Finding, Getting, Understanding

The user journey for the GDPR's right to access

Dominik Pins

Fraunhofer Institute for Applied Information Technology, Human-Centered Engineering and Design, Dominik.Pins@fit.fraunhofer.de

Timo Jakobi

University of Applied Science Bonn-Rhein-Sieg, Institute for Consumer Informatics, timo.jakobi@h-brs.de

Gunnar Stevens

University of Siegen, Information systems and new media, gunnar.stevens@uni-siegen.de

Fatemeh Alizadeh

University of Siegen, Information systems and new media, fatemeh.alizadeh@uni-siegen.de

Jana Krüger

University of Applied Science Bonn-Rhein-Sieg, Institute for Consumer Informatics, jana.krueger@h-brs.de

In both data protection law and research of usable privacy, awareness and control over the collection and use of personal data are understood to be cornerstones of digital sovereignty. For example, the European General Data Protection Regulation (GDPR) provides data subjects with the right to access data collected by organizations but remains unclear on the concrete process design. However, the design of data subject rights is crucial when it comes to the ability of customers to exercise their right and fulfil regulatory aims such as transparency. To learn more about user needs in implementing the right to access as per GDPR, we conducted a two-step study. First, we defined a five-phase user experience journey regarding the right to access: finding, authentication, request, access, and data use. Second, and based on this model, 59 participants exercised their right to access and evaluated the usability of each phase. Drawing on 422 datasets spanning 139 organizations, our results show several interdependencies of process design and user satisfaction. Thereby, our insights inform the community of usable privacy and especially the design of the right to access with a first, yet robust, empirical body.

CCS CONCEPTS •**Security and privacy ~ Human and societal aspects of security and privacy ~ Usability in security and privacy**

Additional Keywords and Phrases: Data literacy, usable privacy, usability, GDPR, right to access, user journey

1 INTRODUCTION

The collection and processing of personal data for commercial purposes is an increasingly widespread phenomenon that is becoming more important for companies associated with both digital and non-digital products across all sectors. New, connected "smart" products that collect and exchange detailed personal data—such as smart fitness trackers [41], smart home systems [16], or connected cars [35]—contribute to this trend.

In both data protection law and research of usable privacy, the awareness of [5,12,19,24,50,59,61] and control over [15,23,37,55,62,63,67,73] the collection and use of personal data are understood to be cornerstones of digital sovereignty. For example, the European General Data Protection Regulation (GDPR)

provides data subjects with the right to access data collected by organizations [20]. In this vein, GDPR provides general guidelines for implementing the right to access, such as adequate levels of authentication, pieces of information to provide, and more generally the provision in "concise, transparent, intelligible and easily accessible form, using clear and plain language" in Art. 12 (1) GDPR [20]. Regarding the design of the process of exercising data subject rights such as the right to access, Art. 12 (2) GDPR [20] highlights that the respective data "controller shall facilitate the exercise." Lawmakers thus generally see controllers as responsible for helping their users exercise their right to access.

However, for companies and organizations in general, this new framework raises uncertainty regarding requirements for compliance with the regulation. For the case of the right to access data, organizations need to implement a process for users to claim their data, but it is still unclear how such a subject access request (SAR) process should be designed, how authentication should work, how data should be requested, provided, presented, and explained to customers in a compliant and customer-friendly way.

Usability studies reflect that process design is crucial for users being able to complete an interaction [4,28]. However, little is known about the factors facilitating or hindering SARs in terms of process design. Insights about factors such as user needs and capabilities are highly useful not only from a research perspective, but also for organizations to use data protection as a competitive advantage by optimizing their customer experience [25,31,78]. While there is some research on usability of dashboard solutions for addressing data subject rights, the market adoption of such generally desirable solutions has been low. Vitale et al. designed a prototypical data dashboard that helps support people in organizing their data from different platforms and devices but is focused more on data curation than on aspects of usability due to its prototypic appearance [74]. Raschke et al. also designed, implemented, and evaluated a privacy dashboard intended to enable and ease the execution of data protection rights [51]. Similarly, Pins et al. found that existing dashboards through which companies allow their customers to look into their interaction data with voice assistants are difficult to use [48]. Regarding manual SARs, Alizadeh et al. looked at user satisfaction with SAR data from loyalty card providers; the sample, however, was rather small, which did not allow for a structured analysis in terms of usability [3]. There is even less research that adopts a processual view for assessing handling SARs. The findings from these studies often highlight that many companies are non-responsive or fail to comply with other legal provisions. For example, Urban et al. looked at the responsive behavior of online ad companies and the data they provide [70]. Kröger et al. [38] ran a longitudinal study on app vendors, looking at response behavior, data sharing practices, authentication, and transmission security.

Our work contributes to the design science research [39,53,71] on the implementation of the GDPR. We carry out an empirical study in two-step to analyse in the third step the implications for design. Our goal is to inform designers about process design decisions that facilitate or hinder usable access as well as transparency about data collection and use in light of user needs and capabilities.

2 DESIGN SCIENCE METHODOLOGY

Design is a learning process in which user research is used to derive empirical grounded design guidelines and recommendation. The major difference between commercial design and academic design research is the former aims to solve a specific problem, whereas the latter aims to provide actionable design knowledge [26,54] and summarizing insights and lessons learned in the form of design guidelines at the end of the paper.

In our study, we adopt the Grounded Design (GD) approach [54,65,77] as a specific variant of Design Science Research (DSR) methodology [39,53,71]. The GD Approach comprises three parts: (1) problem diagnosis to understand the design and use context, (2) action taking to design and using artefacts in an iterative way, and (3) evaluation to understand the appropriation of the design artefact.

The problem diagnosis phase present an essential part of the design research where the design context, the problem, its causes, and consequences will be analyzed [39,72]. The design of digital systems poses a challenge that must be equally technically feasible, economically viable, as well as desirable and usable to the users [45]. These characteristics are not necessarily congruent. Hence, an essential part of the design process is to understand the design context from the user's perspective. For this reason, in this phase the GD approach [54] suggest using explorative user research methods, to understand the user's practices covering user's perception but also her actual observable behavior as well as the socio-material context that shapes this behavior. The problem diagnosis provides insight into what the problem is, but not what the solution should be.

To bridge the gap between descriptive and prescriptive knowledge, the second essential part of DSR involves exploring potential solutions of the problems identified in the diagnosis phase [39,71]. Design concepts and guidelines can be developed in various ways. For instance, in a top-down manner they can be deduced from proven kernel theories [27]. GD provide an alternative way [64,65]. Standing in the tradition of grounded theory, GD argues to develop design theories in a button-up manner, abduct them from the explorative user research. In general, the scope of validity and applicability of such design concept and guidelines are typically smaller than the one of behavioral theories such as the theory of planned behavior [2]. For this reason, Gregor und Hevner [27] describe design guidelines as mid-range, nascent design theory.

In this paper we are focusing on the diagnosis phase. Following the recommendations of the GD approach, we carry out an explorative user research. This user research was operationalized by a two-step study about user needs in the implementation of the right to access as per the GDPR. In the first step (cf. section 4) we define from the GDPR, the literature, and heuristic evaluation, an SAR user experience journey which includes five touchpoints. In a second step (cf. section 5), we asked 60 undergraduates to carry out and document their SAR experience, where each touchpoint was evaluated by a modified system usability scale (SUS). After data cleaning, the sample of our study includes 422 SAR evaluation sheets of 59 undergraduates. The findings shed light on the role that the right to access has for privacy control as perceived by users, report user expectations regarding the design of a usable process of SAR and explore potentials for supporting intelligible and transparent design of the data archives provided.

To bridge the gap between what is and what should be, we discuss these findings regarding the implications for design and abduct design guideline that should give design researcher and practitioners an orientation (cf. section 6). These guidelines highlight the need for connecting HCI research with legislation to assess the effectiveness of measures as demanded in Art.12 GDPR and also Art. 25 (1) GDPR (European Parliament and the Council 2016).

3 STATE OF THE ART ON USABLE PRIVACY AND THE RIGHT TO ACCESS

In this section we outline the concepts and protective aims of the right to access as per the EU GDPR. Second, we survey the research on control and awareness in the realm of usable privacy and how our study contributes to research in the field of GDPR data subject rights.

3.1 Data Subject Rights in the European General Data Privacy Regulation

The GDPR introduces new consumer rights. For example, Art. 20 GDPR introduces a new right to request data for transfer to other organizations—the right to portability. This right is intended to increase – but not exclusively – data protection competition among organizations. It obliges data controllers to make personally identifiable information available to the new service provider in a structured, common, machine-readable format as much possible [75]. Overall, Chapter 3 of the GDPR introduces four rights for data subjects:

1. The right for transparent information provides the data subjects with a concise, transparent, intelligible, and easily accessible form of any communication and information related to processing their personal data using clear and plain language.
2. The right to correction and deletion enables the data subjects to have their incomplete or unwanted personal data completed or erased without undue delay.
3. The right to object to automated decision making allows the data subjects to object at any time to the processing of personal data concerning him or her, especially where personal data are processed for direct marketing purposes, including profiling.
4. The right to access personal data grants data subjects the right to claim personally identifiable information collected by an organization in a precise, transparent, understandable, and easily accessible form in clear and simple language.

For each case, many terms are open for interpretation when it comes to implementation. Currently, there is still much to be done to reduce uncertainty about which measures will be judged as sufficiently compliant with concepts such as “understandable,” “transparent,” and “accessible” in court—all the more so as these terms partly overlap, depending on the content [75]. While legislation has been in place for some time, it remains unclear how to design SAR processes that comply with both the GDPR and users’ demand for privacy, for example, in terms of facilitation by the controller and transparency of data collection and use.

According to Article 12 of the GDPR, the controller shall provide information about actions taken regarding the SAR without undue delay and within one month of receipt of the request. Moreover, Article 15 of the GDPR requires that the process of claiming data will result in information being provided “in a concise, transparent, intelligible and easily accessible form” [20].

Criteria for easy access could largely be subsumed under usability criteria and the way to design the process of claiming data. Previous studies show that consumers often do not fully know what data is collected about them, and this in turn may lead to anxiety and concern [44]. Accordingly, consumer trust may be increased by designing products and services with transparency in mind [43]. Creating transparency and intelligibility, however, imposes rather complex and abstract requirements [10,33], especially given that non-programming users often struggle to understand information flow on the internet in general [36].

3.2 Control, Awareness, and the Right to Access in Usable Privacy

Concerning usable privacy research within the HCI community, control [15,23,37,55,62,63,67,73] and awareness [5,12,19,24,50,59,61] about data collection and use are cornerstones of its research; this research often takes a user-centered perspective to design usable solutions.

Traditional fields of research include improving privacy policies [52,58] and password usability [1,7,60]. Particularly with increasing amounts of data being collected and used, studies have also started to investigate the design of supporting users in their struggle to make sense of data [68] by means of adequate and flexible

visualizations [30,51]. For example, Angulo et al. built a data tracking tool that displayed an overview of a user's data disclosures to different online service providers and provided them with the collected data about them [6]. Similarly, awareness and control of data disclosure on the mobile phone interface is a widely explored context [8,18,22]. In this regard, Bentzing et al. compared application designs to investigate how increasing transparency can influence users' privacy-related behavior on mobile phones [11]. Felt et al. [21] conducted an extensive survey of the diverse strengths and weaknesses of design alternatives. Similar studies targeting the support of data and privacy awareness have been conducted in the area of smart home data [33,68] and smart metering [32].

However, relatively few studies so far have explicitly targeted the implementation of provisions and data subject rights provided by the GDPR. These are especially interesting, as the GDPR itself provides that its own implementation should be designed in an "effective" manner [20], which arguably the community of usable privacy researchers is predestined to investigate, given that well-researched yet abstract concepts of awareness and control are also deeply enshrined in GDPR. For example, a usable implementation of the principle of purpose limitation has been investigated using the example of data processing of voice assistants [34]. Moreover, the data subject rights in GDPR are currently being researched, such as the right to data portability [17,76]. Closely related to the right to access data, transparency-enhancing tools have been proposed—mostly following a dashboard approach. For example, similar to the usable privacy dashboard by Raschke et al. [51] mentioned above, Olausson developed a dashboard specifically targeting nurses' work [47]. Tolsdorf et al. [69] qualitatively compared ten implementations of dashboards, comparing their levels of compliance. Still, dashboard implementations are scarce in practice and are often only adopted by big players on the market. Looking at manual SARs, Alizadeh et al. interviewed customers of German loyalty card systems, who were asked to make use of their right to access [3]. The scope of the study, however, is limited to a single organization, focusing on how data is provided and the potential to help users with their privacy practices. With a similar perspective on supporting sense-making and data literacy, Pins et al. [49] designed and tested a prototype that visualizes the interaction with voice assistants based on data of SARs from Amazon Alexa and Google Assistant.

3.3 Research Gap: Effectiveness of the Right to Access from a User Perspective

In general, investigations into the provisions of the GDPR from a consumer perspective in the field of usable privacy are scarce. Existing studies that adopt a consumer perspective on the right to access largely ignore the challenge of getting data in the first place. Instead, studies focus either on the compatibility of the data provided as a result of the SAR with user demands in terms of supporting privacy practices and data literacy [3,49]—or they take the provision of data for granted by building dashboards on top of the data [51]. The study closest to our approach, Urban et al. [70], contacted 39 companies to check for several SAR parameters such as response time, reaction to questions, and the disclosed information in the context of online advertisement. Similar work has been done by Kröger et al. [38] for app vendors who conducted a longitudinal study on several SAR items such as response time, data provided, and security mechanisms. However, these studies do not apply a processual lens, nor do they evaluate the phases they identify in a user-centered way; instead, these studies merely check against legal provisions. To the best of our knowledge, there are no large-scale studies on usability assessments of the implementation of SAR processes. Moreover, there is currently no structured approach for evaluating the design factors that drive or hinder users when conducting SARs. This information,

however, is relevant both from the standpoint of research on usable privacy as well as for assessing the controller's role in facilitating the user in this process, as demanded in the GDPR.

4 GETTING THE DATA: THE USER EXPERIENCE JOURNEY

4.1 Procedure

To learn more about user needs in the implementation of processes for SARs, we started with a problem diagnosis according to the DSR by analyzing the implicit and explicit provisions of the GDPR. For instance, the GDPR defines the right to request an electronic copy of data by users, the need for authentication, as well as the expectation that the data should be provided within 30 days. While such specifications shape the user experience, they do not guide the implementation of an SAR process. For these reasons, three researchers individually carried out 10 SARs from various organizations to understand the procedure and current design practices so as to facilitate the problem diagnosis [71]. During the process each researcher documents his/her insight and creates a protocol of the most important steps. In a second step, we conducted a group discussion in which we compared the results. During this discussion, consensus was reached on what would be the major touchpoints that constitute the SAR User Experience Journey (see Appendix D, Table 15). In sum, the identified touchpoints were experienced and documented in a similar way, deviations and inconsistencies in the procedure and naming of the touchpoints could be cleared up within the discussion. The only open point related to when the user should authenticate because there are several possible options in this regard.

As this definition is based on the expert opinion of the researchers, our definition of the User Experience Journey is related epistemologically to the heuristic evaluation methodology, which involves a small set of evaluators to examine the user experience [46].

4.2 The User Experience Journey

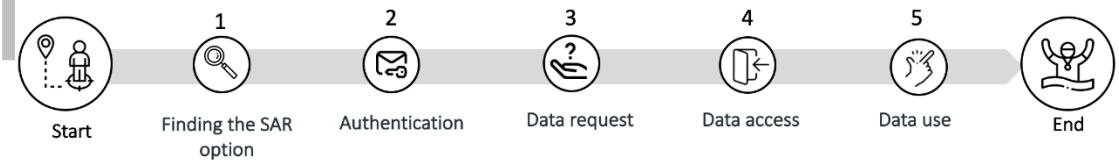

Figure 1: The five major touchpoints in the user experience journey of a subject access request process

As a result of our study, we identified five major touchpoints in the process of conducting an SAR as illustrated in Figure 1:

Finding the SAR option: One key challenge to kick off the process is finding the entry point to conduct an SAR. Arguably, this challenge is strongly influenced by where the user finds the needed information and what communication channels must be used to get that information, e.g., a website, phone, or e-mail. In addition to the challenge of finding the right place, the user experience can suffer if the information is not prominently presented, if it is described in a complex way, or if it uses terms unknown to the user such as "data controller" or "data subject."

Authentication: Users need to provide authorization for the SAR to avoid misuse. This phase is directly informed by the GDPR. Often this task can be managed via an existing account with the service in question. In cases in which, for example, name-password combinations do not pre-exist, other methods must be used such as two-factor authentication; moreover, in non-digital environments (e.g., doctor or authorities), the authentication can be inconvenient, for example, requiring one's passport or other documents for identification.

Data Request: For completing the SAR, the user must express her demand. At first glance, this seems trivial. Yet, this touchpoint will be more complex in practice. Therefore, this touchpoint focuses on the different steps to complete the SAR such as how it is implemented and supported by the organization. Further, user experience will be influenced by the provided feedback such as whether the request was received or what further steps will follow.

Data Access: Providing users with access to the data includes several aspects to consider. For example, once collected and set up, there is the question of how organizations inform users about the availability of the data (post, e-mail, app message, etc.). Moreover, there are different ways of providing data. For example, compressed folders can be downloaded or are directly sent via mail or e-mail. These variations come with different challenges such as requesting large files or managing access control to the files. Additionally, there is no prescribed file format on the part of the GDPR in Art. 15 specifying what this copy of data should look like.

Data Use: Finally, after having received the data, utility of data for users must be subject to special scrutiny because it addresses the regulatory aim of providing users with transparency regarding data collection and use by data controllers. It is this phase that ultimately decides the process's main goal: informing users about potentially complex data.

The touchpoints have some logical order; for example, a data request must be carried out before the data can be provided and accessed. However, the steps do not fully determine the chronological order of the process, but alternative paths are possible. Authentication, for example, may have been completed by logging into a services homepage before finding the option to exercise the right to access.

5 IMPLEMENTATION OF THE USER EXPERIENCE JOURNEY STUDY

In a second step, we conducted a user study, in which we asked the participants to exercise SARs and document their experience regarding the identified user experience journey touchpoints. This step further completed our problem diagnosis by exploring and uncovering usage and design issues in current data request solutions to better understand the problem space and identify user needs. In the following, we provide details about our participants, the adaption of the SUS to measure and document the user experience, the procedure, and ethical considerations in the process.

5.1 Method

5.1.1 Participants

Our participants were HCI and Information Systems master students in a course in Germany that presents lectures about the basics of usable privacy using the organizational implementation GDPR as a case study. The course was conducted between October 2020 and January 2021. We asked the students to exercise their right to access with up to 10 organizations and assess the user experience of the whole process. We give this task quite early (in the second session of the course) to make sure participants would be able to complete the

SAR process and thus their documentation. Because organizations have up to three months to fulfill data subject requests, we asked the students to fill out the touchpoints-related SUS and questions as soon as they completed it. Beyond that, students had two more months (end of March) to submit their fulfilled questionnaires. We did not impose any restrictions on the selection of organizations (see the following ethical considerations), but we did ask them to identify organizations that were potentially interesting to them personally.
Overall, we collected a total of 454 SAR evaluations by 60 participants (min = 1, max = 11, avg. contributed = 7.55). After data cleaning and data preparation (see Section 5.1.5), we included 422 evaluations by 59 participants in the analysis.

5.1.2 Ethical Considerations

Working with the right to access may bring researchers into contact with sensitive information. We took several steps to safeguard the privacy of the participants. Since we were only interested in the process design and its user experience, we refrained for privacy reasons from looking into the data. Instead, we asked for file types and information categories provided, such as “name” and “address” in a “yes/no” manner. To further strengthen the privacy of our participants, we let them freely choose which organizations to contact and exclude any services they were uncomfortable disclosing their relationship with.

While there was a guideline of aiming for 10 SAR evaluations, we did not treat this as a criterion for course completion and it did not have any influence on the grades if less (or more) than 10 documentation sheets were provided. Since the documentary sheet was also important for their overall coursework, it was in the students’ best interest to have these sheets filled in order to serve as a basis for the creation phase when improving the SAR process user experience.

An ethics application was submitted to the Ethics Board of the University of Siegen. From their point of view, no vote is required for this study.

5.1.3 Measuring the User Experience

For measuring the user experience of the user journey, we provided an Excel sheet that featured a modified SUS questionnaire for each for the first four phases and one for evaluating the overall experience. SUS is a “de facto standard” [14] for measuring the usability of systems according to ISO norm 9241-11 [29] and was created by John Brooke [13]. The result of the questionnaire is a score between 0 and 100, where a score of 68 is commonly seen as average and, consequently, values above 68 represent good usability [56]. Its generalizability of use is supported by the fact that SUS has already been used in numerous studies [13,57]. We further decided to use the SUS as it allows comparison of different products and services due to the generic wording of the statements—even with small sample sizes [56]. This was important as it was not clear what different kinds of solutions or ways of requesting the data exist, how many different organizations were contacted, and by how many participants.

For our study—and to evaluate the process of requesting—we made some modifications of the questionnaire statements, as can be seen in Appendix A. For instance, we excluded the question about the potential frequency of use as the GDPR’s data subject rights are limited by law against unreasonably frequent use with organizations being allowed to reject or charge a reasonable fee for requests (Article 15 (3) GDPR, [20]).

5.1.4 Questionnaire Design

For an initial overview, we asked the participants to document some basic information about the organization (e.g., the name and type), the data provided (such as the volume), the time needed to conduct the SAR, the purposes stated for use of data, and the provided file formats.

We used SUS to rate the overall user experience as well as the touchpoints "finding," "authentication," "request," and "access." To allow comparison, the questions were presented as a 5-point Likert scale, with answers ranging from "strongly disagree" to "strongly agree". For each touchpoint, we ask four additional open questions to document details about shortcomings, personal perception, feelings, and ideas for improvements. To reflect the regulatory goal of transparency, we also include the post-SAR activities defined as the "data use" touchpoint. Our aim was to evaluate the extent to which getting and using the data would contribute to users' privacy practices and might ease potentially existing concerns about the data use of organizations. While privacy is a complex and multifaceted issue, for practical reasons we limit this section to the following three topics:

1. The first topic was about the contribution of the SAR process to the perceived data protection compliance of the organization. We see it as a minimum requirement that users perceive the overall process to at least pose a contribution to organizational compliance with the legal GDPR requirements.
2. The second topic was about the data provided contributing to improve the data literacy of the user. With the help of the SAR, the user should be able to examine whether the data collected match her expectations or lead to new insights.
3. Finally, the third topic covered behavioral changes. To find out if SAR could be a tool for privacy awareness and making potentially extensive data collection by organizations more transparent, we asked if the data provided had an impact on the users' attitudes and privacy behavior.

5.1.5 Data Cleaning and Preparation

In total, we collected 454 SAR evaluations by 60 participants (min = 1, max = 11, avg. contributed = 7.55). Of these evaluations, 363 provided by 29 participants (min = 1, max = 10, avg. = 7.62 contributed) were fully complete. For most incomplete sheets, the comment section revealed that data never was received. In other cases, participants left out sections where they were unable to rate the process step in question because they did not go through a process (e.g., when they were identified in person and not via a digital touchpoint or interaction). Given the exploratory character of our study, where such reasoning was provided, we included partially complete sheets for analysis in the sample, resulting overall in 422 SAR evaluations, conducted by a total of 59 students.

5.2 Findings

In the following, we provide an overview of our adjusted sample and insights into the evaluation of the different touchpoints, focusing on the SUS score.

5.2.1 Sample Characteristics

Our sample includes 422 SARs from 139 organizations. For a better overview, we clustered these into 11 organization types based on the participants' responses. Table 1 provides an overview about what

organizations belong to what category. For reasons of space, further individual cases are partly indicated by "...".

Table 1: Overview about the sample

Category	SAR	Organizations		Volume [MB]	
	n	n	name (SAR count)	Mdn	M
Total	**422**	**139**		**1.0**	**1178.9**
Social Media / Messenger	153	16	Facebook (41), LinkedIn (28), Instagram (23), WhatsApp (16), XING (8), Snapchat (7), Skype (6), Telegram (5), Dribble (4), Pinterest (4), Slack (3), Twitter (3), Steam (2), Discord (1), …	1.0	227.1
Multi Service Provider	74	3	Amazon (39), Google (29), Apple (6)	28.0	4708.1
E-Commerce	41	20	eBay (11), Zalando (9), Etsy (2), Lieferando (2), Rebuy (2), About you (1), Adidas (1), Alles Rahmen (1), Daraz.pk (1), DefShop (1), dm drugstore (1), Goodreads (1), …	0.2	1.1
Finance/ Payment	35	20	PayPal (12), Sparkasse (5), Access Bank PLC (1), Allied Bank Limited (1), BBVA Bancomer (1), Binance (1), Deutsche Kreditbank Berlin (1), FBNQuest (1), …	0.7	2.4
Streaming / Entertainment	21	4	Spotify (11), Netflix (8), Blizzard Entertainment (1), Electronic Arts (1)	6.8	646.7
Travel / Tourism	20	13	DB (4), Airbnb (3), Booking.com (3), ADAC (1), Autoscout24 (1), Deutsche Lufthansa AG (1), Flixbus (1), MVV GmbH (1), Nextbike (1), Primera Plus (1), TIER (1), Uber (1), Waze (1)	0.2	0.2
Health / Insurance	17	16	Techniker Krankenkasse (2), AOK Bayern (1), AXA (1), Clue (1), DRK (NORD) (1), Gesetzliche Krankenkasse (1), Family Doctor (1), mhplus (1), Nürnberger (1), …	0.3	1.1
Edu/Learn	17	12	Duolingo (4), University (3), CodeCademy (1), Coursera (1), DataCamp (1), FH Aachen (1), Gymnasium Waldkraiburg (1), Italki (1), Wilhelm von Oranien Schule (1), …	0.2	0.4
Cloud / Mobile Provider	15	11	Yahoo (4), Samsung (2), 1&1 (1), Dropbox (1), idrive (1), Lidl Connect (1), Mega (1), Personal Paraguay (1), Tigo Paraguay (1), Unitymedia (1), Vultr (1)	2.5	55.7
Job / Work	12	7	GitHub (4), Trello (3), Amila (1), Fiverr (1), Indeed (1), Miro (1), OCC Mundial (1)	0.3	33.8
Miscellaneous	17	17	Deutscher Evangelischer Kirchentag (1), Landessportbund NRW (1), McFit (1), openPetition (1), Philippine Embassy, Städtische Bühnen (1), Stadtverwaltung (1), …	0.9	0.9

Reflecting the special position of Google, Amazon, and Apple as providers of multiple and different services and products, we placed them in their own category of "Multi Service Providers."[1] Although Facebook, Instagram, and WhatsApp, are part of the same organization, each of these services implements an independent SAR process, so we have treated them independently.

Most of the SARs targeted social media and messenger organizations, with Facebook as the most frequently contacted organization. Although we gave participants the freedom to choose the organizations for the SAR themselves, the table shows that primarily organizations for online services such as Amazon, eBay, PayPal, Duolingo, or Yahoo were selected. Nevertheless, there also were local organizations in our sample, such as family doctors, universities, banks, or sport and fitness clubs.

[1] Interestingly, no one requested his/her data from another international multi service provider, Microsoft.

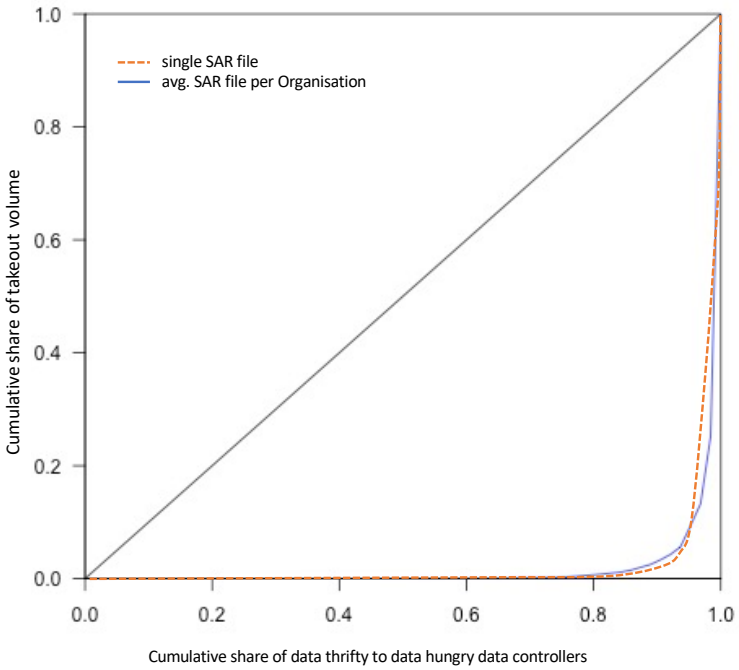

Figure 2: Lorenz Curve showing the high irregular data volume distribution of the sample

Our sample reflects the so-called *data-opolies* [66] as the data volume provided via the SAR is rather small and there are only a few exceptions with large file size. To quantify this phenomenon, we calculate the Lorenz curve and the Gini index [42] as a measure for the concentration of the data-driven economy. Figure 2 shows that the Lorenz curves calculated by the single SAR files and by the average SAR file per organization are quite similar. The same holds for the corresponding Data Gini index (DGI), which is 0.95 in the case of the single SAR files and 0.96 in the case of the average SAR file for the organizations. As this value is near 1.0, it expresses the high irregular distribution of the data volume. For this reason, in Table 1 we provide both statistical average values, the mean, and the more robust median. The mean data volume in our sample was M = 1.178 GB; however, the median is just Mdn = 1 MB. This means that in 50% of the cases, the data provided was 1 MB or less in size.

5.2.2 Overall Experience

Table 2: SUS evaluation for the individual touchpoints presented for the different organizational categories

Category	**Overall**			**Finding**		**Authentic.**		**Request**		**Access**	
	grade	M	SD	M	SD	M	SD	M	SD	M	SD
Total	**F**	**55**	**14**	**45**	**12**	**56**	**11**	**53**	**12**	**51**	11
Streaming / Entertainment	**D**	65	8	49	8	59	6	56	5	49	6
Social Media/Messenger	**F**	59	10	50	10	57	8	56	8	51	11
Education / Learning	**F**	58	11	46	10	54	9	49	9	53	12

Multi Service Provider	**F**	55	11	44	11	56	10	54	10	47	11
E-Commerce	**F**	54	14	45	12	56	13	54	12	55	11
Cloud / Mobile Provider	**F**	53	15	45	9	57	10	53	10	55	10
Travel / Tourism	**F**	49	13	38	12	52	12	48	11	48	15
Health / Insurance	**F**	46	19	39	18	53	13	47	16	45	16
Job / Work	**F**	45	20	40	16	48	18	41	17	49	12
Finance / Payment	**F**	42	21	38	15	53	14	45	18	52	10
Miscellaneous	**F**	52	13	38	11	51	17	48	18	56	9

Note: We applied Bangor et al.'s grading scheme [9], ranging from A to F, where F is the lowest category with an SUS score below 60.

The rating of the overall experience is low across all categories (see Table 2). Applying the grading scheme of Bangor et al. [9], all categories are graded F, which equals "not acceptable". Only Streaming /Entertainment makes it to a D with a scoring of 65. Notably, "finding" the option to start the SAR scores lowest in all sections and has also been rated unacceptable for the best scoring categories. Surprisingly, the overall rating is higher than every single touchpoint except for authentication, which is higher.

The two top-rated categories are providing digital services only (Streaming/Entertainment, Social Media/Messenger), while the lowest-scoring categories are likely to contain highly sensitive data, as defined by GDPR (Health/Insurance, Job/Work, Finance/Payment). The one-way analysis of variance ($F(10, 390) = 8.42$; $p < .001$) showed that the organizational category has a significant impact on the SUS rating. The additional conducted Kruskal-Wallis rank sum test (Chi square = 940.13, $p < .001$, df = 2) showed the same result.[2] As illustrated in Figure 3, Two-Sided t-test (Bonferroni adjusted) showed a significant difference between the SUS score of the Streaming/Entertainment sector and the E-Commerce sector ($t(54.9) = 3.73$; $p < .001$; p.adjusted = .025), the Multi Service Provider sector ($t(41.8) = 4.74$; $p < .001$; p.adjusted = .001), the Travel/Tourism sector ($t(27.3) = 4.41$; $p < .001$; p.adjusted = .008), and the Finance/Payment sector ($t(41.7) = 5.57$; $p < .001$; p.adjusted < .001). There is also a significant difference between the SUS score of the Social Media/Messenger and the Finance/Payment sector pair ($t(32.7) = 4.56$, $p < .001$; p.adjusted = .004).

[2] The Shapiro-Wilk test was performed for the residuals of ANOVA and showed that the distribution of the residuals departed significantly from normality ($W = 0.951$, $p < .0001$). The visual analysis (Appendix C, Figure 9), shows that the distribution is roughly "bell-shaped," and the deviation mainly concerns the outer value range. Because of the results of the Shapiro-Wilk test, however, we perform an additional Kruskal-Wallis rank sum test.

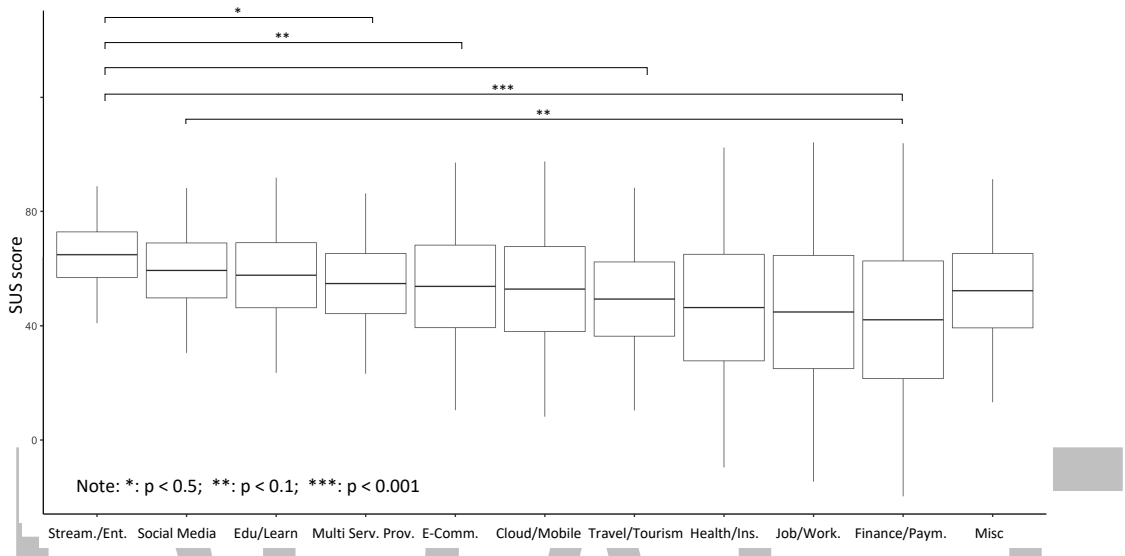

Figure 3 The overall experience SUS scores for the different organizational categories

To add to the understanding of the overall rating, we explored the comment section. Comments stemming from negative SUS ratings were mostly highlighting the negative experience of a specific touchpoint. For instance, the participant of SAR 313 comments that in the case of Amazon, it was the most complicated task of all to find out how to initiate the SAR. Other participants noted that they liked to have more control and that more options should be available to tailor the SAR according to individual needs. Others offered surprising insights into the SAR data and mentioned specific design decisions they liked throughout the SAR process, such as that the process was easy and informative because the touchpoint could be found in the settings. Other participants were torn; on the one hand, they praised the little time needed to get through the SAR process while, on the other hand, they missed information on their GDPR rights or the purpose of extensive data collection.

5.2.3 Finding

Of all touchpoints, the finding touchpoint scored lowest with a SUS of 45 in average (sd = 12; Table 2). The organizations of Finance/Payment had the worst SUS value (avg = 38, sd = 15), while Social Media/Messenger services performed best. But even here, the SUS value of avg = 50 (sd = 10) is still poor and falls into the lowest SUS category F.

A major reason for the SUS score seems to be that it takes a long time to find the place to apply for the SAR (see Table 3). On average, the finding process took between a couple of minutes up to several hours (Mdn = 5 min, M = 3h 09 min). We perform a linear regression ($F(1, 352) = 53.03$, $p < .0001$, $R^2 = .131$, Adjusted $R^2 = .0128$).[3] The results (see Table 4) show that the SUS score is significantly related to the logarithmized time

[3] The visual analysis (Appendix B,

needed to search and find the touchpoint, which itself essentially shows a standard distribution. The regression analysis can be interpreted in such a way that when search time doubles, the SUS score deteriorates by about 3 points. However, the low R^2 indicates that the search time explains only a small part of the variance, such that other factors also come into play for a good user experience.

Table 3 Time needed for the overall experience and at each touchpoint

Touchpoint	n	Mdn	M	min	max
Overall Experience	344	7 min	42h 16 min	< 1 min	5 weeks
Finding	354	5 min	3h 09 min	< 1 min	5 weeks
Authentication	237	1 min	1h 28 min	< 1 min	2 weeks
Data Request	303	2 min	2h 43 min	< 1 min	3 weeks
Data Access	257	3 min	18h 59 min	< 1 min	4 weeks

Note: Some participants gave up during the process or never received the data. We excluded these cases from the calculation

Table 4: Regression results using the finding duration as the criterion of the SUS score

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	r	Fit
(Intercept)	37.81**	[35.34, 40.28]						
Finding Duration (log scaled)	-3.14**	[-3.99, -2.30]	-0.36	[-0.46, -0.26]	.13	[.07, .20]	-.36**	
								R^2 = .131** 95% CI[.07,.20] *Adjusted* R^2 = .128

Note. A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. beta indicates the standardized regression weights. sr2 represents the semi-partial correlation squared. r represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

The challenge of identifying the entrance point is also well reflected in the comments, revealing general search patterns. For example, participants reported searching in their account settings or the privacy policy page on both the website and app of the respective organization. Moreover, participants used search engines both right away and as a secondary option when they were unsuccessful in browsing the website and app. Sometimes, participants even ended up contacting customer service, such as chat or hotline. Across all types of organizations, participants missed clear instructions and/or contact persons for their issue. While personal support is costly for organizations, participants often expected to be able to request their data online automatically to avoid time consuming support cues like simple download options or a self-explanatory

Table 6) shows that the distribution is roughly “bell-shaped.” As the Shapiro-Wilk test for the residuals of regression (W = 0.974, $p < .0001$) indicates a deviation from normality, we also perform an additional linear regression with bootstrapping (with R = 5000 replications). This leads to similar results (Intercept-95% CI [LL, UL] = [34.632, 41.002]; Finding Duration (log scaled)-95% CI [LL, UL] = [-4.189, -2.054])

navigation. Unfortunately, some organizations offered the opposite and were even inconsistent within their own architecture—for example, by linking to non-helpful support pages.
Overall, both positive and negative comments reveal the search strategies adopted and show that the variety of solutions across organizations resulted in inconsistencies and a lack of guidance for users. Where placement of the SAR initialization was similar to previously experienced cases, participants memorized the position and found it easier. SAR 259 therefore highlights the potential benefits of standardization for the finding touchpoint:

"The layout for designing this process was perfectly allied with social media platforms, and I didn't need to think a lot to reach the result I was looking for."
- SAR 259 Google

5.2.4 Authentication

Authentication is necessary to ensure that only legitimate users can request data. On average, the authentication process took between a minute up to an hour (Mdn = 1 min, M = 1h 28 min). The evaluation sheets show that various means for authentication are used (see Table 5): Most frequently the participants authenticated themselves via their user login (63.4%), their e-mail (15.6%) or other digital verification such as Google or Facebook ID (5.4%). In only 3.8% of the cases, users needed physical proof of identification (e.g., showing a passport, doing a Postident verification, or scanning and sending authentication documents via mail). The evaluation sheets show that in 11.8% of cases, other means for authentication are requested by the organizations, which the participants did not specify.

Table 5 The different means of Authentication and their SUS Score

Authentication Means	**Count**		**SUS score**	
	n	%	M	SD
Total	391	100	56	11
User login	248	63.4	58	8
E-mail	61	15.6	53	13
Google ID, Facebook ID or similar	21	5.4	57	8
Passport, Postident or similar	15	3.8	44	10
Other	46	11.8	54	17

The authentication touchpoint has the best SUS score of all touchpoints (see Table 2). The total score of 56 (sd = 11), however, still places it in the lowest category F. Table 5 and Figure 4 shows that authentication via user login or another digital proof of identity reached the highest scores (score = 57, score = 58, respectively), followed by e-mail (score = 53). "Passport, Postident or similar" performed worst (score = 44).

The one-way analysis of variance ($F(4, 361) = 8.58$; $p < .001$) showed that the means of authentication has a significant impact on the SUS rating. The additional conducted Kruskal-Wallis rank sum test (Chi square =

1248.1, $p < .0001$, df = 3) showed the same result.[4] Two-Sided T-test (Bonferroni adjusted) showed a significant difference between the Passport, Postident or similar, and User login pair (t(13) = -5.26; $p < .001$; p.adjusted = .002), as well as using Google ID, Facebook ID or similar (t(22) = -4.29; $p < .001$; p.adjusted = .003), but not between the other pairs. This indicates that digital authentication channels are preferred, while the kind of digital channel that will be used is less important.

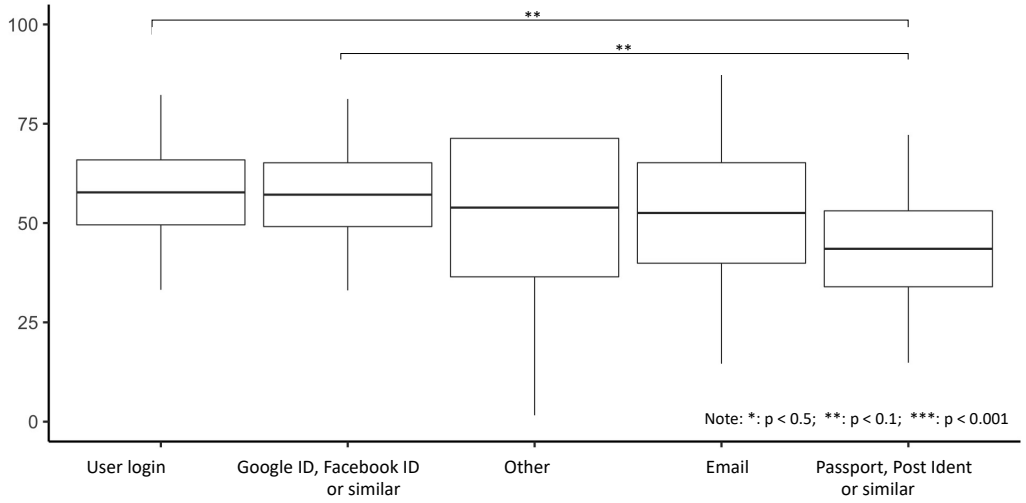

Figure 4 The SUS scores of the authentication touchpoint for the different authentication means

The comment section provides additional insights. In general, the comments either criticized the fact that there was no additional authentication in place or that too much was asked to provide. While authentication by passport was perceived to be secure, having to jump back and forth between devices to scan and provide the ID card produced difficulties, was time consuming, and was thus unsatisfactory. In other cases, such as with market platforms like eBay, the low level of authentication needed, however, raised concerns.

In their comments, participants pointed out that they felt more secure when the authentication procedure includes different modes or devices. A suggestion was to make two-factor authentication mandatory for the process. Finally, participants mentioned that they lacked proper information altogether about what ways of authentication existed in advance. On the contrary, most comments that conveyed positive feelings described a functional download process that used only a simple log-in without any further devices.

Notably, participants were positive when the authentication could be made online via means already known to users such as the user login. However, this expectation was somewhat taken for granted, which means it resulted in neutral (not positive) ratings.

[4] We perform an additional Kruskal-Wallis test because the Shapiro-Wilk test indicates that the ANOVA residuals departed significantly from normality (W = 0.951, $p < .0001$). See also: Appendix C, Figure 10.

5.2.5 Data Request

The data request touchpoint describes the activities of the user to complete the request for the SAR. With a SUS Score of 53, the data request receives the grade F and is at the midpoint of the different touchpoints (see Table 2).

Table 6: Regression results using the data request duration (log scaled) as the criterion of the SUS score

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	r	Fit
(Intercept)	61.41**	[58.94, 63.89]						
Data request duration (log scaled	-2.67**	[-3.45, -1.89]	-0.36	[-0.47, -0.26]	.13	[.07, .20]	-.36**	
								R^2 = .131** 95% CI[.07, .20] *Adjusted* R^2 = .128

Note. A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. beta indicates the standardized regression weights. sr2 represents the semi-partial correlation squared. r represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

On average, the data request touchpoint took between a few minutes up to several hours (Mdn = 2 min, M = 2h 43). We perform a linear regression ($F(1, 299) = 45.28$, $p < .0001$, $R^2 = .131$, Adjusted $R^2 = .0128$).[5] The results indicate that there is a significant impact of the needed time on the SUS rating (see

[5] The visual analysis (Appendix B, Figure 7) shows that the distribution is roughly "bell-shaped." Because the Shapiro-Wilk test for the residuals of regression ($W = 0.957$, $p < .0001$) indicates a deviation from normality, we also performed an additional linear regression with bootstrapping (R = 5000 replications). This leads to similar results (Intercept-95% CI [LL, UL] = [59.034, 63.548]; Data Request Duration (log scaled)-95% CI [LL, UL] = [-3.476, -1.871])

Table 6). However, the low R^2 indicates that the duration explains only a small part of the variance. This means that other factors besides time also play a role for a good user experience.
Besides the largely automated procedures, writing an e-mail for SAR was perceived as a simple way for a data request; however, some participants also expressed concerns about making an SAR request via e-mail for various reasons. In the case of PayPal, for instance, the participant's comment made in SAR 330 identifies security concerns because PayPal, as a financial service, holds sensitive data. Another problem when making an informal request via e-mail is that the participant receives no guidance about how to formulate the e-mail and what the next steps will be after sending them. These unguided processes led to confusion. Especially in this step, there were some comments about not receiving answers or data from organizations.

5.2.6 Data Access

The data access touchpoint is when the user waits for the data to be provided by the organization and the subsequent access to that data. On average, the access process took between a couple of minutes up to several hours (Mdn = 3 min, M = 18h 59 min). The access touchpoint SUS score of 51 places it in the lower midfield of all touchpoint scores.

Table 7: SAR data formats used by organizations

Data Format	**Count**		**Data Volume [MB]**		**SUS score**	
	n	%	Mdn	M	M	SD
Total	372	100	1.0	1178.9	51	11
Human-Readable	79	21.2	0.8	333.3	55	10
PDF	7	1.9	0.4	0.5	66	6
HTML	72	19.4	1.0	369.7	54	10
Machine-Readable	150	40.3	1	535.2	49	10
CSV/XLS	52	14.0	0.2	279.7	51	10
JSON/XML	98	26.3	1.3	670.0	49	10
Paper slips	20	5.4	7 (Pages)	20 (Pages)	54	12
Other	123	33.1	3.0	2883.9	49	12

Our study reveals that various data formats exist in practice (see Table 7). Most frequently, machine-readable data formats are used, such as JSON/XML (26.3%) or CSV/XLS (14.3%). Less common are human-readable data formats such as PDF or HTML. However, about one third (33.1%) of the SARs still used none of these formats. To our surprise, 5.1% of all SAR were answered with paper printouts. On average, in the case of human-readable SAR data formats, the SUS score is higher (score = 55) compared to the machine-readable ones (score = 40.3). The one-way analysis of variance ($F(3, 353) = 6.45$; $p < 0.001$) showed that the data format has a significant impact on the SUS rating. The additional conducted Kruskal-Wallis rank sum test (Chi square = 543.11, $p < .0001$, df = 1) showed the same result.[6] As illustrated in Figure 5, two-sided T-test (Bonferroni adjusted) showed a significant difference between the human-readable versus machine-readable pair (t(157) =

[6] We perform an additional Kruskal-Wallis test because the Shapiro-Wilk test indicates that the ANOVA residuals departed significantly from normality (W = 0.973, $p < .0001$). See also: Appendix C, Figure 11.

4.01; $p < .001$, p.adjusted < 0.001) as well human-readable versus other ((t(185)= 3.85; $p < .001$, p.adjusted < .001), but not between the other pairs.

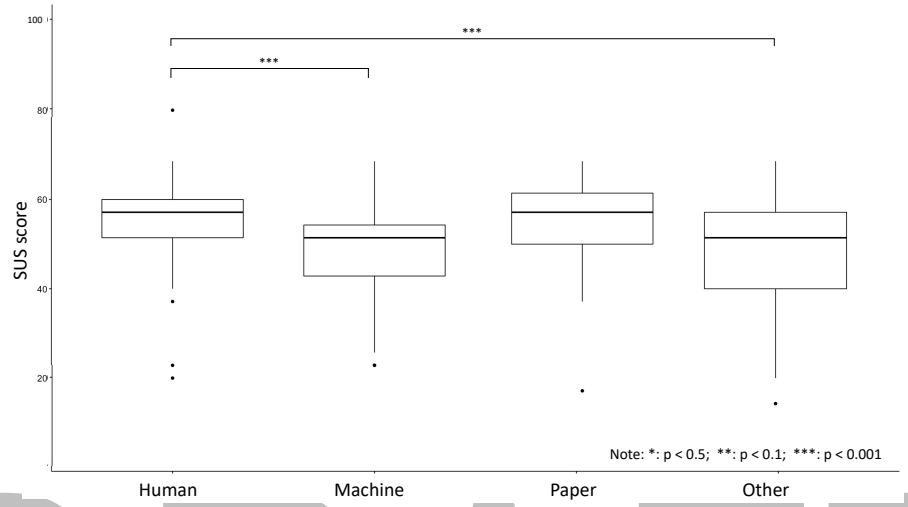

Figure 5 The SUS score for the human-readable data formats (PDF, HTML) is higher than machine-readable one (CSV/XLS, JSON/XML)

In their comments, participants expressed their negative impressions, describing the implementation as "catastrophic" and the access possibilities as cumbersome (e.g., "Simplifying the workflow of data request and receiving the data"). Readability, representation, and data structure was also an issue: "Even as a 'non-layman' it was hardly possible to understand the data".

In general, answers referring to a neutral SUS contained neutral descriptions but also positive reactions to the access points. Participants were satisfied with the access to the data because the representation of the data was simple and reasonable.

5.2.7 Data Use

The data use touchpoint covers the post-SAR activities that include anything the data subject found the data provided to be useful for. We were especially interested in evaluating the perceived contribution of the provided data to the perceived transparency of data collection and use, GDPR compliance, data literacy, and impact on users' privacy behavior.

Perceived contribution to organizational GDPR compliance

The GDPR provides that users of a service collecting personal data shall have a right to access collected data. Hence, we would expect that the overall SAR process contributes to perceived organizational compliance with the legal GDPR requirements. Table 8 shows that, on average, participants agree with this statement (M = 0.68).

Table 8: Contribution of the SAR to the perceived compliance

The SAR contributes to …	**n**	**M**	**SD**
… the GDPR compliance	366	0.68	0.75

The SAR contributes to …	n	M	SD
… the right to object	358	0.51	0.76
… the right of access and data transparency	364	0.42	0.99
… the right to data portability	362	0.36	0.89
… the right to rectification	356	0.33	0.93
… the right against automated decision	348	-0.14	0.90

Note: strongly disagree = -2, disagree = -1, neutral = 0, agree = +1, strongly agree = +2

To a large extent, the SAR process also contributes to a perceived compliance with other data subject rights. The highest contribution is provided to the Right to Object (M = 0.51), while for the Right to Rectification, the impact is lowest (M = 0.33). According to participants, the SAR process does not contribute to the right against automated decision making (M = 0.14). One explanation in this regard could be that while the SAR process does make data collection and processing more transparent, the use of data and its impact on personal life might not be explained sufficiently.

Table 9: Regression results using the SUS score as the criterion of the perceived GDPR compliance

Predictor	b	b 95% CI [LL, UL]	beta	beta 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	r	Fit
(Intercept)	-0.15	[-0.46, 0.17]						
SUS score	0.01**	[0.01, 0.02]	0.27	[0.17, 0.37]	.07	[.03, .13]	.27**	
								R^2 = .072**
								95% CI[.03,.13]
								Adjusted R^2 = .069

Note. A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. beta indicates the standardized regression weights. sr2 represents the semi-partial correlation squared. r represents the zero-order correlation. LL and indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

We perform a linear regression ($F(1, 363) = 28.04$, $p < .0001$, $R^2 = .072$, Adjusted $R^2 = .069$).[7] The results (see Table 9) indicate that the extent of the contribution to perceived GDPR compliance seems to depend on the overall User Experience of the SAR process in question. Nevertheless, the R^2 is quite low, such that the influence seems to be relatively low.

Perceived Transparency of organizational Data Processing

Regarding transparency, we found two major aspects having an impact. The first concerns the design and presentation of the data. Comments showed that participants were sometimes left in doubt about the

[7] The visual analysis (see Appendix B, Figure 8) shows that the distribution is roughly "bell-shaped." Because the Shapiro-Wilk test for the residuals of regression (W = 0.863, $p < .0001$) indicate a deviation from normality, we also performed an additional linear regression with bootstrapping with (R = 5000 replications). This leads to similar results (Intercept-95% CI [LL, UL] = [-.507, .209]; SUS score-95% CI [LL, UL] = [.009, .021])

organization actually sharing all data stored about them. Such doubt was especially present when users thought that they would have contributed a considerable amount of data but did not receive much. In this sense, data protection-friendly policies of organizations could even backfire, if the takeout was not designed accordingly—for example, by providing transparency about what kinds of data might have been deleted or anonymized—and thus would not be present in the SAR. SAR 83 is a vivid example of such doubts:

> *"I guess they only provided that data to give the impression they are GDPR compliant, but I feel they are hiding data because I have used the platform plenty of time to hire teacher services, but the document they provided me has no info about that"*
> SAR 83 Italki.

The second issue that emerged concerning transparency involves the principle of purpose limitation—one of the key principles of the GDPR in ensuring that users have means to understand the ins and outs of what may be done with their data. These purposes must be provided as part of an answer to an SAR. In our sheet, we provided an arbitrary list of commonly used purpose statements (see Table 10). First, the rather low mentioning of "Other" purposes points to the fact that the few purposes we chose do seem to apply to a large number and variety of organizations in our sample. On the one hand, this "standardization" tendency may be interpreted as something good for consumers, to provide resemblance and comparability. On the other hand, from the perspective of transparency for users, it seems unlikely given the variety of organizations that types of processing can be subsumed under merely nine different purposes while at the same time helping users understand the extent of the data use practices of the 139 different organizations in our sample.

Table 10: Data processing purposes provided

Purpose	Count		Answers		
	n	%	Unclear [%]	Yes [%]	No [%]
Transfer to Third Parties	406	96.2	31.0	42.6	26.4
Pricing	403	95.5	22.1	23.1	54.8
Fraud Prevention	405	96.0	22.0	44.2	33.8
Legal Obligations	408	96.7	19.9	52.9	27.2
Contract Execution	410	97.2	18.8	41.2	40.0
Client Communication	400	94.8	16.2	60.8	23.0
Provision	407	96.4	15.7	52.6	31.7
Marketing	403	95.5	14.4	57.3	28.3
Personalization	402	95.3	13.4	62.9	23.6
Other	216	51.2	30.6	35.6	33.8

The purposes that participants were able to exclude most often were those of individual pricing (No = 54.8%) and contract execution (No = 40.0%). At the same time, these purposes also were among the least popular (Pricing: 23.1%, Contract: 41.2%). Given the variety of organizations, this finding is not surprising as participants did not have a customer relationship with many organizations or did not have a contract. On the other side, the

purposes of personalization (Yes = 62.9%), client communication (Yes = 60.8%), and marketing (Yes = 57.3%) were most commonly identified.

Furthermore, even though organizations must inform their users or clients regarding the purposes of collecting data via the SAR, there was nevertheless a notable absence of transparency regarding the processing purposes. In terms of a lack of clarity, personalization (unclear = 13.4%) and marketing (unclear = 14.4%) achieve the best values. Overall, however, these values show that participants were unable to consistently rule out the existence of any of the purposes we asked for. For about one out of three SAR, it remained unclear if data was transferred to third parties. Almost every fourth SAR sheet (unclear = 22.1%) reported lack of clarity about the existence of individual pricing.

Perceived Benefits and Data Insights

Data awareness and data literacy are important prerequisites to make informed privacy decisions. It is not yet clear to what extent SARs can have a positive effect on this; the answers in our sample show mixed results (see Table 11). Participants agree that received data contributes to data awareness by helping them to understand what an organization knows about them (M = 0.64). In addition, participants on average perceive the data to be easy to understand (M = 0.47). The contribution to data literacy, in contrast, is far lower (M = 0.11). One reason for this discrepancy may stem from the fact that although participants were able to browse their data, they were hardly supported in using them. As a result, participants had a hard time figuring out what to do with it (M = -0.55). One factor in this regard may also be the fact that participants found the data format difficult to use (M = -0.19).

Table 11: The perceived value of data received

The received data …	n	M	SD
… helps me to know what they know about me	360	0.64	0.91
… was easy to understand	363	0.47	1.07
… had improved my data literacy	360	0.11	0.96
… includes information which I did not expected	358	0.07	1.14
… does not provide a benefit for me	358	-0.03	1.07
I learn a lot from the received data	360	0.01	1.05
The data format makes it easy to use it	357	-0.19	1.01
I know what I can do with the data	363	-0.55	1.01

Note: strongly disagree = -2, disagree = -1, neutral = 0, agree = +1, strongly agree = +2

Attitude Behavioral Change

In principle, the SAR can change the user's attitude toward the organization in both directions. On the one hand, by creating transparency and making the data available, trustworthiness can be created, and privacy fears can be reduced. On the other hand, however, trustworthiness can also be eroded and privacy concerns fostered if (a) the SAR is poorly implemented or (b) users find out about a surprising extent of data collecting and use.
Table 12 shows, however, that the SAR has almost no impact on attitudes of trust toward the organization. One reason could be that the SAR is not yet well implemented, as indicated by the SUS, but it is not so bad that trust

is lost as a result. In addition, the SAR does not include unexpected data (M = -0.03), which neither increases nor lowers privacy concerns. In addition, both effects could balance each other out, such that overall trust is neither increased nor decrease on average.

Table 12: The impact of the SAR on the attitude toward the organization

The SAR …	n	M	SD
… changes my feeling of being valued as customer	363	0.07	0.86
… *increases* my privacy concerns	358	0.01	0.90
… *lowers* my privacy concerns	358	-0.54	0.66
… changes my attitude towards the organization	362	-0.05	0.91
… *improves* my trust in the organization	363	-0.22	0.95
… *decreases* my trust in the organization	363	-0.33	0.82

The low impact of the SAR on attitudes toward the organization, trust, and privacy concerns are also reflected at the level of behavioral change (see Table 13). As there is no change in attitudes, it seems that participants do not see a need to change their behavior in terms of privacy settings (M = -0.43), information disclosure (M = -0.54), or data deletion practices (M = -0.55).

Table 13: The impact of the SAR on privacy behavior

The SAR motivates me …	n	M	SD
… to take some action	365	-0.22	0.96
… to change my privacy settings	359	-0.43	1.04
… to change my information disclosure behavior	350	-0.54	0.81
… to change my approach towards using the right of rectification	354	-0.55	0.91

6 DISCUSSION

Our study demonstrates the large variety of ways to implement an SAR process. Following the DSR [39,54,71], we now continue with exploring and highlighting major challenges and potential solutions according to the user experience journey and design of the SAR procedure. We close our discussion with a summary of our insight in the form of a set of design guidelines to provide actionable design knowledge in accordance to DSR [26,54]

6.1 The User Experience Journey as Guiding Map for Implementation

Our initial step was to understand the procedure of conducting an SAR and current design practices to identify different touchpoints the user has to go through. We already assumed that our five identified touchpoints follow a logical order; for example, a data request must be carried out before the data can be provided and accessed. However, the steps do not fully determine the chronological order of the process. The touchpoint of authentication, in particular, may occur in other or even multiple positions. For example, users who are permanently logged in (e.g., Facebook or Google) may skip the authentication or login step right at the beginning. In connection with the touchpoint data access, authentication might be necessary to ensure that only a legitimate person can request and access or open the file. With our study, however, we were able to validate

the five touchpoints as coherent steps within the user journey, which may have to be adjusted in sequence for individual cases.

6.2 Finding instead of Searching

An important challenge for any user journey is to offer users a good start. Our findings show how this holds true for the SAR process, where the finding touchpoint had the lowest SUS score with 45. Notably, participants started searching in many cases via search engines immediately—even before turning to the website or organizational app. The importance of search engines for finding the entrance point is also highlighted by participants as a “last resort,” in case they could not find anything on the website or in the app.

As noted by participants, organizations should provide a prominent place on their website to make it easier to find the SAR option. From a usability perspective, the SAR option should be located close to the account or privacy settings or as a separate contact option, especially when there is no automated procedure. Additional access points could be provided as part of the privacy policy. This way, the mere existence of the SAR and the right to access could be presented as part of regular use of a service. Most importantly, the finding touchpoint would perhaps benefit from standardization in terms of positioning.

However, organizations must clearly take care that such a prominent option will not disrupt everyday use. Regarding this design tension, we need further studies into when and why users might want to initiate an SAR by themselves.

6.3 Authentication: Dealing with Tension between Usability and Security

Users must authenticate before they can get their data. Our study shows that there is a tension in authentication between usability and security. Depending on what kind of data the organization collects, participants expected stronger security measures (e.g., for market platforms or financial service providers), while in other cases participants expected more convenient measures.

Our study indicates that authenticating via the normal login procedure based on one’s existing username and password was the most comfortable option. However, this practice also represents a security risk, as when user accounts get hacked. Without additional authentication, attackers could easily get hold of all data about a user. For more security, two-factor authentication (e.g., for financial service providers) might be the better option.

Another practice was that people authenticated via phone or via e-mail. For security reasons, this practice is problematic, as this could also easily be misused by hackers. Organizations should at least implement some additional authentication (e.g., date of birth or place of residence).

In contrast to such lax security measures, our study also demonstrates that some organizations asked for physical proof of identification and/or requested scans of ID via e-mail. These designs were criticized for being time consuming, complicated, and for having to provide even more data to an organization.

One option to deal with the tension between usability and security would be the introduction of a trusted third party to whom the user identifies himself once and who then identifies himself to other data controllers on behalf of the user. Further research, however, is needed to specify and regulate such a distributed authentication in detail.

6.4 Guiding the Data Request Process

The major design challenge implementing the data request touchpoint is helping users to formulate the SAR request. Our study reveals that participants desire more guidance through the procedure, particularly in the case of making an SAR request via e-mail. At first glance, writing an informal e-mail requesting one's data seems easy. However, our study shows that participants were unsure about what information to include and how to explain to organizations that they are referring to the right to access, which requires organizations to answer. Moreover, there was a lack of feedback on the status of the SAR—for example, whether the e-mail had arrived and was being processed and whether the SAR was even acknowledged.

Organizations should provide more guidance to users throughout the process, informing them how they can request their data and letting them choose what data they want to request and in what format they want to receive it. In addition, the availability of data and general changes in the state of the SAR process should be actively communicated.

The comments also show that the "data request" experience was often affected by the experience of the previous touchpoints. This insight highlights that it is not enough to optimize the user experience of each individual touchpoint, but that a good user journey must take care of a coherent process design.

6.5 Making Data Accessible – Both for Human and Machines

Setting up all data for answering the SAR appropriately often takes time. While little waiting time was preferred, delays are understandable, especially in non-automated processes. The design challenge here is to ensure that users are informed about the data becoming available and providing the data via easily accessible channels. Moreover, data should be usable by humans and machines for further processing. The SUS score of 51 shows that these challenges are often not yet being met in practice.

As with the request touchpoint, the comments suggest that there is a negative impact on the user experience when users lack feedback, for example, on data having become available or delays with the respective reasons. Hence, organizations should specify an approximate time when the data will be made available or show a status in the settings where a user can check for updates.

Our study also raises awareness about the tension between human-readable and machine-readable data format. While machine-readable data format offers more possibilities for further processing, in terms of user experience the human-readable format was preferred. For instance, the SUS score was highest when data was provided in PDF (score = 66) or HTML format (score = 54). The data formats received make it likely that many organizations will respond to the SAR as well as a request in the form of the right to portability. While not illegal per se, from a UX perspective, the design for different use cases cannot avoid making compromises. We call, however, for providing the data in the machine-readable json-format together with an additional HTML-viewer to allow for human exploration (see touchpoint data use). In addition, when zip files with a complex structure are provided, the SARs should provide guidance and explanation of the directory and a clear labelling of the files to ensure comprehensibility.

6.6 From Data Provisioning to Data Literacy Support

Our study shows that in general, based on the data received, users perceived organizations to be compliant with GDPR regarding the right to access. In addition, the SAR did help participants see what data had been collected by organizations. Nevertheless, participants reported that most of the data was in line with their

expectations without any bad surprises or greater impact on their privacy behavior. In fact, they tend to feel safer.

Our study further indicates that the SAR currently is a necessary, but not sufficient, condition for privacy awareness and data literacy. Concerns about data completeness repeatedly emerged. Additionally, participants complained that it was unclear what they can do with the data and were not motivated to take action. As a result, they did not feel empowered but were disenchanted as they often received raw data instead of graphics and statistics.

This shows that we need more than the right to get our data; we also need support to make sense out of it. As noted, for example, by Pins et al [49], what is needed is a data storytelling, so that users can gain added value from it, allowing them to take actions like deleting or correcting the data. A promising approach is data visualizations which increase user understanding, promote transparency, enable risk assessment, and foster personal data literacy.

6.7 Towards a Usable Process Design for SARs

As a conclusion drawn from the previous sections, we derive a set of guidelines for designing the SARs User Experience Journey that covers five touchpoints. Table 14 summarizes these guidelines. These guidelines should serve as an orientation for practitioners who must adapt them to the specific context of their organization. These guidelines, however, cannot replace a rigorous evaluation of the usability of the specific organizational implementation of the SAR process.

Table 14 Lessons Learned and the Resulting Design Guidelines for the SAR User Experience Journey

Touchpoint	Lessons Learned	Guidelines
Overall Journey	Guiding the process and giving feedback	• Design the SAR as a multi-step user experience journey with finding, authentication, data request, data access, and data use as the key touchpoints. • Provide the user feedback concerning where s/he is in the SAR process and the next steps that need to be taken.
Finding	Support finding instead of searching	• Do not hide behind complex menu items or use misleading terms. • Place the SAR option prominently in the account or privacy area of your website or app. • Make it clear how to initiate an SAR.
Authentication	Dealing with tension between usability and security	• Ensure that authentication is done in a detour. • Protect the user by asking for more than just username and password.
Data Request	Guiding the data request process	• Disclose what you need from the user and how s/he should proceed. • Inform the user about the format in which they will receive the SAR and, at best, let them choose the right one among different formats.

Touchpoint	Lessons Learned	Guidelines
Data Access	Making data accessible—both for human and machines	• Keep the user in the loop about the request status and when the SAR is available. • Provide data in machine-readable format (e.g., json) for portability and processing and additionally in a user-friendly format (e.g., HTML) for explaining and exploring the data immediately.
Data Use	From data provisioning to data literacy support	• Give support and explanations on the provided data for transparency about processing and sense-making. • Point out which data might be missing because they have been anonymized or pseudonymized. • Highlight potentially risky data and explicitly show why this data is stored and how the user can make settings and changes for permissions and collection.

6.8 Limitations and Future Research

In the following, we outline several key limitations of our research and suggest some open questions to address in follow-up research. First, the participants of our study largely were HCI students in a graduate course on Usable Security and the GDPR, thus forming a rather homogenous group (e.g., potentially being more tech-savvy than average users). Further, we did not address whether the course content taught may have had an impact on response behavior and scoring with regard to the SUS evaluation. Consequently, our sample is biased toward corresponding organizations selected, as shown in the large "Education/Learning." The selection of organizations by the participants may further impact the results; for example, the Gini index should be considered as an estimation. These biases should be kept in mind when seeking to generalize findings. Additionally, participants had to document a process which sometimes took several weeks, which may result in inconsistencies and inaccuracy in remembering the duration of the activities. For this reason, the time specifications given here should only be understood as rough, subjective estimates. Despite these limitations, our study is the first to be able to paint a broader picture of user-centered usability assessments of the right to access.

Regarding future research, our work provides a first orientation about how to structure the user experience journey. As a general guideline, users would benefit when being able to reidentify patterns used in other SAR processes. The touchpoints identified should be investigated in more detail in future studies to gain more insight into critical incidents and design patterns to improve the user experience. This is especially true for the finding touchpoint, where the low SUS score indicates a substantial need for action.

Future research should also study what people do—and want to do—with their data. Our study shows that the data received can promote users' data awareness, but it left many questions open and was perceived as not improving data literacy. This raises the question of how we can support users in understanding and making sense of their data. In addition, we should study how users in daily life can be motivated to make use of their rights to access their data.

7 CONCLUSION

Data awareness and control over personal data and its use are cornerstones for usable privacy protection. Likewise, the right to access as per the GDPR is one of the key mechanisms for consumers to be able to control data collection and use by organizations. For implementing this right, Article 12 of the GDPR provides general guidelines, such as adequate levels of authentication, which information to provide, and more generally the provision is to be made in "concise, transparent, intelligible and easily accessible form, using clear and plain language" [20]. However, there is no clear guidance about how an SAR process should be designed in terms of a user journey.

In this paper, we adopt the Grounded Design approach [54,65] to provide an empirically grounded problem diagnosis [71,77] about the implementation of the SAR process today. Doing so, we first introduced a user journey for the right to access and, secondly, conducted an evaluation of different design decisions. We discovered that the whole procedure lacks structure and guidance. Difficulties were already arising at the beginning, when users experienced difficulties finding the SAR option, and there were difficulties at the end with a lack of support in making use of the requested data. Addressing the problems identified, we discuss design opportunities. Intending to make this design knowledge actionable, we provide a set of guidelines for practitioners. As mid-range theory propositions, guidelines present a nascent design theory [27], which are of practical relevance in giving organizations an orientation about how to implement the GDPR in a manner that is faithful to the idea of strengthening user rights. It also informs consumer organizations about the current state of the GDPR implementation and where the consumer needs help to pursue their rights in an effective way. Future work should test our nascent design theory, providing instantiation validity [40] by evaluating how practitioners adopt our guidelines and whether this would improve the usability of the SAR process compared to current implementations.

Beyond the legal discourse on the GPDR, our study shows that this discourse would benefit from a serious incorporation of human factors. Research on usable privacy can play a major role in informing practical solutions and providing data protection authorities with a user-oriented view. We show that user needs, mental models, and capabilities can play an important role in both the process of drafting legislation and in supporting data protection authorities to evaluate the facilitating role of data controllers when implementing the right to access.

ACKNOWLEDGEMENTS

We thank the students of the Usable Privacy and Security course in winter semester 20/21 for their active participation.

REFERENCES

1. Anne Adams, Martina Angela Sasse, and Peter Lunt. 1997. Making passwords secure and usable. In *People and Computers XII*. Springer, 1–19.
2. Icek Ajzen. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50, 2: 179–211.
3. Fatemeh Alizadeh, Timo Jakobi, Jens Boldt, and Gunnar Stevens. 2019. GDPR-Reality Check on the Right to Access Data: Claiming and Investigating Personally Identifiable Data from Companies. In *Proceedings of Mensch Und Computer 2019* (MuC'19), 811–814. https://doi.org/10.1145/3340764.3344913
4. Keith J. Anderson. 2001. Internet Use Among College Students: An Exploratory Study. *Journal of American College Health* 50, 1: 21–26. https://doi.org/10.1080/07448480109595707

5. Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. Usable transparency with the data track: a tool for visualizing data disclosures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1803–1808.
6. Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15), 1803–1808. https://doi.org/10.1145/2702613.2732701
7. Salvatore Aurigemma, Thomas Mattson, and Lori Leonard. 2017. So much promise, so little use: What is stopping home end-users from using password manager applications?
8. Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. 2013. Little brothers watching you: Raising awareness of data leaks on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 12. Retrieved August 11, 2015 from http://dl.acm.org/citation.cfm?id=2501616
9. Aaron Bangor. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. 4, 3: 10.
10. Victoria Bellotti and Keith Edwards. 2001. Intelligibility and accountability: human considerations in context-aware systems. *Human–Computer Interaction* 16, 2–4: 193–212.
11. Jan Hendrik Betzing, Matthias Tietz, Jan vom Brocke, and Jörg Becker. 2019. The impact of transparency on mobile privacy decision making. *Electronic Markets*. https://doi.org/10.1007/s12525-019-00332-3
12. Data Brokers. 2014. A call for transparency and accountability. *US Federal Trade Commission*.
13. John Brooke. 1996. SUS: a "quick and dirty" usability scale. In *Usability evaluation in industry*, Patrick W. Jordan (ed.). Taylor & Francis, London ; Bristol, Pa.
14. John Brooke. 2013. SUS: a retrospective. *Journal of Usability Studies* 8: 29–40.
15. Kelly Caine and Rima Hanania. 2012. Patients want granular privacy control over health information in electronic medical records. *Journal of the American Medical Informatics Association* 20, 1: 7–15.
16. Nico Castelli, Corinna Ogonowski, Timo Jakobi, Martin Stein, Gunnar Stevens, and Volker Wulf. 2017. What Happened in my Home?: An End-User Development Approach for Smart Home Data Visualization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 853–866. https://doi.org/10.1145/3025453.3025485
17. Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez. 2018. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review* 34, 2: 193–203. https://doi.org/10.1016/j.clsr.2017.10.003
18. William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N. Sheth. 2014. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)* 32, 2: 5.
19. European Data Protection Board. 2018. *Guidelines on Transparency under Regulation 2016/679 Rn.9*. WP29.
20. European Parliament and the Council. 2016. *REGULATION (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*.
21. Adrienne Porter Felt, Serge Egelman, Matthew Finifter, Devdatta Akhawe, and David Wagner. 2012. How to Ask for Permission. In *HotSec*.
22. Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth*

Symposium on Usable Privacy and Security, 3. Retrieved December 15, 2014 from http://dl.acm.org/citation.cfm?id=2335360
23. Simone Fischer-Hübner, Julio Angulo, and Tobias Pulls. 2014. How can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used? In *Privacy and Identity Management for Emerging Services and Technologies*. Springer, 77–92. Retrieved March 20, 2015 from http://link.springer.com/chapter/10.1007/978-3-642-55137-6_6
24. Nina Gerber, Benjamin Reinheimer, and Melanie Volkamer. 2018. Home sweet home? Investigating users' awareness of smart home privacy threats. In *Proceedings of An Interactive Workshop on the Human aspects of Smarthome Security and Privacy (WSSP)*.
25. Maximilian von Grafenstein. 2019. Co-Regulation and the Competitive Advantage in the GDPR: Data protection certification mechanisms, codes of conduct and the "state of the art" of data protection-by-design. In *Research Handbook on Privacy and Data Protection Law: Values, Norms and Global Politics, Edward Elgar Publishing*, G González-Fuster, R van Brakel and P De Hert (eds.). Edward Elgar Publishing.
26. Shirley Gregor. 2006. The nature of theory in information systems. *MIS quarterly*: 611–642.
27. Shirley Gregor and Alan R. Hevner. 2013. Positioning and presenting design science research for maximum impact. *MIS quarterly*: 337–355.
28. Human Factors International. 2001. *HFI Helps Staples.com Boost Repeat Customers by 67%*. Retrieved September 14, 2021 from https://humanfactors.com/downloads/staples.pdf
29. ISO. 2018. DIN EN ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. Retrieved February 14, 2020 from https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en
30. Timo Jakobi, Edna Kropp, Gunnar Stevens, and Mats Schmal. 2017. Providing smartphone data visualizations to support Privacy Literacy.
31. Timo Jakobi, Max von Grafenstein, and Thomas Schildhauer. 2021. Data Privacy: A Driver for Competitive Advantage. In *The Machine Age of Customer Insight*. Emerald Publishing Limited.
32. Timo Jakobi, Sameer Patil, Dave Randall, Gunnar Stevens, and Volker Wulf. 2019. It's About What They Could Do with the Data: A User Perspective on Privacy in Smart Metering. *ACM Trans. Comput.-Hum. Interact.* 9, 4: 43. https://doi.org/10.1145/3281444
33. Timo Jakobi, Gunnar Stevens, Nico Castelli, Corinna Ogonowski, Florian Schaub, Nils Vindice, Dave Randall, Peter Tolmie, and Volker Wulf. 2018. Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4: 28. https://doi.org/10.1145/3287049
34. Timo Jakobi, Gunnar Stevens, Maximilian von Grafenstein, Dominik Pins, and Alexander Boden. 2020. User-friendly formulation of data processing purposes of voice assistants: a user perspective on the principle of purpose limitation. In *Proceedings of the Conference on Mensch und Computer*, 361–372.
35. Timo Jakobi, Gunnar Stevens, and Anna-Magdalena Seufert. 2018. Privacy-By-Design für das Connected Car: Architekturen aus Verbrauchersicht: Eine nutzerorientierte Diskussion. *Datenschutz und Datensicherheit - DuD* 42, 11: 704–707. https://doi.org/10.1007/s11623-018-1029-7
36. Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, 39–52. Retrieved January 7, 2016 from https://www.usenix.org/conference/soups2015/proceedings/presentation/kang
37. LMU Munich Kranz, Julia Gallenkamp, and Arnold Picot. 2010. Exploring the Role of Control – Smart Meter Acceptance of Residential Consumers. *AMCIS 2010 Proceedings*. Retrieved from http://aisel.aisnet.org/amcis2010/315

38. 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*, 1–10.
39. Bill Kuechler and Vijay Vaishnavi. 2008. On theory development in design science research: anatomy of a research project. *European Journal of Information Systems* 17, 5: 489–504.
40. Roman Lukyanenko, Joerg Evermann, and Jeffrey Parsons. 2014. Instantiation validity in IS design research. In *International Conference on Design Science Research in Information Systems*, 321–328.
41. Jochen Meyer, Anastasia Kazakova, Merlin Büsing, and Susanne Boll. 2016. Visualization of Complex Health Data on Mobile Devices. In *Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care* (MMHealth '16), 31–34. https://doi.org/10.1145/2985766.2985774
42. Branko Milanovic. 1997. A simple way to calculate the Gini coefficient, and some implications. *Economics Letters* 56, 1: 45–49. https://doi.org/10.1016/S0165-1765(97)00101-8
43. Timothy Morey, Theodore "Theo" Forbath, and Allison Schoop. 2015. Customer Data: Designing for Transparency and Trust. *Harvard Business Review*. Retrieved April 5, 2019 from https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust
44. Timothy Morey and Allison Schoop. Customer Data: Designing for Transparency and Trust. 11.
45. Roland M. Müller and Katja Thoring. 2012. Design thinking vs. lean startup: A comparison of two user-driven innovation strategies. *Leading through design* 151: 91–106.
46. Jakob Nielsen. 1995. How to conduct a heuristic evaluation. *Nielsen Norman Group* 1: 1–8.
47. Michaela Olausson. 2018. *User control of personal data: A study of personal data management in a GDPR-compliant grahpical user interface*.
48. Dominik Pins, Alexander Boden, Britta Essing, and Gunnar Stevens,. 2020. "Miss Understandable": a study on how users appropriate voice assistants and deal with misunderstandings. In *MuC '20: Proceedings of the Conference on Mensch und Computer*, 349–359. https://doi.org/10.1145/3404983.3405511
49. Dominik Pins, Timo Jakobi, Alexander Boden, Fatemeh Alizadeh, and Volker Wulf. 2021. Alexa, We Need to Talk: A Data Literacy Approach on Voice Assistants. In *Designing Interactive Systems Conference 2021*, 495–507.
50. Stefanie Pötzsch. 2009. Privacy awareness: A means to solve the privacy paradox? In *The future of identity in the information society*. Springer, 226–236. Retrieved August 11, 2015 from http://link.springer.com/chapter/10.1007/978-3-642-03315-5_17
51. Philip Raschke, Axel Küpper, Olha Drozd, and Sabrina Kirrane. 2017. Designing a GDPR-Compliant and Usable Privacy Dashboard. In *IFIP International Summer School on Privacy and Identity Management*, 221–236.
52. Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia M. McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, and Florian Schaub. 2014. *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*. Social Science Research Network, Rochester, NY. Retrieved December 5, 2014 from http://papers.ssrn.com/abstract=2418297
53. Markus Rohde, Peter Brödner, Gunnar Stevens, Matthias Betz, and Volker Wulf. 2017. Grounded Design-a praxeological IS research perspective. *Journal of Information Technology* 32, 2: 163–179.
54. Markus Rohde, Peter Brödner, Gunnar Stevens, Matthias Betz, and Volker Wulf. 2017. Grounded Design-a praxeological IS research perspective. *Journal of Information Technology* 32, 2: 163–179.
55. 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*, 340–351.

56. Jeff Sauro. 2011. Measuring Usability with the System Usability Scale (SUS) – MeasuringU. Retrieved September 14, 2021 from https://measuringu.com/sus/
57. Jeff Sauro. 2011. *A practical guide to the system usability scale: background, benchmarks & best practices*. Measuring Usability LLC, Denver, CO.
58. 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.
59. Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. 2016. Watching them watching me: Browser extensions impact on user privacy awareness and concern. In *NDSS workshop on usable security*.
60. Richard Shay and Elisa Bertino. 2009. A comprehensive simulation tool for the analysis of password policies. *International Journal of Information Security* 8, 4: 275–289.
61. Dayana Spagnuelo, Ana Ferreira, and Gabriele Lenzini. 2018. Accomplishing Transparency within the General Data Protection Regulation. In *5th International Conference on Information Systems Security and Privacy. To appear*.
62. Sarah Spiekermann. 2005. Perceived Control: Scales for Privacy in Ubiquitous Computing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.761109
63. Felix Stalder. 2010. Autonomy and control in the era of post-privacy. *Open* 19: 81–83.
64. Gunnar Stevens. 2009. Understanding and Designing Appropriation Infrastructures: Artifacts as boundary objects in the continuous software development.
65. Gunnar Stevens, Markus Rohde, Matthias Korn, Volker Wulf, V. Pipek, D. Randall, and K. Schmidt. 2018. Grounded design. A research paradigm in practice-based computing. *V. Wulf; V. Pipek; D. Randall; M. Rohde*: 139–176.
66. Maurice E. Stucke. 2018. Should We Be Concerned About Data-opolies? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3144045
67. Omer Tene and Jules Polonetsky. 2013. Big Data for All: Privacy and User Control in the Age of Analytics. Retrieved May 13, 2014 from http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11&section=20
68. Peter Tolmie, Andy Crabtree, Tom Rodden, James Colley, and Ewa Luger. 2016. "This has to be the cats": Personal Data Legibility in Networked Sensing Systems. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 491–502.
69. Jan Tolsdorf, Michael Fischer, and Luigi Lo Iacono. 2021. A Case Study on the Implementation of the Right of Access in Privacy Dashboards. In *Privacy Technologies and Policy* (Lecture Notes in Computer Science), 23–46. https://doi.org/10.1007/978-3-030-76663-4_2
70. 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*. Springer, 61–79.
71. John Venable. 2006. The role of theory and theorising in design science research. In *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006)*, 1–18.
72. John Venable. 2006. A framework for design science research activities. In *Emerging Trends and Challenges in Information Technology Management: Proceedings of the 2006 Information Resource Management Association Conference*, 184–187.
73. Michele Vescovi, Bruno Lepri, Christos Perentis, Corrado Moiso, and Chiara Leonardi. 2014. My data store: toward user awareness and control on personal data. 179–182. https://doi.org/10.1145/2638728.2638745

74. Francesco Vitale, Janet Chen, William Odom, and Joanna McGrenere. 2020. Data Dashboard: Exploring Centralization and Customization in Personal Data Curation. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 311–326. https://doi.org/10.1145/3357236.3395457
75. Paul Voigt and Axel von dem Bussche. 2018. *EU-Datenschutz-Grundverordnung (DSGVO): Praktikerhandbuch*. Springer-Verlag, Berlin Heidelberg. Retrieved March 26, 2019 from https://www.springer.com/de/book/9783662561867
76. Janis Wong and Tristan Henderson. 2018. How Portable is Portable?: Exercising the GDPR's Right to Data Portability. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 911–920.
77. Volker Wulf, Markus Rohde, Volkmar Pipek, and Gunnar Stevens. 2011. Engaging with practices: design case studies as a research framework in CSCW. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 505–512.
78. Justyna Żywio\lek and Eva Nedeliaková. 2020. Personal data protection as an element of competitive advantage. *System Safety: Human-Technical Facility-Environment* 2, 1.

APPENDIX

Appendix A: Modification of the SUS Questionnaire exemplified by the Step "Finding the SAR option"

SUS Questionnaire [13]	Modified questions for Step "Finding the SAR option"
I think that I would like to use this system frequently	--
I found the system unnecessarily complex	I found finding the SAR option unnecessarily complex
I thought the system was easy to use	I thought finding of the SAR option was easy
I think that I would need the support of a technical person to be able to use this system	I think that I would need the support of a technical person to find this SAR option
I found the various functions in this system were well integrated	--
I thought there was too much inconsistency in this system	I thought there was too much inconsistency in making this SAR option accessible
I would imagine that most people would learn to use this system very quickly	I would imagine that most people would learn finding this SAR option very quickly
I found the system very cumbersome to use	I found finding of this SAR option very cumbersome to use
I felt very confident using the system	I felt very confident finding this SAR option
I needed to learn a lot of things before I could get going with this system	I needed to learn a lot of things before I could find this SAR option

Appendix B: Visualization of the data used in the conducted regression analysis

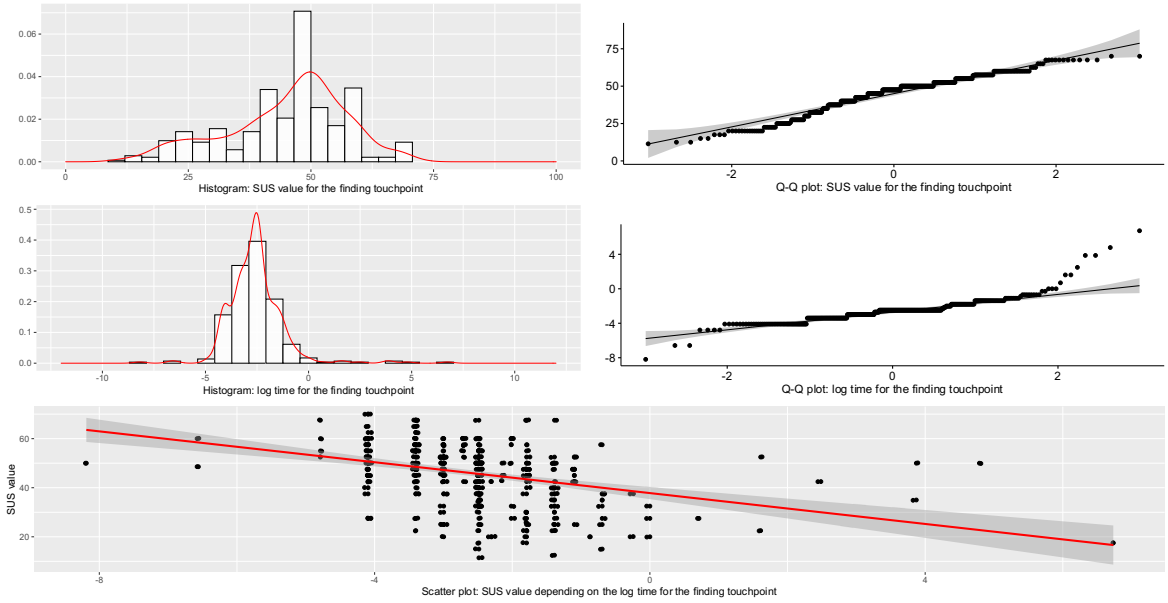

Figure 6 Plots of the SUS values of the finding touchpoint (top), the log time for the finding touchpoint (center), and the relation between both (bottom)

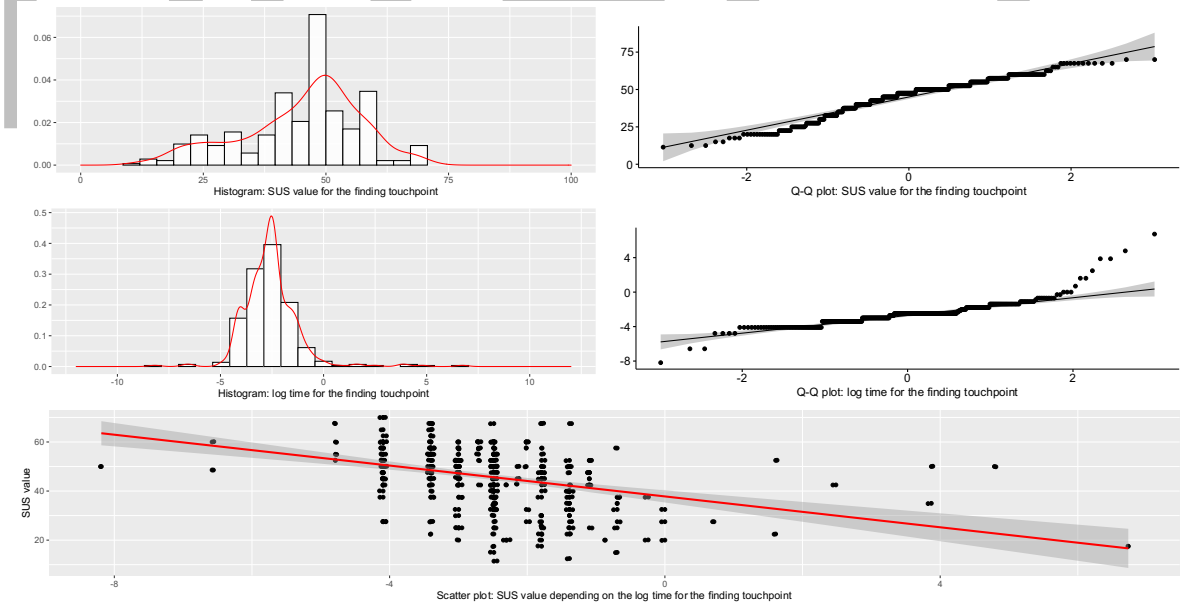

Figure 7 Plots of the SUS values of the data request touchpoint (top), the log time for the data request touchpoint (center), and the relation between both (bottom)

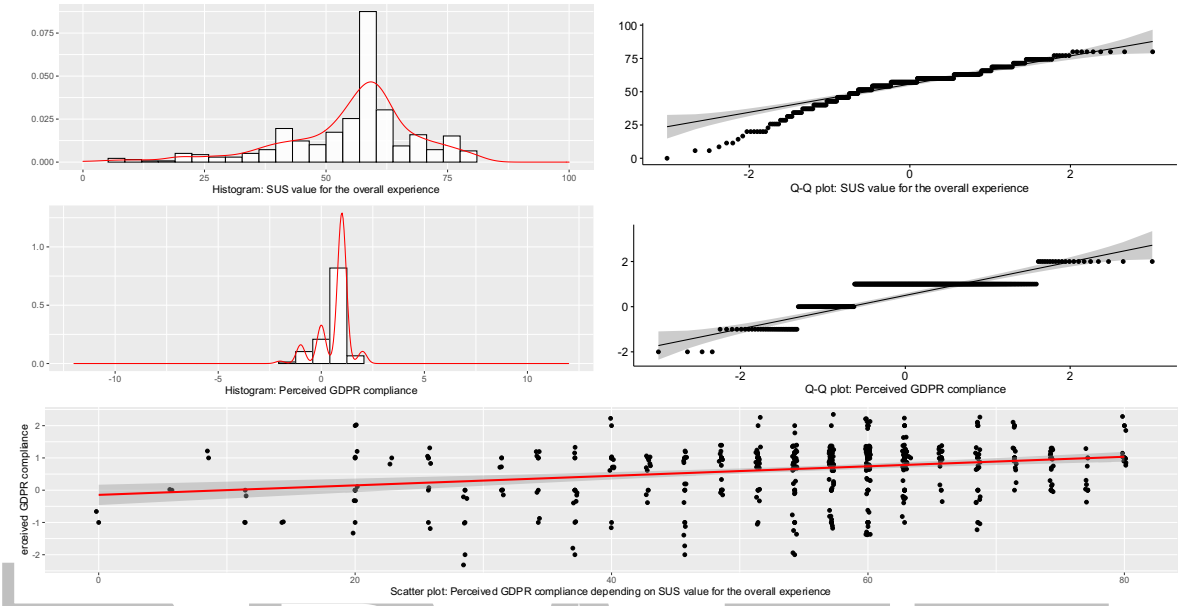

Figure 8 Plots of the SUS values of the overall SAR experience (top), the perceived GDPR compliance (center), and the relation between both (bottom)

Appendix C: Visualization of the data used in the conducted ANOVA

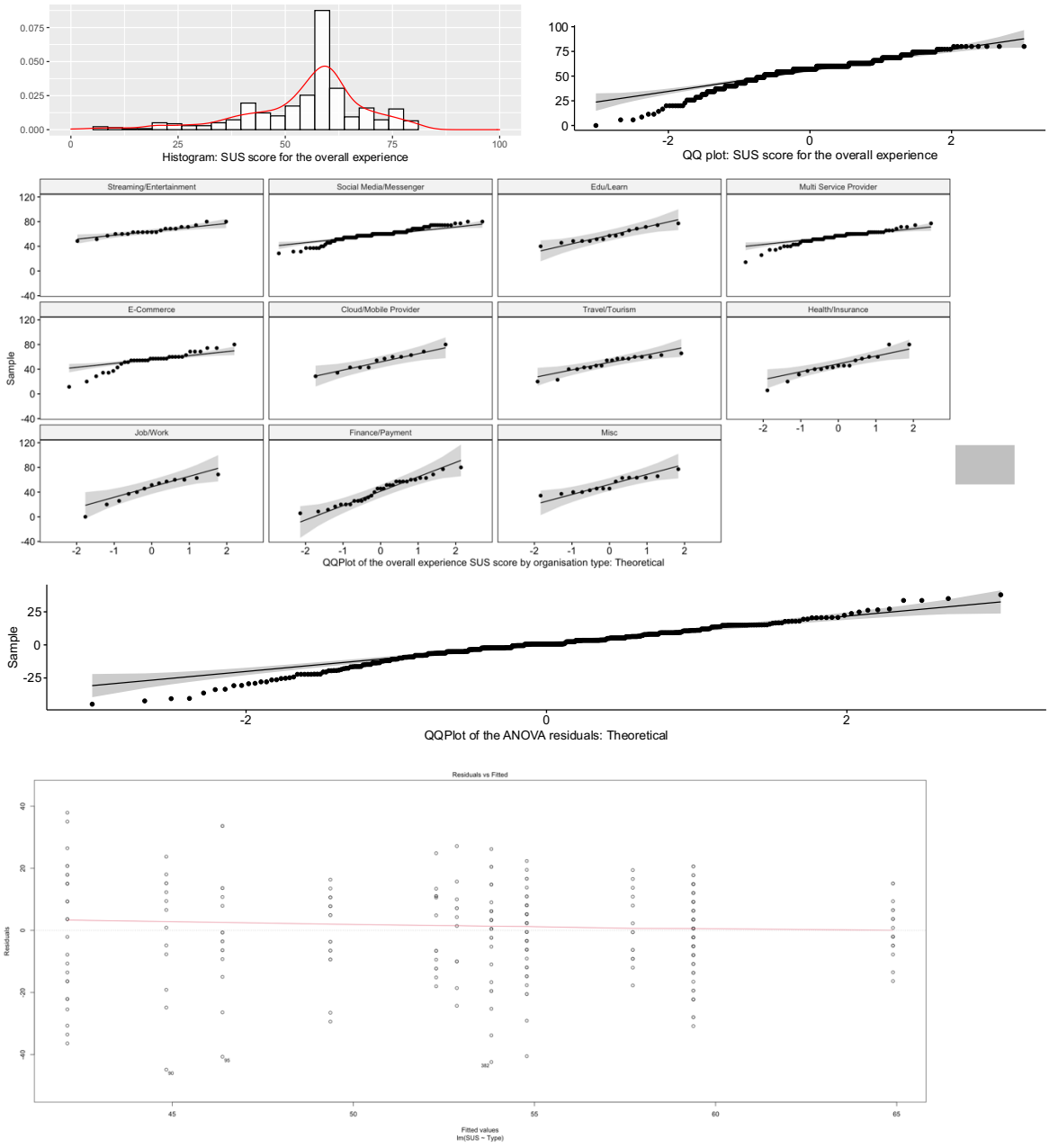

Figure 9 Plots for one-way ANOVA analyzing the overall experience SUS score by the organization type: the deviation from normality of the SUS score (top), by each group (center-top), the deviation from normality of the ANOVA residuals (center-bottom) and the relation relationships between residuals and fitted values to check the homogeneity of variances (bottom)

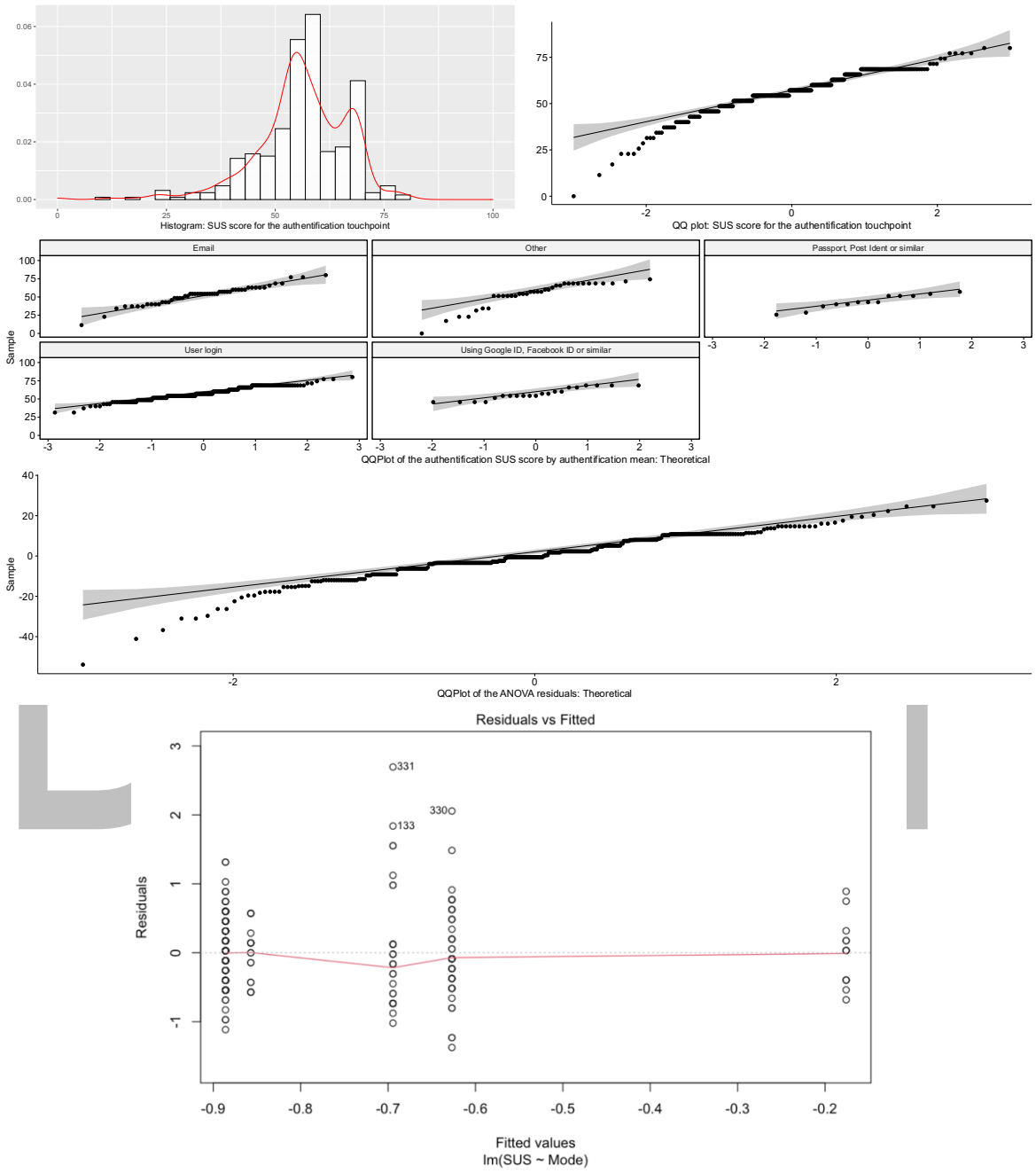

Figure 10 Plots for one-way ANOVA analyzing the authentication SUS score by the authentication mean: the deviation from normality of the SUS score (top), by each group (center-top), the deviation from normality of the ANOVA residuals (center-bottom) and the relation relationships between residuals and fitted values to check the homogeneity of variances (bottom)

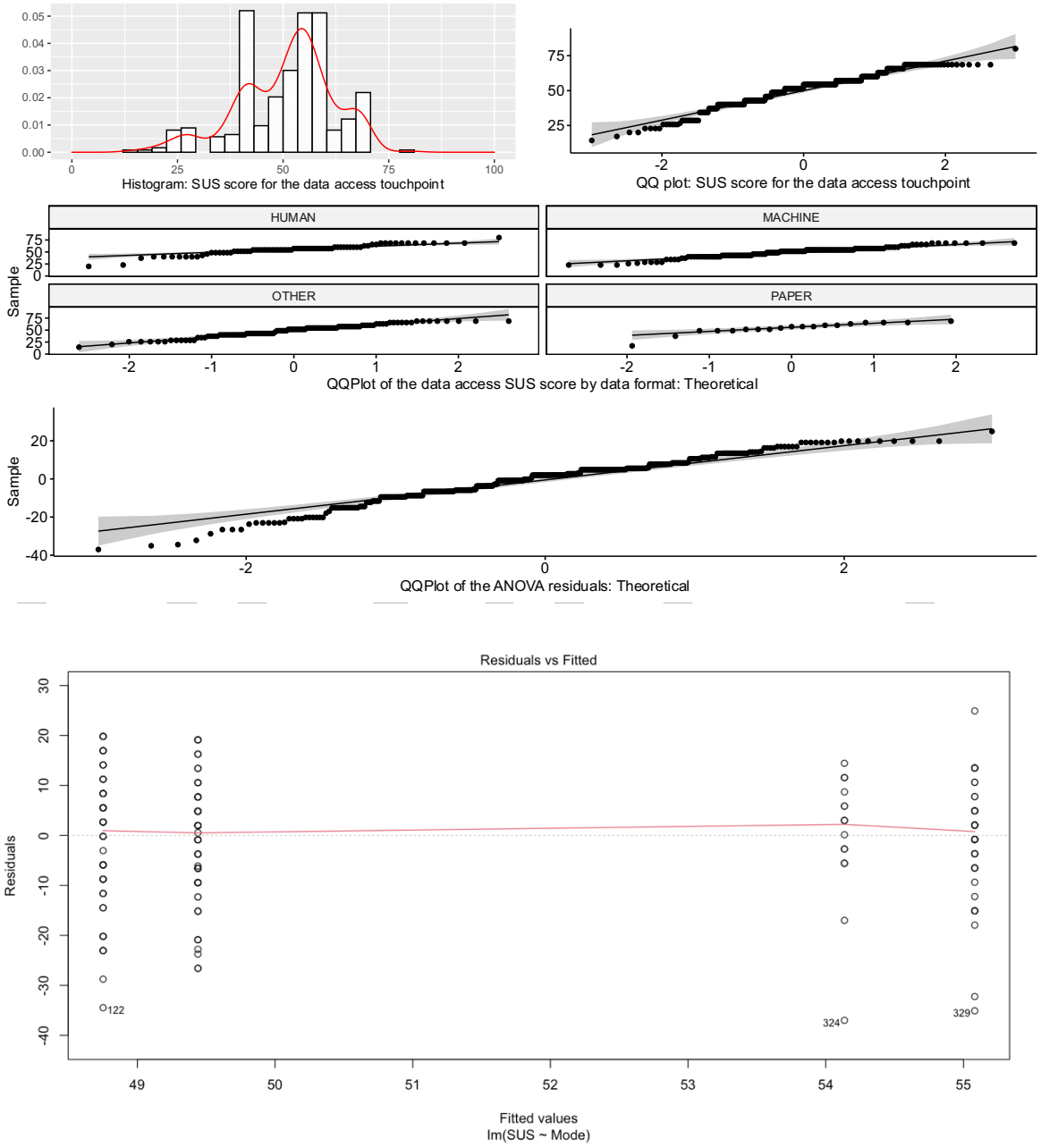

Figure 11 Plots for one-way ANOVA analyzing the data request SUS score by the data format: the deviation from normality of the SUS score (top), by each group (center-top), the deviation from normality of the ANOVA residuals (center-bottom) and the relation relationships between residuals and fitted values to check the homogeneity of variances (bottom)

Appendix D: Heuristic Evaluation for defining an SAR user journey

Table 15 The consulted organizations for the heuristic evaluation and the amount of conducted SAR by the involved authors, see column on the left. On the right is shown the identified and discussed key touchpoints, labelled by the final touchpoints.

Organizations	Identified Key Touchpoints and final labels
AMAZON (2)	1. Finding the SAR option 2. Authentication 3. Data Request 4. Data Access 5. Data Use
Google (3)	1. Finding the SAR option 2. Authentication 3. Data Request 4. Data Access 5. Data Use
Facebook (2)	1. Finding the SAR option 2. Authentication 3. Data Request 4. Data Access 5. Data Use
Telekom (1)	1. Finding the SAR option 2. Data Request 3. Authentication 4. Data Access 5. Data Use
Apple (2)	1. Authentication 2. Finding the SAR option 3. Data Request 4. Data Access 5. Data Use