Generating Practices: Investigations into the Double Embedding of GDPR and Data Access Policies

JUSTIN PETELKA, University of Washington, USA
ELISA OREGLIA, King's College London, UK
MEGAN FINN, University of Washington, USA
JANAKI SRINIVASAN, International Institute of Information Technology, Bangalore, India

The right of access, found in the EU's GDPR and similar data protection regulations around the world, requires corporations and other organizations to give people access to the data they hold about them. Such regulations create obligations for data controllers but leave flexibility on how to achieve them, resulting in variation in how Data Subject Access Requests (DSARs) are implemented by different corporations. To understand the various practices emerging around DSARs and how requesting data influences the way people think of data protection laws, we asked participants in India, the UK and USA to make 38 DSARs from 11 different companies. Using the metaphor of the policy-design-practice "knot" [33], we examine DSARs as a case of the co-constitutive links between policy, design, and practice. We find that the DSAR process was not linear and participants employed many work-arounds. The challenges they encountered in the overall DSAR process negatively affected their perceptions of data protection policies. Our study suggests that researchers have to be flexible in adapting research methodology to understanding emerging practices, and that there is a need for more collaborative experimentation with DSARs before standardizing the process.

CCS Concepts: • **Social and professional topics** → **Governmental regulations**; • **Human-centered computing** → *Empirical studies in collaborative and social computing*; • **Applied computing** → **Law**.

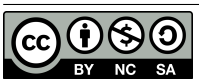
Additional Key Words and Phrases: data subject rights, right of access, GDPR, CCPA, data protection law

ACM Reference Format:
Justin Petelka, Elisa Oreglia, Megan Finn, and Janaki Srinivasan. 2022. Generating Practices: Investigations into the Double Embedding of GDPR and Data Access Policies. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 518 (November 2022), 26 pages. https://doi.org/10.1145/3555631

1 INTRODUCTION

On January 6 2022, the French Commission Nationale de l'Informatique et des Libertés (CNIL), an independent authority that oversees data protection in France, fined Google and Facebook for "non compliance with French legislation" on cookies. CNIL found that "the websites facebook.com, google.fr and youtube.com offer a button allowing the user to immediately accept cookies. However, they do not provide an equivalent solution (button or other) enabling the Internet user to easily refuse the deposit of these cookies. Several clicks are required to refuse all cookies, against a single one to accept them." [16] CNIL found the companies in violation of Article 82 of the French Data Protection Act (FDPA). Notably, neither Article 82 nor any provision of the General Data Protection Regulation (GDPR) specifically state that the number of clicks required to opt in and out of cookies must be equivalent.

Authors' addresses: Justin Petelka, jpetelka@uw.edu, University of Washington, Seattle, USA; Elisa Oreglia, elisa.oreglia@kcl.ac.uk, King's College London, London, UK; Megan Finn, University of Washington, Seattle, USA, megfinn@uw.edu; Janaki Srinivasan, janaki.srinivasan@iiitb.ac.in, International Institute of Information Technology, Bangalore, India.



2573-0142/2022/11-ART518
https://doi.org/10.1145/3555631

Both the FDPA and GDPR set overall policy objectives for protecting personal data but both are intentionally vague about implementation, which allows different companies and technology providers to fulfill these goals in flexible ways. For example, Article 7 of the GDPR, related to Conditions for Consent, states that the request for consent should be clear, easily accessible, and as easy to withdraw as it is to give. In their interpretation of this article, CNIL fined Google and Facebook for what it determined to be a violation of this general principle. The formulation of consent expressed in Article 7 exemplifies both the EU and European states' recent policies regarding personal data protection: individuals must have control over their personal data, from understanding what is collected, the ability to object to such collection, to see what data are held about them and to port their data from one service to another.

Given the lack of specific guidance for implementation in GDPR policies, it is unsurprising that "data controllers," the GDPR's label for organizations that collect and/or process personal data, have interpreted GDPR in different ways and implemented principles like accessibility in different ways. In the coming years, EU citizens and others will likely see a number of adjustments in accepted practices as the implications of both data regulations and the measures to implement them stabilize [63].

The shift from policy and compliance to practice is important because it engages more directly with a fundamental judgement in the GDPR's data access provision: that it is beneficial for individual users of different data systems to know about and control their data. Additionally, the shift to practice necessitates that we engage across people, institutions (including policies), and technology – an approach honed by CSCW researchers [33]. We contribute to a long CSCW tradition of studying how sociotechnical systems are constructed through the collaboration of humans and technical interfaces by asking: How do different people experience the emerging right to data access? Through individual experiences, are these laws achieving their goal of enabling people to not just be, but also feel, in control of their data? How do people with and without the right to data access understand their personal data and the laws that are protecting them?

Our goal is to articulate what practices are emerging in this novel policy-technology environment enabled by Data Subject Access Requests (DSARs), and how these practices and their attendant experiences enhance our understanding, discussion, and analyses of the efficacy of data protection laws. We looked at these practices using the heuristic of the "knot" of design, practice and policy [33] to understand data protection laws as entangled with (instead of isolated from) policy and design. The policy-practice-design "knot" asks that we attend to how policy enables and relies upon particular designs and social practices, pointing out that policy is always embedded in instantiations of design and the practices that policy itself makes possible [33]. We particularly focus on how corporate organizations and software have adapted to comply with right to access policies as an instance of policy-influenced technical design adaptations.

We use these lenses to explore how the presence of right of access provisions and the absence of specific implementation requirements shape the practices of requesting, receiving, and interpreting data. To do so, we recruited participants in India, the United Kingdom (UK) and United States (US) to make 38 data requests from 11 different companies. In this paper, we present our participants' experiences requesting data, provide insights into how policy, design, and practice intertwine using the policy-knot [33], and recommendations that may have simplified or clarified the process for our participants. In the following sections, we will discuss related work and methods, followed by our findings organized around the experience of making a data access request and interpreting the results, and conclude with a discussion on the importance of practice in shaping the perception of usefulness of laws, the benefits of working directly with requestors, and finally, possibilities for how CSCW research(ers) might be brought to bear in helping data protection laws achieve their ambitious goals.

2 RELATED WORK

We examine the co-constitutive links between policy, design, and practice in the instrumentation and exercise of the right of access in relatively new policies. First, we review how CSCW researchers have conceived of the relevance of policy-related research. Next, we describe HCI and legal perspectives examining data protection rights, first considering studies where researchers made DSARs, and then studies that examine other aspects of data protection law, such as cookie banners, consent notices, and other data protection rights.

2.1 Policy and Design

A particular focus of CSCW are the interactions between policy, technology design, and social practice [34]. Building on previous work in infrastructure studies [7, 19, 68], Jackson, Gillespie, and Payette [33] offer the analytical heuristic of the "knot", signaling the intertwined, mutually constitutive, and dynamically bound relationships between design, practice, and policy. The policy-design-practice "knot" has been used to trace the contingencies between policy, design, and practice in a various, such as radio spectrum licensing frameworks [6], understandings of copyright law in online communities [22, 23], or the development of digital libraries [13]. In these applications, the knot metaphor is used to articulate the ways that configurations and tensions among the threads of policy, design, and practice can impact (and complicate) attempts to unravel the ways that each of these constituent parts affect the whole.

Importantly, Jackson, Gillespie, and Payette [33] show that policy is not outside of design and practice, despite of a tendency in CSCW research to treat policy as a standalone research area. They use the example of fair use balancing tests, which define exemptions to US Copyright law, to describe how policy enables and depends both on subjective and discretionary assessment (i.e., how regulators apply the fair use balancing tests in copyright infringement claims) and on past forms of design and practice (i.e., how platforms identify and moderate copyright infringement claims). Jackson, Gillespie, and Payette argue that this "double embedding" of policy: "is more than a simple error or limit, or reflection of the fact that law and policy have somehow failed to 'keep up' with the pace of social and technological change. Indeed, echoing a longer line of legal and policy scholars, we would argue that the kinds of flexibility built into things like reasonable expectation or fair use balancing tests are precisely what allows law and policy to grow and remain relevant over time, and our systems of order and governance to remain supple and 'live,' rather than fossilized remains of a different sociotechnical moment." (pg. 591)

In short, when policy is broadly written, it can become useful in separate or emerging sociotechnical contexts. We use this observation to understand the double embedding of data protection law and its evolution. As discussed in the introduction, design adaptations to policies (offering easy opt-in but not opt-out options) do not always translate into practices (multiple user clicks to opt out) that satisfy regulatory assessment (opting out must be as easy as opting-in).

The benefits of interpretive flexibility in policy builds on concepts from scholars of the social construction of technology who demonstrate that there is variation both in "how people think of or interpret artifacts" and "in how artifacts are designed" [56]. Eventually, Pinch and Bijker argue, this "interpretive flexibility" ends and there is "closure and stabilization" where the controversies about how to interpret artifacts are ended through rhetorical moves or by redefining the problem, after which the form of the artifact is stabilized (pg.44). The concept of "interpretive flexibility" can be enhanced by highlighting "who is doing the interpreting," particularly if they were anticipated by the designers [57]. Though the original formulation of interpretive flexibility under-theorized power differentials amongst users (and thus one way of explaining what uses and meanings of technologies win out when particular interpretations and designs are "closed") [36, 73], we find the

concept useful in studying DSARs as an artifact within a sociotechnical system of laws, technologies, and practices to understand the great variation in how different companies and users interpret DSARs in this early period of adoption. Our findings suggest that we have not reached "closure" on this system yet – an insight that we return to in the conclusion

We see our research as investigating a key moment to understand the flexibility of the right of access after their implementation as DSARs – how is the right of access being interpreted by different corporations as evidence of DSARs, and how will requestors react to these interpretations? From the perspective of the "knot", the right of access has shaped the design of corporate infrastructure. Together with our study participants, we investigated the world of practices in this configuration of design and policy. First, we answer the call to make policy the site of CSCW [34] and HCI research [65] using the right of access as a case study. We also build on that work by expanding on Jackson's ideas about the role of policy in CSCW research and show how the affordances of the law create possibilities for design and suggest ways in which we understand these interrelations. Our aim is to examine the "knot" by looking at new possibilities for practice that GDPR and related policies enable, and the varied ways in which these policies have been implemented in corporate infrastructures.

But before we get to this, we examine how other researchers have similarly answered the call to explicitly integrate policy into research. We focus especially on the work of researchers at the nexus of policy and privacy within HCI.

2.2 HCI, Policy, Privacy, and Data Protection

HCI privacy researchers have been active in evaluating and theorizing the interaction between privacy choices and preferences, privacy laws, and technical artifacts. We build on these existing privacy studies, but are mindful that from the perspective of law, "privacy" and "data protection" are related but often distinct goals. This is particularly true in the EU, where privacy and data protection rights stem from distinct sources of EU law [26, 35]. Gutwirth and DeHert [29] describe this distinction as "two sorts of legal tools, namely normative opacity tools that draw the limits of interferences with individuals [privacy laws], and transparency tools that organize the channeling control and restraint of power [data protection laws]". Jones and Kaminsky restate this distinction as "two different means for addressing privacy disparities" and note that privacy law also tends to employ vaguer standards than data protection law [35]. While debate over whether the right of access is a data protection or privacy law may seem pedantic, as we will discuss below this distinction impacts how both we and our participants assess the practical utility of the right of access.

A large body of HCI work has examined how people engage with privacy and data protection technologies driven by legal mandate. In 2008, McDonald and Cranor estimated that an average US internet user could spend as much as 200 hours per year reading privacy policies in the US [47]. Since then, several researchers have found that the amount and length of privacy policies have since grown in the EU and beyond [17, 39]. Bailey, Parsheera, Rahman, and Sane found that privacy policies in India are poorly drafted and serve more as check-the-box compliance than meaningful control or understanding of corporate data practices [4]. Researchers have found that contemporary data breach notifications mandated by US law are insufficient to spur many people to take proactive measures [46, 76]. Habib et al. conducted a laboratory study to understand the usability of legally-mandated opt-out processes [30]. They tasked participants with opting-out of newsletters, advertising, or requesting deletion of personal data, and found that users struggled to navigate through similar looking choices, labyrinthian menus, and uncertainty in how to format written opt-out requests. However, providing multiple paths to a privacy control made controls easier to find.

Many of these authors concluded that the problem is bad technical design (choices) and suggested that more specificity in policies on what is and is not compliant would help crystallize and popularize effective designs [21, 28, 30, 31, 51, 61, 72]. Several authors have worked to create taxonomies to help system designers think through system design choices, such as design spaces for privacy notices [61, 62] and privacy choices in Internet of Things (IoT) systems [21]. The work by Schaub and his collaborators articulated five design dimensions for privacy notices: timing (when the notice is sent to a user), channel (how the notice is sent), modality (the ways users interact with a notice), and control (how choices are provided in a notice). Feng et al. built upon this model to develop a taxonomy for the privacy choices present in a given privacy notice and used it to identify and refine the privacy affordances for an IoT management application [21].

In addition to design spaces, several authors have used the concept of "dark patterns" to categorize and identify designs that, intentionally or not, thwart or inhibit the ability of a user to exercise their data rights. Gray et al. apply an interaction criticism perspective to common website notices for collecting consent [28]. Their interdisciplinary team found that some "dark patterns" are both easy to identify and thus legally actionable, but others are more ambiguous and assessing their compliance requires multiple perspectives and methods. They conclude that what counts as a "dark pattern" is a subjective assessment, and similar to Spaa [65] and Jackson [33, 34], identify the need for interdisciplinary work across law, design, computer and social science in order to understand the policy-design-practice knot.

We now turn to a discussion pertaining to the specific policy we focus on for this paper: data protection laws, particularly the right of access. We begin by outlining previous studies of the right of access.

2.3 Researching (with) the Right of Access

Emerging from the individual participation principle in the 1960's [38, 54], the right of access obligates organizations that collect personal data to provide individuals with copies of their personal data on request. Organizations often formalize this request process as a Data Subject Access Request (DSAR), or through semi-automated data export processes like "Download My Data" buttons [2]. The right of access has been enshrined in sectoral and increasingly omnibus data protection laws around the world, including the European Union's General Data Protection Regulation (GDPR Art. 15), Brazil's Lei Geral de Proteção de Dados Pessoais (LGPD Art. 18), as well as US state laws in California, Virginia, and Colorado.

Design, HCI, and legal scholars have largely examined the right of access from a perspective of compliance [2, 44, 49] or the completeness and legibility of DSAR outputs [44, 72, 74]. In addition, they have examined the practical exercise of DSARs, and have found barriers that requestors encounter in the process of requesting, receiving, and understanding copies of personal data [2, 9, 12, 37, 44, 49, 72]. Navigating companies' DSAR processes often requires levels of legal and technical proficiency many people do not have [55]. Even if people complete a request, organizations can stall or refuse to comply with requests [3, 49]. When organizations do fulfill DSARs, returned data is often "unorganized," difficult to assess [72], and varies in type and structure between and within corporations [12, 69, 75]. DSARs have also raised security and privacy concerns, including spoofing a requestor's identity [12, 18], security flaws in DSAR procedures and authentication processes [11], and the responsibility and risks of creating archives made of other people's personal data [58]. Singh and Cobbe [64] speculate that DSAR outputs might be used to infer specific technical choices and implementations in organizational data infrastructure. These barriers and risks make it difficult for individuals to exercise and derive value from their right of access. Given these hurdles to requesting and understanding data, Veys et al. [72] conducted co-design sessions with requestors to improve the usability and utility of data downloads.

In addition to evaluation, the right of access has been used as a data collection method in both academic study designs [45, 69, 74, 75] and non-profit campaigns [53]. Wei et al. [74] asked participants to request data from Twitter and used that data to examine Twitter's ad targeting practices. Researchers have also used the right of access to assess the impact of GDPR outside of the EU's boundaries [43, 55]. NGOs also attempt to use DSARs to gain insight into corporate data practices. In 2019, Algorithm Watch and Open Knowledge Foundation Deutchland launched OpenSCHUFA, a campaign to reverse-engineer the credit-scoring algorithm of SCHUFA, Germany's largest private credit bureau. OpenSCHUFA solicited German citizens to request and submit data from SCHUFA in an (unsuccessful) effort to reverse-engineer SCHUFA's opaque credit scoring algorithms [53].

These studies point to the potential utility of the right of access in achieving a broad, societal layer of data protection. This insight has not been lost on EU legal scholars, who argue that collective applications of the right of access can help democratize control of corporate data practices [1, 43–45]. Envisioning DSARs as a "collective endeavor" [44], prior work has proposed having participants delegate access requests to researchers [1], creating data trusts [58], or involving multiple groups in assessing DSAR outputs [44]. This combination of work from HCI, policy governance, and technical measurements suggest that, while the right of access has the potential to support the collective "ecologies of transparency" these data protection regimes aim to create [45], the numerous barriers in the practical implementation of data rights hinders this potential from being achieved.

While this section focused on the right of access, the next section examines prior work that evaluates the effects of other data protection laws, such as the right to opt-out, legal requirements for collecting consent, the use of cookies and trackers, and the length and contents of privacy policies.

2.4 Evaluating the Technical Effects of Data Protection Law

In addition to the right of access, researchers from HCI, design, and policy have devised several ways to assess the effects of data protection laws through technical and often automated measurement. Several researchers have used automated or large-scale methods to evaluate the translation of GDPR requirements into technical infrastructure, such as measuring the impact of data protection laws on public websites and documents. Automated analysis of websites has been used to investigate other aspects of the GDPR beyond DSARs like privacy policies [17, 39, 50], cookies and trackers [17, 59, 70, 71], and HTTPS adoption [17]. Together, these automated studies of technical GDPR compliance suggest that corporations are responding to data protection laws, though how these changes affect user practices are necessarily understated in such work.

Automated studies often frame their object of study as software infrastructure –including privacy policies, cookies, and data. – allowing the researchers to systematically analyze technical infrastructure at scale. However, to systematize and automate research, automated studies make assumptions about technical infrastructure which do not always reflect user experiences. For instance, automated studies of privacy policies make choices about websites that have multiple or hierarchical privacy policies (e.g., parent-child companies). While this is a trivial distinction for machines, multiple privacy policies force citizens to search for and find the policy that applies to their question, effectively hindering the exercise of an informational right. Further, automated studies have procedures to account for websites without privacy policies in their analysis, but the same failure for a human might lead to further investigation, such as contacting an organization's data protection officer or customer service department. Analyzing the effects of GDPR on technical infrastructure is important and automated methods can provide broad snapshots of technical compliance. However, these methods do not capture the gaps between our analytical models of

technical infrastructure and the infrastructure itself, nor the practices that emerge from these gaps, which we intend to analyze in this paper.

3 METHOD

We recruited people using pre-selected services to execute DSARs in different parts of the world and tell us about their experiences. Our goal was to understand how companies respond to data requests in regions with and without legal data access rights, both in a comparative and in-depth, localized perspective. A separate part of our project [55] examined whether international companies facilitate data requests outside the EU. Thus, we designed our study to examine companies subject to different laws and headquartered in different locations. Our research participants made requests from Karnataka in India, Washington state in the USA, and England in the UK (prior to Brexit).

We selected both transnational companies (those that operate in all three countries): Google, Amazon, Facebook, and Spotify); and domestic companies (those that only operate in one of the regions): Monzo and RyanAir in the UK; Alaska Airlines and Venmo in the USA; and RelianceJio, PayTM, and Flipkart in India. We selected these companies because they operate in economic sectors that are data rich and vary in regional business activities: media and technology, retail companies, airlines, and payment companies. For example, we selected airlines because of their long history of utilizing data in business operations [25] and because they operated in only one of the regions we were concerned with but not the others.

3.1 Questionnaire-based Study Design

Our original study design consisted of a series of questionnaires and a follow-up interview. The use of questionnaires or surveys is a common data collection method in studies looking to empirically examine the practical reality of the right of access [2, 44, 49]. First, we assessed participant eligibility in a screening survey. Participants had to be at least 18 years old, use at least two of our pre-selected services, and be willing to request and inspect data from those services. Eligible participants were sent the first questionnaire, which asked participants to request data from these services and record their experiences searching for and making requests; where they looked, who they talked to, where (if at all) they found a request process, and what tasks were involved in making a request. After a request resolved (i.e., data was received or a request was rejected), timed out (more than 60 days passed without a response), or a participant was unable to find out how to request data from a service, we sent participants a second questionnaire, which asked participants to report on their experiences receiving data (or not), and if data was received, what data was included, whether they believed the services presented them with all the data held about them, and how easy or difficult it was to request, obtain, and understand data.

We followed up these questionnaires with a 30 minute in-person interview to ask participants more details about their experiences requesting and analyzing data. After an initial pilot test where two members of the research team requested data from services they use, we recruited participants using convenience and snowball sampling from our personal networks and word of mouth. We informed participants briefly about the right of access and the goals of the study, but we did not provide them with any guidance on where this process might begin or use pre-defined templates in requests as others have [3, 9]. Instead, we encouraged participants to reach out to the study team if they had any questions about the process, and to submit any relevant screenshots they were comfortable sharing with us.

Though it was our intention to conduct this protocol in all three regions simultaneously, we found it difficult to retain participants. More than 30 participants completed the initial screening survey, but only 7 participants, all in India, completed our questionnaire-based protocol. Most participants dropped out of the study after being asked to request data. After an informal review

of papers that engage with DSARs, we found that this is not uncommon in studies involving the right of access: both small-scale studies [44, 72, 74], and large-scale projects [53] report on the difficulties in supporting people unaffiliated with the study through exercising the right of access and maintaining their interest for the duration of the study.

3.2 Interview-based Study Design

In response to these challenges, we revised our protocol for US and UK participants. We replaced the two questionnaires with two remote interviews via Zoom to record people's experiences (1) making DSARs and then (2) analyzing data that they received. We structured the interviews to capture data congruent with the initial survey protocol. Like our questionnaire-protocol, we briefed participants about the right of access and the goals of the study, but did not tell participants where to begin, nor employed pre-defined template language when corresponding with companies unless asked. During the interviews, we used a coaching variation of a think-aloud protocol [52]: as participants made sense of the request process and data, we would periodically ask them to reflect on their experiences or provide additional context for their actions. If participants expressed frustration with the process or they were unable to proceed to the next step after 4 – 5 minutes of trying, the interviewer would remind the participant that it was not a test and they were free to ask any questions – which they did. Each interview lasted about one hour, though some interviews were more complicated and thus took longer.

Using this "coaching" protocol, we recruited 6 participants from the US and 6 from the UK, and asked them to make DSARs to two different services. Notably, all participants completed the "coaching" protocol of our study, which corroborates Olmsted-Hawala et al.'s finding that coaching think-aloud protocols yields higher task completion rates [52]. Bowyer et al. also report high participant retention rates using a similar interview-based study [9]. Given these results and our experiences, we believe our improved retention rate is due to the interviewer's interaction with participants and the provision of live guidance for navigating DSAR processes and data, but this requires further research. While coaching participants surely influenced how participants engaged with the process, higher task completion rates were a worthwhile tradeoff for understanding the practical obstacles and concerns of our participants.

3.3 Data Collection

Combining data from both our questionnaire and interview protocols, we recruited a total of 19 participants (see Figure 1): 7 participants from India, 6 from the US, and 6 from the UK, and asked them to request data from at least two services. Data collection was conducted in overlapping phases as surveys and interviews between September 2019 and March 2021. Crucially, all UK access requests were completed (and requested data accessed) before the UK left the EU on January 31st, 2021, and all US requests were completed after the California Consumer Privacy Act (CCPA) came into effect on January 1st, 2020.

3.4 Data Analysis

We first used an inductive thematic coding strategy [10] to analyze the questionnaire responses, interview transcripts, interview field diaries, available service correspondence, and UI screenshots to understand DSARs in the context of the policy-knot; what practices are emerging from these policy-influenced technical designs? This qualitative coding took place in two phases. First, all four authors coded one interview from each region using independent, inductive codes. We then met to discuss codes, merging, deleting and changing some codes and creating hierarchies of codes and themes. We then re-coded all the material (interviews and field diaries from interviews) so that each interview was coded by a minimum of two and a maximum of three researchers. We

ID	Gender	Age	Education Attained	Region	Protocol	Service 1	Service 2
P1	man	30-39	Post-graduate	US	Interviews-only	Venmo	Spotify
P2	woman	20-29	Post-graduate	US	Interviews-only	Facebook	Venmo
P3	woman	20-29	Post-graduate	US	Interviews-only	Google	Facebook
P4	woman	30-39	Post-graduate	US	Interviews-only	Alaska Air	Amazon
P5	man	20-29	Post-graduate	US	Interviews-only	Amazon	Spotify
P6	man	30-39	Post-graduate	US	Interviews-only	Google	Alaska Air
P7	woman	20-29	Undergraduate	UK	Interviews-only	Monzo	RyanAir
P8	woman	30-39	Post-graduate	UK	Interviews-only	Spotify	Google
P9	woman	20-29	Post-graduate	UK	Interviews-only	RyanAir	Amazon
P10	man	60-69	Post-graduate	UK	Interviews-only	Google	Amazon
P11	woman	30-39	Post-graduate	UK	Interviews-only	Facebook	Spotify
P12	woman	20-29	Post-graduate	UK	Interviews-only	Facebook	Monzo
P13	man	20-29	Undergraduate	India	Questionnaires	Amazon	Spotify
P14	man	30-39	Undergraduate	India	Questionnaires	Google	Facebook
P15	woman	20-29	Undergraduate	India	Questionnaires	Google	Flipkart
P16	man	20-29	Undergraduate	India	Questionnaires	Flipkart	PayTM
P17	woman	20-29	Undergraduate	India	Questionnaires	Amazon	Facebook
P18	woman	20-29	Undergraduate	India	Questionnaires	Facebook	Flipkart
P19	woman	30-39	Post--graduate	India	Questionnaires	Pay TM	Reliance Jio

Fig. 1. A table of our participants and their demographics.

used multiple coders in an interpretivist tradition of qualitative coding to establish the reliability of our findings. We did not calculate inter-coder reliability, but rather used both agreements and disagreements among coders to generate themes to explore further in regular meetings and data analysis [48]. Finally, the first author examined the questionnaires and transcripts and, as much as possible, recorded each step of a participant's path in requesting and receiving data from a service (see Appendix - Figures 4 and 3).

Our convenience and then snowball sampling (Table 1) yielded a group of educated people, all fluent in English even when they were not native speakers, mostly with graduate education, and often (though not always) with some technical background and awareness of GDPR rules. While our participants did not know the specifics of data protection laws in the region in which they requested data, some participants demonstrated high level understandings of some aspects of data protection. During interviews, participants checked their settings to remove third party access and adjust their privacy settings, even as they expressed frustration with how even constant vigilance was not sufficient to guarantee their privacy goals (P12, P6). Our participants included several people with some computer programming skills and/or a professional interest in online privacy. One participant even set up a GDPR-compliant database at their job (P3).

Close observation of participant data, practices, and computing environments creates both opportunity and responsibility. We worked closely with participants in minimizing data capture of potentially sensitive personal data, including existing device files, browsing history, and sensitive information. Before each interview, we offered participants tips for data minimization, such as clearing out their downloads folder or creating a new dedicated folder, using private browsing, or hiding file explorer sidebars. Participants shared their screen in interviews only if they were confident no sensitive information was displayed. When there was sensitive information, participants described what was on their screen, such as listing the categories of data in their DSAR outputs. However, both participants and the interviewer were uncertain about what was going to come

next in the process; on a few occasions the interviewer proactively turned off screensharing when potentially sensitive information became visible and notified the participant.

There are several limitations to our approach. In this study, we observed how individual requestors interacted with corporate request processes. We could not observe or record how companies process requests internally, so we cannot speak about the intentions behind specific request process implementations, nor can we speak with first-hand knowledge about journeys of data within corporate data infrastructures [5]. The number of participants who carried out DSARs allowed us to reach preliminary findings and conclusions about emergent data practices in a period of legal flux. However, we did not endeavor to exhaustively list all possible practices and designs emerging from the right of access. Additionally, data from participants in India are from self-reported questionnaires, rather than direct observation with US and UK participants. When compared with data from the US and UK (see Table 4), we cannot rule out potential underreporting of practices in India. Without observing requests happen in real-time in India, we were unable to distinguish between companies who do not provide DSAR processes and requestors who are unable to locate existing or obfuscated processes.

Our goal was to understand different user practices and expectations of DSARs, if and how they change in different countries, and how individual expectations and experiences are incorporated into the process. Qualitative studies on the practical experience of the right of access have mostly been characterized by a small number of participants [3, 44, 49]. The focus of our qualitative work was on how the DSAR process worked in practice from the perspective of users. Only the recent work of Bowyer et al. [9] has directly observed participants as they work through the request process within their own computing environments. However, Bowyer's work prescribes specific steps like beginning requests at the privacy policy or using pre-approved templates when corresponding with data controllers. This is best practice, as it shuttles participants through the most effective path through requesting data. However, what happens when participants do have this advice, or live in areas without a formal right of access?

Instead of prescribing steps for participants, we were intentionally vague in prescribing a specific process. This flexibility precluded us from assessing the rate at which services comply with formal GDPR requests; on the other hand, it allowed participants to travel down paths that surprised them and us, which destabilized our expectations of accessing data as researchers and enabled us to support non-expert participants in this complex, trying process. As the European Commission considers mandating technical interfaces to allow, for example, an increases in the use of data portability [20], more small, exploratory studies that engage closely with participants are necessary to avoid fossilizing technical solutions [33] before we even understand emerging problems.

4 FINDINGS

Thinking in terms of the policy-design-practice knot, our study focuses on the user practices that emerge from the GDPR's influence on corporate designs. The requirements of GDPR are articulated in the policy, but the specific details of how this change manifests in the black box of corporate software infrastructures are hard to divine; our only view into this is through their forward-facing interfaces, our participants' practices, and the results of the DSAR process. Thus, we examined the double-embeddedness [33] of the right of access in our participants' interactions with the company and participant reactions to DSARs.

With the exception of Amazon in India, companies operating in the USA, India, and UK (Amazon, Google, Spotify, and Facebook) responded to DSARs with data regardless of the requestor's location. These transnational companies employed semi-automated processes, all fashioned as a "download my data" services. However, most DSAR processes were largely *visual*, *de-coupled* from a company's

Participant	Region	Service	Did participant visit the privacy policy?	Did participant check account or profile settings?	Did participant check the website footer?	Did participant consult FAQs or other help documentation?	Did participant use a search engine?	Where did participant try to request data?	Was the participant able to locate where to make a request?
P7	UK	Monzo	NO	NO	YES	YES	YES	Email	YES
P12	UK	Monzo	YES	NO	YES	NO	NO	Chat	YES

Fig. 2. This table shows the different paths participants took requesting data from Monzo.

main experience (i.e., accessible via settings menus) and available *on-demand* [61]. Transnational corporations often fulfilled requests instantly, though some took up to 15 days.

By contrast, domestic companies, or those operating in only the USA (Venmo, Alaska Airlines), UK (Monzo, Ryanair), or India (Reliance Jio, PayTM, Flipkart) were more varied in their responses. Most domestic companies that returned data did so within 23 days. However, Alaska Airlines was a severe outlier, returning data to both participants and authors over 18 months after the original request, long after data collection was completed.

When we designed our interview protocol, we imagined a sequential process to execute a DSAR as others have [2, 44, 49]: 1) user requests data from company; 2) user receives data; 3) user reads or interprets data and reflects on the process. This framing presents the process of requesting data as a linear sequence. However, we found that a line does not adequately describe the experiences of participants. They struggled to figure out how or where to request data; adjusted or re-requested data because their initial attempt yielded un-workable or inaccessible data; re-visited documents when requesting data to check their understanding; did not receive data, could not open their data, and received different sets of data from the same company; or compared data returned from different companies. These challenges were magnified by language issues for people who were not native speakers of English or who used services in different languages. How people thought of their data protection rights also influenced how they approached DSARs and how they reacted along the process - insisting to company representatives they had the right to request data and get answers, or doubting (correctly, in some cases) whether or not data protection laws applied to them.

For instance, Figure 2 shows the different ways our participants requested data from Monzo. This is a slice of our larger participant table (the full table is available in the Appendix, Tables 3 and 4[1]), but this snapshot is emblematic of the multiple options, challenges and occasional dead-ends that DSARs involved. While it is possible to describe requesting and analyzing data in a linear fashion, it is important to note that this is an analytic convenience, flattening the practical experience of the twists, turns, folds, and breaks involved in requesting data.

4.1 Practices in Requesting Data

With our analytical considerations in mind, request processes begin by requesting data from a service. The next sections describe the complexities that users encountered when trying to request data.

4.1.1 Understanding local laws. Though most participants expressed some trepidation about how difficult requesting data might be, EU participants were more confident that such a request was possible. This was unsurprising, given that India does not have a national level data protection law (though such a law has been debated since at least 2017 [66], and the USA's CCPA did not apply to our participants who were not California residents. Participants had all heard of GDPR

[1]Note: Since we did not directly observe participants in India requesting data (see Methods), we could only include the tasks participants reported in their questionnaires. This table leaves unreported activities blank.

and similar laws but were not sure whether these laws applied to them. This legal uncertainty added an extra layer of anxiety in locating where to make a request: As P6 requesting from Alaska Airlines in the US state of Washington said: "So if I can figure out which data protection laws apply to me, I can figure out what rights I've got." The practice of "figuring out" rights in non-EU locations involved a variety of methods, including using search engine results, contacting customer support, or consulting the privacy policy. In the case of the US, the privacy policy was one of the few places that detailed whether a participant had California data rights. However, language in privacy policies was often vague, as P4 in the US noted about Alaska Airlines: "If I'm in a certain jurisdiction, which it doesn't tell me if I am, I might have a legal right to obtain confirmation of whether they have personal data about me." For non-EU residents, assessing whether they had a right to access or could exercise a request was a necessary and often opaque prerequisite to making a request.

4.1.2 English as the Language of IT (Policy). English was the default language of all the services we selected, but several of our participants used non-English versions of services according to what they needed or where they had originally signed up for the service. This unearthed a series of inconsistencies in translations, with text that appeared in English not appearing in other languages, incorrect translations, or wrong or non-existent words. These combined to create a sense of incompetence, helplessness and of being second-class citizens among participants. As P8 in the UK mentioned in discussing a Google data request: "Even if they send that page explaining it's so vague (...) And if you go further, I did not understand. And then (it) was in English and very difficult vocabulary and legal vocabulary" Participants discussed practices such as constantly switching languages and selecting specific languages for specific purposes: "For configuration, English. By daily activity, usually Bahasa." (P14, an Indonesian participant in India). Some mentioned that English is the default IT language, part and parcel of using technology, and thus the most useful when looking for help online: "I think with stuff like that, I often stick to English because I'm used to thinking and talking about stuff like that in English, so then when it translates, when it gives me the Dutch version, then either I don't always get it, or then sometimes if I want to Google something...when I use the Dutch terms, there isn't as much information or there aren't as many blogs." (P9 in the UK). This points to a set of practices that grow around a dominant language (English) and that lead to the interpretation of the local (legal) context through the lens of English language tools.

4.1.3 Locating Request Processes. A common challenge across companies and regions was to locate where and how to request data from companies. Some participants looked in Privacy, Security or Settings menus on corporate websites. Interestingly, at least 15 of our 38 requests (40%), mostly to transnational companies, were completed without viewing a privacy policy. After searching through formal corporate resources, most quickly gave up and turned to search on Google, even, ironically, when trying to request data from Google itself. But even Google searching was not a panacea as P17 in India described: "I was not able to make request actually, I wasn't able to find, actually I searched [on Google] for everything there, every option I was searching, but yeah, I did not get that option of making a request."

In addition to search, a company's customer service played an important role in exercising the right of access. In response to data requests, some customer service representatives directed participants to the privacy policy, told our participants they did not have access to the data being requested, or that the data being requested was already visible within their account information. In India, company representatives sometimes confused DSARs and Know Your Customer requests. Requesting from PayTM in India, P19 said "I felt that the customer care people, they're just unaware about personal information or data or something." P6 in the US expressed similar dissatisfaction

with Alaska Airline's customer service representatives after being directed to send an email to their Data Privacy Office from chat: "They don't mean it this way, but [asking me to send a separate email] is like giving me the middle finger." Company privacy policies were an important resource for some people, but participants also identified search engine results, customer service resources, blog posts, and help documentation as equally important resources.

4.1.4 Navigating Intermediaries. The usability requirements in GDPR and CCPA allow companies a great deal of flexibility in determining how to comply with legal obligations [73]. Some companies chose to process requests for data using intermediaries. The airlines in our research (Alaska Airlines and Ryanair) employed OneTrust, a third-party privacy compliance management service. Both companies provided links for requesting data in their privacy policy, which led participants to a request form on OneTrust's website. After submitting the form and verifying their identity and email address, participants were instructed to wait for an email response. The introduction of a third party required users to interact with multiple organizations or fulfill multiple authentication steps, which added extra time and hurdles to the process, such as visiting web pages outside of the company, or assessing whether to bring questions to the company or the intermediary.

In addition to third party intermediaries, participants also had to navigate the parties operating the infrastruture that supports the right of access. P7 in the UK was unable to request data from Ryanair through OneTrust due to breakdowns in these intermediaries. After OneTrust notified P7 their data was available, OneTrust sent a one-time verification code via email, which P7 had to input into OneTrust's website before they could access their data. However, P7 did not receive their verification code. P7 and the interviewer engaged in collaborative troubleshooting over Zoom, and eventually discovered there was service disruption to Gmail. P7 commented: "And I don't even know, I think if I contact [RyanAir's] customer service, I think it would be hard to explain exactly what is happening because it's not their websites. So they should contact OneTrust and then get back to me. So I suppose it will be a long process." Direct (OneTrust) and indirect (Google) intermediaries in request processes create additional complexity for requestors to navigate, and our participants were not confident companies would be helpful.

But the use of intermediaries did not make or break requests. RyanAir (through OneTrust) responded to other requests in 3 days. By contrast, Alaska Airlines (also using OneTrust) responded to requests 18 months after they were made. What is important is acknowledging the direct (OneTrust) and indirect (e.g., Google) intermediaries in request processes. These intermediaries can facilitate request processes but also create additional points of failure, creating uneven and lengthy experiences for requestors.

4.1.5 Selecting Data to Request. Once participants located where and how to make a request, participants were frustrated by the lack of feedback about what the process entailed in terms of data formats, timeline, and outputs. Amazon, Facebook, and Google offered numerous choices in requesting data, such as format, service, and time period (e.g., data from the last year, etc). Instead of reassuring participants of the thoroughness of the process, offering options made participants uncertain about how their selections would impact their request. Facebook offered participants the option of HTML or JSON data formats, but many participants were unfamiliar with one or both format types. P2 in the US offered that "it might be cool if there were examples like, this is what a JSON format would look like, what I'm trying to now imagine, what that looks like for my Facebook data, and I have no idea how to conceptualize what that would look like versus an HTML format." Requestors had no idea what to expect would come out of their data and thus struggled to envision outputs. Google, Facebook, and Amazon also presented confusingly similar options to both "View" and "Download" data to all requestors, echoing prior work that found companies offer similar-looking choices for opting-out of marketing communications [30]. Clarifying what these

optional choices mean for exported data before users initiate a request could increase a requestor's sense of control and understanding of the process.

4.2 Practices of Receiving Data

If a service provided data, users next accessed their data. We discuss the many practices, challenges, and workarounds associated with accessing data in this section.

4.2.1 Accessing Data. Out of the companies that returned data, all but Monzo and Venmo provided data to participants via a link that typically expired after 5 to 10 days. These restrictions are good security practice, but many participants simply forgot to access or download their data within the given time. This was less of a problem for companies who responded to requests quickly; participants who were unable to access or download their data before the link expired simply re-requested the data. Once the valid link was visited, participants were sometimes asked to authenticate their identity a second time before they could access their data. However, when services took more than a few days to comply with requests these minor difficulties in accessing requested data created additional steps and time. These experiences underscore some of the practical difficulties in gaining access to data. But gaining access to data was not the same thing as actually getting the data.

4.2.2 Downloading Large Data Files. While both GDPR and CCPA obligate companies to provide a copy of all personal data upon request, participants were often surprised by or unable to accommodate the time and resources required for large files. Many participants requested data through trial and error, which led participants to choose options that led to unusably large data. P6 in the US initially requested all of their data from Google, including a copy of their Google Drive and Gmail. The resultant export file was over 130GB, spread out over 6 different download links. Once P6 identified what went wrong, they initiated a second request to Google, but unchecked the box that would include their Google Drive and Mail data. This resulted in a much more manageable download.

4.2.3 Downloading Multiple Files. In addition to file size, participants were often shocked at the sheer number of files in their request outputs and the practical difficulties in retrieving those files. Data from Spotify contained less than 10 individual files; data from Google, Facebook, and Amazon all contained more than 30, sometimes more than 100 individual files. Facebook and Google made these files relatively easy to download all at once. By comparison, participants had to individually click on and save 50 to 80 files from Amazon. Reflecting on retrieving data from Amazon, P4 in the US said: "The fact that this is, I have to download 60 different files. Like, oh come on, just give me one zip file ... I get that maybe somebody only wants to see one particular thing, but if there was some option on here where I could just download all of it. Also, if there was maybe some explanation on what these things are and how they're used, that would be cool. It's not necessarily what I like the least, but that is something that would help the experience." While P4 indicates that this was not the most difficult part of the request process, each participant that downloaded from Amazon mentioned that the ability to download files as a batch would have been appreciated.

4.2.4 Opening Data. Companies provided a variety of formats in their downloaded data, a finding consistent with prior work [3, 42, 44, 72, 75]. No data is "raw" [8, 27], but many DSAR outputs given to users by companies resembled the "raw data" from a database migration. In the case of RyanAir, the output consisted not of machine-readable outputs but a Microsoft Word document with screenshots of booking reservations viewed through operations management software. This is not to say there was no variation. Notably, Monzo returned a 21 page PDF, 8 pages of which were dedicated to explaining their DSAR report. This included the what, how, and why of Monzo's data

collection practices and data sharing partners, with the remaining pages listing categories of data such as payments, transactions, and devices used.

Some participants did not recognize the data format at all and were only able to open the file with the help of the interviewer: "And then in my data, it has a bunch of JSON files, which I'm not familiar with, and I haven't actually tried to open those. So I don't know if that's something I can even... If I have an app that will read that... There was nothing in [the data] about how to open or what a JSON file is." (P1, Spotify and Venmo, US). Trouble with data formats was not unusual; P3 in the US did not unzip their data folder before inspecting the contents, which led to errors interacting with the data. P11's (UK) DSAR output became corrupted because of inadvertent syncing to a cloud provider when retrieving it from an encrypted storage application. Accessing data is often presented in a straightforward way, however these retrieval processes can break down when integrated into a participant's computing ecosystem.

4.3 Interpreting Data

Finally, users who received data attempted to read their data and discussed their feelings about their data in relation to the rest of the DSAR process.

4.3.1 Deciphering Data. Explanations of data via README documents or company websites were greatly appreciated by participants when they were available and seemed to improve the perceived utility of the DSAR process. However, these explanations were few and far between, and often consisted of a short paragraph which participants found less than elucidating. For instance, Spotify's website notes that their Inferences.json file "includes a list of market segments with which you are currently associated", but participants noted that this does not explain what a market segment is, how it is used, or how it came to be associated with their account. Understanding data was difficult even for participants with knowledge or experience working with data, perhaps the ideal candidates for a DSAR. Several of our technically-savvy participants tried (and failed) to reverse engineer the steps that led to certain documents or data. As P4, a data scientist requesting from Amazon in the US, mentioned: "To derive insights from this data … I would have to recreate their code base. [To do this]… I imagine you would need the power of all of Amazon, of that many engineers, or that many developers, or that many data scientists." For our technically-minded participants, this limitation felt frustrating, as described by P11 in the UK: "I'm a person who researches the digital. And so the fact that I was so taken aback just goes to show, I think that even incredibly digitally literate people still are woefully uninformed [about corporate data collection]." Across the board, we were surprised at how people with technical education or experience working with data or software still struggled to interpret DSAR outputs.

4.3.2 Data's Value to Participants. When participants imagined possible uses of the data, they did not feel that inspecting their data was worth the effort of the DSAR process. After participants examined their data, we asked them whether they could imagine themselves performing DSARs in the future on their own. Some participants provided speculative use cases. P6 in the US mentioned they might request data from their work's cloud service to save personal documents or projects; P2 in the US mentioned potentially downloading an archive of their Facebook account if they were to delete their account; P7 in the UK suggested they might request data from other financial institutions to compare their DSAR output with Monzo's to understand differences in data collection practices. While many participants suggested, as P5 in the US does, that it was "good to know that I can [request data]," few participants were able to identify a concrete reason to perform a DSAR on their own other than hypothetical archives or out of curiosity. P8 in the UK summed up this position: "No, I can't see (doing this again) because I don't have a reason why. Maybe just for curiosity but I don't know. Then what I would do with that, I would have to have for myself a

plausible reason to ask them. There's too much struggle and it's difficult to read." Many participants cited either the difficulty of the process or the inability to make sense of the data as reasons why they would not make a request on their own in the future.

4.3.3 Feelings of Powerlessness. Participants mentioned that these practical difficulties and interpretative obfuscations compounded a sense of powerlessness over their data, as P3 in the US suggests: "Unless you're in a space where you're able to do something about the fact that, yes, all my locations are being tracked, and now I'm gonna be thinking about it all the time, but do I have the power to stop it? Do I have the power to say no? Because one or the other the app is always gonna be using my location. Just gives me a little bit of anxiety." Some were very skeptical that the process could be empowering: "And I suspect that when I look at that... Again, I won't say it's deliberately obfuscating, but it's also... It doesn't clarify anything. So it makes me think, well, maybe it's a tool to let me think that they're being more transparent than they might actually be." (P10, Amazon and Google, UK). As P10 suggests, DSARs can be perceived as disingenuous overtures to transparency when returned data reveals little of how corporations collect and use data.

4.3.4 Coaching. Many participants, especially those without technical expertise, mentioned that the interviewer's intervention greatly facilitated their progress through the study. In addition to being supported in task completion, many participants found it helpful to have a dialogue with a knowledgeable "confidante." As P12 describes: "when you're [the interviewer] asking me what I'm interested in, I'm actually then talking through: well is that the case with this? Is that the case with that? So that was really helpful to actually get someone who knows a little bit more about how to analyze the data." Throughout the interviews, the interviewer also encountered unfamiliar systems and data. Establishing a dialogue between interviewer and participant opened space for collaborative sensemaking over DSARs processes and data. When P8 in the UK requested Spotify data, both participant and interviewer were stumped by the values of Inferences.json, a file with a list of strings all containing the prefix "day-part". When P11 in the UK received their data from Spotify, they leveraged their work experience in advertising and explained to the researcher what this meant in advertising practice: "the term "day part" is a Nielsen term. It's how TV buyers buy television or media planners plan media", so inferences labeled "day-part-late-night" indicate that an advertisement was shown to a participant late at night (which has an associated cost). P11 was also able to speculate how the data in Inferences.json may have been created, thus providing useful context for subsequent discussions among the research team and with other interviewees.

4.3.5 Assessing the right of access. With the results of the DSAR in hand, participants questioned the overall effectiveness of the right of access. After requesting data from Google and Amazon in the UK, P10 looked back on their experience, commenting: "GDPR makes it looks like, "Oh, here, I've got all this control," but if you actually look at the control it's offering you, it doesn't actually explain [how choosing control options affect tracking and data collection]. ... [T]his is just Brussels imposing on us, craziness that we have to comply with. And it doesn't save you any effort and it doesn't do anything to [protect you from companies]." Our participants' experiences and reflections suggest that the usability of "legal tech" [60] such as DSARs are entwined with a requestor's assessment of how empowering and effective data protection laws are broadly. Since emergent policy and technical discussions will shape the future of personal data control, we found that the nexus of laws, technical implementations, and practices are not achieving their intended goals.

5 DISCUSSION

The perspective of the policy "knot" allows us to observe how law is entwined with design and practice. Consistent with prior work [2, 9, 44, 53, 72, 72], our qualitative analysis revealed the sheer difficulty participants experienced in requesting, receiving, and analyzing the outputs of DSAR requests – not the intended implication of data access rights. Many studies propose a deterministic relationship between legal rights and practical experiences (i.e., data protection rights are only available to requestors in jurisdictions with data protection laws), but here we show that even citizens without rights of access were able to receive data upon request. Using the policy-design-practice "knot"Using the policy-design-practice "knot" [33], the experiences of our participants lead us to five arguments for future CSCW research involving the right of access, DSARs, and individual requestors' practices and experiences.

First, since the DSAR process is not linear for users, employing linear paths as analytical frames should be appreciated as an analytical convenience. Second, user frustration and disappointment by the DSAR process can undermine the perceived utility of data protection laws. Third, we need more experimentation in how to make these policies meaningful to different groups of people. We advocate for resisting calling for more policy specification or DSAR standardization for now to avoid locking in ineffective designs [32]. Fourth, our findings underscore that DSARs should be understood as part of collective data protection endeavors, a view which creates opportunities for CSCW researchers to deepen collaboration between researchers and research communities. Fifth, our research method involved coaching our participants which provides a helpful model for thinking about the potentialities of the right of access as a collective data protection program. We discuss these points in detail, and then offer design implications for corporate request processes which we hope will improve not just request processes and the readability of returned data but, most importantly, the efficacy of the right of access on the ground.

5.1 DSARs are Not Linear

DSARs are not linear experiences: despite having a clear beginning and end, the process itself can loop, split, and break. Along with other researchers working on DSARs and other corporate responses to data protection law [2, 9, 17, 30, 44, 49, 69], we analytically segmented DSARs into a sequence of tasks. This scheme was necessary to organize the interview process and report our findings in a readable way, but it obscured how our participants' experiences began, folded, curved, split, looped, and broke at multiple points throughout this process. These insights challenge our segmentation of DSARs. Requests may not always be one-time events; 5 of our 19 participants requested data from a service multiple times. Many researchers begin their studies at the privacy policy [2, 44], yet 15 of the 38 requests in our study resulted in data despite never referring to a privacy policy.

When researchers imagine these experiences as linear processes, they risk both misrepresenting the practical difficulties participants experience in requesting data and unintentionally excluding the communities that are most likely to be helped by data protection laws [40, 41, 53, 67]. Engaging with close observation and collaboration with participants as they exercise their rights within their computing environments is an opportunity to reflexively interrogate our assumptions about how the law works on the ground and generate deeper insights into data protection policy-knots [55].

5.2 Usability Informs Perceived Utility of Data Protection Laws

The difficulties experienced by our participants shaped how they perceived the data they received and whether data protection laws were achieving (or not) their intended goals. Our participants struggled to request, retrieve, assess, or interpret data. After completing their requests, many

participants expressed feelings of ambivalence or disempowerment when their data did not reveal more details into how companies generate, collect, and share their data. Many were disillusioned by the lack of transparency, though a few companies provided documentation of varying quality. The hurdles participants faced left many feeling like DSARs are "not worth the effort." Participants did not see the point of requesting data, and the DSAR process itself had left some feeling a sense of hopelessness over their data ownership and their power to control it in any way. Few participants could see a reason they might do a DSAR in the future.

5.3 Experimentation Instead of Closure

Companies have responded to right of access obligations in a myriad of different and inconsistent ways, yet few of our participants felt confident or empowered after performing their DSAR, even (and perhaps especially) when they received data. When confronted with usability issues such as those we describe above, it is tempting to advocate for specification of compliant design patterns, standardization, and more policy specification [9, 21, 28, 30, 31, 51, 61, 72]; actions that might stabilize the DSAR practices leading to closure in experimentation. However, we believe that the intentional lack of technical specifications in data protection laws is practical and, at this point in time, necessary. We are in a profound moment of variable interpretation, where different entities are figuring out how to comply with the law and serve their interests. Flexibility in interpreting law allows laws to remain relevant over time [33]. Instead of closing in around ineffective request processes, we need to experiment with how DSARs might achieve their ambitious goals of providing people with a base level of personal data protection. Rather than closure, we need more experimentation both in how DSARs are instrumented, and how DSAR outputs might be utilized to support data protection.

5.4 DSARs as Collaborative Data Protection

Delineating between GDPR as a privacy law or a data protection law becomes vitally important when we try to assess what DSARs (and other data subject rights) are useful for. The GDPR and the right of access in the EU legal tradition are data protection laws [35]. However, if we imagine that DSARs are meant to protect individual privacy, we may come away feeling disillusioned by the impact of the GDPR and other data protection laws as our participants did. As a privacy protection, only exercising a DSAR does little to accomplish the goals of privacy law: to prohibit corporate behavior or check for noncompliant data flows [29, 35]. Our assessment is that DSARs should be understood as a data protection endeavor rather than an individual privacy choice [21, 30]. A productive avenue of inquiry lies in assessing DSARs for their ability to establish "rules of the game" for data processing [35].

Thinking of the right of access as a data protection endeavor allows us to revisit a critical problem: DSARs are often individually fraught or ineffective (but not always, see the work of Maximillian Schrems [14, 15]). As both prior work [1, 44] and our participants found, DSARs are the product of a dizzying array of technical, organizational, and user practices that no one requestor or researcher could confidently or reasonably re-create. EU legal scholars echo our assessment that DSARs hold more potential as a collective endeavor [1, 42, 44, 45]. The concept of "data altruism" [20] where stakeholders collaborate to request data [53] or including different stakeholders in the DSAR process [1, 44] are both examples of how DSARs might be used to create social controls for corporate data practices.

5.5 "Coaching" DSARs

Coaching people through DSARs is (1) an example of engaging with the right of access collectively and (2) has benefits for participants and researchers alike. Corroborating [52]'s finding that

"coaching" think aloud protocols lead to higher rates of task completion, all participants who began the request process in the "coaching" version of the study completed their requests. One possibility for engaging with DSARs as a collective data protection endeavor is to work directly with communities to understand and exercise their data protection rights and answer their situated questions about technology, surveillance, and data. In prior work with DSARs, most participants (our study included) skewed towards technically-savvy and well-educated participants [2, 44, 72, 74]. Coaching participants through requesting data may be an effective way to engage with broader sets of communities.

We see DSARs, but also other data protection rights, as areas of opportunity for CSCW researchers. CSCW researchers often work with social media data [74] and are concerned with how to ethically work with research participants, including issues of reciprocity and consent [24]. Working directly with communities to support data protection rights is one such avenue for reciprocity. CSCW researchers are also in an excellent position to fashion additional resources (beyond corporate policy documents that requestors may not even need to read) that support communities through the processes of requesting, opening, and understanding not only their data, but the information economics and corporate logics that are impressed upon this data.

Finally, the interviewer, as lead researcher, also benefited from this "coaching" arrangement, learning from interviewees about the DSAR process as the protocol was designed to do, and about interaction with systems and data that only specific interviewees understood. This approach, like other participatory and ethnographic research methods, can engage a broad range of participants, generate rich insights into the lived experiences of exercising data rights, and support collaborative sensemaking between researchers, communities, advocacy groups, policy makers, and technologists.

6 IMPLICATIONS FOR POLICY AND DESIGN

DSARs enable novel practices for users. Our findings about user practices can feed back into policy and design. Streamlining and simplifying request processes or organizing data may indeed improve the usability of DSARs. However, the frustration our participants expressed at how little they were able to know or do with their data suggests that this alone is not sufficient. Following the 'knot', we offer design and policy considerations that might help increase the perceived effectiveness of DSARs without specifying technical arrangements.

6.1 Policy Consideration 1: Information Beyond the Privacy Policy

Studies investigating request process or data rights broadly often position a company's privacy policy as the primary source of information [2, 9, 17, 44, 69]. Privacy policies contain legal descriptions of data practices and are important for regulators and researchers assessing compliance, but few participants began their request with the privacy policy. Many participants (mostly those requesting from transnational companies) successfully requested data without consulting a privacy policy. Instead, participants relied on a bevy of alternative resources such as customer service representatives, chat applications, service-specific forums and help documentation. For participants, engaging with privacy policies was resource intensive; they are dense in legal jargon, distributed among products and services, and often privilege English technology terms. Increasing the number of resources (such as FAQs) that point towards a corporation's DSARs page could help users request their data more easily [30]. Companies should also expect requests to come through their customer service channels, prepare representatives to identify and forward requests as appropriate, and have contingencies ready to account for disruptions to direct and indirect intermediaries.

6.2 Policy Consideration 2: Companies and Civil Society Organizations Should Support Exercising Data Rights

As we described, requesting data was an act of ongoing sensemaking, troubleshooting, and trial and-error. Before requesting data, participants were uncertain of what they were requesting, whether they were doing it correctly, and what the output of a request would be because there was little information from corporations. This created numerous problems after participants requested data; some participants did not click the access link before it expired, some people did not download the data once they accessed it, others found that the data was untenably large or numerous. Civil society organizations, CSCW researchers, and companies could help by crafting and providing additional information before data is requested, such as details on what documents participants will need to authenticate their identity, what steps to expect after a request is made, feedback on how different format and quality options change a data export, or explicitly explain how to interact with different data formats.

Companies also need to provide accessible and complete explanations of data so users do not have to develop their own theories of what their personal data means. Data subjects are not helpless; participants tried to "reverse engineer" what data means or speculate on how data controllers use their data. This speculation engendered suspicion and doubt among participants; expending effort to explain categories, criteria, and uses would go a long way towards creating a sense of control and trust for users.

7 CONCLUSION

We applied the metaphor of the policy-design-practice knot [33] to the right of access as a case study to understand what practices are emerging from requesting data. For the right of access and other data protection rights to achieve their ambitious goals, more research is needed that carefully observes the lived experiences and practices involved in exercising data protection rights. In addition, we call for more experimentation with how DSARs are instrumented and used. CSCW researchers have the tools, theories, and perspectives to investigate, improve, and support DSARs and other data protection rights.

The fines CNIL levied on Facebook and Google demonstrate that data protection authorities in Europe are ensuring data protection rights are easy to exercise in practice. But these individual data protection rights, the crown jewel of the EU's strategy to level the digital playing field between corporations and society, were not highly valued by our participants. Focusing on existing practices, understandings and challenges are valuable for lawmakers and researchers to ensure adherence with both the letter and the spirit of data protection law.

ACKNOWLEDGMENTS

We would like to thank all of our participants, including those who did not complete the study with us. This paper would not have been possible without their time, attention, and feedback. Particular thanks goes out to the student participants at IIIT Bangalore; Dr. Jef Ausloos for his generosity sharing his original protocol with our team; and Dr. Jaime Snyder and our CSCW reviewers for their thoughtful comments and helpful suggestions.

We would also like to thank previous members of the study team, whose contributions were invaluable to this research, including recruiting participants, conducting interviews, and analyzing privacy policies: A. Janani, Nayana Dhavan, Meg Young, Tasha Lin, Hemica Saxena, and Divya Seth. This work was funded by the University of Washington's Strategic Research Fund.

REFERENCES

[1] Hadi Asghari, Thomas van Biemen, and Martijn Warnier. 2021. Amplifying Privacy: Scaling Up Transparency Research Through Delegated Access Requests. *arXiv:2106.06844 [cs]* (June 2021). arXiv:2106.06844 [cs] http://arxiv.org/abs/2106.06844

[2] Jef Ausloos and Pierre Dewitte. 2018. *Shattering One-Way Mirrors. Data Subject Access Rights in Practice.* SSRN Scholarly Paper ID 3106632. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3106632

[3] Jef Ausloos and Michael Veale. 2020. Researching with Data Rights. *Technology and Regulation* 2020 (2020), 136–157. https://techreg.org/index.php/techreg/article/view/61

[4] Rishab Bailey, Smriti Parsheera, Faiza Rahman, and Renuka Sane. 2018. *Disclosures in Privacy Policies: Does 'Notice and Consent' Work?* SSRN Scholarly Paper 3328289. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3328289

[5] Jo Bates, Yu-Wei Lin, and Paula Goodale. 2016. Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows. *Big Data & Society* 3, 2 (Dec. 2016), 2053951716654502. https://doi.org/10.1177/2053951716654502

[6] Nicola J. Bidwell, Roberto Cibin, Conor Linehan, Laura Maye, and Sarah Robinson. 2021. Being Regulated: Licence to Imagine New Technology for Community Radio. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 154:1–154:27. https://doi.org/10.1145/3449228

[7] Geoffrey Bowker and Susan Leigh Star. 1999. *Sorting Things Out.* MIT Press. https://mitpress.mit.edu/books/sorting-things-out

[8] Geoffrey C. Bowker. 2008. *Memory Practices in the Sciences* (illustrated edition ed.). The MIT Press, Cambridge, Mass.

[9] Alex Bowyer, Jack Holt, Josephine Go Jefferies, Rob Wilson, David Kirk, and Jan David Smeddinck. 2022. Human-GDPR Interaction: Practical Experiences of Accessing Personal Data. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3501947

[10] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[11] Luca Bufalieri, Massimo La Morgia, Alessandro Mei, and Julinda Stefa. 2020. GDPR: When the Right to Access Personal Data Becomes a Threat. In *2020 IEEE International Conference on Web Services (ICWS)*. 75–83. https://doi.org/10.1109/ICWS49710.2020.00017

[12] Matteo Cagnazzo, Thorsten Holz, and Norbert Pohlmann. 2019. GDPiRated – Stealing Personal Information On- and Offline. In *Computer Security – ESORICS 2019 (Lecture Notes in Computer Science)*, Kazue Sako, Steve Schneider, and Peter Y. A. Ryan (Eds.). Springer International Publishing, Cham, 367–386. https://doi.org/10.1007/978-3-030-29962-0_18

[13] Alissa Centivany. 2016. Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 926–940. https://doi.org/10.1145/2818048.2820069

[14] CJEU Judgement. 2015. Maximillian Schrems v Data Protection Commissioner, C 362/14. https://eur-lex.europa.eu/legal-content/en/TXT/PDF/?uri=uriserv%3AOJ.C_.2015.398.01.0005.01.ENG

[15] CJEU Judgement. 2020. Data Protection Commissioner v Facebook Ireland Ltd and Maximillian Schrems. https://curia.europa.eu/juris/document/document_print.jsf?docid=228677&text=&dir=&doclang=EN&part=1&occ=first&mode=req&pageIndex=0&cid=10480014

[16] Commission Nationale de l'Informatique et des Libertés. 2022. Cookies: The CNIL Fines Google a Total of 150 Million Euros and Facebook 60 Million Euros for Non-Compliance with French Legislation. https://www.cnil.fr/en/cookies-cnil-fines-google-total-150-million-euros-and-facebook-60-million-euros-non-compliance

[17] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. 2019. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. *Proceedings 2019 Network and Distributed System Security Symposium* (2019). https://doi.org/10.14722/ndss.2019.23378 arXiv:1808.05096

[18] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte, and Ken Andries. 2019. Personal Information Leakage by Abusing the {GDPR} 'Right of Access'. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*. 371–385. https://www.usenix.org/conference/soups2019/presentation/dimartino

[19] Paul N. Edwards, Steven J Jackson, Geoffrey C Bowker, and Cory P Knobel. 2007. *Understanding Infrastructure: Dynamics, Tensions, and Design.* NSF Grant. Ann Arbor, MI. https://deepblue.lib.umich.edu/handle/2027.42/49353

[20] European Commission. 2020. Data Protection as a Pilalr of Citizens' Empowerment and the EU's Approach to the Digital Transition - Two Years of Application of the General Data Protection Regulation. https://doi.org/10.1163/2210-7975_HRD-4679-0058

[21] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. 2021. A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445148

[22] Casey Fiesler. 2014. Copyright and Social Norms in Communities of Content Creation. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW Companion*

'14). ACM, New York, NY, USA, 49–52. https://doi.org/10.1145/2556420.2556821

[23] Casey Fiesler, Jessica L. Feuston, and Amy S. Bruckman. 2015. Understanding Copyright Law in Online Creative Communities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 116–129. https://doi.org/10.1145/2675133.2675234

[24] Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4, 1 (Jan. 2018), 2056305118763366. https://doi.org/10.1177/2056305118763366

[25] Philip L. Frana. 2018. Telematics and the Early History of International Digital Information Flows. *IEEE Annals of the History of Computing* 40, 2 (April 2018), 32–47. https://doi.org/10.1109/MAHC.2018.022921442

[26] Gloria González Fuster. 2014. *The Emergence of Personal Data Protection as a Fundamental Right of the EU*. Springer Science & Business.

[27] Lisa Gitelman (Ed.). 2013. *Raw Data Is an Oxymoron*. The MIT Press.

[28] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA, 1–18. https://doi.org/10.1145/3411764.3445779

[29] Serge Gutwirth and Paul De Hert. 2006. Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power. In *Privacy and the Criminal LAw*. Vol. 18. Intersentia. https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6200

[30] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2020. "It's a Scavenger Hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376511

[31] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. 2021. Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 63. ACM, New York, NY, USA, 1–25. http://doi.org/10.1145/3411764.3445387

[32] Woodrow Hartzog. 2018. *Privacy's Blueprint: The Battle to Control the Design of New Technologies*. Harvard University Press.

[33] Steven J. Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The Policy Knot: Re-Integrating Policy, Practice and Design in Cscw Studies of Social Computing. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 588–602. https://doi.org/10.1145/2531602.2531674

[34] Steven J. Jackson, Stephanie B. Steinhardt, and Ayse Buyuktur. 2013. Why CSCW Needs Science Policy (and Vice Versa). In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*. ACM Press, San Antonio, Texas, USA, 1113. https://doi.org/10.1145/2441776.2441902

[35] Meg Leta Jones and Margot E. Kaminski. 2020. *An American's Guide to the GDPR*. SSRN Scholarly Paper ID 3620198. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3620198

[36] Hans K. Klein and Daniel Lee Kleinman. 2002. The Social Construction of Technology: Structural Considerations. *Science, Technology, & Human Values* 27, 1 (Jan. 2002), 28–52. https://doi.org/10.1177/016224390202700102

[37] Jacob Leon Kröger, Jens Lindemann, and Dominik Herrmann. 2020. How Do App Vendors Respond to Subject Access Requests? A Longitudinal Privacy Study on iOS and Android Apps. In *Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES '20)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/3407023.3407057

[38] Martha K. Landesberg, Toby Milgrom Levin, Caroline G. Curtin, and Ori Lev. 1998. *Privacy Online: A Report to Congress*. Technical Report. United States Federal Trade Commission. https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

[39] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. 2020. The Privacy Policy Landscape After the GDPR. *Proceedings on Privacy Enhancing Technologies* 2020, 1 (Jan. 2020), 47–64. https://doi.org/10.2478/popets-2020-0004

[40] Mary Madden. 2017. Privacy, Security, and Digital Inequality. *Big Data & Society* (2017), 125.

[41] Mary Madden, Michele Gilman, Karen Levy, and Alice Marwick. 2017. Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans. *Washington University Law Review* 95, 1 (2017), 53–126. https://heinonline.org/HOL/P?h=hein.journals/walq95&i=59

[42] René Mahieu. 2021. The Right of Access to Personal Data: A Genealogy. *Technology and Regulation* 2021 (Aug. 2021), 62–75. https://doi.org/10.26116/techreg.2021.005

[43] René Mahieu, Hadi Asghari, Christopher Parsons, Joris van Hoboken, Andrew Hilts, Masashi Crete-Nishihata, and Siena Anstis. 2021. Measuring the Brussels Effect through Access Requests. In *BILETA Conference 2021*. https:

//hcommons.org/deposits/item/hc:38231/

[44] René Mahieu, Hadi Asghari, and Michel van Eeten. 2018. *Collectively Exercising the Right of Access: Individual Effort, Societal Effect.* SSRN Scholarly Paper ID 3216615. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3216615

[45] René Mahieu and Jef Ausloos. 2020. Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access. https://doi.org/10.31228/osf.io/b5dwm

[46] Peter Mayer, Yixin Zou, Florian Schaub, and Adam J. Aviv. 2021. "Now I'm a Bit Angry:" Individuals' Awareness, Perception, and Responses to Data Breaches That Affected Them. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*. https://www.usenix.org/conference/usenixsecurity21/presentation/mayer

[47] Aleecia M. McDonald and Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society* 4, 3 (2008), 543–568. https://heinonline.org/HOL/P?h=hein.journals/isjlpsoc4&i=563

[48] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 72:1–72:23. https://doi.org/10.1145/3359174

[49] Clive Norris, Paul de Hert, Xavier L'Hoiry, and Antonella Galetta (Eds.). 2017. *The Unaccountable State of Surveillance: Exercising Access Rights in Europe.*

[50] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating Their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, NY, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376321

[51] Sean O'Connor, Ryan Nurwono, Aden Siebel, and Eleanor Birrell. 2021. (Un)Clear and (In)Conspicuous: The Right to Opt-out of Sale under CCPA. In *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*. ACM, NY, NY, USA, 59–72. https://doi.org/10.1145/3463676.3485598

[52] Erica L. Olmsted-Hawala, Elizabeth D. Murphy, Sam Hawala, and Kathleen T. Ashenfelter. 2010. Think-Aloud Protocols: A Comparison of Three Think-Aloud Protocols for Use in Testing Data-Dissemination Web Sites for Usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, NY, NY, USA, 2381–2390. https://doi.org/10.1145/1753326.1753685

[53] OpenSCHUFA. 2019. OpenSCHUFA: The campaign is over, the problems remain. https://openschufa.de/english/

[54] Organisation for Economic Co-operation and Development. 2013. *The OECD Privacy Framework.* Technical Report. https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

[55] Justin Petelka, Megan Finn, Janaki Srinivasan, and Elisa Oreglia. Forthcoming. "A Mirror, Not a Glass Door": Legal Code and Software Code in Practice. In *Just Code: Power, Inequality, and the Political Economy of IT*. Johns Hopkins Press.

[56] Trevor Pinch and Wiebe E. Bijker. 1987. The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In *Social Construction of Technological Systems. New Direction in the Sociology of Technology*. Vol. 14. 17–50.

[57] Pablo-Alejandro Quinones, Stephanie D. Teasley, and Steven Lonn. 2013. Appropriation by Unanticipated Users: Looking beyond Design Intent and Expected Use. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, NY, NY, USA, 1515–1526. https://doi.org/10.1145/2441776.2441949

[58] Anouk Ruhaak. 2020. Data Trusts in Germany and under the GDPR. *Algorithm Watch* (2020), 19.

[59] Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. 2019. Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (Asia CCS '19)*. ACM, NY, NY, USA, 340–351. https://doi.org/10.1145/3321705.3329806

[60] Kristin Bergtora Sandvik. 2019. Is Legal Technology a New "Moment" in the Law and Development Trajectory? https://antipodeonline.org/2019/12/04/legal-technology-law-and-development/

[61] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing Effective Privacy Notices and Controls. *IEEE Internet Computing* 21, 3 (May 2017), 70–77. https://doi.org/10.1109/MIC.2017.75

[62] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub

[63] Paul M. Schwartz. 2019. Global Data Privacy: The EU Way. *NY University Law Review* 94, 4 (2019), 771–818. https://heinonline.org/HOL/P?h=hein.journals/nylr94&i=789

[64] Jatinder Singh and Jennifer Cobbe. 2019. The Security Implications of Data Subject Rights. *IEEE Security Privacy* 17, 6 (Nov. 2019), 21–30. https://doi.org/10.1109/MSEC.2019.2914614

[65] Anne Spaa, Abigail Durrant, Chris Elsden, and John Vines. 2019. Understanding the Boundaries between Policymaking and HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, NY, NY, USA, 1–15. http://doi.org/10.1145/3290605.3300314

[66] Srikrishna Committee on a Data Protection Framework for India. 2018. *A Free and Fair Digital Economy Protecting Privacy, Empowering Indians.* Technical Report. Ministry of Electronics and Information Technology. https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf

[67] Janaki Srinivasan, Savita Bailur, Emrys Schoemaker, and Sarita Seshagiri. 2018. The Poverty of Privacy: Understanding Privacy Trade-Offs From Identity Infrastructure Users in India. *International Journal of Communication* 12, 0 (March 2018), 20. https://ijoc.org/index.php/ijoc/article/view/7046

[68] Susan Leigh Star and Karen Ruhleder. 1996. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research* 7, 1 (March 1996), 111–134. https://doi.org/10.1287/isre.7.1.111

[69] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2019. A Study on Subject Data Access in Online Advertising After the GDPR. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology (Lecture Notes in Computer Science)*, Cristina Pérez-Solà, Guillermo Navarro-Arribas, Alex Biryukov, and Joaquin Garcia-Alfaro (Eds.). Springer International Publishing, Cham, 61–79. https://doi.org/10.1007/978-3-030-31500-9_5

[70] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2020. Measuring the Impact of the GDPR on Data Sharing in Ad Networks. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (ASIA CCS '20)*. ACM, NY, NY, USA, 222–235. https://doi.org/10.1145/3320269.3372194

[71] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)Informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. ACM, NY, NY, USA, 973–990. https://doi.org/10.1145/3319535.3354212

[72] Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman, Nathan Reitinger, Michelle L. Mazurek, and Blase Ur. 2021. Pursuing Usable and Useful Data Downloads Under GDPR/CCPA Access Rights via Co-Design. In *Seventeenth Symposium on Usable Privacy and Security ({SOUPS} 2021)*. 217–242. https://www.usenix.org/conference/soups2021/presentation/veys

[73] Ari Ezra Waldman. 2021. *Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power.* Cambridge University Press, Cambridge, United Kingdom ; NY, NY.

[74] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, and Blase Ur. 2020. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *29th USENIX Security Symposium (USENIX Security 20)*. 145–162. https://www.usenix.org/conference/usenixsecurity20/presentation/wei

[75] Janis Wong and Tristan Henderson. 2019. The Right to Data Portability in Practice: Exploring the Implications of the Technologically Neutral GDPR. *International Data Privacy Law* 9, 3 (Aug. 2019), 173–191. https://doi.org/10.1093/idpl/ipz008

[76] Yixin Zou, Abraham H. Mhaidli, Austin McCall, and Florian Schaub. 2018. "I've Got Nothing to Lose": Consumers' Risk Perceptions and Protective Actions after the Equifax Data Breach. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*. 197–216. https://www.usenix.org/conference/soups2018/presentation/zou

A APPENDIX

A.1 Service Response Table

Participant	Region	Operation	Service	Did the service return a copy of requested data?	What format was the returned data?	How was data returned?	How long did the service take to return data?
P1	US	Domestic	Venmo	Partial	PDF	Downloadable in account settings	23 days
P1	US	Transnational	Spotify	Yes	CSVs	Expiring link via email	2 days
P2	US	Transnational	Facebook	Yes	Multiple	Expiring link via email	< 1 day
P2	US	Domestic	Venmo	Partial	CSV	Email attachment	20 days
P3	US	Transnational	Google	Yes	Multiple	Expiring link via email	< 1 day
P3	US	Transnational	Facebook	Yes	Multiple	Expiring link via email	< 1 day
P4	US	Domestic	AlaskaAir	No	N/A	N/A	~ 18 months **
P4	US	Transnational	Amazon	Yes	Multiple	Downloadable through website	5 days
P5	US	Transnational	Amazon	Yes	Multiple	Downloadable through website	5 days
P5	US	Transnational	Spotify	Yes	CSVs	Expiring link via email	3 days
P6	US	Transnational	Google	Yes	Multiple	Expiring link via email	< 1 day
P6	US	Domestic	AlaskaAir	No	N/A	N/A	~ 18 months
P7	UK	Domestic	Monzo	Yes	Password-protected PDF	Email attachment	18 days
P7	UK	Domestic	Ryanair	Partial	N/A, unavailable at interview	N/A	3 days
P8	UK	Transnational	Spotify	Yes	CSVs	Expiring link via email	4 days
P8	UK	Transnational	Google	Yes	Multiple	Expiring link via email	< 1 day
P9	UK	Domestic	Ryanair	Yes	Word document with screenshots	OneTrust website	2 days
P9	UK	Transnational	Amazon	Yes	Multiple	Downloadable through website	15 days
P10	UK	Transnational	Google	Yes	Multiple	Expiring link via email	< 1 day
P10	UK	Transnational	Amazon	Yes	Multiple	Downloadable through website	< 1 day
P11	UK	Transnational	Facebook	Partial	N/A, corrupted on download	Expiring link via email	< 1 day
P11	UK	Transnational	Spotify	Yes	CSVs	Expiring link via email	6 days
P12	UK	Transnational	Facebook	Yes	Multiple	Expiring link via email	< 1 day
P12	UK	Domestic	Monzo	No	N/A	N/A	N/A
P13	IN	Transnational	Amazon	N/A*	N/A	N/A	N/A
P13	IN	Transnational	Spotify	Yes	CSVs	Expiring link via email	8 days
P14	IN	Transnational	Facebook	Yes	Multiple	Expiring link via email	< 1 day
P14	IN	Transnational	Google	Yes	Multiple	Expiring link via email	< 1 day
P15	IN	Domestic	Flipkart	N/A	N/A	N/A	N/A
P15	IN	Transnational	Google	Yes	Multiple	Expiring link via email	< 1 day
P16	IN	Domestic	Flipkart	N/A	N/A	N/A	N/A
P16	IN	Domestic	PayTM	N/A	N/A	N/A	N/A
P17	IN	Transnational	Amazon	N/A	N/A	N/A	N/A
P17	IN	Transnational	Facebook	Yes	Multiple	Expiring link via email	< 1 day
P18	IN	Transnational	Facebook	Yes	Multiple	Expiring link via email	< 1 day
P18	IN	Domestic	Flipkart	N/A	N/A	N/A	N/A
P19	IN	Domestic	PayTM	N/A	N/A	N/A	N/A
P19	IN	Domestic	RelianceJio	N/A	N/A	N/A	N/A

Fig. 3. This table shows the different ways that services responded to requests for personal data.
* N/A values mean participants were unable to request data.
** Participants notified the study team of Alaska Airline's response 18 months after their initial request. Data collection was already complete by this time, but we noted this response.

Received January 2022; revised April 2022; accepted August 2022

A.2 Participant Request Table

Participant	Region	Operation	Service	Did participant visit the privacy policy?	Did participant check account or profile settings?	Did participant check the website footer?	Did participant consult FAQs or other help documentation?	Did participant use a search engine?	Did participant consult a 3rd party source (e.g., blogposts, news articles)?	How did participants try to request data?	Was the participant able to locate where to make a request?	Where did participant make their request (if applicable)?	Did the service offer a separate option to "View" data instead of "Download"?
P1	US	Domestic	Venmo	YES	YES	YES	YES	YES	NO	Contact Us form	YES	settings (website only)	NO
P1	US	Transnational	Spotify	NO	YES	YES	NO	NO	NO	Download Data button	YES	privacy settings	NO
P2	US	Transnational	Facebook	NO	YES	NO	NO	NO	NO	Download Data button	YES	settings	YES
P2	US	Domestic	Venmo	YES	YES	NO	YES	YES	NO	Contact Us form	YES	email	YES
P3	US	Transnational	Google	NO	NO	NO	YES	YES	NO	Download Data service	YES	Google Takeout	YES
P3	US	Transnational	Facebook	NO	YES	NO	NO	NO	NO	Download Data button	YES	settings	YES
P4	US	Domestic	AlaskaAir	YES	NO	YES	NO	NO	NO	OneTrust form	YES	privacy notice->form	NO
P4	US	Transnational	Amazon	YES	YES	YES	YES	NO	NO	Download Data button	YES	help documentation	YES
P5	US	Transnational	Amazon	NO	YES	YES	YES	NO	NO	Download Data button	YES	help documentation	YES
P5	US	Transnational	Spotify	NO	YES	YES	YES	NO	NO	Download Data button	YES	privacy settings	NO
P6	US	Transnational	Google	NO	NO	NO	YES	YES	YES	Download Data service	YES	Google Takeout	YES
P6	US	Domestic	AlaskaAir	YES	NO	NO	YES	YES	NO	OneTrust form	YES	privacy notice->form	NO
P7	UK	Domestic	Monzo	NO	NO	YES	YES	YES	NO	Email	YES	help->contact form/email/chat	NO
P7	UK	Domestic	Ryanair	YES	YES	YES	YES	YES	NO	OneTrust form	YES	privacy notice->form	YES
P8	UK	Transnational	Spotify	NO	YES	NO	NO	NO	NO	Download Data button	YES	privacy settings (website only)	NO
P8	UK	Transnational	Google	NO	YES	NO	YES	NO	NO	Download Data service	YES	Google Takeout	YES
P9	UK	Domestic	Ryanair	YES	YES	YES	NO	NO	NO	OneTrust form	YES	privacy notice->form	YES
P9	UK	Transnational	Amazon	NO	YES	YES	YES	YES	NO	Download Data button	YES	help documentation	YES
P10	UK	Transnational	Google	NO	YES	NO	YES	YES	NO	Download Data service	YES	Google Takeout	YES
P10	UK	Transnational	Amazon	NO	YES	NO	YES	YES	NO	Download data button	YES	help documentation	YES
P11	UK	Transnational	Facebook	NO	YES	NO	YES	YES	YES	Download data button	YES	settings	YES
P11	UK	Transnational	Spotify	YES	YES	NO	YES	NO	NO	Download Data button	YES	privacy settings (website only)	NO
P12	UK	Transnational	Facebook	NO	YES	NO	YES	NO	NO	Download Data button	YES	settings	YES
P12	UK	Domestic	Monzo	YES	NO	YES	NO	NO	NO	Chat	YES	privacy notice->contact form/email/chat	NO
P13	IN	Transnational	Amazon	YES				YES		Chat	NO	N/A	YES
P13	IN	Transnational	Spotify	YES	YES		YES	YES		Download Data button	YES	privacy settings	
P14	IN	Transnational	Facebook	YES	YES		YES			Download Data button	YES	settings	
P14	IN	Transnational	Google		YES		YES			Download Data service	YES	Google Takeout	
P15	IN	Domestic	Flipkart	YES	YES					Phone, Email	NO	N/A	YES
P15	IN	Transnational	Google		YES					Download Data service	YES	Google Takeout	YES
P16	IN	Domestic	Flipkart	YES						Email, Phone	NO	N/A	YES
P16	IN	Domestic	PayTM							Chat, Phone	NO	N/A	YES
P17	IN	Transnational	Amazon		YES						NO	N/A	
P17	IN	Transnational	Facebook	YES	YES			YES		Download Data button	YES	settings	
P18	IN	Transnational	Facebook	YES	YES		YES	YES	YES	Download Data button	YES	settings	
P18	IN	Domestic	Flipkart	YES	YES		YES	YES		N/A	NO	N/A	
P19	IN	Domestic	PayTM				YES			Email, Phone	NO	N/A	
P19	IN	Domestic	RelianceJio	YES			YES	YES		Email, Phone	NO	N/A	

Fig. 4. This table documents the actions participants took while requesting data from each service.