Introducing Data Trustees: A Vignette-Based Study Approach to Get Users in the Loop

Leona Lassak*
Ruhr University Bochum

Hanna Püschel*
TU Dortmund University

Tobias Gostomzyk
TU Dortmund University

Markus Dürmuth
Leibniz University Hannover

Abstract

Data sharing as a central component of society, must be done securely. One promising approach is the concept of data trustees, which is gaining importance in legal research and policy making. Data trustees bridge the gap between data use and protection, offering the potential for a privacy-preserving data economy. Data trustees could have applications in a variety of sectors, including medicine, mobility, agriculture, the Internet of Things (IoT), and online domains, where the tension between privacy and data use is particularly pronounced. In this work, we present the conception of a vignette-based study of end-user acceptance of and preferences for data trustees to provide insights into different design options for data trustees. Early dissemination of the concept of data trustees is critical to ensure that expertise from different domains is incorporated into the development of viable solutions.

1 Introduction

Data sharing is vital for societal communication, action, and progress. Secure data sharing requires innovative approaches, including the concept of "data trustees." This concept is gaining relevance, evident in various regulatory developments. The European Data Governance Act (DGA) created a framework to facilitate data sharing through regulations on so-called "providers of data sharing services" and "data altruism organizations" which could include date trustees. There is also an overlap with authorized agents introduced by California's Consumer Privacy Act (CCPA). But data trustees offer even more potential. They unite data usage and data protection, two demands seemingly contrasting. According to regulators, in the future, data trustees shall play a key role in the data economy by facilitating the aggregation and sharing of data while at the same time providing a high level of data protection and giving individuals control over their data [1, 2, 9, 11, 16]. However, this development is still in its early stages.

Conceivable areas for the use of data trustees include traditional fields like medicine, mobility, and agriculture. Also interesting are the more end user-centric Internet of Things (IoT) and online sectors. In fact, in the field of medical research, data trustee models are partially already in practice [17]. In the automotive industry, meanwhile, a concept has been developed to enable economic use and progress through the exchange of vehicle data [13]. On the internet, privacy-protecting usage of data for personalized advertising plays a major role, as shown by, so far unsuccessful developments such as Google's Privacy Sandbox [8].

In this work, we present the conception of a large-scale vignette-based study on end users' acceptance and preferences on the novel topic of data trustees. This is a work in progress and we are currently in the process of collecting data from a representative sample in Germany. However, we believe that it is important to spread the word about new legislative developments in the security and privacy community early on to ensure expertise from all areas can influence those developments to ultimately lead to better, more usable solutions. Prior examples of regulatory developments like the cookie banners of the General Data Protection Regulation (GDPR) have proven to not only be ineffective but, if done incorrectly, can even be harmful to end users in terms of privacy, usability, and agency over their data [7, 18].

2 Research Objective and Contribution

The acceptance and trust of users in the use of newly introduced technical concepts are crucial for their ability to succeed. The past has shown how neglecting the user perspective in the development process can lead to serious obstacles or even the complete failure of new approaches. As data trustees

USENIX Symposium on Usable Privacy and Security (SOUPS) 2023.
August 6–8, 2023, Anaheim, CA, USA

are still in their theoretical development phase, we think our user-focused research on factors that influence the acceptance of data trustees can contribute to their successful implementation. As per the nature of new areas of research, our study is designed in an exploratory manner. Our research interest thereby surrounds the factors influencing the end-user acceptance of data trustees in different scenarios and fields of application. Concretely we study their acceptance in the fields of medical, mobility, IoT, and online data.

Fields of Application The aforementioned areas are all characterized by (partly different) weaknesses in current data processing practices. Thus, data trustees in these areas may also vary. In the online and IoT area, a wealth of data is generated through the use of social networks, search engines, and the sensors of the devices themselves. Many of those affected are not aware of the extent of those data collection practices [5, 14, 18]. If users do read the data protection regulations, they are often unable to grasp the extent and risks of the planned data processing or accept cookie banners without reading the data protection regulations [5, 6, 14, 18]. The possibility of designing data trustees as so-called "Personal Information Management Systems" (PIMS) could give users back control over their personal data. They could also exercise other data subject rights such as rights of access, rectification, erasure, data portability, and objection in the interest of the individual, and assert claims for erasure or damages [11]. At the same time, in areas such as medical research, there are problems regarding data collection, data use, and data access [2–4, 10, 15]. Here, data trustees could act as a trust center, pseudonymize, and anonymize data, and/or mediate data access to various third-party actors. Through neutral data averaging, data can be made more systematically usable.

3 Methodology

As mentioned previously, we utilize the concept of vignette studies. In a representative online survey (n=1000), participants get presented with a description of potential data trustees in different randomly combined configurations, called "vignettes" or scenarios. Each scenario differs on different factors, in our case 9. Every factor has a varying number of factor levels. Factors and their levels are shown in Table 1. An example vignette for the medical context is shown in Appendix A.

The factors and scenario texts were developed through a systematic and iterative process. Drawing from legal literature and research on user-centric privacy-enhancing technologies, we compiled a comprehensive list of potentially influential factors for data trustees' acceptance. We categorized and refined these factors through multiple discussion sessions with the research team and sought input from external practitioners and data privacy experts. In a workshop with ten experts from the usable security and privacy domain, we validated our list, and no additional factors were identified. Therefore, we consider our list to be exhaustive and concise.

Table 1: Factors and their respective factor levels that scenarios are created from. *Note:* Factors with * are not shown at all in some scenarios to assess their influence.

Factor	**Factor Level**
Provider	Government
	Corporation
	NGO
Data Type	Non-anonymized raw data
	Anonymized data sets
	Only non-personal data
Data Processing	Only storing
	Aggregation
	Analysis
Storage Location*	Servers in Germany
	Servers in Europe
	Servers worldwide
Receiver	Research institutions
	Private enterprises
	Everybody who is interested
	Law enforcement agencies
Access Type	Data sets transmitted to receiver
	Data sets remain with trustee
Benefits for Users	Monetary compensation
	Individualized services
Certification*	Certified
Monitoring*	Governmental institution
	Public auditors

Survey Structure The survey begins with a consent form for each participant. After demographic questions (gender, age, educational attainment), participants receive the instruction that in the following sections of the survey, they should imagine being in the situation described to them by the vignette. We ensured not to use the term "data trustee" explicitly, to not induce trust just based on the term. Upon receiving their first scenario description, participants are then asked to choose whether they would use the trustee described in the scenario and to assess its usefulness. This process is repeated with two scenario descriptions. In the second part of the survey, we then explicitly describe the concept of data trustees to participants. In subsequent multiple-choice and open-ended questions, we inquire about the positive or negative influence of all factors on their acceptance of data trustees. This is followed by a section with the standardized privacy questionnaire IUIPC [12], as well as a few questions about their familiarity with data trustees and other privacy-protecting measures. Finally, participants are asked for further basic demographic information (income, IT background). We received ethics clearance for the study from Leibniz Universities' ethics board. All personal data is only stored anonymously.

A Example Vignette

Imagine you are at the doctor's office and the doctor stores the following data about you:

- Name, address, date of birth, sex
- Medical check-ups, regularity of check-ups, illnesses
- Emergency information (allergies, illnesses, blood type...)

Your doctor asks you if you are interested to allow a non-profit service provider access to this data. The service provider grants third parties access to the data under the following conditions. This option is voluntary.

- The service provider receives anonymized data and analyzes it.
- The data is only stored on servers in the EU.
- Access to the data is granted to research institutions and private companies.
- You receive monetary compensation for your data.
- The certified service provider gets monitored for compliance with the regulation by public auditors.

Are you interested to give the service provider access to your data?

Figure 1: Example of a scenario description (here: medical) with exemplary factor levels marked in colors corresponding to their factors (c.f. Table 1)

References

[1] Clara Beise. Datensouveränität und Datentreuhand. *Recht Digital*, pages 597 – 604, 2021.

[2] Aline Blankertz. Vertrauliche Datentreuhand: Wie die Datentreuhand effektiv Daten schützen und sichern kann. *Datenschutz und Datensicherheit*, pages 789–793, 2021.

[3] Benedikt Bucher, Anna Christine Haber, Horst Karl Hahn, Harald Kusch, Fabian Prasser, Ulrich Sax, and Carsten Schmidt. Das Modell der Datentreuhand in der medizinischen Forschung. *Datenschutz und Datensicherheit*, pages 806–810, 2021.

[4] Benedikt Buchner. Widerrufbarkeit der Einwilligung. *Datenschutz und Datensicherheit*, page 831, 2021.

[5] Gordian Konstantin Ebner. Information Overload 2.0? Datenwirtschaftsrecht IV: Die Informationspflichten gem. Art. 3 Abs. 2 Data Act-Entwurf. *Zeitschrift Datenschutz*, pages 364 – 369, 2022.

[6] EU-Kommission. Special Eurobarometer 487a "The General Data Protection Regulation". 2019.

[7] Matthias Fassl, Lea Theresa Gröber, and Katharina Krombholz. Stop the Consent Theater. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[8] Google Privacy Sandbox. Privacy Sandbox for the Web. www.privacysandbox.com, as of June 29, 2023.

[9] Simon Kempny, Heike Krüger, and Martin Spindler. Rechtliche Gestaltung von Datentreuhändern, Ein interdisziplinärer Blick auf "Data Trusts". *Neue Juristische Wochenzeitschrift*, pages 1646 – 1650, 2022.

[10] Wolfgang Kerber and Louisa Specht-Riemenschneider. Designing Data Trustees - A Purpose-Based Approach. 2022.

[11] Jürgen Kühling, Florian Sackmann, and Hilmar Schneider. Datenschutzrechtliche Dimensionen Datentreuhänder: Kurzexpertise, 2020.

[12] Naresh Malhotra, Sung Kim, and James Agarwal. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15:336–355, 12 2004.

[13] German Association of the Automotive Industry. ADAXO: Automotive Data Access – Extended and Open VDA concept for access to in-vehicle data. 2021.

[14] Robert Rothmann and Benedikt Buchner. Der typische Facebook-Nutzer zwischen Recht und Realität: Zugleich eine Anmerkung zu LG Berlin v. 16.01.2018. *Datenschutz und Datensicherheit*, pages 342–346, 2018.

[15] Louisa Specht-Riemenschneider and Aline Blankertz. Lösungsoption Datentreuhand: Datennutzbarkeit und Datenschutz zusammen denken. *Multimedia und Recht*, pages 369 – 370, 2021.

[16] Louisa Specht-Riemenschneider, Aline Blankertz, Pascal Sierek, Ruben Schneider, Jakob Knapp, and Theresa Henne. Die Datentreuhand: Ein Beitrag zur Modellbildung und rechtlichen Strukturierung zwecks Identifizierung der Regulierungserfordernisse für Datentreuhandmodelle. *Multimedia und Recht Beilage*, pages 25–48, 2021.

[17] Unabhängige Treuhandstelle der Universitätsmedizin Greifswald. https://www.ths-greifswald.de, as of June 29, 2023.

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