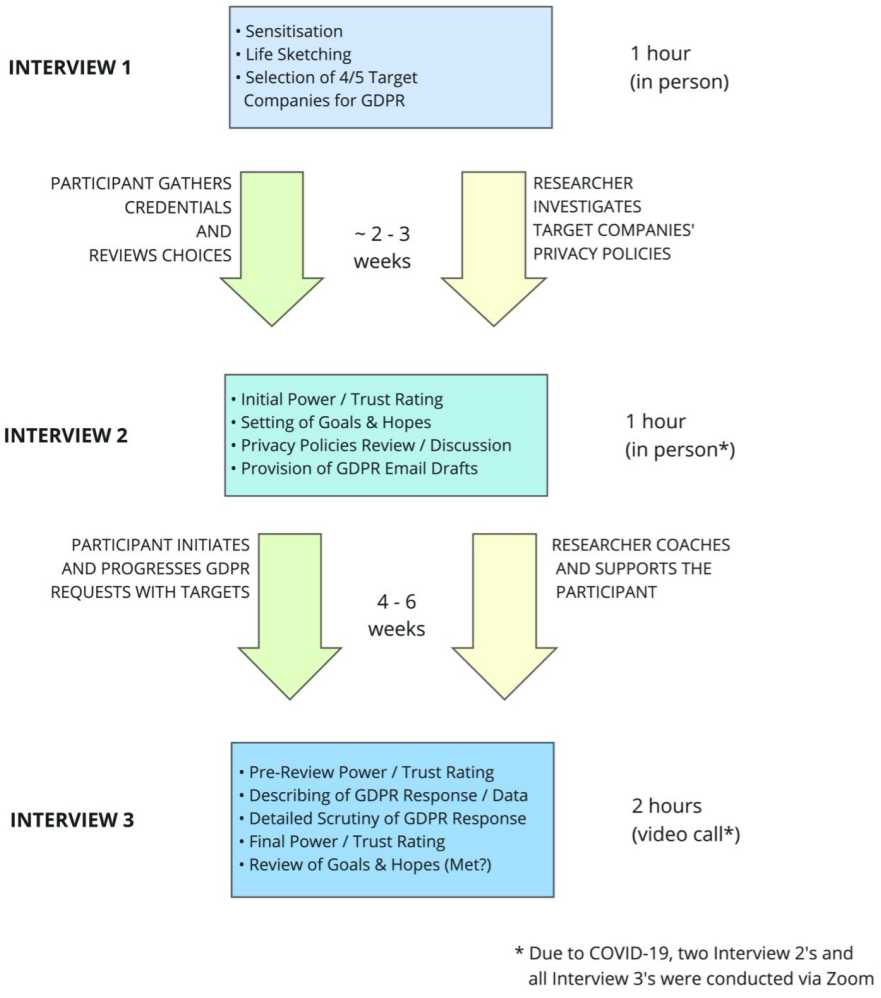

Figure 1: Figure 5.1: A Journey Map of Each Participant's Study Progression

rights, potential data-holding organisations, potential types of data and potential uses for GDPR-obtained data. Baseline data was collected on participants' hopes and motivations, their current understanding of personal data, data access, data control, and power as it relates to data. Using a sketch interviewing (Hwang, 2021) technique, participants mapped out their 'data lives' (e.g. Figure 5.2), annotating key organisations that they have relationships with, types of data those companies might hold, and feelings about such data use and storage by each holder. Each participant selected 3-5 candidate companies to target with GDPR requests.

2. **Interview 2: Privacy Policy Reviewing, Goal Setting and GDPR Request Initiation** [1 hour, in person]. To stimulate reflective thinking and measure impacts, participants were asked to discuss and score their initial feelings of trust and power with each company. Participants then viewed key sections of privacy policies on a screen with the researcher, to identify each company's statements on collection and use of personal data. Participants then initiated an email GDPR request for each company, which had been prepared using a tried-and-tested template generated by personaldata.io (Wiki.personaldata.io, no date). Interview 2 took place in person, except for P10 & P11 whose interviews took place over Zoom due to the COVID-19 pandemic.

3. **Interview 3: Detailed GDPR Response Review** [2 hours, online video call] Having allowed sufficient time for GDPR requests to conclude (there is a legal duty to reply within 30 days), a deep dive into the specifics of each GDPR experience took place. Participants' personal data was not collected by the researcher, only described verbally by the participant; screen sharing was used to show excerpts to the researcher where the participant wished to do so. Participants were asked a structured set of questions about the completeness and value of any data returned, as well as to provide new evaluations of trust and power, whether their hopes had been met, and any general feelings about the experience. Answers were recorded in a screen-shared spreadsheet, which was also used to structure the discussion (for a sample see the supplemental materials of (Bowyer *et al.*, 2022)).

Interviews were audio and video recorded, then auto-transcribed using Google Recorder/Zoom, producing a 370,000-word corpus. Transcripts were split up and recombined across participants into six topic areas - digital life, company-specific discussions, general discussions, power, trust, and hopes/goals. These transcript topics were then analysed through reductive coding cycles to produce thematic findings [5.4]. Quantitative data from interview spreadsheets was also summarised and analysed [5.5]. Sketches, recordings, screenshots and field notes were referenced throughout thematic analysis to aid interpretation of the transcripts.

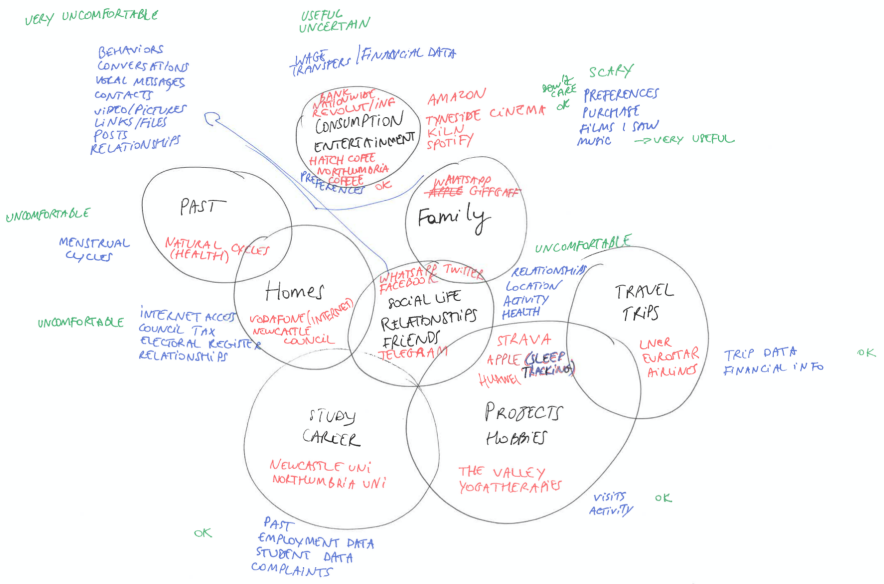

Figure 2: Figure 5.2: An Example Life Sketch from Interview 1, with Data Handling Companies in Red, Data Types in Blue, and Feelings in Green

1.3 GDPR Request Outcomes

1.3.1 Interview 1: GDPR Target Selection

Initially eight participants chose 5 target companies and three chose 4 to request data from. One participant (P9) withdrew from the study due to COVID-19 after Interview 1. Five participants withdrew a chosen company upon further consideration. Reasons for withdrawing chosen targets included having one's personal data mixed with other household members (Netflix), the account being in someone else's name (Morrisons), not wishing to impact active customer support matters (LNER), and inability to contact the provider by email (ifun.tv, see below). One participant selected Newcastle University, which was vetoed by the research team to avoid conflicts of interest. Hence, 41 out of a possible 52 GDPR subject access requests were made (to 28 distinct data holders) as shown in Table 5.1:

Type of Company	Company Names[a]
Major Internet Companies	Apple (3), Amazon, Facebook (4), Google (5)
Hardware Companies	Apple (3), Huawei, Google(5), Philips Hue (smart lightbulb manufacturer)
Online Platforms/Websites	Airbnb, Bumble (dating site), Check My File, Credit Karma, Direct Line, last.fm, LinkedIn
Social Networks & Dating	Facebook (4), Instagram, LinkedIn, Bumble (dating site)
Software/App Manufacturers	Freeprints, Niantic (creators of Pokémon Go), Natural Cycles (a menstrual tracker), Revolut, Spotify
Transport Companies	Tyne Tunnels, Nexus (Tyne & Wear Metro), LNER
Retailers & Loyalty Schemes	Amazon, Tesco, Sainsbury's, Nectar
Telcos	Virgin Media, Three
Sports Clubs	Sunderland AFC

[a] Where a company was chosen by more than one participant, the number of participants choosing that company is shown in brackets.

[PRODUCTION TODO replace table with text]

Table: Table 5.1 - Types of Data Holding Organisation Targeted for GDPR Requests by Study Participantsa

To ensure fairness and consistency, the aim was that all GDPR requests be sent by e-mail by the participant to the data-holder's identified Data Protection Officer, requesting both a subject access request (Information Commissioner's Office, 2021a) and a data portability request (Information Commissioner's Office, 2021b) be initiated, and specifically enumerating and asking for all those datapoints that the company stated in its privacy policy, as well as others which the GDPR entitles individuals to obtain. To identify these datapoints, company privacy policies were analysed and the necessary information was compiled in personaldata.io's semantic wiki ('List of target companies for GDPR requests', no date). This has a feature to generate bespoke GDPR request emails, which were adapted then provided to participants (Bowyer *et al.*, 2022, supplemental materials). Facebook, Apple, Huawei and Philips Hue do not offer a contact e-mail address, so the email text (shortened where length restrictions applied) was pasted into a contact form. In one case, entertainment website ifun.tv, the only available means of contact was via WeChat, resulting in the participant (a Chinese citizen) choosing not to contact ifun.tv due to fear of Chinese government surveillance. Through analysis of companies' privacy policies and with reference to GDPR rights, a taxonomy of the types of personal data that could be returned was constructed, using terms from those privacy policies and GDPR legislation: there are five types of personal data, as shown in Table 5.2:

Type of Personal Data	Description	Examples
Volunteered Data	Data that the data subject has directly provided to the company through upload, contact or form completion.	Personally Identifiable Information (PII), contact details, user-generated content, photos, files, profiles, settings, communication history, financial information, security credentials, surveys/forms.
Observed Data	Data that has been indirectly or automatically collected about the data subject through product/service use or customer/staff interaction.	App usage information, behaviour on website, search/browse history, location tracking/tags, activity/health tracking, technical/device information, network/telco/ISP information, cookies & pixel trackers, staff observations, customer interaction notes.
Derived Data	Inferred data or profiles that have been created through algorithmic or human analysis of volunteered, observed or acquired data.	Interest profiles, advertising demographics, market segmentation, customer categorization, product/service recommendations, internal customer codes.
Acquired Data	Data that has been obtained or purchased from external sources such as civic records, reference agencies, advertisers or third parties.	Public records and information from internet searchs, reports or reviews from individuals, electoral roll data, credit checks, fraud checks, criminal record checks, e-mail/interest lists from advertisers, information shared between affiliates, sister companies or partner organisations.
Metadata	Information about how the other four categories of data have been handled, including storage, processing, uses, decision-making and external sharing.	Names of third parties data has been shared with, details of where data is stored and when/where it has exited the EU, explanations of how data has been used in automated or human decision making, legal bases for storage and processing.

[PRODUCTION TODO replace table with text]

Table: Table 5.2 - Types of Personal Data Potentially Accessible from Data Holders via GDPR Rights

1.3.2 Interview 2: Privacy Policy Review and Goal Setting

Participants reviewed and discussed privacy policies for their chosen target companies and were asked to define hopes and expectations for each GDPR request (see Table 12). 74% of goals expresses related to participants wanting to have greater insight and control into their personal data ecosystems; most commonly a desire to see the breadth and depth of data collection by companies, to understand what was being inferred and how personal data was used, and to use such information to better assess trustworthiness of those companies. Such goals were often motivated by curiosity or suspicion, or a desire to shed light on specific incidents or answer specific questions. In some cases participants wanted not just to learn and acquire knowledge but to take control of or delete held data. In contrast, 26% of goals related to gaining personal benefit from one's obtained data: motivators included the desire to reflect on past data to gain self-insight, as well as goals relating creativity, fun, and nostalgia.

At the conclusion of interview 2, participants were provided with the emails and instructions to start their GDPR requests, which progressed as illustrated in Figure 5.3. Eight requests resulted in no data being obtained, due to either data holder non-responsiveness, inability to access the right account or satisfy ID requirements, or confirmation being received that there was no data to supply. 32 requests (80%) resulted in at least some data being returned; 10 of these directed the participant to use a publicly-available download dashboard such as Google Takeout, and the rest resulted in data being made individually available.

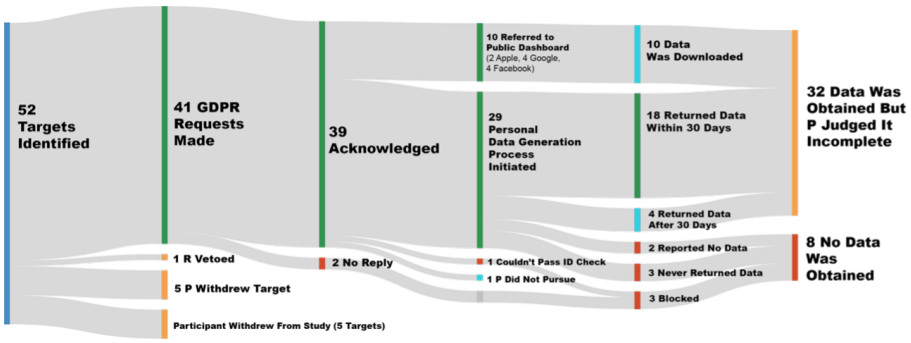

Figure 3: Figure 5.3: Sankey Overview of Participants' GDPR Requests

Of these, one was mailed as printouts, another was mailed on CD-R, and the rest were delivered by e-mail (sometimes involving a secured online website to download). While 22 companies supplied bespoke data packages, 4 did not return it within the 30 days the legislation specifies (note: requests took place within the context of a global pandemic so response rates may not be typical). Following discussion, participants judged that all 32 requests receiving data had failed to return all requested data (across all five of the categories in Table 5.2).

1.3.3 Interview 3: Reviewing the GDPR Response

Once each participant's GPDR requests had reached a conclusion point (as described above), they were invited to discuss the GDPR response in detail. Participants were asked to describe (and optionally show) the data they had received, then to evaluate the data holder's response for each data type, according to multiple metrics designed to assess the perceived quality of the GDPR request handling and the subjective value of any returned data. All questions were posed from the perspective of (a) the data that providers said they collect and process in their privacy policies, and (b) the rights that the GDPR specifies, to ensure discovery of missing data or unfulfilled rights would be considered objectively. Participant responses were considered quantitatively (see Table 5.3 below) and qualitatively (see 5.4).

Type	Valued? [b]	Returned?	Complete?	Accurate?	Understandable?	Meaningful?	Usable?	Useful?
Derived	82%	39%	10% (dk:13%) [c]	20% (dk:20%) [c]	40% (p:40%) [d]	40%	0%	20%
Acquired	81	49	16 (dk:16)	50 (dk:25)	75 (p:0)	50	25	17
Metadata	73	4	0 (dk:7)	0 (dk:0)	0 (p:100)	0	0	0
Volunteered	57	53	55 (dk:0)	92 (dk:0)	72 (p:20)	72	52	58
Observed	48	33	18 (dk:12)	57 (dk:30)	61 (p:20)	57	52	61

[a] Percentages represent the proportion of "Yes" answers to each question, per data subtype, from all those *where a judgement was given*.
[b] Participants were asked whether this category of data from each provider would be valuable *if they were to receive it*.
[b] dk = don't know (percentage of cases where participants felt unable to assess data accuracy or completeness).
[c] p = partially (percentage of cases where data was judged *partially* understandable).

[PRODUCTION TODO replace table with text]

Table: Table 5.3 - Presence and Quality Assessments of GDPR Responses by Data Type (as Percentagesa)

Table 5.3 shows quality assessments for each data type, with rows descending by subjective value. Notably, the kinds of data participants value most (derived, acquired and metadata) were less frequently returned, especially metadata (returned in 4% of cases). Where data was returned in these categories, it suffered from poor data quality, often judged as incomplete, inaccurate, unuseable and not useful (although acquired data was largely understandable). At 53%, even the most returned category, volunteered data, was lacking. Where it was returned, accuracy (92%), meaningfulness (72%) and understandability (72%) were high. Observed data was least valued and also rarely returned or complete (yet judged to be of moderate quality). Looking below organisation level at the perspective of individual data categories, data was only judged to be complete in 22% of cases. In 62% of cases, personal data specified in privacy policies to be collected was not returned, despite the legal obligation.

Data collected in these interviews also allowed comparison of different companies. This analysis is included in the Additional Reference Information.

At the conclusion of the final interview, participants were reminded of the specific hopes and anticipated data uses they had expressed at the start of their journey and asked about how well each goal had been met. These answers were recorded and combined to produce percentage values showing in how many cases goals were fully met, partially met, or not met at all, as shown in Table 5.4.

Participants felt their goals were not fully met in 78% of cases, and 54% were not met at all. Specific shared problem areas included (1) the desire to understand what providers infer from held data (7 participants), which was unmet in 73% of cases and only fully met in 7% of cases; and (2) the desire to delete one's data, which was a stated goal in 10 cases but was only met in one of them. Four wholly unmet hopes were to investigate specific incidents (GDPR responses were often delivered as a one-off package without any kind of backchannel or opportunity to ask questions), to secure data, to check accuracy, and to move data to another service.

Hope or Goal	Distinct instances of this goal	Distinct participants	Specific companies in mind for this goal, if any	Was this hope met? Unmet?	Partially met?	Fully met?
GOALS RELATING TO ACCOUNTABILITY AND CONTROL (74%):						
Understand the breadth and depth of what data is collected	24	7	Amazon, Apple, CheckMyFile, Credit Karma, Facebook, Google, LNER, Nectar, Philips Hue, Spotify, Tesco, Three, Virgin Media	42%	17%	42%
Understand what is inferred about you from your data	15	7	Amazon, Apple, Direct Line, Google, Instagram, last.fm, LNER, Spotify, Tesco, Three	73	20	7
Assess provider trustworthiness	12	6	Apple, Credit Karma, Direct Line, Facebook, Freeprints, Nectar, Niantic, Sunderland AFC, Tesco, Three	42	42	17
Remove your data & control/limit its use	10	3	Bumble, ifun.tv, Instagram	90	0	10
See inside 'black box' algorithms & processes	9	4	Amazon, Facebook, Google, Tesco	56	11	33
Understand how and why your data is used	6	5	Direct Line, Google	50	33	17
Investigate specific questions or incidents	4	4	AirBNB, Three, Credit Karma, Instagram	100	0	0
Learn about data use and how to be safer online; educate others	3	2		0	33	67
Secure data about you and identify risks and leaks	2	2	Apple, Facebook	100	0	0
Check accuracy of data about you	1	1	CheckMyFile	100	0	0
Move your data to another service	1	1		100	0	0
Test your data rights	1	1		0	100	0
GOALS RELATING TO USING DATA FOR PERSONAL BENEFIT (26%):						
Reflect on past activities & gain insights	14	5	AirBNB, Apple, Google, last.fm, LNER, Tesco, Virgin Media	57%	36%	7%
Find patterns/habits & track goals	6	5	last.fm, Nectar, Spotify, Tesco	17	50	33
Combine data from many sources for deeper insights	3	2	Philips Hue, Google	33	67	0
Play with, create, hack & remix your data	3	3	Google	67	0	33
Nostalgia, fun & inspiration	3	3	Spotify, Niantic	33	33	33
Keep your own data archive	2	2	last.fm	0	50	50
OVERALL	**18 goal types**	**10 people**	–	**54%**	**24%**	**22%**

[PRODUCTION TODO replace table with text]

Table: Table 5.4 - Participants' Hopes, Imagined Data Uses and Goals for GDPR, as well as Resultant Outcomes

1.3.4 Perceived Power and Trust

Repeating scoring questions were used to examine how participants' feelings towards the data holders changed throughout the process: Participants were asked to assess trust from 0 (total distrust) to 10 (total trust), and to assess their perceived power on a scale of -5 (total provider power) through 0 (balanced power) to +5 (total individual power). Explanations and reasoning for initial ratings and for any changes were uncovered through questioning. By repeating the same question at different times, longitudinal comparisons could be made. Many participants' attitudes did change as a result of the experience [Figure

5.4] for both perceived power (45% of cases) and trust (66% of cases). For those with changed attitudes, the change was often negative: in 63% of cases where participants perceived a change in individual power, that change was a loss in individual power, and in the majority (52%) of cases, participants felt more distrustful of GDPR targeted companies after completing the process (constituting 79% of cases where a change in trust was perceived). However, it is important to note that in some cases GDPR had a positive impact; in 17% of cases participants felt their perceived power had increased, and in 14% of cases participants felt more trusting of providers after GDPR.

Figure 4: Figure 5.4: Longitudinal Distribution of Net Changes in Participants' Perceived Power and Trust Scores

Looking deeper into these datapoints, changes in attitude could be attributed both to the impact of reviewing the privacy policy as well as to the experience of the GDPR process and the discursive review of GDPR responses. Figures 5.5 and 5.6 show snapshots of power and trust ratings at different points in the process which illustrate these impacts. Looking to explain these changes qualitatively, it was found that privacy policies often contradicted participants' expectations, resulting in discomfort. In two cases (Philips Hue and last.fm) privacy policy review revealed that the service relationship was with a completely different company than the participant thought, which was disturbing to them. LinkedIn's privacy policy was noteworthy as being exceptionally clear, reassuring and trust-enhancing to the participant, largely due to its 'easy read' text sidebars but also good use of examples. However it does appear that simplifying privacy policies can go too far: Google's privacy hub (which includes video explainers) was considered easy to understand but necessarily broad (given their breadth of services) and thus over-simplified, raising uncertainty about generalisations

made, and in some cases increasing distrust.

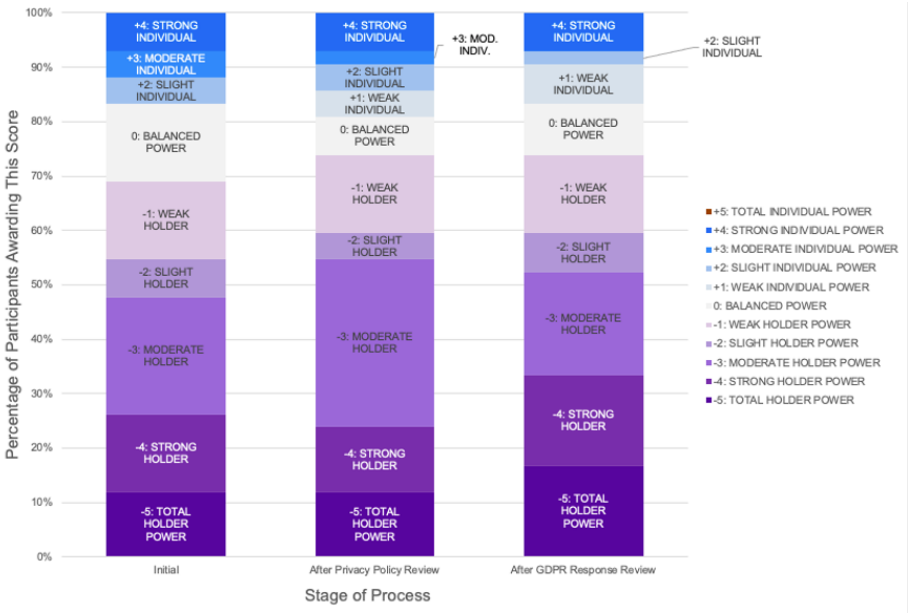

Figure 5: Figure 5.5: Perceived Power Balance Between Individual and Data Holder at Different Stages of the GDPR/Study Process

Considering the process as a whole, participants' attitudes were impacted particularly by the *"hassle"* (P11) they experienced in getting through the data access process, and from the realization that what seemed at first glance to be a thorough response, when scrutinised more closely in Interview 3 and viewed through the lens of the privacy policy promises and one's GDPR rights to the five categories of data, was in fact quite poor.

1.4 Thematic Findings

As described in 5.2, the topic-focused transcripts were carefully examined to identify themes and findings, a process involving over 200 person-hours of iterative data analysis (Huberman and Miles, 2002) of the interview transcripts. In this section the findings from that analysis are presented and summarises, with the three main themes being introduced in 5.4.1, then each theme is further detailed in sections 5.4.2 to 5.4.4, including participant quotes.

1.4.1 Themes & Subthemes

The findings are presented through three main topics: Insufficient Transparency, Confusing and Unuseable Data, and Fragile Relationships.

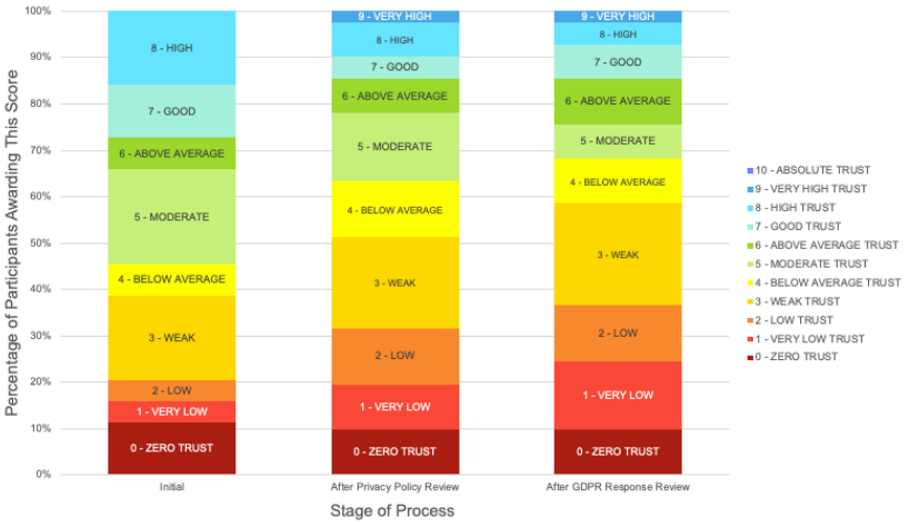

Figure 6: Figure 5.6: Participants' Perceived Trust in Provider at Different Stages of the GDPR/Study Process

Theme 1 (see 5.4.2 and summary in Table 5.5 below) describes the **Insufficient Transparency** that participants experienced in privacy policies and through the GDPR process; Organisations appear evasive over data when responding to GDPR, leaving people "in the dark" even after making GDPR requests.

Table 1: Table 5.5 - **Theme 1 - Insufficient Transparency**. Subthemes & Participant Quotes.

Subtheme	Description	Quote
A Desire for Awareness and Understanding	Participants want to see, know and understand the data held about them. There was particular interest to see data collected or inferred about them without their involvement, and to understand how data is used and shared and how that might affect them.	*"[Companies have more power] because they're making decisions about things and you don't know how they're making those decisions."* (P5)

Subtheme	Description	Quote
Non-Compliance Without Consequence	Many providers failed to provide data on time or at all. In 100% of cases, returned data was incomplete, and many viewed this as non-compliance. Data holders' freedom to disobey legislation was attributed to a lack of enforcement and seen as an exertion of power.	*"I am surprised at Google's unwillingness to provide me with all of the data . . . they haven't provided me with all of my data. And that's not legal."* (P7)
Inadequate Data Responses	Participants judged data holders to be unhelpful, GDPR procedures to be painful and ineffective, and returned data to be lacking in coverage and in quality. Their questions remained unanswered; after GDPR they were still *"in the dark"* (P4). There was widespread disappointment and a view that GDPR did not confer any power to the individual.	*"It's kind of disappointing because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn't] seem to be doing that."* (P1)

Theme 2 (see 5.4.3 and summary in Table 5.6 below) explains how participants received **Confusing & Unuseable Data** from data holders; When presented with their data, people struggle to understand it and relate it to their lives and are not able to make use of it.

Table 2: Table 5.6 - **Theme 2 - Confusing & Unuseable Data**. Subthemes & Participant Quotes.

Subtheme	Description	Quote
The Search for Personal Value in Data	Participants found the large volumes of data that were sometimes returned overwhelming, and wanted summaries and breakdowns to understand it, as well as tools to help them make sense of and explore the (often technically formatted) data. Data that spanned a period of time was judged particularly meaningful as it could serve as a window into past memories and would allow for trends and changes over time to be observed.	*"[It's] almost too much [. . .] for a normal person to be able to process and understand [. . .] It could do with a document detailing, like, 'this is what is in here'."* (P1)
Unuseable Data Formats	Participants anticipated receiving data in formats they could explore, visualise, mashup and play with, but in fact often received data that lacked explanations. Data was often arranged in ways that were more reflective of internal systems than being optimised for use or understanding. Both useable data files and explanations of how to use it are needed.	*"They did give me the data, but not how it fitted together. It's like being given the bricks to a house, and then they're like 'Here's your house'. It doesn't really mean anything when it's just bricks, if you don't know how to put it together."* (P5)

Subtheme	Description	Quote
The Liability of Data You Can't Delete or Control	Having understood that the amassed data about them could be exploited by businesses or third parties, participants wanted to see their data so that they could understand the extent of that capability, and wanted to be able to restrict its use and, in some cases, delete their data. No clear pathways to take such control were offered, nor was this control practical to achieve; the retaining of data against their wishes was seen as a liability and a lack of control.	*"[Companies did not] tell me what they are doing with [my data]... And sometimes I think my willingness to give a company data might be quite intrinsically linked with what they're gonna do with it."* (P7)

Finally Theme 3 (see 5.4.4 and summary in Table 5.7 below) examines the **Fragile Relationships** that individuals have with data holders; Companies' data practices, and in particular their privacy policies and GDPR response handling, can be impactful to customer relationships, carrying a risk of damaging trust but also the potential to improve relations. These three themes are

Table 3: Table 5.7 - **Theme 3 - Fragile Relationships**. Subthemes & Participant Quotes.

Subtheme	Description	Quote
Power and Enforced Trust Through Data Holding	Participants feel that the sacrifice of (or the giving of permission to collect) personal data is a necessary cost in order to get the valued benefits of the services they want to use, something they are pressured to do and have no choice about. Such sacrifice is seen as the giving up of power, as participants lack access and control to that data. In the face of providers making decisions based on data and processes that they could not observe, participants felt powerless. This amassing of data was sometimes seen as surveillance, and some saw great potential for misuse and abuse of it.	*"For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it's basically a proxy to have power over me."* (P8)
Perceptions of Data Holders	Participants entered the study with varying perceptions of providers' integrity, influenced by reputation, business model and size. Participants' various observations reveal a strong link between their perceptions of providers' data handling practices and the trust they hold in those same providers.	*"When I like the company already, I'm more willing to give them my data."* (P2)

Subtheme	Description	Quote
Changed Perspectives Through Scrutiny	In general, the more that participants found out about data-centric practices through the process of scrutinizing privacy policies and making data access requests, the more they distrusted providers. Failure to explain or provide complete data was harmful to trust. Conversely, where providers were more transparent or participants did obtain interesting data insights, trust was increased.	*"If someone's not completely open with you, then you're like, 'What are you hiding?', which means you trust them less."* (P4)

1.4.2 Theme 1: Many Companies are Evasive and People are "Still in the Dark"

A Desire for Awareness and Understanding

As Table 12 shows, in the vast majority (62%) of cases, participants wanted to see, know and understand what data was held about them and how it was used. For example, P11 wanted to know what data was collected by train company LNER when he bought tickets, so that he might judge whether it was appropriate:

"I'd be interested to understand what data they have [. . .] Is it just the patterns of my spending on trains, or is it a bunch of other stuff that they're using for advertising to me?"—P11

Beyond the data that participants had directly volunteered [Table 5.2], most data was currently unknown to participants. In particular they wanted to gain awareness of what data might have been collected without their knowledge.

"The bit that concerns me is where I don't know what data is being taken by companies. If I'm registering for a library or something, I know [what] data I'm giving to them, but what I don't know is all the other stuff that they're recording"—P9

Participants were equally unaware of what holders might infer from the data they had collected. P4 wondered if Philips could use data from his smart home lighting to deduce his sleep and TV-watching routines. P7 had received targeted advertisements relating to pregnancy that she felt weird about because she did not understand why she had been targeted in this way. P5 raised concern about how data inferences could affect decision making, surmising that the data holder had greater power than him because *"they're making decisions about things and you don't know how they're making those decisions"*. Sharing of personal data

is also insufficiently visible to participants; two participants (P3,P4) targeted GDPR requests to credit-check websites (Credit Karma, CheckMyFile) - P4 wanted to get *"a picture of what other companies can currently expose"*.

Non-Compliance without Consequence

As detailed in 5.3.2, few requests resulted in a timely provision of requested data (44% or 68% depending whether referral to a download portal is excluded or included in the count). Many data holders responded late or not at all; such actions are objectively a breach of legislation. However, participants were broadly unsatisfied even when they did receive a GDPR response. In 100% of cases where data was obtained, it was considered incomplete, and this was usually seen as further failure to comply. Participants had reviewed their GDPR rights in Interview 1 (though, as expected (Rughiniș *et al.*, 2021), most were already aware), and so several participants saw this apparent non-compliance relative to their understanding of their rights as a poor quality of response, for example:

"I feel more concerned now, [. . .] what they've given me seemed reasonable. But then comparing against what we asked them for, what I'm legally [entitled to], it's a fraction."—P5

For some participants, sceptical from the start, such poor responses were consistent with their expectations; P6 found the incompleteness of Facebook's response *"alarmingly unsurprising"*. Others had expected compliance:

"I am surprised at Google's unwillingness to provide me with all of the data. . . they haven't provided me with all of my data. And that's not legal."-—P7

Many participants, reflecting on a feeling of having less power than they had initially thought, felt that the prevalence of non-compliance showed that too much power relative to the authorities, that a lack of pressure is being applied by regulators and that *"there needs to be more enforcement"* (P11). P6 revised his view of Facebook's power versus his own because he felt that after review he now could clearly see *"which [data] they are prepared to share and which they aren't"*. P11 also framed the selectivity of responses as an exertion of power:

> *"It seems like there's a lot of derived data about things like purchases and stuff [that I would expect] that just isn't there. So they're free to not give me the data. That, to me, suggests [that despite GDPR] they retain an awful lot of power."*-—P11

Inadequate Data Responses

While in some 22% of cases participants did meet their goals through GDPR (see Table 12), when it came to the desire for greater awareness and understanding [5.4.2], this want was largely unmet. Only volunteered data such as basic personal information or user-generated content was usually returned complete; this was often viewed as mundane and uninteresting, and the focus on these data types in returns was viewed as evasiveness. Facebook, P6 observed, *"give you that kind*

of descriptive boring data which is mainly all publicly available anyway" and had omitted *"the stuff that I would consider valuable to them"*.

In general, the data responses did not provide the answers participants sought. Many reported *"still"* not knowing what they wanted to find out. P4 said he remained *"in the dark"* (P4). P7 stated that *"even though I did the process correctly, I still didn't get that much back"*. Concerns held by participants from the outset remained unaddressed, as in P11's case:

"I still am quite concerned about how much data organisations have, particularly how they link that other data and how data is bought and sold, and I haven't really got any answers on that."

It was not just the data returned, but the process itself, that participants were dissatisfied with; requesting and achieving data access was time-consuming and difficult. *"Jumping through hoops"* was a phrase used independently by four different participants (P4, P5, P7 & P11) to describe the experience. Some found data holders obstructive and unhelpful:

"I feel like they give you a response that [makes it so] you cannot proceed intentionally"—P10

Participants recognised that they had received help and coaching, and that the processes were so tedious that without that, they may not have persisted. P1 suggested that without the provided template, it would be *"a lot harder to get meaningful data out"*, and P7 attributed her sole successful request to the guidance she had received in progressing it. P5, having experienced problems with expiring links, delayed responses and missed emails, had been surprised at *"how difficult it was just to get my data, and the fact that I had to ask them about six different times"*.

Not all requests were this painful, some were handled smoothly. As P11 put it, *"Some companies make it dead easy to get, but then the data is not massively useful.[. . .] Other companies make it a pain in the neck to get it."* Overall the view of GDPR data access was one of disappointment. Participants found GDPR ineffective: P10 said *"Frankly, [GDPR] doesn't have as much influence as I expected"* and P1 commented that:

"It's kind of disappointing, because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn't] seem to be doing that."

1.4.3 Theme 2: People Struggle to Understand, Use and Control Their Data

The Search for Personal Value in Data

Prior to receiving data, participants had anticipated discovering insights about their own lives by browsing and reflecting on their personal data, consistent with personal informatics literature (Li, Forlizzi and Dey, 2010). However, there was a

comprehension gap between the useful information they imagined and the actual data returned; data was typically delivered as a bundle of technical files, which were hard to understand and often delivered without explanation. Some felt (in line with *effective access* [2.1.4]) that they lacked the necessary skills or tools to make the data understandable or useable *"for a non-techie person"* (P11). When the researcher guided P7 to jsonlint.com, an online formatter, she found her JSON-formatted data more understandable. P2 made the point that data holders must be using tools themselves to make sense of people's data: *"They're not just looking at a JSON file, so I would like to have the same visualisation [as them]."*

There was a sense that by sending people individual data files, data had been removed from the environment in which it has meaning, and that the returned data excluded necessary context for interpretation. This was often manifested in the form of internal codes and abbreviations that individuals could not understand. P4 stated of his experience looking at smart-lightbulb data from Philips Hue, that there was *"just so much of it that it's impossible to know [what it all means]... You'd have to spend a few hours going through this and being like, 'OK, what does that line mean, and that symbol, and that code?' "*. This lack of context also materialised as a failure to explain decision-making processes: P5 reflected, when looking at driving scores from a car insurer that uses a mobile app to monitor her driving, *"I could see the data; it was the score that was weird for me. Like, it doesn't tell you how it's calculated."* P1 noticed that although some companies did make some effort to explain the returned data, this varied substantially across providers. He said that *"it would be nice if these companies had a standardised model of how this information is presented to people, so it [could] be easily understood"*.

One of the greatest obstacles to understanding that participants faced was being faced with a large volume of information and no means to quickly digest or navigate it; either very large files, or complex hierarchies of nested directories containing many separate files. It is clear that there is a need for *summaries* so that participants can quickly get a handle on what is - or is not - present. Returned data *"could be valuable if you knew what the hell [was] in there"* (P4). P1 described one of his data responses as *"almost too much [...] for a normal person to be able to process and understand."* He said that it *"could do with a document detailing, like, 'this is what is in here' "*, and described the disparity across responses as being *"either like death by thirst or death by drowning [...] It would be better to drown, but still not ideal"*. Ultimately it is clear that in general, returned data was not presented in a way that is optimised for understanding.

Another question that our findings were able to shed some light upon, in service of RQ1, was to consider what precisely makes data valuable to individuals. This is especially important given participants did identify the potential to gain personal benefits from their data (as seen in the second set of goals in Table 5.4). An idea that came up again and again was that data is most valuable when it *spans a period of time* and can be related to events in the individual's life over

that period. This could potentially provide new insights to participants.

P2 for example hoped to see, or be able to construct, breakdowns and charts that would help him examine his food shopping habits. Through the GDPR process, P10 accessed details of her spending on micro-transactions in the mobile game Pokémon Go that had not been available to her through the app. P11 wanted to derive insights about his train travel by examining the geography, cost, journey length and patterns of his past journeys through data he hoped to receive from LNER. Long-duration data offers the potential ability to identify trends and changes in one's own behaviour over time.

It was these historical parts of their data that participants found most meaningful, offering as it does a means of remembering, with data potentially serving as a *"window into your past"* (P11). P5 saw value in perusing music-listening data *"just because it's cool to look back on stuff that you've done and you don't necessarily distinctly remember it"*. Generally the longer period the data covered, the more valuable it was deemed to be:

"I would actually be interested in last.fm, partly because the data goes back to 2008 . . . Spotify only goes back about four or five years and not everything I listen to is on Spotify."—P11

P6 saw the data accumulated by service providers as potentially forming part of a valuable background context to understanding life events in his past: *"I would like to [. . .] build a picture, not just like, 'I remember going to Reykjavik', but if there's other data around that time [I could] sort of paint a biography of myself"* and described some of his data as *"a kind of personal history that has been quantified and sort of datafied"*.

Unuseable Data Formats

This personal value that captured data has the potential to offer shows that it is all the more important that participants be able to understand and make use of their data. Our participants found that the format in which data was returned often meant that it was not only difficult to understand, but difficult to use as well. Using data meant different things to different participants, with imagined uses including budgeting, record-keeping/archiving, or using the data for creative or fun purposes. Some participants (e.g. P5) saw value in potentially combining data from multiple sources, though this did not turn out to be practical. Participants did not know what data to expect, and generally imagined returned data being more useful than it turned out to be:

"I think . . . you could do some interesting mashups, but I don't really know what with until I've got the data. It depends on the data; I'm sure there could be some cool uses of it."—P4

Once data was received, participants struggled to interpret and understand it to a sufficient extent to be able to identify the useful data or meaningful information they had hoped for. Returned data formats and response structure

were extremely varied. Some reported that there was not sufficient machine-readable data to make use of the data. For example, P4 received a Microsoft Word document full of pasted screenshots from an internal portal as part of his response from his ISP Virgin Media, and said that its usefulness *"depends on what you want to get out of it, really. If you want to view the data they have about you, it's quite useable. If you want to do something automated[analytical], then it's not"*. P11 found a similar returned screenshot from an internal system to be *"completely non-understandable"*. In other cases, the opposite problem occurred, with data being too technical for the participant to use. P10 said of JSON data: *"For normal people who don't understand programming, I feel it's just, there's no use at all."* P7 felt she lacked the technical proficiency to make use of the returned data:

"They have provided it in formats where I can see that, if I were a developer, I could do things with it, [. . .] but if I was not that sort of person, it might be quite difficult to understand"—P7

In P5's case, she saw the potential to use the data but felt that what was missing was additional explanation or guidance on how to interpret it:

"They did give me the data, but not how it fitted together. It's like being given the bricks to a house, and then they're like 'Here's your house'. It doesn't really mean anything when it's just bricks, if you don't know how to put it together."—P5

P11 highlighted a problem with his Tesco shopping data that was not just a matter of formatting or skill, but the granularity or focus of the data itself:

"As a technical person, having a CSV of data is quite useful, potentially, but actually what can I do with that if it's Tesco's internal systems data?"

On the face of it this finding seems to contradict the previous one—i.e. demands for both more technical and less technical data might seem contradictory. However, what we can infer is that participants collectively need *both* useable technical data *and* easy-to-read information summaries - and that those summaries should cover both the relatable life information encoded within the data *and* the information *about* the data, what it means and how to use it. This idea is explored further in (Bowyer, 2021).

The Liability of Data You Can't Delete or Control

Having recognised that potential value of data relating to their lives, before or during this research, several participants were concerned about personal data being held. P10 for example said with reference to dating site Bumble: *"Since I found my partner [and therefore no longer need a dating site] I deleted my account and I've been wondering, 'Are they still keeping my data at the back?'"* and with reference to both Instagram and Bumble, expressed a desire to have her data deleted and expected GDPR to play a role in the enforcement or verification of that deletion, something she could not otherwise be sure of. P8 considered the holding of sensitive data to be a liability that she was only willing to tolerate

while she was actively using a service, and this was part of her motivation for targeting Natural Cycles:

"I now use a different one, but I used, for about a year [their] app to track my menstrual cycle. [It was my] main contraception method, so that's things that this company probably has. Now that I'm not using it any more, I don't know if they delete the things or not"

Many participants expressed a desire that data be held only for a short time, and questioned the default practice of data being kept beyond the period where it was needed to deliver a service:

"The thing that concerns me is that I haven't used Tesco online for at least four or five years, so why are they hanging on to my IP address from five years ago?"—P11

He went on to spell out the liability he saw in such apparently mundane data being held, the liability coming from the duration of the data: *"10 years of worth of shopping records... how much would that be worth to a health insurance company, and would [Tesco] succumb to the temptation to sell that on?"* P10, a Chinese citizen, identified long-term sources of personal data as an enabler for future privacy violations, saying that *"in China, [there is a trend] that as soon as someone becomes famous, people begin digging [through] all their past experiences".*

Most participants described the ability to delete or enforce the deletion of their data as having control over it, and given the current practical lack of such a capability felt that they had insufficient control over data holding. One of the first steps participants identified in gaining control of their data was simply an ability to see it, for accountability, so that they might check the accuracy, security and breadth of collected data and flag any unforeseen concerns. They felt that a deeper understanding might lead to an increased sense of individual safety and data control and facilitate them to make changes in data habits or choice of service provider:

"I want to understand how much they're keeping. And what they're doing with it. I'm hoping that by knowing that, I might change my behaviour about all the data I accidentally create."—P7

In this participant's case, this hope was unsatisfied, and upon looking back at her experience she remarked:

"I guess that's one of my criticisms of GDPR in general - that although I can understand what data a company holds about me, there's no obligation for them to tell me what they are doing with it.. And sometimes I think my willingness to give a company data might be quite intrinsically linked with what they're gonna do with it."—P7

In fact, that legal right does exist through GDPR, but as we can see it was not delivered in practice. What participants want is to feel aware and in control of

their data; this must begin with better data legibility and explanations of data use, accompanied by clear pathways to enable data correction or deletion.

1.4.4 Theme 3: Poor GDPR Handling Can Damage the Forced & Fragile Trust Relationship

Data Holders Enforce an Uneasy Trust

The lack of visibility and control over personal data discussed in the previous section, combined with a sense of being in the dark [5.1] about data practices, caused participants discomfort before, during and after the GPDR process. This stemmed from a sense of finding themselves facing uncertain risks they feel powerless to change. Many participants, discussing their relationships with providers, expressed a range of emotions ranging from curiosity to anxiety and distrust:

"I'm curious... I wonder what they've got on me. [...] If it's anything other than the barest minimum that is necessary for them to do their job [...] then I get creeped out by that."—P11

Participants felt most uneasy about the amount of *"intimate"* (P1,P2) data that providers collect. P1 was uncomfortable about Facebook having information about his social circles. P2 said he felt *"quite vulnerable"* that his Google search terms *"say pretty much everything you have done... the most intimate things you were thinking about"*. P11 singled out ISPs as having the potential to track everything their customers look at online, noting that *"I don't think you've got much choice about that."*

Participants also felt that some data holders held so much data that it had begun to resemble surveillance, such as in the case of P1, who used *"an absurd amount of [Google's] services"* and reflected that *"if I'm driving somewhere, I've got Google Maps open, so they know exactly where I'm going, they know how fast I'm going, they know what I'm listening to while I'm driving..."*. Participants saw the potential for abuse, fearing this kind of deeply personal knowledge could be *"used against"* them (P2). P11 felt that Apple had enough data to *"screw me over"*, and P5 considered that her car insurer Direct Line uses data to *"judge"* her, noting that *"it's not like I can contest the data and say 'Actually, no, I disagree!'"* In a more extreme illustration, P10 shared her fears that data collected by WeChat and Weibu (Chinese services similar to Facebook Messenger and Twitter respectively) would be at risk of abuse against citizens by the Chinese government. In some cases participants were able to identify concrete instances where providers had exploited the personal knowledge they held: in P6's view, Facebook use their knowledge of their users' friendships and relationships to *"hook your attention"* and prevent users deactivating accounts in a *"disingenuous"* manner.

Whether or not data is used nefariously against individuals, thinking about the potential for this caused participants to associate the mass collection of

personal data as an acquisition of power over them: *"[Companies that] know a lot about everyone will inherently be able to have power either through persuasion or manipulation"* (P1). P7 saw the **holding of data** as the source of holders' power: *"when I think about other people having my data [. . .] the control isn't sitting with me."*. Others identified the ability of data holders to **deny or limit access to data** as their key source of power:

"If you're not getting what you perceive to be yours back in completion [sic] then you're not in control of your own data and you have fairly little power over it."—P1

The view of data holders having more power in the service relationship (mirroring reports described in 2.1.2) was reflected in participants' evaluations of power balance: in 69% of relationships participants felt that the data holder had more power than them (rising to 74% after GDPR), whereas in only 17% of cases (unchanged by GDPR) did participants feel they themselves had more power.

Several participants equated power over a person's data with power over the individual. When asked to define power in the context of data, P8's unprompted comments aligned with my Case Study One findings describing power over data as a proxy for individual participation or inclusion in decision making [Bowyer *et al.* (2018); 4.2.2; 4.4.1]:*"For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it's basically a proxy to have power over me."*

A key dynamic to understand the value exchange within these relationships is that individuals sacrifice their data in exchange for value - that value being the capabilities offered by the services. All 11 participants expressed the idea that the sacrifice of data is something that they have grown to tolerate in exchange for some benefit. P6 tolerates data collection by travel agents because *"they might help me pick a better deal next year."* P11 said he was happy for Tesco to collect data in order to *"profile me to try to sell me more cheese, fine, whatever,"* though expressed caution that he doesn't *"know what else they're doing with it,"* and more generally was *"deeply concerned"* about unseen data trading. The benefit can be convenience too; P10 had logged into Pokémon Go with her Facebook account, knowing that implied data collection by Facebook, *"because it's much easier"*.

Participants often felt this sacrifice was something they had no choice about, but they did not like it. Unease over the trade-off being made surfaced most often in the context of recommendations; generally, participants valued data-derived suggestions provided they were *"relevant"* (P1, P8) and not too *"intrusive"* (P1, P6). It is clear that data sacrifice is only tolerable up to certain limits: P10 said of Niantic found the sacrifice acceptable provided that *"they don't sell where I live or my daily routine"*; however while Niantic's privacy policy promises data is not sold, it does appear that some level of personal location information *is* accessible in some form to third-party advertisers (Varghes, 2019). P8 said that relevant music recommendations were *"very useful"* but found Amazon shopping

recommendations *"very scary"* because *"I don't want to see that I'm predictable"* and felt that *"if someone out there knows [what I want] before you [it's] like taking agency away from me."*

Permission to collect and use data is knowingly provided by individuals to data holders, but the mechanisms to do so are considered inadequate: P2 felt that permission giving options are *"not granular enough"*, and in P11's view *"it's not a negotiation at all, it's all or nothing."* Worse, some participants feel that permission is coerced from them: P10 observed that Niantic *"pressure you into"* giving continuous access to your location data by tying it to the availability of in-game benefits such that *"you don't want to lose out"*.

Such lack of choice or coercion led to feelings among participants of resignation about data collection, seeing it as a Hobson's choice:

"I feel like it's inevitable that if you want to access their services at all, in any normal kind of way, that you automatically have to give them your data."—P7

Ultimately, participants felt that their data was *"revealing"* (P2, P3, P11) a lot of information about them, and so their only real option to maintain their privacy was to prevent data collection in the first place by not using that service at all (P1, P2, P3, P7, P10, P11), and living with the subsequent lack of service capability.

Perceptions of Data Holders

The discussion of attitudes towards data holders through this study allows some insights to be drawn as to how data holding service organisations are perceived, particularly with regard to data handling:

Factors such as reputation, size and business model were often a major contributor to participants' impressions of companies. For example, P2 described feeling *"more at ease"* with Apple, due to their hardware-oriented business model, than with Google, who *"make money through data"*; in general, where there was a lack of clarity around how a company makes money, or that model was clearly based exploiting sacrificed personal data, there was a greater suspicion, while trust was higher in those companies that offered a paid service:

"One of the main things was there [are] no ads. [Natural Cycles is] a paid service, so there's no, like, 'you don't have to pay but we use your data to make money'."—P8

Returning to Apple, P2 also noted that they *"position themselves as a defender of privacy rights"* and along with P11 (another participant who had targeted Apple) held a more favourable view as a result. P10, on the other hand, had been influenced by a documentary she had seen, becoming suspicious of Apple's control over her hardware.

While attitudes to Apple were generally positive, Facebook—which has, and continues to be, the subject of much negative media attention over its apparently cavalier attitudes towards personal data—was held in much lower regard. P6 said

Facebook had *"in every shape or form, shown themselves not to be trusted"*, an opinion formed from *"high profile news stories where they have done unscrupulous things and are very willing to just hand over data"*. P9 reported feeling *"slightly dubious"* about Amazon as a result of *"[press coverage] about their ethics that may or may not be true, and just the size of them... and just the level[amount] of data, as well"*. Clearly expectations around data handling are a strong influence on attitudes toward service providers, though sometimes other factors play a role, such as with P8, who was comforted not just by Natural Cycles' payment model, but the values they project: *"This is woman-empowerment-orientated [sic], so in that sense I think I do put my trust there as well."*

As well as these more impression-based influences, it also became clear that participant's direct experience of interacting with a company affects their feelings toward that provider. P1 found that *"in the same way that Amazon is quite janky [unreliable and awkward to use], Google feels fairly polished and so I trust them more."*. As well as customer/user experience, a perception of receiving a valuable service creates trust: P4 said of Google that *"the amount I trust them is in line with the utility I get from them"*. In the context of data sacrifice, high levels of trust do have an effect on customer behaviour:

"When I like the company already, I'm more willing to give them my data"—P2

Changed Perspectives Through Scrutiny and Transparency

Longitudinal examination of participant's feelings of perceived individual trust and power across their GDPR experience allows the impact of the experience to be analysed. As illustrated in Figure 5.5 and 5.6, trust in data holders tended to diminish through the data request process. Some of this distrust arose from examination of privacy policies, for example in P5's case who commented that Spotify *"shouldn't need to know that much about me, they should just give me music"*. In most cases the most noticeable decline in trust occurred between Interview 2 and 3 (when the participant experienced the GDPR process) or within Interview 3 (where the returned data was examined), showing that both the quality and coverage of the data return, as well as the execution of the data request process, often have a detrimental effect on trust. Individuals' perceived power, however, did not undergo a corresponding change:

"They've not given me everything back that I thought they'd be collecting, which makes me trust them less. So power-wise, I don't think [anything]'s changed, but trust, I think it has."—P1

The absence or sparsity of derived and acquired data and of metadata around sharing and handling [5.3] noticeably damaged trust. P1 directly attributed his reduced trust scores to what he saw as non-compliance [5.4.2] through failure to return all data categories. P5 lowered her evaluation of trust in Spotify further still upon completion of Interview 3 *"because they didn't say anything about what they're doing with my data or where it's going"*. P8's trust score Natural Cycles was similarly reduced *"because I think it's hard to get any sensitive data, and it's not really made clear what they're using it for"*.

It is also clear that poor GDPR handling in itself can damage trust, independently of the data returned. P2 reduced his trust score for AirBNB *"because of the way they've handled [the data request], and the way they've made it hard for me to read the data"*. Similarly, P7 downgraded her score for LinkedIn *"because I feel like they have my data and [they've] not bothered to find my data, and that makes me feel like I shouldn't trust them quite as much"*.

Participants want greater transparency than the current processes are currently providing, and the failure to do so is a direct cause of distrust:

"If someone's not completely open with you, then you're like, well 'What are you hiding?', which means you trust them less."—P4

Despite the impacts on trust, both using GDPR access rights and the wider process of scrutiny and discussion surrounding that process within this study had a positive impact on participants' awareness, offering *"insights into how big companies are actually handling these requests"* (P7) and how to practically use one's data rights, showing that transparency (even in the hampered form of current GDPR handling) has an education benefit. Participants had initially expressed wishes to gain insight into data handling practices in order to increase accountability and inform their decision making on provider loyalty and privacy settings. GDPR offered the potential to compare data expectations with reality—for example P11 was initially *"curious to find out if [Apple's] marketing claims match their reality around privacy"*. While such broad goals were generally unmet, several participants found the process thought-provoking and reported feeling more aware about what data they were enabling their providers to gather. P4 felt the process *"got me thinking about, like what other things could I try, and what other sources of personal data are there"*. P8 reflected that *"it's a skill and a kind of knowledge that I think everyone should [have]. I don't think it [should be] normal that I felt so clueless"*. Some commented on the value of understanding GDPR itself through the experience:

"[I] think the exercise was useful in that I understand what a GDPR request can do and what it cannot do. And there's a lot it cannot do. And I think it might seem that it gives you a lot of power, but really, it doesn't."—P2

While considering the negative impacts of the GDPR experience on trust some realised the potential trust-engendering impact that a more transparent response could have brought:

"I think the lack of transparency in a lot of these processes has not helped, you know, if Tesco had [. . .] plain English processes for getting the data and you've got the data in a plain English way, that would do a lot to bolster trust."—P11

In a small number of cases, this was witness in practice, with a good GDPR response actually increasing participants' trust in certain providers. For example, P5 reflected that she may have been *"a little harsh"* in her initial judgement of Instagram and said she *"actually really liked what they sent. . . in comparison to the three others, I was genuinely, I opened Instagram's one and I was like*

'this is really cool.'". P10 was very impressed with the response from Niantic and after GDPR she trusted them very highly *"because they replied really fast, the data provided is very detailed, and their attitude towards this whole issue is very positive,"* concluding that they are *"a really nice company"* and even indicating an increased willingness to spend money on their product. P6 trusted Sunderland AFC because *"they were really kind of upfront and ... I got the data from them first, [...] no messing about, the format they gave me just made sense"*.

In these comments, we can see an indication that, although the data requests often did not live up to the hopes of the participants, positively engaging with the process was influential and did affect participants' outlook. In particular, close attention was paid to the willingness of companies to be transparent and forthcoming, with GDPR representing an opportunity to test organisations on their data practices and assess their integrity and competence as holders of their data.

1.5 Discussion

This study examined the GDPR's effectiveness in improving individuals' access and control over their personal data. The participants' experiences support the existence of a power imbalance over data [2.1] and suggest GDPR largely fails to empower individuals: both objectively (to the extent possible by this limited sample), in that most companies do not comply fully (either by returning insufficient and inadequate data, or by failing to return data on time or at all), and subjectively, in that returned data was often difficult to understand, impractical for use, and raised new questions and concerns. The findings also indicate that swift, transparent, and easy-to-use GDPR procedures can positively impact an individual's perception of an organisation. In light of these findings, this discussion offers insights on how the personal data landscape might be redesigned through policy [5.5.1] and business practice [5.5.2], and how individual action can have important impact too [5.5.3]—all in pursuit of the human-centric empowerment goals described in 5.1 as well as 2.2 and 2.3):

1.5.1 Implications for Policymakers: Compliance, Quality and Ongoing Access

Despite significant and obvious GDPR-motivated investment by service providers in dashboards, processes and bespoke data package production, the findings (while limited by the small number of participants) indicate that inadequate compliance with the GDPR is common. The findings are consistent with literature too: the participants' issues with completeness and compliance echo those first reported within the GDPR's first year (Ausloos and Dewitte, 2018), suggesting completeness and compliance have not improved over this period. However in this study, the focus was on the effectiveness and experience of engaging with GDPR procedures from the individual's perspective. Participants' experiences were overwhelmingly of disappointment and frustration, with their hopes rarely

met. They found that data holders often did not engage meaningfully with the process, and that the responses typically excluded or obscured data that could have provided them with the insights into their data privacy and the organisation's data practices that they sought. Evaluations of perceived power compared to data holders largely remained the same or worsened after accessing data through GDPR, and participants were not confident in the capabilities of the legislation to shift the balance of power. The process was perceived by some as a *"box-ticking exercise"* that was both frustrating and time-consuming and did not ultimately help them. Even though in 7% of cases participants did feel empowered by the GDPR, *all* participants receiving data were in practice left with the prospect of additional time-consuming and sometimes technically-skilled work to take advantage of or interpret their returned data. This suggests that to improve the situation, policymakers need to make changes towards:

1) Better Compliance Through Enforcement of Complaints. At present, enforcement of the GDPR is uneven; each country has its own DPA (for example in the UK, this is the Information Commissioner's Office or ICO) and complaints are rarely pursued for individual cases. Instead, cases are processed by specific DPAs in a form similar to a class action lawsuit. This means that individuals have little impact when they do raise a complaint, and many GDPR complaints "become lost or resulted in lengthy delay" (Burgess, 2021), or may even be erroneously dropped (Lomas, 2020). Until individuals have a clear and effective means to issue complaints (Baker, 2018) that result in enforcement action (or a clear threat of it), it is likely that individuals will continue to have little recourse other than to repeat the request and hope similarly dissatisfied individuals will act on their behalf. Data holders must be held to account when they do not deliver the full set of data that they report possessing, or when they fail to do so within the legally obligated time frame.

2) Policies to Enforce Better Quality Responses. Many participants received data in frustrating formats, including screenshots, printouts or files that were too technical or littered with acronyms. Data was provided in formats too technical to understand, or not technical enough to be useable [5.4.3], showing a demand for both human-readable information summaries and machine-readable data files, where most providers typically provide only one or the other. Policymakers could provide suggested data formats or even propose new standards; this would help data portability, improve effectiveness (Gurstein, 2011) and legibility (Mortier *et al.*, 2014), can reduce costs through common tooling and catalyse the building of tools to interpret and understand data. Such standards are emerging (Morgan, 2020) as they are a technological necessity for data unification, but lack adoption. We note that the European Data Protection Board has published new guidelines [TODO ADD REFERENCE] that could help somewhat to improve GDPR responses, though these do not offer new standards, and will not be as effective as legally-mandated policy changes.

3) Policies to Enforce Data Access as Ongoing Support, not One-Time Delivery. A radical redesign of policy is needed to give people the practical

outcomes they desire and, according to the GDPR itself, deserve. Data access needs to be seen as more than the one-off delivery of data files. People need understanding of their data and of its handling. Not only that, they need a timely, up-to-date and ongoing view of the changing picture of how they are seen in data, and the occasional 'snapshots' with a 30-day delay that GDPR delivers can never deliver this. Giving people an ongoing awareness and understanding of their data is the measure by which compliance should be assessed. The explanations GDPR mandates are not forthcoming; of the 119 hopes expressed by participants (see Table 5), 70 (59%) related to acquiring greater understanding of data practices. 38 (54%) of these were unmet, and a further 15 (21%) were only partially met. By mandating data holders to support individuals with not just the delivery of data, but assistance to understand (and potentially make use of) that data, policies could become more impactful, not least because such understanding is critical to inform judgements around consent, loyalty and compliance.

1.5.2 Implications for Data Holders: Earn Trust by Opening Up Data and Enabling Users

While this study, and the GDPR itself, might seem adversarial to data holders given the goal to reduce their power by imposing new procedures, the findings emphasise the role of personal data in consumer relations. Data holders are likely aware of the paramount role of personal data in decision making, but may not be aware of individuals' perceptions about this. The findings suggest that failure to satisfy users who are concerned about the collection and usage of their personal data risks harms to consumer trust and confidence, at least for those users, and perhaps for others they might influence. In turn, however, this presents opportunities to use the mechanisms of the GDPR for customer loyalty and building better relations.

In 52% of cases, following the process of examining privacy policies and engaging in GDPR data requests resulted in a decrease in reported trust in the data holder. While such impacts may for now be minimal, as only a small proportion of users read privacy policies (Steinfeld, 2016) and—one can assume—an even smaller number conduct GDPR requests, this is likely to change as issues around data privacy and trust continue to take centre stage in global geopolitics (Véliz, 2020; Zuckerman, 2021). Furthermore, the emergence businesses focused on *'getting your data'* or *'taking control'* ('Whose data is it anyway?', 2019; Dehaye, 2021; CitizenMe, 2021; Gener8, 2021; 'About Us', no date; 'datacy - About Us', no date; 'Ethi', no date; 'Digi.me', no date; 'Exist.io', no date) suggests demand for data access is growing. From the findings, there are three positive takeaways for data holders:

1) Data transparency is an opportunity to increase customer loyalty and trust. GDPR's basic rights provide a starting point for delivering practical data transparency that will allow organisations to demonstrate that they are deserving of trust. By responding clearly and engaging openly and helpfully with GDPR data requests, organisations can demonstrate consistency between

their privacy policy and their actions and demystify to their users the role that data holds in their business model. Research has shown that explanations can *"ease humans' interactions with technology [. . .], help individuals understand a system's function, justify system results, and increase their trust"* (Glavic *et al.*, 2021). This was borne out in our results: in 14% of cases, participants felt more trusting of the service brand as a result of their GDPR experience (sometimes even displacing prior apprehensiveness or distrust), citing reasons such as speedy, hassle-free responses, clear and understandable data, providers being upfront and open with data, and staff who exhibited a positive attitude to the request.

2) Data transparency is an opportunity for improved and re-imagined customer relations around data. Beyond the opportunity to improve trust, the mechanisms of data transparency suggested by the GDPR provide individuals with new capabilities for data curation and involvement. By offering individuals the ability to engage in empowering data interactions, data holders have the opportunity to improve engagement with their organisation and their services. If organisations view personal data as a shared resource to be curated and co-owned by the individuals that contributed it, there may be correspondingly shared benefits: for the individual, a sense of agency, influence and negotiability (Mortier *et al.*, 2014); and for the service provider, an incentive for individuals to generate and share more data, an increased likelihood of individuals correcting inaccurate data, and more reliable and human-centric forms of ongoing consent closer to dynamic consent (Kaye *et al.*, 2015) than today's ineffective models of informed consent (Luger and Rodden, 2013).

3) New customer demands indicate untapped business opportunities. As the 500-member-strong MyData Global organization (MyData.org, 2018) shows, there is growing demand for personal data empowerment. People's personal data is splintered and trapped (Abiteboul, André and Kaplan, 2015; Bowyer, 2018), and they cannot correlate data from different sources in order to reflect upon it, gain insights, and set goals (Li, Forlizzi and Dey, 2010). Due to commercial motivations, service providers generally deliver capabilities within a closed silo, not at the level of one's wider environment (Abowd, 2012). To be better empowered the individual could be the point of integration, the centre of their own Personal Data Ecosystem (PDE) (MyData, 2017). Life-level capabilities (Bowyer, 2021) and the opportunities that well-designed and well-regulated GDPR-type regulations promise in this regard have not yet been exploited. Thorough, complete and timely data access in standard formats, as mentioned above, will be critical to enabling this vision. Growing companies such as CitizenMe (CitizenMe, 2021), Digi.Me ('Digi.me', no date), Mydex (Mydex CIC, 2010), ethi ('Ethi', no date), HestiaLabs (Dehaye, 2021), udaptor (Udaptor, 2021) and exist.io ('Exist.io', no date) as well as larger organisations like BBC R&D (BBC R&D, 2017) and Microsoft (Microsoft, 2021) are already starting to innovate in this space.

1.5.3 Implications for Individuals: Becoming Aware of the Value and Power of Data, and Demanding More

While participants experienced disappointment and frustration in their GDPR journeys, all participants gained new understandings; if not always of their data itself, at least of their target companies' approach to data access requests. This new knowledge was sufficient to re-affirm or challenge existing attitudes or inform judgements—P1, for example, left Facebook after the study. Even an attempt to access data can be educational, and even a cursory look at a provider's 'What data do we collect' privacy policy section can provide pause for thought.

Today, individuals remain largely in the dark about the collection, use and sharing of their data through a combination of perceived complexity and effort combined with a lack of clear benefits. Table 12, alongside the increased control and insight promised by the PDE movement and platforms linked in 5.5.1 and 5.5.2 above, provide a glimpse of what the future may hold: a world where individuals take more control of their data and gain actionable self-insights. Three key messages for individuals can be inferred:

1) Your data is used to represent you and define your user experience. We hand over our data in exchange for access to services, but providers then use it (usually in aggregate) e.g. to inform product design or decide what content you see. This 'innocent' handover of data is in fact giving providers the means by which we are treated and – at times – controlled. Recognizing the crucial role of data (and our limited influence over it) is the first step to pursuing greater agency and control.

2) Your data contains meaningful and valuable data about your life. Data, as participants found, is dry and technical, but they all sought meaning and value within it [5.2.2]. Within provider-held data lies potentially rich information about one's life and past activity – some of which can even be inaccessible through any other means. This highlights both a risk (that others might gain this insight) and a potential benefit (that we could access this insight ourselves). In this context, data deletion without keeping a copy may be inadvisable. To access the value in data, individuals will need to demand data standards, better access and control mechanisms and insight tools.

3) Self-education and awareness enable accountability and informed choices. The findings highlight a lack of knowledge. Transparency is critical to judging 'to what extent the bargain is fair' (Larsson, 2018). It is not always delivered, but GDPR makes it your right; a right that cannot be fully refused. Through challenging poor GDPR responses and demanding better information, individuals can have impact. Providers are ultimately motivated by public demand—one of the reasons download dashboards exist. Through the public pressure of negative attention, companies can be motivated to improve data access (Dehaye, 2018). With patience, GDPR rights can be exploited to force small changes.

1.6 Summation of Case Study Two

Through a longitudinal study of 10 participants lasting three months, this case study has qualitatively, and to a lesser extent quantitatively, evaluated the human experience of using one's GDPR access rights and of living with data-centric service provider relationships.

The findings, while not statistically representative, suggest that people currently lack awareness of held data and its uses by service providers. By guiding participants on a journey of discovery and careful scrutiny, encouraging them to draw their own conclusions about service providers on the basis of companies' own promises, individuals' legal rights, and participants' own hopes (see Table 12), this research has shown that such a journey can be educational and enlightening with regard to increasing awareness, but also can seriously damage brand loyalty and trust in providers if comprehensive and well-explained data is not returned in a supportive and open manner [5.4.4].

The experience of GDPR seems to be an unsatisfactory one for individuals; participants were generally still 'in the dark'. Serious problems with compliance have been highlighted [5.4.2]: Participants received data that was incomplete, impractical for use, and they failed to acquire desired explanations. By its own aim to enhance individuals' rights and control, the GDPR does not succeed. Participants continued to feel a lack of agency and choice, were largely unable to pursue goals such as data checking, correction or deletion, and their perceived sense of power within the provider relationship was largely unchanged by the experience. Nor does the GDPR allow individuals to adequately pursue their own goals related to accountability, self-reflection or creative data exploration [5.4.3]. Individuals cannot be given power over their data through designing better Human-Data Interaction interfaces alone, but only through redesigned policies and business strategies that consider the sociotechnical context (Baxter and Sommerville, 2011; Bowyer, 2021).

In order to bring the human-centric 'personal data ecosystem' concept closer to reality, action must be taken to improve both compliance and quality of GDPR responses. Considering these findings, there is cause for radical policy reform, to move away from 'data access as package delivery' and to provide individuals a more effective and ongoing two-way window into their data [5.5.1], providing ongoing awareness, accountability, and negotiability. Data needs to be expressed to individuals in ways they can understand, as little to no practical impact is currently being achieved by delivery of a one-time snapshot of some technical files; in fact, we have shown such responses can be harmful to customers' perceptions of the data holder in many cases.

For providers, the risk of reputational damage uncovered by this study should motivate them to engage meaningfully with data access requests; but such risk can be averted by redesigning both interfaces and processes to approach data access experiences as an opportunity to educate, and to build trust and loyalty, perhaps even through establishing progressive co-operative data stewardship relationships

that truly *involve* the service user [5.5.2]. While the GDPR experience is often disappointing and frustrating, it can provide insights that help individuals to challenge their assumptions, re-evaluate choices, and in some rare cases, feel empowered to act upon their data. Wider assertion of GDPR rights could demonstrate a desire for data holders to be transparent; without such visible demand, little may change [5.5.3].

Considering RQ1 (the pursuit of a deeper understanding of people's attitudes to everyday data holding and people's wants from that data), this work suggests that people struggle to develop the meaningful relationship with their data that they desire because of the difficulties faced in seeing, accessing and understanding it. They are aware that within data is the potential for value to themselves, but cannot access that value, which in turn causes feelings of resignation, concern, distrust or suspicion towards data holders. What they seek most are two things: sufficient understanding to better judge the value exchange they have signed up for with providers (see goals in top half of Table 12), and good quality insights from data that would allow them to understand themselves better, learn from the past, set personal goals, and harness personal data for their individual benefit (see goals in lower half of Table 12). This duality of needs around data interaction is expanded upon in (Bowyer, 2021).

With respect to RQ2 (the pursuit of a better understanding of the role of that data in everyday service relationships), the findings suggest that personal data, held by providers, as in Case Study One, serves as a proxy for direct user involvement, and is treated as such. Once users have sacrificed their data, or given permission for its collection, they are rarely consulted and most services exclude them from seeing how that data will travel through the organisation and be used in decision making; this is consistent with the *'point of severance'* concept observed by Luger and Rodden (Luger and Rodden, 2013). As a result, the trust relationship between service provider and service user is extremely fragile, highly susceptible to subjective impressions of service brands, and as the findings show, discovery of poor data practices or a lack of transparency around data is sufficient to harm that relationship and, in some cases, even motivate individuals to change provider. As discomfort grows and scrutiny occurs, providers can expect customers to lose trust and loyalty. At the same time, this same data could play a central role in a re-invigorated relationship between a provider and a user, one based upon *earned* trust. It appears that providing easy, clear, data access and showing a willingness to respond to questions and explain data usage to users could be sufficient to allay concerns and instil strong customer loyalty. Of course, this assumes that the openness offered reveals practices the user finds agreeable, so perhaps this in some way explains why some companies that have more commercially-motivated approaches to personal data use (such as Facebook and Google) that many would find disagreeable upon examination, are apparently less willing to engage in transparency and user empowerment around data.

The general principles of earning trust through transparency, and rethinking

data access as a means to involve users in decision making, could be applied in a wide range of service endeavours that are currently very data-centric.

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.

'About Us' (no date). Rita. Available at: https://ritapersonaldata.com/about.html (Accessed: 10 August 2022).

Abowd, G. D. (2012) 'What next, ubicomp?: celebrating an intellectual disappearing act', in *Proceedings of the 2012 ACM conference on ubiquitous computing.* New York, New York, USA: ACM Press, pp. 31–40. doi: http://dx.doi.org/10.1145/2370216.2370222.

Abowd, G. D. and Mynatt, E. D. (2000) *Charting Past, Present, and Future Research in Ubiquitous Computing.* 1, pp. 29–58. Available at: https://www.cc.gatech.edu/fce/pubs/abowd-mynatt-tochi-millenium.pdf.

Alizadeh, F. *et al.* (2019) 'GDPR-reality check on the right to access data', in *ACM international conference proceeding series.* New York, New York, USA: ACM Press, pp. 811–814. doi: 10.1145/3340764.3344913.

Arfelt, E., Basin, D. and Debois, S. (2019) 'Monitoring the GDPR', in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, pp. 681–699. doi: 10.1007/978-3-030-29959-0_33.

Ausloos, J. (2019) 'GDPR Transparency as a Research Method', *SSRN Electronic Journal*, (May), pp. 1–23. doi: 10.2139/ssrn.3465680.

Ausloos, J. and Dewitte, P. (2018) *Shattering one-way mirrors-data subject access rights in practice.* Available at: www.irissproject.eu https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3106632.

Ausloos, J. and Veale, M. (2020) 'Researching with Data Rights', *Technology and Regulation*, pp. 136–157.

Baker, J. (2018) 'What's a GDPR complaint? No one really knows'. Available at: https://iapp.org/news/a/whats-the-definition-of-a-gdpr-complaint-spoiler-alert-no-one-knows/.

Baxter, G. and Sommerville, I. (2011) 'Socio-technical systems: From design methods to systems engineering', *Interacting with Computers.* OUP, 23(1), pp. 4–17. doi: 10.1016/j.intcom.2010.07.003.

BBC R&D (2017) 'Human Data Interaction - BBC R&D'. Available at: https://www.bbc.co.uk/rd/projects/human-data-interaction.

Bowyer, A. (2018) 'Free Data Interfaces: Taking Human- Data Interaction to the Next Level', *CHI Workshops 2018.* Available at: https://eprints.ncl.ac.uk/273825.

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.
Bowyer, A. (2021) 'Human-Data Interaction has two purposes: Personal Data Control and Life Information Exploration'. Available at: https://eprints.ncl.ac.uk/273832#.
Bowyer, A. *et al.* (2022) 'Human-GDPR interaction : Practical experiences of accessing personal data', *CHI '22*.
Bufalieri, L. *et al.* (2020) 'GDPR: When the right to access personal data becomes a threat'. doi: 10.1109/icws49710.2020.00017.
Burgess, M. (2021) 'Why Amazon's £636m GDPR fine really matters', *Wired*. Available at: https://www.wired.co.uk/article/amazon-gdpr-fine.
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.
CitizenMe (2021) 'Become a Citizen and unlock the value of your data'. Available at: https://www.citizenme.com/for-citizens/ (Accessed: 23 August 2021).
Clarke, N. *et al.* (2019) 'GDPR: an impediment to research?', *Irish Journal of Medical Science (1971-)*. Springer, 188(4), pp. 1129–1135.
Comandè, G. and Schneider, G. (2021) 'Can the GDPR make data flow for research easier? Yes it can, by differentiating! A careful reading of the GDPR shows how EU data protection law leaves open some significant flexibilities for data protection-sound research activities', *Computer Law & Security Review*. Elsevier, 41, p. 105539.
Cormack, A. (2021) 'Thinking with GDPR: A guide to better system design', *Information Services & Use*, 41(1-2), pp. 61–69. doi: 10.3233/isu-210107.
Council of the European Union (2015) 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)'. Brussels. Available at: http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.
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.
'datacy - About Us' (no date). Available at: https://www.datacy.com/personal/about-us (Accessed: 22 March 2019).
Dehaye, P.-O. (2018) 'Post-hearing questions by Senator Blumenthal to Mark Zuckerberg'. Available at: https://wiki.personaldata.io/wiki/Item:Q1800.
Dehaye, P.-O. (2021) 'HestiaLabs'. Geneva, Switzerland. Available at: https://hestialabs.org/en/ (Accessed: 23 August 2021).
'Digi.me' (no date). Available at: https://digi.me/ (Accessed: 23 August 2021).
'Ethi' (no date). Available at: https://www.ethi.me/.
'Exist.io' (no date). Available at: https://exist.io/ (Accessed: 23 August 2021).
'Facebook - Data Policy' (no date). Available at: https://www.facebook.com/about/privacy (Accessed: 9 August 2021).
'Facebook–Cambridge Analytica Data Scandal' (2014). Available at: https://en.wikipedia.org/wiki/Facebook\T1\textendashCambridge_Analytica_

data_scandal.
Gellman, B. (2013) 'Edward Snowden, after months of NSA revelations, says his mission's accomplished', *The Washington Post*, 23. Available at: http://www.washingtonpost.com/world/national-security/edward-snowden-after-months-of-nsa-revelations-says-his-missions-accomplished/2013/12/23/49fc36de-6c1c-11e3-a523-fe73f0ff6b8d_story.html%5Cnhttp://www.washingtonpost.com/world/national-security/edward-.
Gener8 (2021) 'Gener8'. Available at: https://gener8ads.com/ (Accessed: 23 August 2021).
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/XXXXXXXXX.Boris.
Gonscherowski, S. and Bieker, F. (2018) 'Who You Gonna Call When There's Something Wrong in Your Processing? Risk Assessment and Data Breach Notifications in Practice', in *IFIP international summer school on privacy and identity management*. Springer, pp. 35–50.
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.
Hamon, R. *et al.* (2021) 'Impossible Explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario', in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 549–559.
'HDI Lab, Heerlen' (2020). Available at: https://hdilab.com/.
'HDI Network Plus, University of Glasgow' (2018). Available at: https://hdi-network.org/.
Huberman, M. and Miles, M. B. (2002) *The qualitative researcher's companion*. Sage.
'Human Data Interaction Project at the Data to AI Lab, MIT' (2015). Available at: https://hdi-dai.lids.mit.edu/.
Human, S. and Cech, F. (2021) 'A human-centric perspective on digital consenting: The case of GAFAM', *Smart Innovation, Systems and Technologies*, 189, pp. 139–159. doi: 10.1007/978-981-15-5784-2_12.
Hwang, E. (2021) 'Sketching Dialogue : Incorporating Sketching in Emphatic Semi-structured Interviews for HCI'.
Information Commissioner's Office (2018) 'Your data matters - Your rights'. Available at: https://ico.org.uk/your-data-matters/.
Information Commissioner's Office (2021a) 'Your right of access'. Available at: https://ico.org.uk/your-data-matters/your-right-to-get-copies-of-your-data/ (Accessed: 23 August 2021).
Information Commissioner's Office (2021b) 'Your right to data portability'.
Kasirzadeh, A. and Clifford, D. (2021) *Fairness and Data Protection Impact Assessments*. Association for Computing Machinery (1), pp. 146–153. doi: 10.1145/3461702.3462528.
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.

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.
Li, I., Forlizzi, J. and Dey, A. (2010) 'Know thyself: Monitoring and reflecting on facets of one's life', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 4489–4492. doi: 10.1145/1753846.1754181.
'List of target companies for GDPR requests' (no date). Available at: https://wiki.personaldata.io/wiki/Item:Q2369 (Accessed: 22 September 2021).
Lomas, N. (2020) 'UK's ICO faces legal action after closing adtech complaint with nothing to show for it'. Available at: https://techcrunch.com/2020/11/05/uks-ico-faces-legal-action-after-closing-adtech-complaint-with-nothing-to-show-for-it/.
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.
McCarthy, J. and Wright, P. (2004) 'Technology as experience', *Interactions*, 11(5), pp. 42–43. doi: 10.1145/1015530.1015549.
Microsoft (2021) 'Project Bali'. Available at: https://www.microsoft.com/en-us/research/project/bali/ (Accessed: 23 August 2021).
Morgan, J. (2020) 'Making your Solid Apps interoperable with ShapeRepo.com'. Available at: https://medium.com/@JacksonMorgan/making-your-solid-apps-interoperable-with-shaperepo-com-8da512936073.
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.
MyData (2017) 'Declaration - MyData.org'. Available at: https://mydata.org/declaration/ (Accessed: 8 November 2019).
MyData.org (2018) 'MyData - Who we are'. Available at: https://mydata.org/about/.
Mydex CIC (2010) 'The Case for Personal Information Empowerment : The rise of the personal data store', *World*, pp. 1–44.
O'Donnell, B. (2020) 'Zoom, the office and the future: What will work look like after coronavirus?' Available at: https://eu.usatoday.com/story/tech/columnist/2020/09/07/zoom-work-from-home-future-office-after-coronavirus/5680284002/.
'Privacy' (no date). Available at: https://privacy.linkedin.com/ (Accessed: 9 August 2021).
'Privacy - Apple (UK)' (no date). Available at: https://www.apple.com/uk/privacy/ (Accessed: 9 August 2021).
'Privacy & Terms – Google' (no date). Available at: https://policies.google.com/ (Accessed: 9 August 2021).
Quinn, P. (2021) 'Research under the GDPR–a level playing field for public and private sector research?', *Life Sciences, Society and Policy*. Springer, 17(1), pp. 1–33.
Rughiniș, R. *et al.* (2021) 'From social netizens to data citizens: Variations of GDPR awareness in 28 european countries', *Computer Law & Security Review*.

Elsevier, 42, p. 105585.
Savage, A. and Hyde, R. (2014) 'Using freedom of information requests to facilitate research', *International Journal of Social Research Methodology*. Routledge, 17(3), pp. 303–317. doi: 10.1080/13645579.2012.742280.
Spagnuelo, D., Ferreira, A. and Lenzini, G. (2019) 'Accomplishing Transparency within the General Data Protection Regulation.', in *ICISSP*, pp. 114–125.
Steinfeld, N. (2016) '"I agree to the terms and conditions":(how) do users read privacy policies online? An eye-tracking experiment', *Computers in human behavior*. Elsevier, 55, pp. 992–1000.
Symons, T. *et al.* (2017) 'Me, my data and I: The future of the personal data economy', *DECODE (DEcentralised Citizen Owned Data Ecosystems) Report*, (732546), p. 88. Available at: https://media.nesta.org.uk/documents/decode-02.pdf%0Ahttps://decodeproject.eu/publications/me-my-data-and-ithe-future-personal-data-economy%0Ahttps://media.nesta.org.uk/documents/decode-02.pdf%0Ahttps://decodeproject.eu/publications/me-my-data-and-ithe-fu.
The European Parliament and the Council of the European Union (2016) 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data'. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=ES.
Udaptor (2021) 'Udaptor Assistant'. Available at: https://udaptor.io/assistant.html (Accessed: 23 August 2021).
Varghes, S. (2019) 'Pokémon Go was a warning about the rise of surveillance capitalism'. Available at: https://www.wired.co.uk/article/the-age-of-surveillance-capitalism-facebook-shoshana-zuboff.
Véliz, C. (2020) *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. Transworld Publishers Limited, p. 224. Available at: https://b-ok.lat/book/11000161/b53144.
Waldman, A. E. (2020) 'Data Protection by Design ? A Critique of Article 25 of the GDPR', 1239(2019), pp. 147–168.
Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, 265(3), pp. 94–105. doi: 10.1145/329124.329126.
'Whose data is it anyway?' (2019). 04: UBDI. Available at: https://www.ubdi.com/blog/whose-data-is-it-anyway (Accessed: 31 March 2021).
Wiki.personaldata.io (no date) 'Subject Access Request Template'. Available at: https://wiki.personaldata.io/wiki/Template:Access (Accessed: 21 September 2021).
Wong, J. and Henderson, T. (2018) 'How Portable is Portable ? Exercising the GDPR ' s Right to Data Portability', *Acm*, pp. 911–920.
Woolgar, S. (2014) 'Configuring the User: The Case of Usability Trials', *The Sociological Review*, 38(1_suppl), pp. 58–99. doi: 10.1111/j.1467-954x.1990.tb03349.x.
Wright, P. and McCarthy, J. (2008) 'Empathy and experience in HCI', *Conference on Human Factors in Computing Systems - Proceedings*, pp. 637–646. doi: 10.1145/1357054.1357156.

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.