Balancing Privacy and Data Utilization: A Comparative Vignette Study on User Acceptance of Data Trustees in Germany and the US

Leona Lassak*, Hanna Püschel†, Oliver D. Reithmaier‡, Tobias Gostomzyk† and Markus Dürmuth‡
* Faculty of Computer Science, Ruhr University Bochum,
leona.lassak@rub.de
†Institute of Journalism, TU Dortmund University,
{hanna.pueschel, tobias.gostomzyk}@tu-dortmund.de
‡ Faculty of Electrical Engineering & Computer Science, Leibniz University Hannover,
{oliver.reithmaier, markus.duermuth}@itsec.uni-hannover.de

***Abstract*—In times of big data, connected devices, and increasing self-measurement, protecting consumer privacy remains a challenge despite ongoing technological and legislative efforts. Data trustees present a promising solution, aiming to balance data utilization with privacy concerns by facilitating secure data sharing and ensuring individual control. However, successful implementation hinges on user acceptance and trust. We conducted a large-scale, vignette-based, census-representative online study examining factors influencing the acceptance of data trustees for medical, automotive, IoT, and online data. With $n = 714$ participants from Germany and $n = 1036$ from the US, our study reveals varied willingness to use data trustees across both countries, with notable skepticism and outright rejection from a significant portion of users. We also identified significant domain-specific differences, including the influence of user anonymity, perceived personal and societal benefits, and the recipients of the data. Contrary to common beliefs, organizational and regulatory decisions such as the storage location, the operator, and supervision appeared less relevant to users' decisions. In conclusion, while there exists a potential user base for data trustees, achieving widespread acceptance will require explicit and targeted implementation strategies tailored to address diverse user expectations. Our findings underscore the importance of understanding these nuances for effectively deploying data trustee frameworks that meet both regulatory requirements and user preferences while upholding highest security and privacy standards.**

I. Introduction

In times of big data, personal health and fitness monitoring, and connected devices, increasing amounts of personal and non-personal data are generated and processed. To protect consumers' privacy, researchers and lawmakers are continually working on new technologies and regulations, but they often face significant challenges. For manufacturers and developers – whether in automotive, smart devices, or digital services – privacy still often remains an afterthought [13]. Despite legislation like GDPR, research consistently proves that data processing based on consent mostly fails to effectively enable sovereign privacy decision-making [18], [31], [34], [59], [71]. While data is overused in domains such as the online advertising and tracking industry, other domains suffer from limited access to data [5], [8], [9], [38], [63]. However, the potential benefits of increased data access especially for the public good are substantial [2], [37], [42], whether in advancing medical research [32], [52], developing and training ethical AI [27], or supporting research on climate change [23].

Innovative approaches are needed to resolve this conflict between data protection and data usage and enable privacy-preserving data utilization. One such approach are "data trustees." Data trustees are intended to reconcile the tension between safeguarding data and harnessing its potential. On one hand, they should facilitate the aggregation and exchange of data, driving innovation and progress in the data economy. Simultaneously, they should offer stringent data protection, empowering individuals with control over their data [4], [37], [43], [64]. As these developments are still in their early stages, various forms of data trustees with different tasks and objectives are conceivable. Concepts range from so-called "Personal Information Management Systems" (PIMS) [7], [15], [41], [53], [71], which grant consent and exercise rights on behalf of data subjects to "data trustees" as institutions acting as an intermediary point of trust, independently congregating data, conducting analyses, and/or mediating data access for third parties [2], [6], [38], [65]. The topic has received increasing public and scientific attention through recent legislative developments [5], [37], [38], [72]. In 2022, the European Data Governance Act (DGA) created a framework to facilitate data sharing through regulations for "data sharing service providers" and "data altruistic organizations," which could include said trustees. Similarly, the California Consumer Privacy Act (CCPA) introduces the concept of "authorized representatives." Pilot-initiatives are currently testing various potential applications for the concept [19], [21].

Unfortunately, in the past, privacy-related legislative developments, although well-intentioned, have repeatedly failed to achieve their intended goals, often putting more burden on users than providing benefits [18], [70]. Increasingly complex data processing procedures, long and complex data protection agreements, and the abundance of requests for consent are only

Network and Distributed System Security (NDSS) Symposium 2025
24-28 February 2025, San Diego, CA, USA
ISBN 979-8-9894372-8-3
https://dx.doi.org/10.14722/ndss.2025.241982
www.ndss-symposium.org

a few examples of such counterproductive developments [18], [31], [59], [68], [70]. Learning from these past mistakes in tech and privacy legislation, where user perspectives were merely an afterthought, can help formulate better regulations in the future. Thus, investigating user acceptance and trust in newly introduced technical concepts (such as data trustees) *before* legislative frameworks are finalized is crucial for their success. In this paper, we therefore explore users' views on data trustees, by addressing the following research questions:

RQ1: What are users' general thoughts and opinions about data trustees?

RQ2: What factors influence users' willingness to use data trustees? Specifically:

a) How should data trustees be configured to increase acceptance?

b) How does the type of processed data impact user acceptance?

c) Do perceptions of data trustees vary between different countries?

To answer these questions, we conducted a large-scale, vignette-based, census-representative online study investigating end users' acceptance of and preferences regarding data trustees in Germany and the United States (US). We investigated four areas where data trustees can find application in the future: *medical* data, *automotive*/mobility data, Internet of Things (*IoT*), and data generated *online*.

Our study with $n = 714$ (Germany) and $n = 1036$ (USA) participants showed that the overall willingness to use a data trustee is heterogeneous in both Germany and the US. A substantial portion of users (about 30 %) reject the use of data trustees altogether. At the same time, a substantial number of participants recognized their societal potential. US participants were generally more open to the idea, expressing fewer privacy concerns but highlighting expected (monetary) benefits. A regression analysis revealed that the anonymity of the transmitted data, the perceived benefits – both for oneself and society – and the recipient of the data are decisive factors for acceptance. In contrast, organizational and regulatory aspects such as the storage location of the data or the identity of the operator and the institution that supervises the data trustee influenced users' decisions less. In conclusion, our findings suggest significant efforts are needed to align data trustee implementations with user expectations and to build trust. Acceptance remains limited, necessitating comprehensive measures to address user concerns and enhance trust in this emerging institution.

II. Background

In the following sections, we define the concept of data trustees and provide legal background. Additionally, we report details about existing data trustees and summarize literature about users' general willingness to share their data.

A. Data Trustees

The development of data trustees is still in the early stages. As of now, there is no uniform understanding or clear definition of the term "data trustee" and various forms with different tasks and objectives are subsumed under it. Generally, a data trustee mediates access to data provided or held by a data subject, following contractually agreed or legally prescribed data governance regulations in the interests of third parties [64].

Data Trustee Regulations The EU recently advanced the discourse on the topic of data trustees through the provisions of the *Data Act* [67] and the *Data Governance Act* [66]. These legislations address so-called "data intermediary services" and "data altruistic organizations," designed to facilitate data sharing.[1] Data intermediaries constitute one of the means through which the EU regulation aims to enhance the usability of existing data stocks. The law also intends to strengthen market confidence for data trustees by preventing undesirable competitive developments early on. The overarching goal is to create an ecosystem based on the shared use of data, reducing the (data) concentration on and market power of a few tech giants [14], [29] and giving European companies, authorities, and scientists access to large amounts of high-quality data. In 2017, "data trusts" were discussed in the UK to facilitate data sharing for the development of artificial intelligence between organizations that have data and organizations that want to use data to develop AI systems [27]. In the US similarly, the California Consumer Privacy Act (CCPA) defines "authorized agents" which could handle matters relating to personal data in the interests of consumers (i.e., submitting requests for erasure or opting out of personal data sales) [12]. Additionally, issues of data access and data sharing for smaller tech companies, non-profit organizations, or data science are being discussed [29].

Data Trustees in Europe Data trustees are primarily discussed in the legal literature but overlooked in the security and privacy domain. As one of the first, Specht-Riemenschneider et al. published a detailed essay on data trustees, defining the term and identifying regulatory requirements [65]. Specht-Riemenschneider and Blankertz also dissected potentials and risks of data trustees in four application domains (medical, product passes, agriculture, PIMS) where increased data sharing is desirable [64]. Based on statistical and economic principles, Kempny et al. explored data trustees' legal structure, design options, and practical requirements [37]. Blankertz extended on this, discussing technical and organizational measures for establishing trust [5]. Beise reported on data trustees as a measure of establishing data sovereignty, suggesting they could help individuals exercise their data rights and make more informed consent decisions [4].

Data Trustees in the US Houser and Bagby were the first to discuss "data trusts" in the US in depth [29]. They describe data trusts as a solution for increased data sharing that could act on behalf of larger groups to increase bargaining power with data users and regain control over the use of their data. Additionally they could store different types of data for various purposes [29]. Regan published a design for a public trustee and privacy protection regulation, to address the shortcomings of current information privacy legislation in the US [56]. She specifically highlights information asymmetries about the flow of personal data, a lack of transparency in data sharing, and a lack of knowledge about short and long-term impacts and costs for individuals. In light of increasing self-measurement and tracking (i.e., step count, sleep quality),

[1] See Article 10 ff. and Article 16 ff. Data Governance Act.

Kang et al. advocate for professional intermediaries ("privacy guardians") that manage the increasing amounts of data to mitigate the privacy risks caused by this self-surveillance [35]. Peppet made a similar point about privacy agents [53]. Lastly, it is debated whether data fiduciary duties should be imposed on data processors [73]. While seemingly related, this differs from our understanding of data trustees as they are intended to be intermediaries between the current data processors and the users but are not the data processors themselves.

Applications of Data Trustees As indicated before, data trustees have various conceivable applications. In medical research, trustee models such as the data trustee *CenTrust* of the German Bundesdruckerei [10], or the Center for Cancer Data by the Robert Koch Institute [11] have been in practice for multiple years. In Finland, Australia, and the UK, models that enable data sharing for medical research are used [21], [28], [51]. The automotive industry proposed *ADAX0*, a framework for automotive data access aiming to facilitate economic utilization and advancement through the exchange of vehicle data [24]. The common European Mobility Data Space (EMDS) is intended to facilitate the access, consolidation and exchange of data for more efficient, safe, sustainable, and resilient transport [16]. Initiatives such as the Willis Tower Watson (WTW) data trust pilot cover general consumer data [29]. Further conceivable areas include IoT or agriculture [64]. Even within the online sector, safeguarding privacy through data utilization has received attention, i.e., in the form of Google's Privacy Sandbox [25].

B. Willingness to Share Data

To our knowledge – we are the first to explore the factors influencing the willingness to use and acceptance of data trustees. Thus there is no explicit related literature. Instead, we gathered results from studies on the willingness to share data in adjacent contexts and with similar methodologies.

In a large-scale consumer study in Switzerland, Ackermann et al. used a vignette study design to investigate contextual factors for consumers' willingness to share data with companies [1]. They tested the influence of the type of data requested, the data usage purpose, the companies' industry sector, the type of compensation, and the level of anonymization. Anonymization yielded the most effective single influence factor on willingness to share, and sharing was liked better if the type of data requested matched the companies' core business. They also confirmed that incentives generally increase sharing willingness unless the data is perceived as sensitive. Leon et al. explored sharing willingness with online advertisers testing different levels of data retention, access to collected data, and scope of use of the data in a full-factorial design with 2900 online participants. More restrictive data-retention and scope-of-use policies increased the willingness to share data while possibilities to modify and review the data showed no influence [45]. Using a 3x4-factorial design with 80 participants, Ziefle et al. identified that the probability of being identified had the largest (negative) influence on the willingness to share data (49%), followed by the type of data (31%). Least important were the benefits of sharing the data (20%) [74]. Based on focus group findings, Rainie and Duggan implemented an online survey with 461 US adults, testing under which conditions sharing information is most accepted. They found that sharing health information with a health website was relatively acceptable while sharing data for free social media and sharing the data collected by smart thermostats were consistently disfavored [55].

Willingness to Share Medical Data Others studied the willingness to share data for specific purposes such as medical data. Nicholas et al. for example found that participants feel more comfortable sharing health information such as sleep or mood logs than personal data [50]. In a cross-country online survey with 8000 participants, Karampela et al. showed that about 30% of users would not share health data under any circumstances. The remainder in their study expressed willingness to share data for scientific purposes (22%), the public interest (12%), in return for compensation (14%), or individualized services (8%) [36]. Seltzer et al. investigated people's willingness to share data for research and their preferences regarding the use of data. Patients in an emergency department were asked whether they would donate 19 different types of data to health researchers. 65% of participants said they would be willing to share at least one of the digital data types listed in the survey. The willingness to pass on digital data after death was greater for all data types [62]. Grande et al.'s conjoint experiment with 3500 US participants showed that the type of data being shared had more influence on the willingness to share than the recipient of the data or the data usage purpose. Cluster analysis revealed similar groupings to Karampela et al. 10% of participants were universally opposed to sharing digital data under any circumstances, 30% were averse to sharing health data, and another 45% uncertain [26]. Kacsmar et al. analyzed users' perceptions of different data collaboration scenarios [33] showing that acceptance depends on the type of collaboration companies have among each other. In addition, participants find data sharing more acceptable if they are explicitly informed or have more control over whether their data is used by third parties. Ayalon et al. examined various aspects of the design of privacy-sensitive apps in the healthcare sector, such as COVID-19 contact tracing. Here, privacy-related attributes were identified as less important than other factors such as monetary and health incentives or the accuracy of the service [3]. Further studies include the willingness to share data from electronic health data records [39], and data from wearable health/activity trackers [58].

Willingness to Share IoT & Car Data Other domains are studied less. A market research survey on IoT device usage identified that 75% of consumers may be willing to share IoT data in exchange for discounts or payments [30]. A small study by Rickert et al. identified relationships between trust, the transparency of data processing, and the willingness to share IoT device data [57]. Schomakers et al. investigated Internet users' privacy preferences for data sharing, such as the motives, barriers, and conditions for privacy in data markets. The level of anonymization influenced the willingness to share data most, followed by the type of data [61]. For automotive data, Pugnetti et al. found that compensation, premium discounts, and specialized services can increase the willingness to share driving data with insurers. Data types not traditionally associated with insurance decreased sharing willingness [54].

III. Method

To study users' perceptions of data trustees and identify factors influencing the willingness to use them, we conducted online surveys in Germany between May and June 2023 and the US in January 2024. The following sections detail the conceptualization of vignettes and the survey structure, present demographic information and ethical considerations.

A. Vignette Design

For our research, we use vignette studies, presenting participants with descriptions of potential data trustees in various, randomly combined configurations, called "vignettes" or scenarios [20]. Each scenario is configured by combining variations (*factor levels*) of dimensions (*factors*), e.g., the storage location of data. While regular survey questions usually treat factors in isolation [20], vignettes more closely reflect the complexity of real-world decision-making, and thus offer greater external validity. The method helps to pinpoint which factors *actually* influence users' decisions and has been used in similar research on various topics [1], [40], [69].

Developing Concise Scenarios To ensure we capture all potentially relevant aspects of data trustees in our vignettes, we developed factors and factor levels in an iterative process. Drawing from legal literature [2], [6], [38], [64], [65] and research on user-centric privacy-enhancing technologies as well as research on data sharing [1], [45], [69], we first compiled a comprehensive list of potentially influential factors for data trustees' acceptance. With the entire research team, consisting of two legal experts and three security and privacy researchers, we then categorized and refined these factors through multiple discussion sessions over the course of six months. Additionally, we sought input from three external practitioners and data privacy experts. Validating our list in a workshop with ten usable security and privacy domain experts revealed no additional factors. Therefore, we consider our list to be exhaustive and concise.

Factors & Factor Levels Concretely, we evaluated the following eight factors with corresponding factor levels. We explain each factor and the reasons for considering it in detail. Table II in Appendix A shows the concrete wording for all factors.

Operator ***(Government*** | ***Business*** | ***NGO)*** Data trustees are still largely theoretical, with no established business model. The literature suggests that the organization operating the data trustee is critical to its perceived trustworthiness [47]. Establishing data trustees as a *government* entity integrated into public services would be the easiest way to handle their funding [6], [38], [41]. However, their success would then also depend on the overall trust the population has in its government. *Private* data trustees operating as regular companies with economic interests could bypass these political connotations. While selling data for profit presents its own legal challenges, B2B models in which third parties pay for access to high-quality, large-scale data appear viable and worth exploring [41], [64]. Non-profit organizations (*NGOs*) could form a potential middle ground between those two opposing positions, being independent of government control without pursuing economic goals.

User Anonymity ***(Raw data*** | ***Anonymized*** | ***Non-Personal)*** Users' perceived anonymity typically influences their technology acceptance [1], [74]. Additionally, different levels of anonymity come with varying legal implications. We distinguish *raw data* (non-anonymized), *anonymized* data sets, and *non-personal* data. Since raw data is considered personal and can lead to the identification of individuals, it falls under the GDPR, which mandates special protection.[2] However, the GDPR does not apply to non-personal or anonymized data, as these (in theory) pose no risk to individual privacy.[3] Anonymized data can theoretically be re-identified, leading to de-anonymization whereas non-personal data lacks any personal reference and thus impacts users' privacy the least.

Processing ***(Store*** | ***Aggregate*** | ***Analyze)*** In practice, data trustees can take on various roles, ranging from simply acting as intermediaries (*store*), to *aggregating* data from different sources, or even performing their own analyses (*analyze*[4]) [64], [71]. These data processing practices may influence users' perception of privacy invasion and affect the quality and nature of the data that can be made available to third parties.

Storage Location ***(GER/US*** | ***[EU]*** | ***Worldwide*** | ***[none])*** The storage location of data may not be users' primary concern, but it has substantial implications for applicable data protection laws, potentially affecting the level of trust users perceive. Therefore, we tested whether mentioning the storage location affects users' acceptance at all (*none*) and whether geographical proximity to their home country plays a role, comparing storing data in one's home country (*GER/US*) versus *worldwide*. For German respondents, we included the *EU* as an additional location.

Recipient ***(Research*** | ***Business*** | ***Law enforcement*** | ***Public)*** The core idea behind data trustees is to increase access to data for third parties, fostering a trusted exchange between science, industry, and society. Even before the concept emerged, the automotive industry had long debated whether car-generated data should be made accessible to companies, public institutions, or scientific research [38]. Similar discussions continue about which stakeholders should have access to medical data [64]. Existing regulations, such as the Federal Cancer Registry Act, already allow data to be made available for research purposes to public and private institutions and to individuals upon request.[5] Additionally, anonymized research datasets are publicly accessible on the registry's website.[6] Based on this, we selected the following recipients: *research* institutions, *private* businesses, *law enforcement* agencies, and the general public (*"everyone"*).

Access Type ***(Transmission*** | ***View-Only)*** We distinguish between data sets that are transmitted to the recipient (*transmission*) and data that remains with the trustee, with access granted only through a limited number of requests (*view-only*). These methods are inspired by the processes of the Federal Cancer Registry, which offers both a transmission option and a more secure alternative where data is provided in a controlled physical or virtual environment supervised by the Cancer Registry Data Center[7] [11], [64].

[2] See Article 2.1 of the GDPR.

[3] See Article 2.1 and Recital 26 of the GDPR.

[4] *Note*: For this factor, participants were informed that third parties would only have access to the results of the analyses.

[5] See § 8 para. 1 BKrebsregisterG.

[6] See § 8 para. 10 BKrebsregisterG.

[7] See § 8 para. 6 sentence 1 BKrebsregisterG.

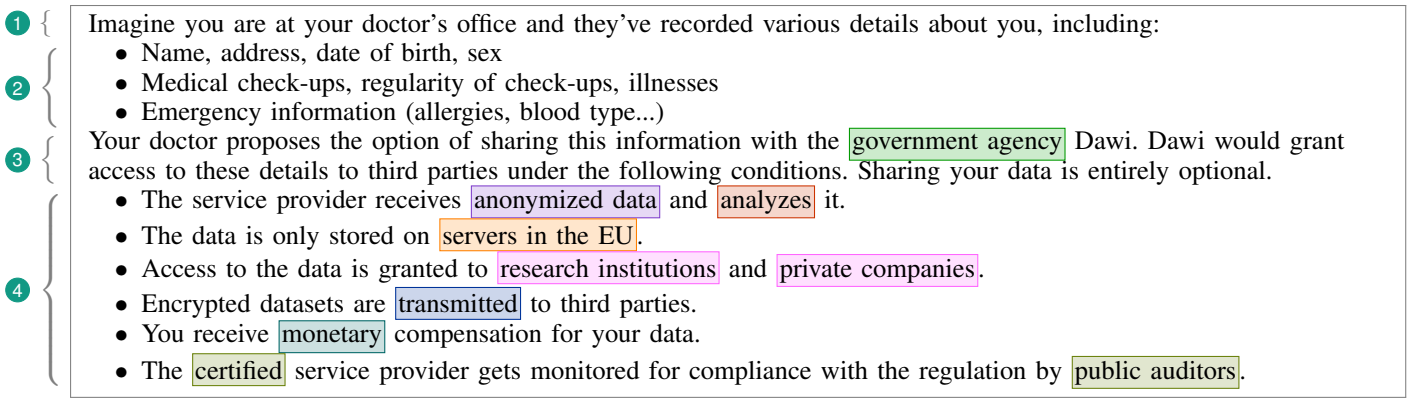

Fig. 1. Exemplary medical vignette with filled-in factor levels. Colors correspond to factors and were not displayed in the survey.

Benefits ***(Money*** | ***[Personal]*** | ***[none])*** Given the limited adoption of data trustees, creating incentives to promote data sharing is essential. Potential incentives to enhance data-sharing willingness have already been extensively researched [3], [30], [54], [74]. Our primary focus was therefore to determine whether offering incentives at all influences the acceptance, particularly in light of the heterogeneous findings on their effectiveness. We compared not mentioning any benefits (*none*) to receiving a monetary compensation (*monetization*), which is the most straightforward incentive. However, monetization raises ethical concerns about pressuring lower-income groups to compromise their privacy for financial gain [64]. Therefore, we also aimed to explore other incentives. Unfortunately, these are highly domain-specific and challenging to establish scientifically, which is why we only tested personal benefits (*personal*) for medical data, saying that using the trustee would benefit one's personal health.

Monitoring ***(Certificate*** | ***Government*** | ***Business*** | ***[none])*** Monitoring is crucial to mitigate risks associated with data trustees. Regulatory frameworks like the Data Governance Act (DGA) already mandate measures of oversight, such as a binding registration procedure for data intermediation services.[8] Thus, we investigate *private* auditors and *governmental* supervision. The literature further suggests that certification can strengthen trust and increase perceived transparency of data trustees, for example using "privacy labels" [60], [64]. Many regulations incorporate certifications or data protection seals, believing they will improve transparency and help ensure compliance.[9] We also explore whether *certification* as an alleged trust anchor actually delivers the expected benefits.

Scenario Descriptions To reflect the different potential applications of data trustees, we investigated four domains: ***medical*** data, ***automotive*** data, internet usage (***online***) data, and data generated through Internet of Things (***IoT***) devices. An exemplary vignette is shown in Figure 1. We introduced each scenario with a short description of the setting, i.e., instructing participants to 1 imagine being at their doctor's office or the car dealership to buy a new car. We then proceeded to list a few 2 domain-specific data points that may already be collected by the respective entity (i.e., their personal details and medical history by their doctor or their location and driving behavior by their car). 3 We then explained that a new service called "Dawi" could handle their data instead of the individual providers. Dawi would grant third parties access to the data under certain conditions and using it is voluntary. Dawi is a fictional name for the data trustee which we chose to omit the term "data trustee" itself to prevent biases evoked by trust-related wording, keeping it as neutral as possible. 4 Lastly, we listed the experimentally manipulated characteristics of Dawi. The full set of scenario descriptions for each domain can be found in the extended version [44].

[8]See Art. 10 DGA.

[9]For example Article 42 GDPR and recital 100 of the GDPR.

B. Survey Structure

Our survey is centered around the vignettes. Thus, after a short introduction, a consent form, and a few basic demographic questions (**D1** – **D3**), participants were directly introduced to the first vignette. We asked participants to imagine being in the situation described to them. Vignettes were drawn at random. Following, participants rated the likelihood of using the described data trustee service on a 7-point Likert scale (**S1**) and assessed how they perceived the services' usefulness for themselves and for society (**S3**). We repeated the same questions with a second vignette.

The second part of our survey then explicitly introduced the concept of data trustees, starting with a definition and explanation. Participants were instructed to answer questions independently of the specific vignettes previously presented, as we aimed to assess their general thoughts on the concept, regardless of domain or configuration. In multiple-choice and open-ended questions, we inquired about the positive or negative influence of all factors on participants' acceptance of data trustees (**G1.1** – **G1.7**). We also asked them to name the most influential factors (**G4** – **G5**) and factors we might have missed (**G2** – **G3**). The standardized privacy questionnaire IUIPC [48] assessed participants' privacy perceptions (**I**). We ended the survey with questions about their familiarity with data trustees (**P1** – **P2**) and privacy-protecting measures (**P3**) as well as further demographic information (**D4** – **D5**). The full survey can be found in Appendix B and was approved by our ethics board. All personal data was stored anonymously.

C. Data Collection and Sample Description

We used the panel provider *Bilendi* (formerly Respondi) to distribute our survey to participants. Prior to launching the survey, we thoroughly pilot-tested the implementation and

TABLE I. DEMOGRAPHIC INFORMATION FOR GERMAN AND US SAMPLES. PERCENTAGES ARE REPORTED WITHIN EACH SAMPLE.

	Germany ($n = 714$)		**United States** ($n = 1036$)		**Total** ($n = 1750$)	
Age (Mean)	48.3		47.5		47.8	
Gender						
Male	340	47.6%	504	48.6%	844	48.2%
Female	370	51.8%	526	50.8%	896	51.2%
Non-Binary	3	0.4%	4	0.4%	7	0.4%
Missing	1	0.1%	2	0.2%	3	0.2%
Education						
Low	179	25.1%	388	37.5%	567	32.4%
Average	387	54.2%	180	17.4%	567	32.4%
High	145	20.3%	467	45.1%	612	35.0%
Missing	3	0.4%	1	0.1%	4	0.2%
Income						
Low	407	57.0%	425	41.0%	832	47.5%
Average	161	22.5%	272	26.3%	433	24.7%
High	35	4.9%	154	14.9%	189	10.8%
Missing	111	15.5%	185	17.9%	296	16.9%
IT Background						
No	533	74.6%	656	63.3%	1189	67.9%
Yes	165	23.1%	359	34.7%	524	29.9%
Missing	16	2.2%	21	2.0%	37	2.1%

comprehensibility with the panel provider, 11 researchers from our social circles, and 100 soft-launch participants. After sanitizing the data, in total, we collected 714 responses in Germany and 1036 in the US. Our samples are census-representative according to age, gender, and educational background in both Germany and the US. The exact numbers are shown in Table I.

D. Analysis

Quantitative Analysis We refined our dataset by excluding participants who failed the attention check (GER: 86 | US: 188) and those faster than 40% of the median response time (GER: 75 | US: 75) as recommended by the panel provider. Using linear regression models with factors as independent variables, we clustered our analysis by domains and countries, fitting eight models to cover all combinations and extract scenario-specific effect estimates. We also performed country-specific one-way ANOVAs with Tukey-corrected post-hoc tests to capture differences in agreement between domains.

Pre-testing of agreement ratings revealed significant differences between countries and scenarios. Hence, we report separate models for domain and country, which allows to i) identify culture-specific factors influencing the willingness to use data trustees, and ii) respect possible differences in perception or sensitivity of data from different domains. Model assumptions were upheld, with only slight violations of normality, which regression models can robustly handle. Models explained 3-16% of variance in Germany and 1-9% in the US.

Qualitative Analysis We iteratively coded open-ended questions (**G2** – **G5**) of the German dataset using an inductive coding strategy. We applied the German codebook to the US data but flexibly added codes if needed. Coding was conducted by three researchers: two with interdisciplinary backgrounds in information security, psychology, and law, and one with a background in law. Researchers one and two coded the German dataset, researchers two and three coded the US data. Initially, each researcher independently coded the first 25% of responses. Subsequently, researchers met to discuss and finalize a codebook for the remaining responses. Finally, we extensively reviewed and refined the codebooks, achieving full agreement for both countries' datasets. Due to sparse responses to questions **G2** and **G3** we only report selected examples.

E. Limitations

Our samples were representative for the population in Germany and the US. However, due to the nature of online studies, viewpoints of individuals who do not use the Internet may be excluded, slightly limiting the overall generalizability of the results. The findings also may not generalize to populations in other cultures, as statistical generalization is inherently limited to the sample population. While we took great effort in crafting comprehensive vignettes and dimensions by consulting various experts, we cannot entirely rule out having overlooked potentially influential aspects. While vignette studies are a great tool to elicit initial ideas and indicate directions regarding broad influential factors, they cannot provide a complete interpretation. Further analyses specified to the domains are necessary to investigate influences in depth. Additionally, vignette studies, while attempting to simulate real-life situations by prompting participants to imagine themselves in the described scenarios, only approximate real-life behaviors. Therefore, responses in real-world settings may differ from those in study conditions. Lastly, participants may have been dissuaded from expressing opinions that go beyond the presented vignettes.

IV. RESULTS

To answer RQ1, we begin by reporting users' general perceptions and impressions of data trustees, including their overall willingness to use. Following that, we investigate factors individually, and discuss their influences on participants' acceptance of data trustees, to answer RQ2. We enrich our statistical results with qualitative findings from open answers, reporting them as follows: "*Quote*" (**M**$_{\mathbf{G}}$15), where "M" represents the domain (**M**edical, **A**utomotive, **O**nline, **I**oT), "G" the sample (**G**ermany, **U**S), and the number is the participant code. The full codebook is provided in Appendix C.

A. General Perceptions of Data Trustees (RQ1)

As data trustees are a new concept, we first report users' general opinions, independent of specific factors. Figure 2 shows the code frequencies in the open answers to provide a high-level impression of the relative importance of each topic.

Familiarity with Data Trustees Unsurprisingly, 71% of all participants reported having never used a data trustee before (US: 60%, GER: 86%), and an additional 18% were unsure (US: 23%, GER: 11%). The familiarity with the concept differed between countries: While only 14% of German participants had heard of data trustees before, almost twice as many US participants said so (27%). The remainder were either unsure (US: 19%, GER: 20%) or had never heard the term (US: 54%, GER: 66%). While as expected, the overall reported familiarity was low, we were surprised

that still around 15% to 30% expressed some familiarity, especially in the American sample.

Overall Willingness To Use Reporting the overall mean willingness to use data trustees provides an initial impression of participants' general acceptance of the concept. Note however that these results are an aggregation across all experimental vignettes including those that influence acceptance significantly. Hence these results do not represent the agreement to an ideally configured data trustee.

Across all domains, the willingness to use was significantly higher for US than for German participants ($p < .001$). The German sample gave an average rating of $m_{\text{GER}} = 3.1$ while US participants rated their willingness at $m_{\text{US}} = 3.8$. Within samples, ratings were relatively similar between domains. In Germany, online data received the highest ranking ($m_{www} = 3.3,\ sd_{www} = 2.0$) but automotive ($m_{car} = 3.2,\ sd_{car} = 2.1$) and medical data ($m_{med} = 3.1,\ sd_{med} = 2.1$) were rated only marginally lower. For IoT however, the agreement was substantially lower with only $m_{IoT} = 2.8$ ($sd_{IoT} = 1.9$), which is significantly different from automotive ($t(705) = -2.47, p < .05$) and online data ($t(701) = -3.23, p < .01$). In the US, ratings were homogeneous, with automotive ($m_{car} = 3.9,\ sd_{car} = 2.1$) and online data ($m_{www} = 3.8,\ sd_{www} = 2.0$) being slightly higher and medical ($m_{med} = 3.7,\ sd_{med} = 2.1$) and IoT data ($m_{IoT} = 3.7,\ sd_{IoT} = 2.1$) being slightly lower. These differences are not statistically significant.

As shown in Figure 3, a substantial portion of participants in both Germany and the US answered they would "definitely not agree" to use the data trustee. This tendency towards the extreme low end is particularly prominent in the German sample. The reluctance to use was similarly strong across domains however for IoT we observed the strongest disapproval (40.5%) out of all domains. The remainder of answers roughly followed a normal distribution, except in the US, where many participants also selected the highest possible agreement – almost three times as many as in Germany.

General Impressions Some participants expressed rather high-level reasons for their (un)willingness to share data with a data trustee. As the acceptance ratings suggested, a group of users (the "***Refusers***") seems entirely unwilling to use the data trustee. This group was bigger for German than for US participants but similar reasoning was identified nevertheless. While some generically expressed just *not* being *interested* (GER: 45 (3%) | US: 82 (4%)), the majority mentioned privacy and trust issues, i.e., calling it "*Shady and untrustworthy*" (O_U14). The expressed distrust covered different dimensions: 1) some do not trust the data trustee as an actor itself, suspecting it to have malicious intentions, 2) others question the data trustees' ability to protect their data but do not suspect it to purposely act malicious, and 3) some believe security technology is inevitably flawed making it impossible to protect the data, even if the data trustee has the best intentions.

On the opposite end of the spectrum, the ***Accepters*** were in favor of the concept, some again without explicitly explaining why, i.e., only stating the data trustee seems *interesting* (GER: 33 (2%) | US: 156 (7%) | "*I like the idea*" (O_U259)). This group was substantially larger for US participants, indicating a more positive and open attitude. Others expressed *trusting* the described service because it sounds legitimate

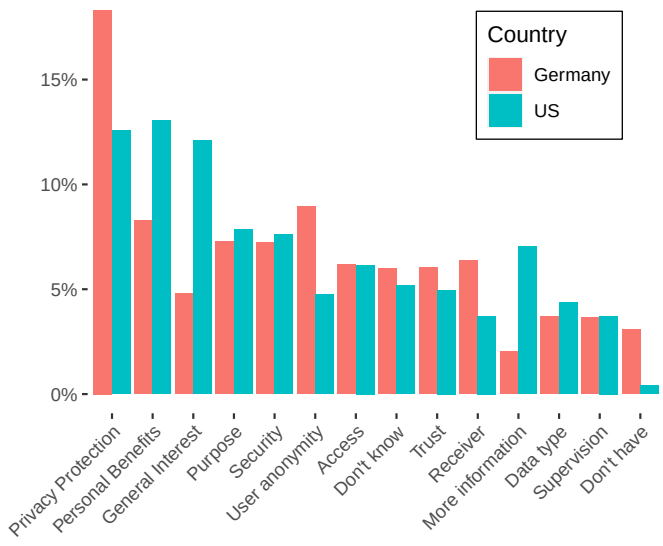

Fig. 2. Frequencies of codes from open answers for Germany and US. Codes with $< 5\%$ occurrence are excluded in this graphic. Refer to Appendix C for the full codebook.

(GER: 33 (2%) | US: 44 (2%)). Some claimed to agree because they had *nothing to hide* or out of resignation, thinking the data was already being used anyways, i.e., "*In the end, they will do it whether I agree or not*" (M_G295).

A substantial portion of participants – the ***Undecided*** – were simply not sure about their opinion on the concept (GER: 97 (6%) | US: 113 (5%)) or they explicitly requested *more information* on the topic before being able to make an informed decision ("*I would have to read up on that*" (O_G220)). Similar user clusters have been identified in related work on specific domains, such as data sharing for medical treatments [26]. Many participants also mentioned aspects of ***security*** (GER: 117 (7%) | US: 165 (8%)) either positively noting that the data is encrypted and it "*seems to be safe*" (O_G316), or saying Dawi is *not secure* due to fear of data theft and misuse or distrust in security technology overall "*technology is not without fail*" (I_U927). A small group of participants demanded to be in charge of what happens with their data exactly (***Sovereignty***: GER: 36 (2%) | US: 63 (3%)) either just wanting to "know" what is happening with the data and to whom it is going ("*I want to know where my data is sold to*" (I_U747)), or wanting to have access to mechanisms for controlling third-party access ("*I have no control what is being done with the data*" (O_G348)).

Perceived Utility In question **S4**, we additionally investigated participants' perceived utility of data trustees. The utility for society was rated mediocre in both countries however it was higher on average in the US ($m_{soc} = 3.96,\ sd_{soc} = 1.98$) than in Germany ($m_{soc} = 3.51,\ sd_{soc} = 1.88$). Additionally, across samples and domains the utility for society was always deemed higher than the utility for oneself (GER: $m_{soc} = 2.93,\ sd_{soc} = 1.98$ | US: $m_{own} = 3.55,\ sd_{own} = 2.08$). In the US, utility ratings were consistent across all domains, participants seemingly did not differentiate their assessment by data type at all. German participants however showed more nuance in their answers. Overall, medical and automotive

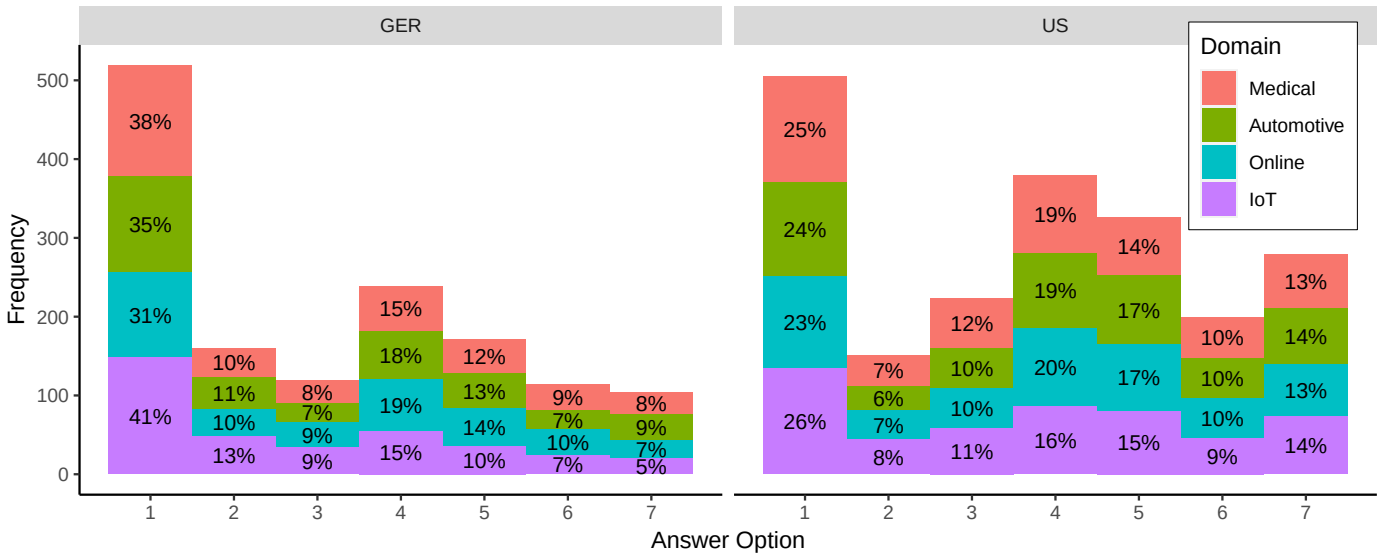

Fig. 3. General willingness to use the data trustee. Combined results for vignettes one and two. Percentages are calculated within the respective population. Answer options ranged from 1-*Very unlikely* to 7-*Very likely*.

data received the highest utility for society ratings, IoT the lowest. The discrepancy between perceived utility for oneself and for society was particularly strong for medical data (difference of 0.73) but similarly noticeable for automotive and IoT data. The online domain shows a smaller divide between personal and societal utility which is less surprising in light of their overall lowest ratings, which also explains the lower willingness to use. Some participants also spoke about the *utility* of using the data trustee in the open answers (GER: 45 (3%) | US: 126 (6%)). Another group which was bigger in Germany explicitly pointed out the potential *societal benefits* (GER: 73 (4.5%) | US: 44 (2%)) of a data trustee. This group was overrepresented for the medical domain as participants seem to have a clearer understanding of the utility of medical data than for the data from other domains.

B. Factors Influencing Willingness to Use (RQ2)

In the following, we report the descriptive analysis of our general questions (**G1.1** – **G1.7**, see Figure 4). Additionally, we report the statistical influence of experimental factors on the agreement. Refer to Table IV in Appendix D for the full regression results. We substantiate these findings with the open answers or contrast them when necessary. This integrated reporting allows to develop a more in-depth understanding of the findings and their implications for data trustees in practice.

1) User Anonymity and Data Processing Practices:

Impact of User Anonymity Anonymity considerably shaped perceptions. When asked directly, data being anonymized received positive ratings in both Germany and the US (GER: ⊖ 21%, ⊕ 53% | US: ⊖ 22%, ⊕ 50%). Conversely, non-anonymized raw data encountered resistance from German participants with 70% perceiving it negatively while US participants appeared rather indifferent answering equally positive (34%) and negative (36%). In the regression, US participants significantly preferred medical data being transmitted *anonymously* over it being transmitted as *raw* data ($\beta = 0.28, p < .05$). Germans significantly preferred *non-personal* online ($\beta = 0.52, p < .001$) and IoT data ($\beta = 0.28, p < .05$), as well as *anonymous* online data ($\beta = 0.04, p < .01$). Anonymity also played a noticeable role in participants' open answers. However, the topic was substantially more present among the German responses than among those of US participants. Across all domains, almost double as many German participants expressed a wish for anonymity than Americans did (GER: 145 (9%) | US: 82 (4.5%)). In both countries, the code was given most often for online data and least for medical data.

Impact of Data Processing We compared only *storing* the data, *aggregating* from various sources, and *analyzing* but only providing third parties access to analyses of the data. When asked directly, opinions about only storing the data were homogeneous. In the German sample, about equal parts answered that only storing the data would positively (28%), or negatively (36%) influence their acceptance. Slightly more participants leaned towards the positive side in the American sample, where 42% expressed positive and only 30% negative attitudes. In the US, aggregating and analyzing the data yielded very similar ratings to only storing (*aggregate*: ⊖ 27%, ⊕ 46% | *analyze*: ⊖ 30%, ⊕ 45%). In the German sample, aggregating or analyzing the data also yielded similar ratings, however in this case slightly worse than only storing (*aggregate*: ⊖ 56%, ⊕ 18% | *analyze*: ⊖ 50%, ⊕ 24%).

Our regression only identified significant influences in the medical domain for the German sample. *Analyzing* the data compared to only *storing* it yielded significantly higher agreement ratings ($\beta = 0.28, p < .05$). As we consider 'analyzing' to be more privacy invasive than just storing the data, this finding is counterintuitive. We identified two potential explanations: First, the mere storage of data might be seen as meaningless or pointless. This view is supported by our open-ended responses, where many participants indicated that a clear purpose for data collection influenced their willingness to use the data trustee.

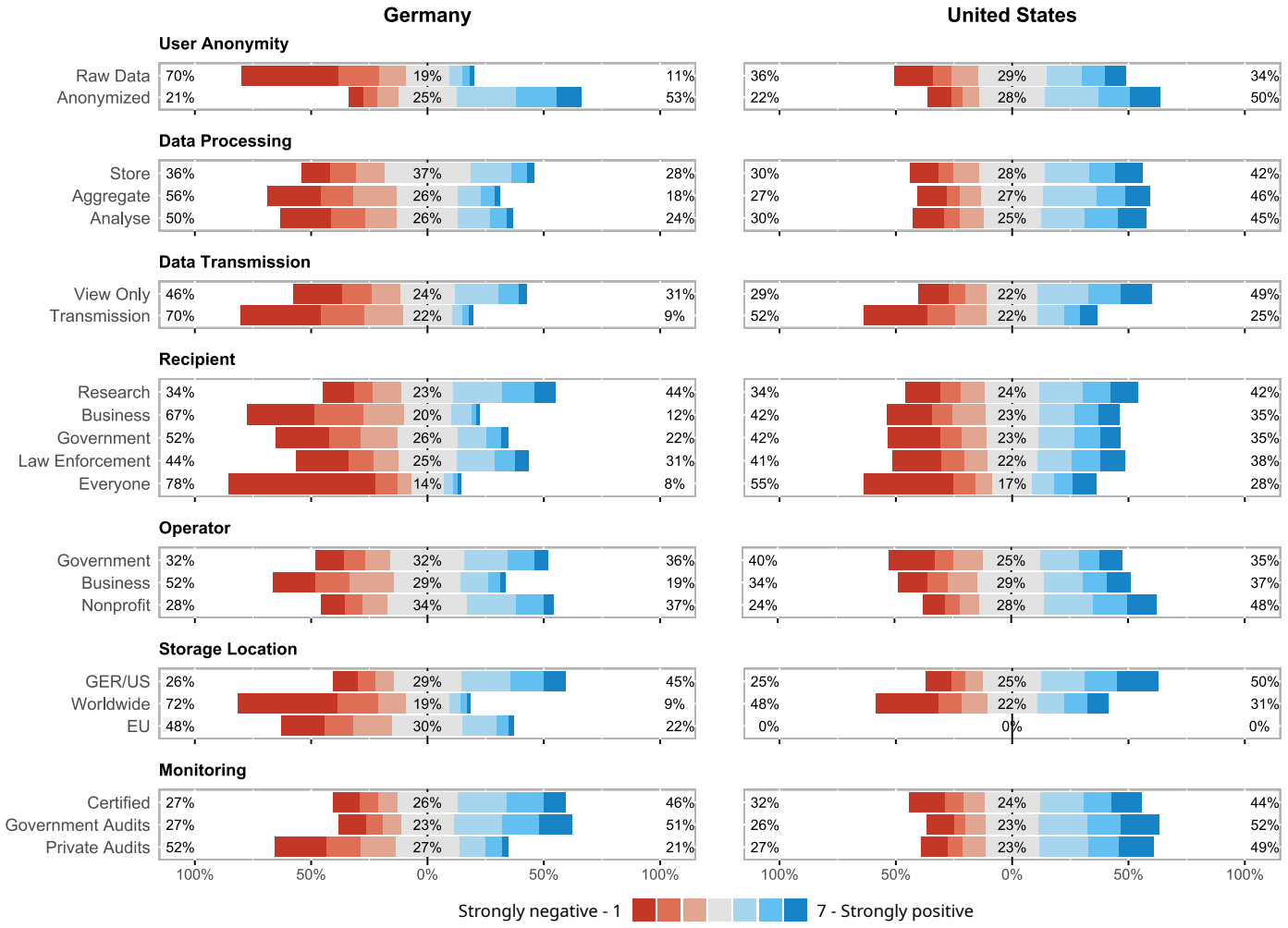

Fig. 4. Positive and negative influence of factors on willingness to use a data trustee (**G1.1**-**G1.7**).

Secondly, participants may have been reassured by the phrasing "third parties only get access to analyses," which might be perceived as a more privacy-preserving option.

Summary Anonymity plays a crucial role in shaping how users perceive data trustees, with many preferring anonymized and non-personal over raw data. This preference is generally stronger in Germany, while U.S. participants are more indifferent. Overall, the findings highlight the importance of user anonymity in fostering trust and acceptance of data trustees, especially for sensitive data types like medical and online data.

2) Third Party Access:

Impact of Access Management For the access factor, we compared *view-only*-access where the data stays at the trustee at any time to data sets being transmitted to the third party receivers (*transmission*). In both Germany and the US, transmission to third parties was rated (very) negatively (GER: − 70%, + 9% | US: − 52%, + 25%) whereas only allowing view access at the trustee yielded more positive reactions. While for Germany the results were balanced between positive and negative ratings (− 46%, + 31%), US participants indicated a clear positive tendency (− 29%, + 49%). Despite these partially strong averse ratings, the only significant influence in the regression analysis was German participants preferring *view-only* access over *transmission* for medical data ($\beta = -0.23, p < .05$). From a data protection standpoint, it is sensible to prefer that only the data trustee holds the data. However, meaningful analyses often rely on combining data from multiple sources, raising doubts about the practicality of avoiding data transmission to third parties.

Participants also referenced access in the open responses – mostly connoted negatively. Several participants explicitly expressed disliking data being accessed by third parties overall. While they appeared to be accepting of the data trustee itself holding the data, giving access to, especially unknown third parties, was disapproved of (GER: 83 (5%) | US: 94 (4%) | "*I don't want my data to be passed on*" (M_G441)). Participants, however, never distinguished between the specific *type* of access. They either disapprove of third-party access entirely or they approve of it but without specifying the type of access they would prefer. Entirely disallowing third-party access, however, would render data trustees useless, as its intended purpose is precisely to serve as a trustworthy intermediary between you and potentially untrustworthy third parties.

Impact of the Recipient Out of the five types of recipients (*research*, *government*, *law enforcement*, *businesses*, and "*everyone*"), research institutions were evidently favored in both Germany and the US. However, positive and negative ratings were fairly balanced, with only a small positive tendency (GER: − 34%, + 44% | US: − 34%, + 42%). In other words, while research institutions are the least disliked out of all recipients, solely sharing data with them will not convince users who are fundamentally opposed to data sharing. Contrary to that, the *government* as a receiver was noticeably disliked in Germany – over half of the participants indicated a negative influence, while only 22% perceived it positively. The US sample, however, was equally comfortable with government receivers as with research institutions (− 42%, + 36%). This is surprising, considering that people in the US generally have less favorable opinions of the government than people in Germany. Access for businesses was disfavored even stronger by the Germans (− 67%), while Americans again gave similarly positive and negative ratings (− 42%, + 35%). Lastly, providing access to the general public was met with the strongest disapproval in both countries. However, the group of opposing voices in Germany (− 78%) was still substantially larger than in the US (− 55%). To our surprise, access for *law enforcement* was perceived relatively positive. The number of positive ratings was higher than for businesses and other governmental organizations in both countries (GER: − 44%, + 31% | US: − 41%, + 38%), almost as high as for research institutions. This might indicate an openness to making data available for aiding crime prevention.

Despite strong (negative) sentiments towards specific types of receivers (at least in the German sample) our regression analysis only yielded few significant results. '*Everyone*' having access was significantly disfavored for online data in Germany ($\beta = -0.35, p < .05$) and for medical data in the US ($\beta = -0.27, p < .05$). Additionally, in the medical domain, *law enforcement* was disliked significantly compared to *research* institutions ($\beta = -0.41, p < .001$). While this finding is not unexpected in itself, it is surprising that we only see this influence in the medical domain. Related research found that data sharing is usually preferred if the shared data has a logical connection to the receivers' core business [1]. For medical data, this perceived mismatch might be particularly strong as it is unintuitive how someone's medical record is relevant for law enforcement.

The open answers reflected similar themes. More German than US participants commented on the recipient (GER: 103 (6%) | US: 81 (4%)). If mentioned at all, private companies, law enforcement agencies, and everyone getting access (GER: 51 (3%) | US: 33 (1.5%) | "*everyone who is interested can get access - big downside*" (O_G281)) was unanimously disapproved of. Few positively highlighted when access was limited to research institutions (GER: 18 (1%) | US: 21 (1%)).

Summary Users generally seem to trust the data trustee as an institution but have significant concerns about third parties accessing data through it. While US participants rated most recipients similarly, Germans preferred research institutions over government and business access, indicating concerns about privacy and misuse. Making data publicly available was unanimously disapproved.

3) Operational Aspects:

Impact of the Operator Participants' judgments differ only slightly when comparing the *government*, private *businesses*, and *NGOs* as potential operators. In the US, NGOs received the most positive ratings out of all operator types, with double as many positive (48%) as negative ratings (24%). Germany showed a similar tendency though much less pronounced (− 28%, + 37%). In contrast, the German sample disfavored businesses (− 52%, + 19%) which was not reflected in the American data (− 34%, + 37%). The government evoked equally positive and negative sentiments in both countries (GER: − 32%, + 36% | US: − 40%, + 35%). Our regression analysis found no significant influence of the operator across domains in both countries and the topic was scarcely mentioned in the open-ended responses. These combined findings suggest that the identity of the data trustees' operator and potential preconceived notions about their intentions play a secondary role in users' decision to trust and use the data trustee. This also indicates the potential to consider various configurations of data trustees.

Impact of the Storage Location Unsurprisingly, most participants favored storing data in their home country or were at least indifferent to the matter (GER: − 26%, + 45% | US: − 25%, + 50%). Interestingly, Germans also showed a certain mistrust towards storing data in other EU countries (− 48%, + 22%), despite GDPR applying to all EU member states. Storing globally elicited adverse reactions, especially in Germany, where nearly $\frac{3}{4}$ rated it negatively; almost 50% even expressed *strong* disapproval (GER: − 72%, + 9% | US: − 48%, + 31%). This contrasts with the regression analysis, which did not provide statistical support for this strongly negative view. Neither Germany nor the US showed any significant influence of the storage location in any domain. In the open answers, very few addressed the storage location at all (US $< 2\%$ | GER $< 3\%$). Those who did mention the storage location reflected similar trends as observed in the direct questions – preferring storage in one's home country while disapproving of global storage. We interpret these findings as a reflection of the privacy paradox, where individuals articulate strong opinions when asked directly, but these opinions do not translate into actual behaviors in real life.

Impact of Monitoring & Certification Participants' views on monitoring were relatively similar in both countries, except for private sector audits. Here, Germans expressed a clear negative tendency with nearly twice as much disapproval compared to Americans (GER: − 52%, + 21% | US: − 27%, + 49%). This finding aligns with Germans' generally more cautious stance towards private sector involvement across different stakeholders, potentially reflecting different cultural attitudes towards the industry. Certification and governmental supervision, however, received positive ratings in both countries with approximately half of the participants rating favorably (*certification*: GER: − 27%, + 48% | US: − 32%, + 44% | *government*: GER: − 72%, + 51% | US: − 26%, + 52%). Once again, the regression model revealed no significant influences and the topic was scarcely mentioned in the open answers. Some participants positively pointed out the mechanisms for independent monitoring and supervision of the data trustee (GER: 45 (3%) | US: 81 (3%)) – in Germany, particularly if it was done by a government authority.

Summary The findings suggest that the type of organization serving as a data trustee and the data storage location may not be as important to users as commonly believed. Although participants express preferences, such as favoring NGOs over businesses or desiring data stored domestically, these opinions do not significantly influence their decisions, indicating limited practical importance. Furthermore, while certification and monitoring are generally viewed positively, they appear to provide little real assurance, suggesting they might not be as effective for building trust as anticipated.

4) Impact of Incentives:

Prior research repeatedly shows that when users are asked *explicitly*, (monetary) incentives supposedly increase the willingness to share data [3], [30], [54], [74]. We were thus particularly interested in whether this would also be reflected in the implicit measurement through the regression, rather than only appearing when users are explicitly asked. In fact, *monetary* benefits significantly increased the acceptance for medical and automotive data in Germany (*medical:* $\beta = 0.39, p < .01$ | *automotive:* $\beta = 0.26, p < .05$) and for medical data in the US ($\beta = 0.29, p < .01$). Additionally, *personal* medical benefits such as improved treatments increased the willingness to use the data trustee significantly, at least in Germany ($\beta = 0.31, p < 0.05$).

The open answers show substantial differences between the US and Germany. In the US sample, personal benefits were mentioned in 50% more cases than in the German sample and it constitutes the most prominent theme overall, even slightly exceeding privacy protection, whereas in Germany, personal benefits only take third place (GER: 134 (8%) | US: 282 (13%)). Comparing the domains, benefits were mentioned most for automotive data in both countries. Interestingly, for the medical domain, benefits were mentioned least frequently which seemingly contradicts the regression result. The majority of codes were focused around monetary incentives (GER: 84 (5%) | US: 143 (7%) | "*I won't give data without compensation*" (A_G189)) however others also mentioned personal benefits such as increased 'data availability,' i.e., not having to transfer medical records between doctors manually or improved speech recognition quality of voice assistants. Lastly, incentives mentioned also included societal benefits which we already discussed closely in Section IV-A.

Summary (Monetary) incentives increase the willingness to share medical and automotive data, especially in Germany. While US participants emphasized personal benefits as key motivators for using data trustees, Germans were less convinced, particularly when they had privacy concerns, indicating cultural differences in data-sharing motivations.

5) Impact of Demographic Factors:

Demographic factors partially influence the willingness to use data trustees. Unsurprisingly, higher *privacy concern* (IUIPC scores) negatively predicted agreement. In Germany, medical ($\beta = -0.22, p < .001$), IoT ($\beta = -0.19, p < .01$), and online data ($\beta = -0.16, p < .01$) yielded this significant effect, implying that the more privacy-concerned people are, the less they will agree to use data trustees in these domains. Interestingly, the American sample showed no such effect, despite the average privacy concern (IUIPC scores) being similar in both countries (GER: $m = 0.98$, US: $m = 0.97$). This reinforces the impression that for Americans, the concept of a data trustee is less connected to privacy decisions and concerns than for German participants.

Age showed a significant negative influence for IoT ($\beta = -0.22, p < .001$), automotive ($\beta = -0.14, p < .05$), and online data ($\beta = -0.23, p < .001$) in Germany and for IoT ($\beta = -0.24, p < .001$), medical ($\beta = -0.18, p < .001$), and automotive data ($\beta = -0.21, p < .001$) in the US. However, the effect sizes are relatively small, meaning a 60-year-old will, on average, be slightly less likely to use a data trustee than a 20-year-old, but the overall influence is negligible. Moreover, age effects are often proxies for natural developments over time such as attitude changes or knowledge acquisition. Nonetheless, this finding mostly aligns with related work, where older age usually correlates with less willingness to share data [1], [54].

In the US, *women* were less willing to share medical data than men ($\beta = -0.13, p < .01$). Related work is discordant on that matter. While some studies generally suggest that women are less willing to share data, especially sensitive information like medical records [1], [46], [74], others identified the opposite [54], or no influence of gender at all [26].

Having an *IT background* positively impacted the willingness to use in the US (*medical*: $\beta = 0.35, p < .001$ | *IoT*: $\beta = 0.39, p < .001$), which may reflect professionals' technical expertise and ability to identify data trustees' potential to mitigate problems that exist in current data processing practices. Similarly, a better understanding of the potential benefits of data sharing might increase the willingness to share. Additionally, technically savvy individuals are typically also more open to new technological developments and more able to adequately assess the risks and benefits compared to the average user. Further effects of income or education, were deemed as statistical artifacts due to their small effect sizes.

Summary Although demographic factors, like age, gender, and IT background show some statistical influence, there is no clear pattern, and the effect sizes are small. Thus, these findings should not be overinterpreted and are unlikely to be relevant in practice. We only observed clear differences for privacy concerns (IUIPC), further underscoring the differing relationship data trustees and privacy concerns seem to have in Germany compared to the US.

V. Discussion & Recommendations

We studied users' perceptions and willingness to use data trustees. Next, we highlight major learnings about this new concept and derive recommendations for the successful development of data trustees in practice.

A. General Perceptions

Data trustees are a recent development, primarily in the early discussion or pilot testing stages. Hence, few participants reported familiarity. The majority expressed some openness to

the idea, highlighting aspects of trust and perceived personal and societal utility. Nevertheless, about $\frac{1}{3}$ of users expressed strong opposition, regardless of the data trustees' specific configuration due to overarching privacy concerns and fear of (government) surveillance when data is aggregated in a single institution. On one hand, this finding is not surprising, as new (technological) advancements are often accompanied by concerns and preconceptions which may evolve as awareness grows. However, our analysis indicates that these averse user clusters will likely persist even after data trustees gain prevalence as users argued quite strongly. Related work on data sharing supports these findings. Karampela et al. for example showed that 30% of users are unwilling to share health data under any circumstances [36]. Grande et al. revealed that 10% of users are *universally* opposed to data sharing and another 30% strongly averse, independent of the data type [26].

Implication: Establishing data trustees as trusted institutions will require significant effort, as most users are unlikely to adopt them without active persuasion. Current services that collect large amounts of data often fulfill a clear personal need for users (e.g., Instagram for socializing, Google Maps for up-to-date navigation). However, that is not the case for data trustees. Identifying motivators for businesses to establish B2B data trustees will likely be easier than convincing end users.

B. Differences between Domains

Participants seemed to have a relatively holistic understanding of *medical* data. Despite its sensitivity, they were more open to sharing it than other types of data, which supports related work [36], [46]. In other domains, however, users' understanding of the implications of data collection seemed limited. This was reflected by poor model fit and few significant factors in the online, automotive, and IoT data sets. Such results typically indicate that participants decided based on uninformed "feelings" due to a lack of understanding instead of making informed rational decisions. The willingness to share *IoT* data was the lowest in both countries and the desire for privacy appeared to be a dominating factor, weighing heavier than for other data types. Users seemed to perceive IoT data as particularly sensitive but, unlike in the medical field, do not recognize a clear purpose for collecting the data. In contrast, *automotive* data received the overall highest acceptance ratings. Yet, the regression had the worst model fit across all domains. Participants seemed to have little understanding of how the data their cars generate potentially affects their privacy. Despite sensitive location data being shared (typically one of the most privacy-invasive data types [74]), not even the general privacy concern (IUIPC) showed impact on the acceptance ratings here. A recent study by Mozilla supports these findings, reporting particularly low awareness of data collection in the automotive sector compared to other domains [13].

Implication: Data trustees should respect the different perceptions and mental models users hold towards different data types and even legislation should possibly be designed in a domain-adjusted manner. Additionally, data trustees should create a logical context for the data collection as consumer research shows that users are more willing to share data with companies when the shared data has a logical connection to its receiver (i.e., shopping behavior with a shopping website and financial information with a bank) [1].

C. Cultural Differences

Comparing the results between Germany and the US revealed clear differences in how data trustees are linked to privacy concerns. Despite similar levels of privacy concern in both countries, the IUIPC only showed significant influence on the acceptance in the German sample. Similarly, Germans emphasized personal privacy, anonymization, and security and expressed aversion to third-party access in their open answers. In contrast, the US population seemed generally more open to data sharing. Additionally, the acceptance of private companies was higher in the US. Germans expressed aversion towards private businesses as operators, for monitoring purposes, and as data recipients and partially disapproved of the data economy. Other studies indicate similar geographic and ethnic relations. Kim et al. for example showed that individuals with Asian ethnic backgrounds are more willing to share health information than other ethnicities [39]. Overall, these findings reflect the pervasiveness of privacy as a concept in the German culture which is reinforced by the omnipresence of privacy regulations like GDPR in Germans' everyday lives.

Implication: For data trustees, it is hence crucial to identify culture-specific factors and respect local differences. It is unlikely that the same legislation can be applied in diverse cultural contexts, as varying societal norms and values shape individuals' perceptions of privacy and data sharing. Furthermore, engaging with local stakeholders and understanding the unique needs and concerns of different populations can lead to more effective communication strategies, fostering trust and encouraging adoption.

D. Identifying Critical Factors

The following details key factors and recommendations for legislators to support successful data trustee adoption.

Guarantee Privacy and Anonymity Anonymity is crucial for users. They prefer sharing anonymized and non-personal datasets and frequently cited privacy concerns as a reason for disliking the idea of data trustees. The concept inherently carries risks as aggregating large amounts of data in a single institution also increases the chance of de-anonymization [22], [49]. This risk increases the more data is collected, so paradoxically, the better the data trustee gets the greater the threat to users' privacy becomes. It is thus crucial to implement adequate technical and organizational measures to mitigate risks associated with centrally storing user data. Additionally, initially establishing data trustees only for non-personal data might motivate users to try the concept. This would also simplify implementation, as fewer regulations apply (e.g., GDPR does not apply to non-personal data.)

Demonstrate Utility The study demonstrated two essential aspects: 1) many recognize the altruistic component of data trustees, even if it is not explicitly mentioned, and 2) many users consider the use of data trustees despite no obvious conceivable benefits. Unfortunately, public debates often overlook these positive aspects and instead focus heavily on addressing privacy and security concerns. However, data collection is not inherently problematic. It only becomes an issue when exploited by greedy or mal-intended actors and when data is not protected properly. As the success of data trustees heavily depends on people's intrinsic motivation to share their data,

they must communicate the deeper purpose of data collection and demonstrate how they benefit society. In domains where users struggle to grasp the (societal) potential of data analyses, communicating clearly is particularly important. The discussion should also avoid focusing solely on privacy concerns and counter the prevailing notion that data collection is primarily driven by financial gain.

The Question of Monetary Benefits While financial incentives increased the acceptance for medical and automotive data and were frequently mentioned in the open answers across all domains, their influence was not as pronounced as commonly believed. Notably, significant influences were not observed for online and IoT data. Some participants even expressed distrust towards monetary incentives, rejecting the use of their data for profit and expressing general skepticism towards the data economy. This motivates the question: Is focusing on monetary compensation even meaningful? So far, literature seldom investigated the specific amounts users envision as adequate compensation for their data. Most studies vaguely talk about unspecified "*compensation*" or "*discounts*" [1], [30], [36], [54], [74]. However, if users' monetary expectations exceed realistic limits, this mechanism may ultimately be impractical anyway. Before making a definite statement about the effectiveness of monetary incentives, research should explicitly investigate what people consider adequate monetary rewards.

Gradually Expand Third-Party Access Users tend to trust the data trustee as an institution more than the stakeholders who would gain access through it. While all participants strongly rejected the idea of making data publicly accessible, Americans rated most other potential recipients similarly. In contrast, German users clearly favored research institutions over governments and businesses, reflecting mistrust and concerns about misuse by government and commercial actors. Therefore, introducing data trustees gradually, initially limiting access to research organizations, before extending access to less trusted institutions could help build users' trust.

Recognize Limited Impact of Organizational Factors Literature suggests that organizational aspects like storage location, operator, and oversight are key to building trust in data trustees [60], [64]. These aspects are also prominent in regulations, discussing measures like "privacy labels," data protection seals,[10] and public registries.[11] However, in our regression analysis, these factors played a minor role. While participants expressed strong opposition to global data storage or private operators, we found no significant influence on the acceptance. The same applies to monitoring. Despite positive feedback from participants, these measures did not show measurable impact. This suggests that while users hold strong opinions about certain operational aspects, these factors are not necessarily decisive and other factors outweigh them in the decision-making process. Hence, trust-building measures alone will likely be insufficient to establish data trustees. This does not imply that such measures should be entirely disregarded, rather they are not central to the success of data trustees from the end users' perspective. Positively the findings actually suggest flexibility in implementing various types of data trustees without being limited by specific user expectations about operators.

[10] Art. 42 GDPR and recital 100 of the GDPR.

[11] Art. 17 DGA.

(Un)importance of Individual Control Mechanisms The aim of data trustees and personal information management systems is to give users autonomy over their personal data. Previous studies have shown that offering users the ability to manage their data autonomously, for example through explicit consent or the option to delete data, generally increases the acceptance of data sharing [33], [45]. Some participants in our study also requested measures to independently manage the data shared with the data trustee, even though we did not explicitly ask about these aspects. Some requested a comprehensive overview of the trustees' data processing activities, others even called for detailed control and decision mechanisms that allow them to precisely determine who can access their data, for how long, and for which purposes. However, besides the technical challenge of implementing such mechanisms, it is questionable whether such options would actually see widespread use in practice. In reality, even when mechanisms to build autonomy are available, users often do not take advantage of them. For example, Farke et al. investigated Google's privacy dashboards, which provide transparency about users' profiles. Only about a quarter of their participants said they would change their behavior after becoming aware of the data collection [17]. Similarly, cookie banners – also intended to foster users' autonomy – have largely failed to enable true sovereign data decision-making as most users still agree to cookies for convenience.

The extent to which data trustees should provide users control mechanisms requires careful consideration. While offering such tools is generally beneficial, they are unlikely to be the primary motivator for using data trustees. Critically, designing user autonomy tools must avoid overwhelming users; excessive information or overly complex features may lead users to ignore or even reject the provided mechanisms.

VI. Conclusion

In a large-scale vignette-based study, we examined users' perceptions of and willingness to use data trustees in Germany and the US. Overall, we found that the acceptance is heterogeneous, with many users rejecting data trustees altogether. This highlights the necessity of addressing user concerns and expectations to build trust for this new concept. Countries aiming to implement data trustees should provide clear information on the specific purposes and public interests data trustees serve, and highlight the benefits for users. Additionally, data trustees must address privacy threats as users prioritize data anonymity, especially in Germany. Users also expressed disapproval of numerous third-party recipients. In contrast, organizational factors like the identity of the operator, storage location, and the supervisory authority have little impact on the decision to use a data trustee. Ultimately, understanding and addressing these user concerns is crucial for the successful implementation and acceptance of data trustees in any context.

Acknowledgment

This work was supported by the PhD School SecHuman – "Security for Humans in Cyberspace" by the federal state of NRW, Germany, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CASA – 390781972. Special thanks go to the "Stiftung Datenschutz," who provided valuable scientific input and fruitful collaboration.

References

[1] K. A. Ackermann, L. Burkhalter, T. Mildenberger, M. Frey, and A. Bearth, "Willingness to Share Data: Contextual Determinants of Consumers' Decisions to Share Private Data With Companies," *Journal of Consumer Behaviour*, vol. 21, no. 2, pp. 375–386, 2022.

[2] T. Arlinghaus, K. Kus, P. K. Rodrigues, and F. Teuteberg, "Datentreuhandstellen gestalten: Status quo und Perspektiven für Geschäftsmodelle," in *HMD Praxis der Wirtschaftsinformatik 58*, vol. 58, 2021, pp. 565–579.

[3] O. Ayalon, D. Turjeman, and E. M. Redmiles, "Exploring Privacy and Incentives Considerations in Adoption of COVID-19 Contact Tracing Apps," in *USENIX Security Symposium*. Anaheim, CA: USENIX, Aug. 2023, pp. 517–534.

[4] C. Beise, "Datensouveränität und Datentreuhand," *Recht Digital*, pp. 597–604, 2021.

[5] A. Blankertz, "Vertrauliche Datentreuhand: Wie die Datentreuhand effektiv Daten schützen und sichern kann," *Datenschutz und Datensicherheit*, pp. 789–793, Aug. 2021.

[6] A. Blankertz, P. von Braunmühl, P. Kuzev, F. Richter, H. Richter, and M. Schallbruch, "Datentreuhandmodelle Themenpapier," 2020.

[7] J. Botta, "Delegierte Selbstbestimmung? PIMS als Chance und Risiko für einen effektiven Datenschutz," *Multimedia und Recht*, pp. 946–951, Aug. 2021.

[8] B. Bucher, A. C. Haber, H. K. Hahn, H. Kusch, F. Prasser, U. Sax, and C. Schmidt, "Das Modell der Datentreuhand in der medizinischen Forschung," *Datenschutz und Datensicherheit*, vol. 45, pp. 806–810, Dec. 2021.

[9] B. Buchner, "Widerrufbarkeit der Einwilligung," *Datenschutz und Datensicherheit*, vol. 45, p. 831, Nov. 2021.

[10] Bundesdruckerei, "Data Trustee Platform With a Trust Center Service on Demand," https://www.bundesdruckerei-gmbh.de/en/solutions/data-trustee, as of December 6, 2024.

[11] Bundeskrebsregister, "Zentrum für Krebsregisterdaten," https://www.krebsdaten.de/Krebs/EN, as of December 6, 2024.

[12] California State Legislature, "California Consumer Privacy Act (CCPA) of 2018," https://leginfo.legislature.ca.gov, as of December 6, 2024.

[13] J. Caltrider, M. Rykov, and Z. MacDonald, "It's Official: Cars Are the Worst Product Category We Have Ever Reviewed for Privacy," Sep. 2023, https://foundation.mozilla.org, as of December 6, 2024.

[14] S. Delacroix and N. Lawrence, "Bottom-up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance," *International Data Privacy Law*, vol. 9 (4), pp. 236 – 252, 2019.

[15] European Data Protection Supervisor (EDPS), "Opinion on Personal Information Management Systems - Towards More User Empowerment in Managing and Processing Personal Data," Oct. 2016.

[16] European Mobility Data Space (EMDS), "Transport Data – Creating a Common European Mobility Data Space," https://ec.europa.eu/info/law, as of December 6, 2024.

[17] F. M. Farke, D. G. Balash, M. Golla, M. Dürmuth, and A. J. Aviv, "Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity," in *USENIX Security Symposium*, ser. SSYM '21. Virtual Conference: USENIX, Aug. 2021, pp. 483–500.

[18] M. Fassl, L. T. Gröber, and K. Krombholz, "Stop the Consent Theater," in *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: ACM, 2021.

[19] Federal Ministry of Education and Research, "Datentreuhandmodelle: BMBF fördert Pilotvorhaben," https://www.bildung-forschung.digital, as of December 6, 2024.

[20] J. Finch, "The Vignette Technique in Survey Research," *Sociology*, vol. 21, no. 1, pp. 105–114, 1987.

[21] FINDATA, "Finnish Social and Health Data Permit Authority Findata," https://findata.fi/en, as of December 6, 2024.

[22] M. Fiore, P. Katsikouli, E. Zavou, M. Cunche, F. Fessant, D. Le Hello, U. M. Aivodji, B. Olivier, T. Quertier, and R. Stanica, "Privacy in Trajectory Micro-data Publishing: A Survey," *Transactions on Data Privacy*, vol. 13, no. 2, pp. 91–149, 2020.

[23] J. D. Ford, S. E. Tilleard, L. Berrang-Ford, M. Araos, R. Biesbroek, A. C. Lesnikowski, G. K. MacDonald, A. Hsu, C. Chen, and L. Bizikova, "Big Data Has Big Potential for Applications to Climate Change Adaptation," in *Proceedings of the National Academy of Sciences*, vol. 113 (39), 2016, pp. 10 729–10 732.

[24] German Association of the Automotive Industry, "ADAXO: Automotive Data Access – Extended and Open," Dec. 2021, https://www.vda.de/adaxo, as of December 6, 2024.

[25] Google, "Privacy Sandbox for the Web," https://www.privacysandbox.com, as of December 6, 2024.

[26] D. Grande, N. Mitra, R. Iyengar, R. M. Merchant, D. A. Asch, M. Sharma, and C. C. Cannuscio, "Consumer Willingness to Share Personal Digital Information for Health-Related Uses," *JAMA Network Open*, vol. 5, no. 1, 01 2022.

[27] W. Hall and J. Pesenti, "Growing the Artificial Intelligence Industry in the UK," *Department for Science, Innovation & Technology*, Oct. 2017.

[28] Health Data Research UK, "Health Data Research Hubs," https://www.hdruk.ac.uk, as of December 6, 2024.

[29] K. A. Houser and J. W. Bagby, "The Data Trust Solution to Data Sharing Problems," *Vanderbilt Journal of Entertainment & Technology Law*, vol. 25, no. 1, pp. 113–180, 2023.

[30] Intel, "Internet of Things and the Smart Home," *Intel Security*, Mar. 2016.

[31] S. Iraschko-Luscher, "Einwilligung - ein stumpfes Schwert des Datenschutzes?" *Datenschutz und Datensicherheit*, vol. 30, pp. 706–710, Nov. 2006.

[32] L. K. Jones, R. Pulk, M. R. Gionfriddo, M. Evans, and D. Parry, "Utilizing Big Data to Provide Better Health at Lower Cost," in *American Journal of Health-System Pharmacy*, vol. 75 (7), 2018, pp. 427–435.

[33] B. Kacsmar, K. Tilbury, M. Mazmudar, and F. Kerschbaum, "Caring about Sharing: User Perceptions of Multiparty Data Sharing," in *USENIX Security Symposium*. Boston, MA: USENIX, Aug. 2022, pp. 899–916.

[34] M. Kamp and M. Rost, "Kritik an der Einwilligung: Ein Zwischenruf zu einer fiktiven Rechtsgrundlage in asymmetrischen Machtverhältnissen," *Datenschutz und Datensicherheit*, vol. 37, pp. 80–84, 2013.

[35] J. Kang, K. Shilton, E. Deborah, J. Burke, and M. Hansen, "Self-Surveillance Privacy," *Iowa Law Review*, vol. 97, p. 809, 2012.

[36] M. Karampela, S. Ouhbi, and M. Isomursu, "Connected Health User Willingness to Share Personal Health Data: Questionnaire Study," *Journal of Medical Internet Research*, vol. 21, no. 11, Nov 2019.

[37] S. Kempny, H. Krüger, and M. Spindler, "Rechtliche Gestaltung von Datentreuhändern. Ein interdisziplinärer Blick auf 'Data Trusts'," *Neue Juristische Wochenzeitschrift*, pp. 1646–1650, 2022.

[38] W. Kerber and L. Specht-Riemenschneider, "Designing Data Trustees - A Purpose-Based Approach." Konrad-Adenauer-Stiftung, 2022.

[39] K. K. Kim, P. Sankar, M. D. Wilson, and S. C. Haynes, "Factors Affecting Willingness to Share Electronic Health Data Among California Consumers," *BMC Medical Ethics*, vol. 18, no. 1, p. 25, Apr 2017.

[40] M. Kowalewski, C. Utz, M. Degeling, T. Schnitzler, F. Herbert, L. Schaewitz, F. M. Farke, S. Becker, and M. Dürmuth, "52 Weeks Later: Attitudes Towards COVID-19 Apps for Different Purposes Over Time," in *ACM on Human-Computer Interaction (CSCW2)*, ser. CSCW '23. New York, NY, USA: ACM, Oct. 2023.

[41] J. Kraemer, "Digitale Selbstbestimmung durch Personal Information Management Systems? Chancen, Hemmnisse und politische Handlungsempfehlungen," *Zu treuen Händen*, 2022.

[42] J. Kühling and F. Sackmann, "Irrweg 'Dateneigentum' - Neue Großkonzepte als Hemmnis für die Nutzung und Kommerzialisierung von Daten," *Zeitschrift für Datenschutz*, pp. 24–30, 2020.

[43] J. Kühling, F. Sackmann, and H. Schneider, "Datenschutzrechtliche Dimensionen Datentreuhänder: Kurzexpertise," *Federal Ministry of Labour and Social Affairs & Institute of Labor Economics (IZA)*, 2020.

[44] L. Lassak, H. Püschel, O. Reithmaier, T. Gostomzyk, and M. Dürmuth, "Balancing Privacy and Data Utilization: A Comparative Vignette Study on User Acceptance of Data Trustees in Germany and the US (Extended)," Feb. 2025, https://leonalassak.com, as of December 6, 2024.

[45] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor, "What Matters to Users? Factors That Affect Users' Willingness to Share Information With Online Advertisers," in *Symposium on Usable Privacy and Security*, ser. SOUPS '13. New York, NY, USA: ACM, 2013.

[46] W. Lesch, G. Richter, and S. C. Semler, "Daten teilen für die Forschung: Einstellungen und Perspektiven zur Datenspende in Deutschland," in *Datenreiche Medizin und das Problem der Einwilligung*. Berlin, Heidelberg: Springer, 2022, pp. 211–226.

[47] M. Lindner and S. Straub, "Datentreuhänderschaft - Status Quo und Entwicklungsperspektiven," *Federal Ministry for Economic Affairs and Climate Actions*, 2023.

[48] N. Malhotra, S. Kim, and J. Agarwal, "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Information Systems Research*, vol. 15, pp. 336–355, Dec. 2004.

[49] A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," in *IEEE Symposium on Security and Privacy*, ser. SP '08. IEEE, 2008, pp. 111–125.

[50] J. Nicholas, K. Shilton, S. M. Schueller, E. L. Gray, M. J. Kwasny, and D. C. Mohr, "The Role of Data Type and Recipient in Individuals' Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study," *JMIR MHealth UHealth*, vol. 7 (4), Apr. 2019.

[51] Office of the Australian Information Commissioner, "My Health Record," 2020, https://www.oaic.gov.au/my-health-record, as of December 6, 2024.

[52] R. Pastorina, C. D. Vito, G. Migliara, K. Glocker, I. Bienenbaum, W. Ricchiardil, and S. Boccia, "Benefits and Challenges of Big Data in Healthcare: An Overview of the European Initiatives," *European Journal of Public Health*, vol. 29, pp. 23–27, 2019.

[53] S. R. Peppet, "Privacy & the Personal Prospectus: Should We Introduce Privacy Agents or Regulate Privacy Intermediaries," *Iowa Law Review Bulletin*, vol. 97, pp. 77–93, 2012-2013.

[54] C. Pugnetti and S. Elmer, "Self-Assessment of Driving Style and the Willingness to Share Personal Information," *Journal of Risk and Financial Management*, vol. 13 (3), Mar. 2020.

[55] L. Rainie and M. Duggan, "Privacy and Information Sharing," *Pew Research Center*, 2016.

[56] P. M. Regan, "A Design for Public Trustee and Privacy Protection Regulation," *Seton Hall Journal of Legislation and Public Policy*, vol. 44, pp. 487–513, 2020.

[57] J. M. Rickert, "The Relationship between Transparency, Consumer Trust and Willingness to Share Data - A Vignette Survey," June 2016.

[58] C. Rising, A. Gaysynsky, K. Blake, R. Jensen, and A. Oh, "Willingness to Share Data From Wearable Health and Activity Trackers: Analysis of the 2019 Health Information National Trends Survey Data," *JMIR Mhealth Uhealth*, vol. 9 (12), Dec 2021.

[59] R. Rothmann and B. Buchner, "Der typische Facebook-Nutzer zwischen Recht und Realität: Zugleich eine Anmerkung zu LG Berlin v. 16.01.2018," *Datenschutz und Datensicherheit*, pp. 342–346, 2018.

[60] A. Ruhaak, "Data Trusts in Germany and Under the GDPR." AlgorithmWatch, 2020.

[61] E.-M. Schomakers, C. Lidynia, and M. Ziefle, "All of Me? Users' Preferences for Privacy-Preserving Data Markets and the Importance of Anonymity," *Electronic Markets*, vol. 30 (3), pp. 649–665, Sep. 2020.

[62] E. Seltzer, J. Goldshear, S. C. Guntuku, D. Grande, D. A. Asch, E. V. Klinger, and R. M. Merchant, "Patients' Willingness to Share Digital Health and Non-health Data for Research: A Cross-Sectional Study," *BMC Medical Informatics and Decision Making*, vol. 19 (1), Aug. 2019.

[63] L. Specht-Riemenschneider and A. Blankertz, "Lösungsoption Datentreuhand: Datennutzbarkeit und Datenschutz zusammen denken," *Multimedia und Recht*, pp. 369–370, 2021.

[64] ——, "Wie eine Regulierung für Datentreuhänder aussehen sollte." Stiftung Neue Verantwortung e.V., 2021.

[65] L. Specht-Riemenschneider, A. Blankertz, P. Sierek, R. Schneider, J. Knapp, and T. Henne, "Die Datentreuhand: Ein Beitrag zur Modellbildung und rechtlichen Strukturierung zwecks Identifizierung der Regulierungserfordernisse für Datentreuhandmodelle," *Multimedia und Recht Beilage*, pp. 25–48, 2021.

[66] The European Parliament and the Council of the European Union, "Regulation (EU) 2022/868 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act)," 2022, https://eur-lex.europa.eu/DGA, as of December 6, 2024.

[67] ——, "Regulation (EU) 2023/2854 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)," 2023, https://eur-lex.europa.eu/DA, as of December 6, 2024.

[68] T. Urban, D. Tatang, M. Degeling, T. Holz, and N. Pohlmann, "A Study on Subject Data Access in Online Advertising After the GDPR," in *Data Privacy Management, Cryptocurrencies and Blockchain Technology*. Springer International Publishing, 2019, pp. 61–79.

[69] C. Utz, S. Becker, T. Schnitzler, F. M. Farke, F. Herbert, L. Schaewitz, M. Degeling, and M. Dürmuth, "Apps Against the Spread: Privacy Implications and User Acceptance of COVID-19-Related Smartphone Apps on Three Continents," in *ACM Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: ACM, 2021.

[70] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(Un)informed Consent: Studying GDPR Consent Notices in the Field," in *ACM Conference on Computer and Communications Security*, ser. CCS '19. London, United Kingdom: ACM, Nov. 2019, pp. 973–990.

[71] F. von Ulmenstein, "Datensouveränität durch repräsentative Rechtswahrnehmung Begriffliche Prägung und normative Gestaltung sogenannter 'Datentreuhänder'," *Datenschutz und Datensicherheit*, pp. 528–534, 2020.

[72] C. Wendehorst, S. Schwamberger, and J. Grinzinger, "Datentreuhand - wie hilfreich sind sachenrechtliche Konzepte?" *Rechte an Daten*, 2020.

[73] N. Wilson and A. Reid, "Data Controllers as Data Fiduciaries: Theory, Definitions & Burdens of Proof," *Colorado Law Review*, vol. 95, 2024.

[74] M. Ziefle and J. Halbey, "Users' Willingness to Share Data on the Internet: Perceived Benefits and Caveats," in *Int. Conference on Internet of Things and Big Data (IoTBD)*, Apr. 2016, pp. 255–265.

APPENDIX

A. Factors

TABLE II. FACTORS, FACTOR LEVELS, AND WORDING IN SCENARIOS. NOTE: FACTORS WITH * HAD "*none*" AS A FACTOR LEVEL, MEANING IN SOME SCENARIOS THE FACTORS WERE OMITTED.

Factors	Description
Operator	
Government	the government agency Dawi
Business	the company Dawi
NGO	the NGO Dawi
User Anonymity	
Raw data	raw data in a non-anonymized format
Anonymous	already anonymized data
Non-personal	only non-personal data
Processing	
Store	stores it
Aggregate	aggregates it with data from different sources
Analyse	analyses the data, third parties can only access analysis
Storage Location*	
GER/US	only on servers in Germany/the US
(Germany only) EU	only on servers in Europe
Worldwide	on servers worldwide
Recipient	
Research	only research institutions
Business	research institutions and companies
Everyone	everyone who is interested
Law enforcement	law enforcement authorities [...] for investigative purposes
Access	
Transmission	encypted datasets are transmitted to third parties
View only	datasets always remain with Dawi
Benefits*	
Monetary	you receive monetary compensation
(Medical only) Individual	improve medical care [...] for your own benefit
Monitoring*	
Certified	Dawi is certified
Governmental control	national supervisory authority
Private audits	private-sector auditing firm

B. Survey Instrument

C – Consent

[We only show an excerpt of the consent form due to space constraints.]
Thank you for your interest in our study!
We will present you with a series of questions regarding potential new approaches to data sharing. The aim of this survey is to gain a comprehensive understanding of users' preferences and desires regarding these new concepts. Your participation can make a valuable contribution to this goal. [...]

S – Scenarios

In the following, we present a fictional scenario. Please try to imagine yourself as being part of the described scenario and consider how you would decide in that situation. *[Note: The same questions were repeated with a second vignette with slightly changed wording in the introduction sentence.]*

Next, we randomly displayed one of the four scenario descriptions with filled-in factor levels.

S1 How likely would you agree to use the service provider Dawi?

Very unlikely	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Very likely
	○	○	○	○	○	○	○	

S2 Which factors influenced your decision the most? Please briefly explain your answer.
Answer: _____

S3 Please complete the following sentence: Most people close to me would think that I should...

definitely **not** use the service provider	(1)	(2)	(3)	(4)	(5)	(6)	(7)	definitely use the service provider
	○	○	○	○	○	○	○	

S4 How useful do you perceive the service of Dawi?

Not at all useful	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Very useful
For yourself	○	○	○	○	○	○	○	
For society	○	○	○	○	○	○	○	

G – General Questions

The service providers we talked about in the scenarios are also called 'data trustees.' Right now, they are mostly just an idea but people are discussing whether and how we could use data trustees in reality. These trustees would act like middlemen between someone who has data (like a doctor or someone using smart devices) and someone who wants to use that data (like research groups, government agencies, or other companies). Some data trustees are already being used, like for cancer or car data.
In the following, we will ask you questions about what you would want from a data trustee. The goal is to learn what is important to you and what would make you feel comfortable using a data trustee.
Important: The following questions have nothing to do with the scenarios you saw earlier in the study.
Next, we present various ways a potential data trustee could be set up. Please indicate how each aspect would influence your judgment of the data trustee.

G1.1 **Operator:** The trustee is operated by...
a government agency • a company/business enterprise • a nonprofit organization

G1.2 **Type of Data:** The trustee receives...
the raw data in a non-anonymized format • already anonymized data

G1.3 **Data Processing:** The trustee...
only stores your data • aggregates your data with data from different sources • analyzes and evaluates your data

G1.4 **Storage Location:** The trustee stores your data...
only on servers in Germany/the US • only on servers in the EU • on servers worldwide

G1.5 **Usage by Third Parties:** ... may use your data.
Research institutions • Private sector/business enterprises • Administrative governmental institutions • Law enforcement authorities (police) • Anyone who is interested

G1.6 **Data transmission:** Your data...
is transmitted to third parties • always remains with the trustee. Third parties can access the data only through a limited number of requests.

G1.7 **Supervision and Transparency:** The trustee...
is government certified • undergoes review by a national supervisory authority to ensure compliance with legal regulations • undergoes review by a private-sector auditing firm to ensure compliance with legal regulations

G2 Please describe what other design options would increase your willingness to use the trustee and why.
Answer: _____

G3 What incentives would increase your willingness to use the trustee?
Answer: _____

G4 Please list the three factors that play the biggest role in your decision to use the data trustee.
1. _____
2. _____
3. _____

G5 Why are these three factors particularly relevant to you?
Answer: _____

A **Attention Check** – Please select the option 'Somewhat Agree.'

I **IUIPC** – Internet Users' Information Privacy Concerns - Scale [48]

P – Prior Knowledge

P1 Have you heard of any of the following terms before participating in this survey? - *Data trustee, data hub, data interchange, PIMS (Personal Information Management Systems)*
○ Yes ○ No ○ Unsure ○ Prefer not to answer

P2 Have you knowingly used a data trustee or something similar before?
○ Yes ○ No ○ Unsure ○ Prefer not to answer

P3 Do you use technical tools to increase your online privacy? Examples of such tools include: Browser plug-ins like *NoScript* or *NinjaCookie*, privacy-friendly browsers like *Brave*, privacy-friendly search engines like *DuckDuckGo*
○ Yes, I am currently using them ○ Yes, I have used them in the past ○ Yes, but only to test out ○ No ○ Unsure ○ Prefer not to answer

P4 How much do you trust the following institutions to handle your data safely?

	Not at all (1)	(2)	(3)	(4)	Completely (5)	Prefer not to answer (6)
Governmental organizations	○	○	○	○	○	○
Business enterprises	○	○	○	○	○	○
Nonprofit organizations	○	○	○	○	○	○

D – Demographics

First, we would like to obtain some general information about you.

D1 What is your age?
Answer: _____

D2 What gender do you identify with?
○ Woman ○ Man ○ Non-binary ○ Prefer not to answer

D3 What is your level of education?
○ Less than High school ○ High school (incl. GED) ○ Some college (no degree) ○ Technical certification ○ Associate degree (2-year) ○ Bachelor's degree (4-year) ○ Master's degree ○ Doctoral degree (incl. professional degrees i.e., JD, MD) ○ Prefer not to answer

D4 My total annual income (gross) is:
○ $< \$20,000$ USD ○ $\$20,000 - \$40,000$ USD ○ $\$40,000 - \$60,000$ USD ○ $\$60,000 - \$80,000$ USD ○ $\$80,000 - \$100,000$ USD ○ $\$100,000 - \$120,000$ USD ○ $\$120,000 - \$140,000$ USD ○ $\$140,000 - \$160,000$ USD ○ $\$160,000 - \$180,000$ USD ○ $\$180,000 - \$200,000$ USD ○ $\$200,000 - \$240,000$ USD ○ $> \$240,000$ USD ○ Prefer not to answer

D5 Do you have experience in computer science, computer technology, or information technology (e.g., through your profession or educational background)?
○ Yes ○ No ○ Prefer not to answer

C. Codebook

TABLE III. CODEBOOK FOR OPEN ANSWERS EXPLAINING THE WILLINGNESS TO USE THE DATA TRUSTEE (**S2**).

	Germany		US		
Codes & Subcodes	n	%	n	%	**Example**
Privacy Protection	297	18.32%	273	12.62%	
Violates Privacy	72	4.44%	121	5.59%	"Violation of my priv.", "don't want to give out my personal data."
Data Protection	51	3.15%	16	0.74%	"data protection"
General refusal to share data	84	5.18%	75	3.47%	"My data is noone's business.", "I fundamentally dislike such things."
Transparent citizen	11	0.68%	1	0.05%	"I'm practically becoming a transparent person."
Nothing to hide	18	1.11%	11	0.51%	"I have nothing to hide"
Resignation	16	0.99%	11	0.51%	"It's already online anyway, and the Internet doesn't forget any data"
Too much data collection	27	1.67%	18	0.83%	"Enough data is already being stored."
Fear of surveillance	18	1.11%	20	0.92%	"I feel surveilled.", "Do not want to be controlled"
Personal Benefits	134	8.27%	282	13.03%	
Undefined benefits	17	1.05%	80	3.70%	"I would have an advantage with it", "Benefits for myself [...]"
Data availability	9	0.56%	33	1.52%	"Medical records always available"
Money	84	5.18%	143	6.61%	"Remuneration", "Für lau' (for free), there is, of course, no data"
None	18	1.11%	9	0.42%	"I see no benefits for myself"
Increased privacy	6	0.37%	17	0.79%	"It will increase my privacy when I access the Internet"
General Interest	78	4.81%	263	12.15%	
All factors	-	-	25	1.16%	"everything", "all of the above"
Interested	33	2.04%	156	7.21%	"I like it", "I have no problem with that", "curiosity"
Not interested	45	2.78%	82	3.79%	"I do not like that", "no interest", "I do not need that"
Security	117	7.22%	165	7.62%	
Security mentioned	6	0.37%	8	0.37%	"Security"
Secure	46	2.84%	76	3.51%	"Seems to be safe", "Data records are transmitted encrypted"
Not Secure	40	2.47%	29	1.34%	"Insecure"
Misuse of data	15	0.93%	38	1.76%	"Fear of data theft", "The risk of data misuse is too high."
Mistrusting technology	10	0.62%	14	0.94%	"Security errors"
Utility/Purpose	118	7.28%	170	7.86%	
Useful	12	0.74%	71	3.28%	"It seems useful to me.", "Utility outweighs potential drawbacks."
Societal benefits	73	4.50%	44	2.03%	"Because it improves medical care.", "help reduce CO2 emissions"
No purpose perceived	22	1.36%	25	1.16%	"Does not have any utility for the general public", "See no sense"
Unclear	11	0.68%	30	1.39%	"Where is the benefit", "Don't know if I need that"
User anonymity	145	8.95%	103	4.76%	
Anonymized	57	3.52%	30	1.39%	"anonymity", "That the data is passed on anonymously."
Not anonymized	54	3.33%	8	0.37%	"Data not anonymized", "Not anonymous"
PII	34	2.10%	65	3.00%	"No personal data", "Name, Address"
Access	100	6.17%	134	6.19%	
Third party access	83	5.12%	94	4.34%	"I don't want my data to be passed on", "Access by third parties"
Limited	17	1.05%	40	1.85%	"Data remain with trustee", "Third parties have no access to my data"
Trust	98	6.05%	107	4.94%	
Trust Dawi	33	2.04%	44	2.03%	"I trust the service provider", "It sounds serious"
Don't trust Dawi	65	4.01%	63	2.91%	"Sounds dubious and not trustworthy.", "Uncertain about the promise."
Receiver	103	6.35%	81	3.74%	
Government agencies	28	1.73%	21	0.97%	"That the police gets the data", "Data transmitted to law enforcement"
Research institutions	18	1.11%	21	0.97%	"Controlled access only for research institutions"
Everyone	51	3.15%	33	1.52%	"Anyone who is interested can get the data. Big minus point"
Private companies	6	0.37%	6	0.28%	"Economic use"
More information	33	2.04%	150	6.93%	
More details needed	21	1.30%	58	2.68%	"I would need more information", "I don't know about that at all."
Unknown	11	0.68%	92	4.25%	"Association is too unknown to me", "I do not know"
Data Type	60	3.70%	95	4.39%	
Car	6	0.37%	10	0.46%	"Auto metadata e.g. consumption, CO2 and fine dust emissions [...]"
Sensitivity	42	2.59%	74	3.42%	"Too private data", "Behavior of individual persons is too private"
Location	12	0.74%	11	0.51%	"Location and movement data", "My movement profile"
Monitoring	59	3.64%	81	3.74%	
Liked monitoring	32	1.97%	72	3.33%	"That it is reviewed annually."
Don't trust private	12	0.74%	63	2.91%	"Private auditing, there. I have no trust"
Governmental	13	0.80%	5	0.23%	"Governmental supervision"
Sovereignty	36	2.22%	63	2.91%	
Voluntariness	5	0.31%	11	0.51%	"It's voluntary"
Control over data handling	10	0.62%	18	0.83%	"Make own decisions", "no control"
Knowledge about data handling	22	1.30%	34	1.57%	"would like to know where data is to be sold to"

Storage Location	41	2.53%	35	1.62%	
Germany/US	14	0.86%	14	0.65%	"Server in Germany"
Europe	11	0.68%	-	-	"Server in the EU [...]"
Worldwide	16	0.99%	20	0.92%	"Storage worldwide", "Is available worldwide"
Operator	17	1.05%	11	0.51%	
NGO	3	0.19%	1	0.05%	"Only NGOs collect the data to sell it"
Private	6	0.37%	3	0.14%	"lack of trust in the private sector"
State	8	0.49%	7	0.32%	"Governmental institution [...]"
Expected Disadvantages	12	0.74%	15	0.69%	
General	6	0.37%	10	0.46%	"There are only disadvantages for me, no advantage!"
Insurance	1	0.06%	2	0.09%	"It shouldn't influence my insurance fees"
Ads	5	0.31%	3	0.14%	"Do not want any advertising calls"
No data economy	26	1.60%	14	0.65%	"Money for data seems suspicious to me."
Don't have	50	3.08%	9	0.42%	"Do not have these devices", "I don't drive a car"
Don't know	97	5.98%	113	5.22%	"I do not know", "Undecided", "Out of the gut"

D. Regression Results

TABLE IV. REGRESSION RESULTS FOR ALL DOMAINS & COUNTRIES. REFERENCE LEVELS ARE IN PARENTHESES. "SCALED" VARIABLES WERE CENTERED AND STANDARDIZED TO THEIR MEAN. SIGNIFICANCE LEVELS: $^{***} = p < 0.001$; $^{**} = p < 0.01$; $^{*} = p < 0.05$.

	Medical		**Automotive**		**Online**		**IoT**	
	GER	US	GER	US	GER	US	GER	US
Operator *(Government)*								
Business	0.01	0.18	0.04	-0.09	-0.05	0.03	-0.00	0.17
NGO	0.11	0.14	0.20	-0.01	0.06	0.04	0.01	0.17
User Anonymity *(Raw)*								
Anonymous	-0.09	**0.28***	0.14	-0.04	**0.40****	0.02	0.19	-0.15
Non-personal	0.26	0.23	0.11	-0.12	**0.52*****	0.03	**0.28***	-0.05
Processing *(Store)*								
Aggregate	0.15	-0.07	-0.20	0.04	0.10	-0.03	-0.16	0.15
Analyse	**0.28***	0.08	0.08	-0.12	-0.06	0.05	-0.19	0.01
Storage Location *(none)*								
GER/US	0.27	-0.12	0.18	-0.05	-0.09	0.10	-0.12	-0.20
EU	0.10		0.20		0.01		-0.20	
Worlwide	-0.04	-0.21	-0.02	-0.06	-0.02	0.02	-0.14	-0.15
Recipient *(Research)*								
Business	-0.10	-0.18	-0.22	0.07	-0.24	-0.20	-0.03	-0.05
Everyone	-0.15	**-0.27***	-0.02	-0.03	**-0.35***	-0.24	-0.11	-0.06
Law Enforcement	**-0.41*****	0.05	-0.13	-0.10	-0.08	-0.05	0.11	-0.10
Access *(View only)*								
Transmission	**-0.23***	-0.06	0.06	-0.02	-0.08	-0.10	-0.12	0.05
Benefits *(none)*								
Monetary	**0.39****	**0.29****	**0.26***	0.11	-0.03	0.19	-0.05	0.00
Personal	**0.31***	0.16						
Monitoring *(none)*								
Government	-0.08	0.09	0.07	-0.09	0.15	0.07	-0.28	0.12
Government + Cert.	0.10	0.25	-0.09	-0.02	-0.17	0.14	0.18	0.25
Private	0.11	0.08	-0.16	-0.20	0.31	0.01	-0.16	0.15
Private + Cert.	-0.12	0.14	0.14	-0.11	-0.12	0.19	-0.25	0.27
IUIPC Score (scaled)	**-0.22*****	-0.06	-0.05	-0.03	**-0.16****	0.00	**-0.19****	0.08
Gender *(Male)*								
Female	-0.11	**-0.25***	-0.10	-0.13	-0.18	-0.04	-0.03	-0.01
Non-Binary	-1.51	-0.19		-0.23	-1.05		-1.08	-0.18
Age (scaled)	-0.12	**-0.18*****	**-0.14***	**-0.21*****	**-0.23*****	-0.10	**-0.22*****	**-0.23*****
Education *(Average)*								
Low	0.04	-0.10	0.07	0.00	0.13	0.23	-0.07	0.03
High	0.17	-0.12	0.28	-0.11	0.16	**0.35***	0.03	0.07
Income *(Average)*								
Low	0.08	0.00	0.11	-0.24	0.00	0.01	-0.07	-0.04
High	0.19	0.04	-0.02	0.12	0.12	-0.10	**-0.77***	0.07
IT Background *(no)*	-0.19	**0.35*****	0.17	0.13	-0.03	0.18	0.07	**0.38*****
R^2	0.23	0.15	0.13	0.10	0.21	0.07	0.18	0.14
Adj. R^2	0.15	0.09	0.04	0.03	0.13	0.01	0.10	0.09
N	302	429	288	410	289	398	308	430