Developing a user interface design to foster trust and maximize consent for data donation in scientific research

Katrine Nordeide Kuiper
6566952

Utrecht University

Primary supervisor: Dr. Paweł W. Woźniak

External supervisor: Dr. Laura Boeschoten

Second examiner: Dr. Almila A. Akdag

A thesis presented for the degree of Master of Information Science in Human Computer Interaction at Utrecht University

Department of Information and Computing Sciences
Utrecht University
The Netherlands
December 13, 2021

Abstract

Through data donation, research participants can voluntarily donate personal data for research purposes. Recent data protection laws, such as GDPR, give the citizens right of access to data stored about them, making data donation in research more relevant than ever. Social scientists at Utrecht University propose to let research participants donate partial data download packages (DDPs) to scientific research projects as a new form of data collection. In this research we present the design process of a prototype for a data donation user interface (UI). For data donation to be a viable option for data collection, participants have to trust that their data is handled fairly and responsibly. To encourage trust, a data donation UI has to present the data pipeline and anonymization of the data in a transparent way. Although anonymization is explained in theory, there is a gap in knowledge on how to communicate anonymization of data to regular users. To address this we applied construal level theory to explain the data handling, and investigated the impact it had on users' willingness, trust and consent level. A within-subjects experiment was performed where participants were asked to donate location data from Google to a fictional research study through the data donation website. Although no significant results were found, there was an overall high trust in the data donation system and a high willingness to donate data. Qualitative analysis revealed that explaining the anonymization was important to participants, and that a majority of the participants would be willing to donate data through a data donation website in the future. Furthermore, participants are eager to contribute to research, and future data donation systems should communicate the data pipeline, provide information about the research project and the impact of participants' contribution to research.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Paweł Woźniak, for his expertise and wisdom through this research project. Discussing with you has advanced my knowledge of research and in the field of Human-Computer Interaction. Thank you also to Dr. Almila Akdag for being the second reader.

Second, I would like to thank the Human Data Science group at Utrecht University for including me in their group and providing me with an interesting research topic for my thesis. I would especially like to thank Dr. Laura Boeschoten for her support and guidance as my external supervisor throughout my thesis project.

A big thank you to my friends, and to fellow students for their support. Thank you to my siblings for cheering me on, and to my parents for always believing in me, and for their encouragement whenever I needed it.

Finally, thanks to everyone who participated in this research project.

Contents

List of Figures

List of Tables

Abbreviations

AI Artificial Intelligence

ART Aligned Rank Transform

CLT Construal Level Theory

DDP Data Download Package

GDPR The General Data Protection Regulation

HCI Human-Computer Interaction

IoT Internet of Things

IS Information System

TET Transparency-enhancing technology/tool

TSEF Total Survey Error Framework

UI User Interface

UX User Experience

1 Introduction

Through the span of a lifetime, an individual produces vast amounts of data almost non-stop every day. The use of Internet of Things (IoT) devices or social media platforms creates digital traces daily about people's smartphone habits, their activity level or the energy consumption of their house. These digital traces have the potential to be used in scientific research and allow researchers to measure and collect data in new ways. Through the means of data donation, participants can voluntarily donate these kinds of personal data files for research purposes. Data donation is already a known concept within the fields of citizen science and medical research, where individuals contribute to research by sharing personal data.

Since 2018, when the European Union (EU) implemented the General Data Protection Regulation (GDPR), data donation has become especially relevant[1]. The regulation gives individuals the right of access to data stored about them, as well as information on how this data is processed through Article 15 [14]. According to this article an organization must provide individuals with the actual data files and an overview of the categories of data in these files. The regulation concerns any organization collecting and storing data about individuals in the EU, and as a result these organizations must offer their customers easy access to copies of their personal data. This further contributes to the possibilities for data donation in scientific research as individuals can access their own *data download packages* (DDP).

There are however some challenges related to DDPs and data donation in scientific research. On the one hand, it is unlikely that participants are willing to share complete DDPs with researchers because of the amount of personal data and the diversity of the data within these files. On the other hand, because of the large amounts of information included in a DDP, parts of the data will not be relevant for the researchers or the research projects to collect. Based on these assumptions, it is desirable to let participants donate partial DDPs to scientific research. Social scientists from the Human Data Science research group at Utrecht University propose a workflow for collecting and analyzing DDPs from data processing entities, such as social media companies or IoT companies [5]. Following this procedure, participants will *(1)* request a DDP from a given site, *(2)*, download the DDP to their own device, *(3)* locally run a script that extracts the relevant data from this package, *(4)* give informed consent to share the transformed data with the research and *(5)* upload the transformed data to the research study (Figure 1).

Building on this framework, the next step is to develop a system which enables participants to donate data through a user interface (UI). However, participants are likely to have privacy and security concerns about donating personal data, and the UI needs to focus on trust and informed consent through the design and development process. A user interface for data donation has to encourage and facilitate trust by presenting the data pipeline of data donation in a transparent and clear way to the research participants.

1.1 Problem statement

The aim of this project is to investigate how to design and develop a data donation UI which maximizes data donation and consent. Through the design and development process we address the gap in HCI research of explaining anonymization techniques in theory and communicating this process clearly to research participants. As the proposed framework on

[1] GDPR was first adopted in 2016, and enforced as a regulation in May 2018 [14].

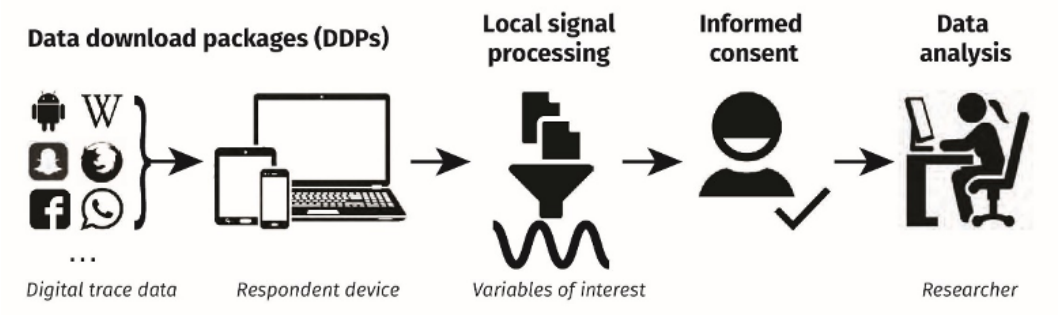

Figure 1: Workflow of how a participant donates data from a data download package to a scientific research project [5].

using DDPs for data collection should be applicable to various different DDPs, this specifically means communicating how the data is anonymized to the *average* person. Building on construal level theory research, we investigate the effects of concrete versus abstract explanations of the data anonymization technique to the participant. The main focus of this thesis is on designing for participant's trust in such a system, and how transparency of the data pipeline can help maximize the probability of participants giving informed consent to publish the data.

To narrow down the scope of the thesis project we focus on a specific use case and a specific DDP, namely location data from Google. For potential research projects concerned with participants' location or travel, location data from Google is a rich source of information. *Google Location History* stores any place a user has visited in order to customize the experience of that specific user. According to Google, the location history setting is an opt-in feature logging the locations a user has visited. This only occurs if the user is logged into their Google account with the Location History setting turned on, as well as location settings being turned on the user's device [18].

As a result of maximizing the trust of the data donation procedure, the consent level would also be maximized as more participants will be prepared to share personal data with researchers. We will evaluate the data donation interface based on levels of consent and users' willingness, trust and confidence in donating data. By investigating various methods of communicating the anonymization and trust related to data donation of location data, we seek to find the best design practices for scientific data donation systems. The results from this research can be generalized for data donation using other DDPs as well. Developing a data donation UI that facilitates trust in users is an important contribution in establishing the framework of digital trace collection, which can set a standard for using DDPs in research and encourage the use of data donation in scientific research.

1.2 Research questions

To address the problem statement, the following research questions have been defined:

> ***RQ1:*** *How can we foster trust in participants when designing a data donation user interface?*
>
> ***RQ2:*** *How do we communicate the anonymization of data to a regular user?*

This thesis paper presents an exploratory research project on the use of data donation in scientific research. In the following section, previous work related to data donation will be explained. The development process of the anonymization script, explanations of the data anonymization pipeline, and the user interface design will be explained in section three of the paper. Section four describes the research methods such as experimental study design, materials, sampling and more. Then section five will go through the quantitative and qualitative results of the research experiment. The results are discussed in section six, as well as the limitations of the research and future work. Lastly, section seven concludes the paper.

2 Related work

In this section previous research related to data donation is presented, such as digital trace collection as well as data donation within the fields of Human Computer Interaction (HCI) and social sciences. To discuss and apply concepts of trust in the UI, past work on the topics of privacy, trust and anonymization are also discussed, as well as construal level theory. Lastly, design concepts related to user interface design are discussed.

2.1 Data donation

The following section includes previous research on data donation. First, we will present the framework of digital trace data collection which serves as the basis for this research project. Then research on data donation within HCI and willingness to share data will be presented.

2.1.1 Digital trace data collection

Social science researchers Boeschoten et al. [5] claim that data download packages (DDPs) from digital platforms can be a new and innovative way for data collection in scientific research, especially within the social sciences. They present a procedure for data donation where DDPs will be voluntarily donated by research participants with their informed consent. Although this type of digital trace collection is proposed to be used in social science research, the procedure has the potential to be used in other research fields as well. The researchers also present a total error framework of the data donation procedure, to critically evaluate whether using the procedure in scientific research does result in the desired outcomes. In a total error framework potential error sources within a survey design are identified, such as bias in research. Consent bias is a type of error where participants will go through the process of data donation, but ultimately decide not to consent to donate their data. In this research we seek to investigate how to minimize this particular bias.

The 2018 EU's General Data Protection Regulation, GDPR, and the subsequent control individuals have over their own personal data is the foundation for the procedure. In Article 15(1) 'Right of access by the data subject' of GDPR every user's right of access to data is stated:

'The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data [...]' European Parliament [14]

Due to GDPR, and specifically Article 15, individuals have the possibility to request and download their personal data from for example social networks, e-commerce sites or music streaming services. The requested data must also be provided to the user in a commonly used electronic format, and the information must be in a human- and machine-readable setup [14]. As a result of GDPR, many companies offer users a central tool where they can easily request and review their data, such as Google's *'Takeout tool'* (Figure 2) [17] and Spotify's *'Download your data'* [34], or from the user's privacy settings (such as Facebook, Twitter and TikTok) [36]. Any company storing and processing personal data about EU citizens has to abide by GDPR, and thus the variety and possibility of data collection from DDPs are virtually endless.

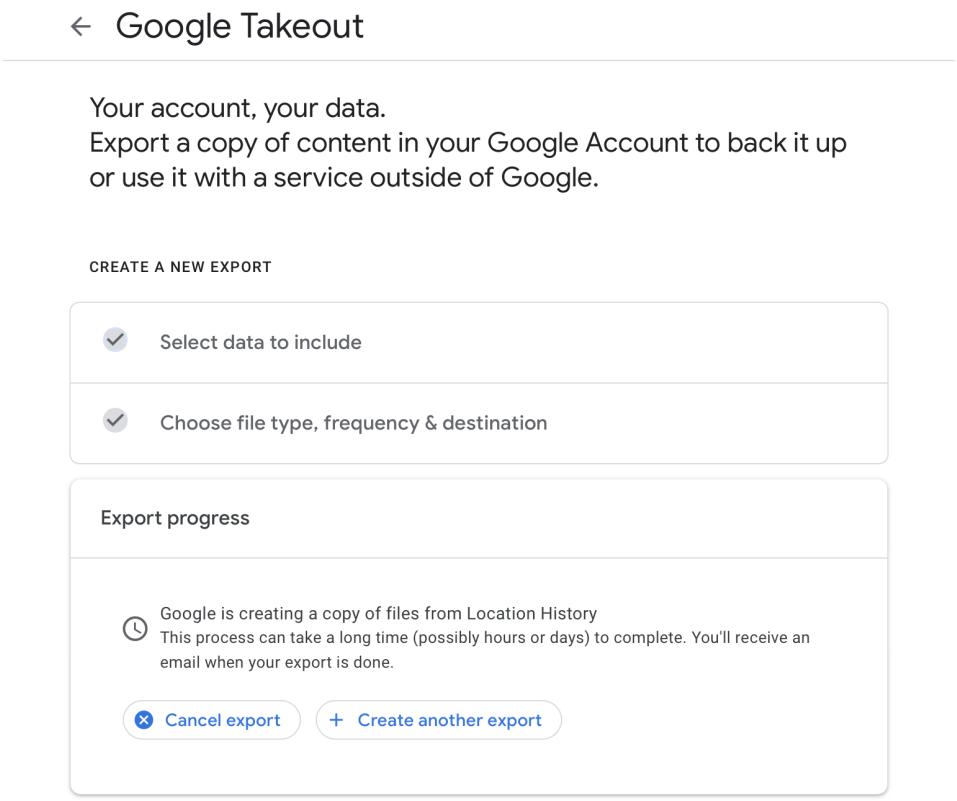

Figure 2: Screenshot of Google Takeout tool, where users can request personal data files for archiving [17].

2.1.2 Data donation in HCI

According to Bietz et al. [4] we can define data donation in scientific research as individuals voluntarily donating personal data, which initially was generated for other purposes, to create a dataset for research. For example, can an individual donate data from a social network by requesting their data (such as proposed with digital trace data collection by the Human Data Science scientists). However, if an organization uses data from users for their own research or advertising it cannot be considered data donation, as it is not a voluntary action of the user. Data donation has particularly gathered attention in medical and health research, as modern fitness trackers and wearables allow people to create large amounts of various personal health data [4, 30].

Within HCI, research has been done on mobile health applications where citizens participate in research by donating health data [11, 26]. The research by Diethei et al. [11] reveals a willingness, especially in crisis, to contribute to research by donating data. The results of the study indicate the importance of communication and transparency when designing such a system. Transparency was also the focus of a study by Maus et al. [26], where transparency and trust within citizen science projects were researched, looking at various use cases and related interface designs for data donation within health research. The research focuses

on how transparency enhancing tools (TET) can be implemented in a data donation user interface, and the effect it has on users' trust. TETs are tools made to give users a detailed overview of how data is stored, processed etc. to give users a better understanding of data processes [21], either before (ex-ante TET) or after (ex-post TET) they make a decision about their data. Although the research by Maus et al. [26] has a limited sample size, the results indicate that information about the institution users donate data to, the data donation procedure, and providing an overview of the donated data are important to the users. Interestingly, ex-post TETs such as privacy notifications to inform about suspicious privacy processes, were not seen as important and might even have an adverse effect on users' trust. Beyond this study by Maus et al. [26] there is little HCI research on transparency and trust related to data donation in scientific research.

Many data donation projects are in part also citizen science projects, a closely related concept. According to Eitzel et al. [13], citizen science is a broad term referring to how the general public, as in non-professional individuals, contribute in some way or form to scientific research. Citizen science is an established concept in scientific fields such as ecology, biology and medical research, but with an increased use of technology in citizen science it is also a growing topic within HCI [31, 38]. However, it should be noted that not all data donation projects are necessarily citizen science projects or vice versa. Citizen science projects also include individuals in other parts of the research besides data gathering, such as in the research design or data analysis [4, 13]. In contrast to data donation used in citizen science projects, the data donation framework proposed in this thesis will not be a citizen science project. The framework is suggested as a new way of collecting data within a scientific research project, where the researchers are still in control of the whole research process. The general public are not a part of the research process other than being participants of a research study. The data donation UI will provide a platform where participants can voluntarily donate data without having to install any applications or programs locally on their devices.

2.1.3 Willingness to share data

Social scientists have researched the willingness of participants to share data for data collection. Keusch et al. [22] researched the willingness to share data through passive mobile data collection from a research application on participants' smartphones. The research showed that participants were hesitant to download an app which passively collected various data from the phone. 35% would agree to passive data collection through a smartphone, while almost 40% were not willing to do so. Keusch et al. [22] suggests giving participants the option to review or delete parts of data in order to give them some control of the data collection. This will give participants the opportunity to share only parts of the data with researchers.

A study by Struminskaya et al. [35] investigates participants' willingness to share data collected using various smartphone sensors such as GPS, camera or fitness trackers. They found that the willingness to share data is generally low among participants, and that it varies depending on the sensor or data type. For GPS location almost 30% would be willing or somewhat willing to share it with researchers. The lowest willingness was found for sharing pictures of themselves (around 18%), whilst the highest willingness was for sharing data from a fitness tracker (around 60%). Privacy and anonymity concerns were mentioned most as reasons for not being willing to share data. One relevant finding was that the inclusion of a summary report featuring feedback to participants seemed to positively

influence the willingness to share data with researchers.

Both of the research studies discussed above are vignette studies where participants are asked for their opinion on sharing data. Although it shows varying willingness to share data, it is interesting to compare these results to users' willingness to share in practice through a data donation website. The chosen use case and development of the data donation UI builds upon the research by Keusch et al. [22] and Struminskaya et al. [35] to investigate how this willingness translates into a practical research experiment.

2.2 Trust and privacy

Personal data is a sensitive matter, and participants are likely to have privacy concerns when donating personal data to scientific research. A fundamental part of the data donation framework and a data donation interface is that the participants not only have trust in the researchers, but also feel that the data donation system is reliable and trustworthy. In the following section, previous research on users' trust and privacy concerns will be discussed.

2.2.1 Users trust

Trust can be defined as the willingness of an individual to depend on another party, based on the other party's attributes (often referred to as trust beliefs) (Rousseau et al. [33] as cited in Mcknight et al. [27]). Although it's been debated whether humans can form trust relationships with technology, trust research in the field of information science shows that users are capable of trusting technology, and that human-to-technology trust relationships do develop [23]. The trust relationships that form between users and technological artefacts are different from human-to-human relationships and measuring users' trust in technology therefore needs to be approached differently. The features of an information system (IS), as well as a user's perception of it, affect how the user's trust in the IS should be measured. According to the study by Lankton et al. [23], users perceive a difference in humanness when interacting with various technologies. This perception influences which trust beliefs should be used when measuring the user's trust in a system. An IS with more system-like features will by the user be perceived as lower in humanness than a system with more human-like criteria. A human-like system could be an artificial intelligence (AI) agent with a personality, or a social media platform where users interact with other users. Users associate human-like trusting beliefs as integrity, ability, or benevolence with human-like systems. In contrast, systems with lower perceived humanness are met with objective trusting beliefs such as the system's reliability, functionality, and helpfulness. However, an IS will never be completely human-like or system-like in its attributes, meaning that both trusting beliefs will be present to some degree for all systems. Nonetheless, one of the two might be more significant, and therefore better suited to use in an evaluation of a user's trust.

Mcknight et al. [27] presents a framework to determine whether technology is trustworthy, as well as how to measure users' trust relationships with technology. Trust relationships with IS have previously been regarded similarly to trust in people, but this leads to human attributes being used to measure the trust relationship, instead of technology attributes. Research on trust in recommendation agents, for example, focuses on the human traits of the IS, as recommendation agents often are built to resemble a human.

Another important aspect to consider when evaluating a trust relationship is the context of use for the system. A user has to trust that an interface will complete a task for them, because the user does not have complete control of the situation when using a specific system

[27]. This means that risk and uncertainty are a part of the trust situations between a user and the technology. Lastly, an important thing to note is that trust in technology can influence the user's decision to use technology in the first place. To measure users' trust in a data donation system it is therefore important to also measure the trusting beliefs of the user before they interact with the system itself. Furthermore, the user should be able to trust that the system is performing the specific tasks of anonymizing the data package and uploading the transformed data to a secure server. The interface will not feature any human-like traits and will therefore also not be evaluated using human-like criteria, but instead using Mcknight et al. [27] trusting beliefs related to technological attributes (reliability, functionality etc.).

2.2.2 Privacy and security concerns

In contrast to investigating users' trust in a specific system, it is also of interest to look at past research on users' decisions concerning privacy and security. The principles of *notice and choice* are seen as important regarding the protection of users' privacy. The idea is that users' privacy is protected by giving them control over their information *(notice)* and giving them various choices regarding this information *(choice)*. According to Cranor [8], this framework is flawed when it comes to current privacy and security concerns of how users' personal data are handled. The biggest issues is that privacy policies and privacy controls which are supposed to give users notice and choice, are not read by anyone and that privacy controls do not have a standard design that is familiar to users.

Previous research has investigated how to design a standard for privacy labels to present privacy information to users in a clear way, similarly to how nutrition labels give people nutritional information about food [8, 20]. The goal of the labels is to make it easier for users to find privacy information and make better decisions regarding personal data. The research by Cranor [8] found that using icons to convey privacy concerns was not seen as intuitive, as users need context to understand them. Privacy icons that are currently used by privacy tools are often not noticed by users. Although icons alone do not work as visualization of privacy concerns, implementing icons along with text in an interface could communicate privacy concerns effectively.

Research by Acquisti et al. [1] discusses users' decision making related to security and privacy concerns. One way to assist users in these decisions is behaviour interventions (also called nudges) that can be designed and implemented in the IS. One type of intervention is soft paternalism, which nudges users to make decisions without limiting their choices. In a user interface, users are nudged towards a direction or choice by every design decision that is included in the user interface. In addition, the context can influence users, for example how a technological artefact looks or feels to the user.

The asymmetric relationship between users and companies behind an IS is a concern when users make privacy decisions. The user has little information about a company's future intentions with their personal data. The user's information could be sold to a different company, but this is not clear at the time when a user is making privacy choices about their data. In addition, users are often not concerned with privacy or security decisions unless they are familiar with someone who has had problems due to not caring about the security of their data. The potential risks related to security are not the primary concern of the user when interacting with a system.

2.2.3 Opting out

A user-study by Habib et al. [19] covers the usability and usefulness of privacy choices on websites. Their findings indicate that users are sceptical of privacy controls, as well as having problems operating them. Examples of privacy controls are cookie consent, or the possibility to delete or edit personal data. The use of design features such as dark patterns and default settings potentially nudges users in a distinct direction. Users are likely to agree to default settings, because they trust that the company knows best.

A research into the opt-out choices available to users by Bannihatti Kumar et al. [3] investigates automatic extraction of these choices in privacy policies. Few users take advantage of opting out of web trackers and making privacy choices. Users also seem to misunderstand the scope of these choices, believing that they opt-out for all tracking instead of just one website. The user study showed that an extraction tool which provides users relevant information and concrete choices helps them become more aware of the privacy choices they have.

2.3 Data anonymization

To encourage participants to donate data, the data in the DDPs have to be anonymized before the data is donated to the research project. Anonymizing the data in the participants' DDPs is therefore a crucial part of the data donation framework presented in this thesis. As stated in the introduction, the anonymization techniques used on the data file that participants upload, will be presented as an element of the user interface design.

An anonymization pipeline consists of input, output and the script used to anonymize the data. We start with inputting the data which contain identifiable information, run an anonymization script on this data source, and output the now transformed, and anonymized data. The process of anonymizing data involves removing any direct or in-direct way of identifying and retrieving individual data points from the initial data input [2, 7, 10]. Direct identifiers can be information about name or address, while age or gender are examples of indirect identifiers. Generalization, suppression, or randomization of data are all examples of various ways to remove identifiers and anonymize data.

Data anonymization techniques are clearly explained in textbooks for people with a background in computer science or similar fields. It is however not clear if this is an efficient way of communicating an anonymization pipeline to a regular user with no particular knowledge in the field. An important aspect of fostering trust when interacting with the data donation UI is to make sure that the users have an understanding of the data handling and privacy of their data. There is a clear gap in HCI research as it is stated in literature how data anonymization works, but not how to effectively communicate it to users to provide transparency and facilitate trust.

In the proposed workflow, the anonymization script runs locally on a user's device and extracts only the data from the DDP which is relevant for the research study. The relevant data will then be anonymized by removing any identifiable data points. Furthermore, the anonymization script should protect the privacy of the participants by de-identifying personal information and any geographical identifiers such as street name, addresses or coordinates. Additionally, anonymization of the data allows researchers to publish the transformed data donated through a data donation interface.

2.4 Construal level theory

To accomplish that participant consent to anonymize and donate data to research projects, it is essential to explain to the participant how the anonymization of the data is accomplished. If the anonymization is communicated in a clear and transparent way so that participants understand the process that happens to anonymize their data, consent and willingness is likely to be maximized.

We propose to apply construal level theory (CLT) to address the research gap of how to communicate the anonymization pipeline to a regular user and research participant. CLT is a psychological theory by Trope and Liberman [37] which proposes that psychological distance has an influence on how people perceive things. People are capable of thinking in abstract or detailed representations, and of events or objects in the future, past, or even imaginary. Psychological distance is the distance to an event or representation that is not a direct experience happening right now [37]. According to CLT, time, space or social distance affects the psychological distance of a representation. This in turn influences the construal level of the mental representations in a person's mind. Higher level construals are abstract representations far removed from a direct experience, which then give a superordinate overview of something. On the other hand, if the psychological distance is smaller, the level of construal will be lower. A lower level of construal will have more concrete, detailed and constant representations. Research shows that attitudes are likely to be positively impacted by an abstract representation, due to the superordinate overview in a higher level of construal [37]. When applied, an abstract, higher level of construal can be related to answering 'why' an action is performed. On the other hand a lower level of construal, a concrete representation, will focus on 'how' this action is performed. Research shows that construal representation can be manipulated by priming 'why' and 'how' when we implement it in an argument or representation [9, 37]. A research study within HCI applied CLT to the feedback design of an eco-driving IS to study changes in the driver's behaviour [9]. The results showed that abstract feedback had a significant effect on driving behaviour, while concrete feedback did not.

An example of communicating the anonymization technique with a higher level of construal could be: *'Anonymization is performed to protect your privacy'*. A lower level construal explanation of the same step could be: *'Anonymization is performed by removing any of your identifiable information from the data'*. To address the gap in knowledge on communication data anonymization techniques to regular users, we propose to apply CLT in the data donation framework. The anonymization will be presented in a higher level and lower level of construal representation, as well as a baseline version with no explanation of the steps in the data anonymization pipeline.

2.5 User interface design for data donation

To design and develop a data donation interface we also need to focus on the user interface design of the system. First design principles of user interface design will be discussed, and then we will make a summary of the design implications from the related work presented above.

2.5.1 Design principles

When designing an interface or application, there are a set of design principles within interaction design that can guide designers to creating good design with a good user experience

[32]. The principle of visibility concerns the fact that users will know what action to perform next if the controls are clearly visible to them. Visibility improves the user experience because information is transparent and communicated to the user, thus removing uncertainty from the users. Another important design principle is feedback, where users receive some kind of feedback information as the result of an action having been completed. Getting a response is important so that a user knows that they can continue with their actions, for example by providing a confirmation or error message when completing an action. Furthermore, implementing constraints in the design of a user interface is another essential design principle. Constraints restrict the users in the actions they can perform, in an effort to avoid mistakes being made by users. Deactivating certain buttons is a common way to implement constraints. The button or option is still visible to the user, but greyed out so that a user cannot choose this option at certain times.

2.5.2 Ephemeral design

In the case of data donation, implementing elements from ephemeral user interface design in the UI might be a way to make users aware of the fact that the data is not stored for a longer period and that the interaction is limited in time.

It should be clear to the user of the data donation interface that the data is handled locally on their computer, and that the interaction with the interface is done when the donation process is completed. Temporal content, which only lasts for a limited time, is a feature of ephemeral interfaces and designs. Döring et al. [12] explores the background of ephemeral design and various ephemeral user interfaces. Ephemeral user interfaces are defined as user interfaces that have at least one element that is only meant to last for a limited time. The inclusion of the disappearing UI element should accompany the intention of the interaction and user experience when using the system. Another characteristic of ephemeral UI is that the material used in the interaction also has a meaning behind it. In addition, the interaction might be purposely designed in a way that gives the users limited control, to make the interaction engaging. Lastly, an ephemeral design has similarities with natural phenomenons as there is a time limit on the interaction.

The concept of ephemeral design has mostly been applied to interaction design which has a strong emphasis on poetic and artistic interaction between the user and the design. An ephemeral UI might also include natural input and output elements like water, plants or soap bubbles. However, applying the concept of ephemerality, including using ephemeral metaphors to a *regular* user interface has not been done before.

2.6 Summary and contribution

The main contribution of this research project is to investigate the effect of the construal level when communicating a data anonymization pipeline to users. In addition, the research project will provide insights from participants on donating data through a data donation user interface, as well as insights from a practical experiment where participants are asked to donate data from personal DDPs.

Based on the research questions defined in the introduction, and the literature from the related work section, we hypothesize that:

> ***H1:*** *The construal level in the explanation of the data anonymization technique will lead to a difference in users' trust in the data donation interface.*

H2: *The construal level in the explanation of the data anonymization technique will lead to a difference in users' willingness to donate data within the data donation interface.*

H3: *The construal level in the explanation of the data anonymization technique will lead to a difference in users' confidence when donating data within the data donation interface.*

3 Design

The following section presents the design process and the materials developed for this research project. First we will present the use case chosen for the data donation procedure, in addition to the process behind the anonymization script and the explanation of the data pipeline. Third we will explain the design process for the user interface, in addition to how the UI was developed as a web application.

3.1 Data donation use case

To research the process of data donation in a practical research experiment, we focus on one specific use case of data donation for scientific research using DDPs. According to previous research on willingness to share data, location data is one of the data types participants are most likely to share with researchers. Writing a script for anonymizing location data is also possible within the scope of this research project. Therefore the use case chosen for the experimental study is data donation of GPS location, more specifically Google location history data. In the experiment participants will be asked to donate their Google location data for a research study on changes in movement patterns during the COVID-19 lockdowns. The focus of this research is to develop a front end design and working prototype of the user interface. The anonymization script will be presented to the user, but will not be implemented to process any data from the user as this is beyond the scope of the research, as well as the collecting and analysing this data is not relevant to the research questions presented in this thesis.

3.1.1 Google Location data

Users of Google's services can request an export of their data stored by the company through Google's Takeout tool (see Figure 2). In this process users are able to download a complete archive of their Google account data, or only select parts of it such as location history in Google maps, YouTube data, or emails and attachments in Gmail. The chosen data is then exported in the requested file type and sent to the user. As the chosen use case is for participants to donate location data, participants can choose to request only the *Location History* data. This is preferable because it will take less time before the participants receive the data file. However, the participant can also select all of the data as the anonymization script is written to only select the relevant location data from the DDP when a user is donating data.

Google Location history data saves every place the user goes with a mobile device if they are signed into their Google account, have Location History turned on and the location tracking is also turned on on the device itself [18]. The Location History feature is an opt-in feature meaning that users have to turn it on for this data to be tracked and stored in their account.

3.2 Anonymization script

In a data donation system based on the proposed workflow by Boeschoten et al. [5], the anonymization script to anonymize the data and extract only the relevant parts of the DDP is a crucial part. To anonymize the location data from Google an anonymization script was written in Python. The script will be implemented in the UI design and presented to the user in the same step as the anonymization and data pipeline is explained.

Two iterations of the anonymization script were developed. In the first version the user is asked to input their primary address(es), in order to identify which addresses to focus on. This would lead to all other locations in the data being ignored and only the relevant data to be extracted and subsequently anonymized. However, for complete anonymization this was changed for the final script and to avoid participants having to input any information. In the final version the code is written to extract the top three most visited places based on the labels *Primary address* or *Secondary address* which are used in the Google location history data file. Only the data with these labels are extracted from the data package. Then, the number of visits and duration of the visits at these places are calculated. This means that there is no personal information included in the final output of the script, and the user does not have to input any information about addresses.

The first version and the final version of the anonymization script are included in Appendix A.

3.3 Communicating the anonymization

In this section we will present the development behind the explanations of the data pipeline. These explanations were implemented in the final working prototype to communicate the data pipeline to participants and how their data was anonymized.

Based on construal level theory by Trope and Liberman [37], the three explanations were different in terms of the level of construals. There are various elements to consider when developing the various test conditions: how the explanation should be implemented in the UI design, the textual content in the explanations and the visual presentation. Both the written explanations of the anonymization technique, and the visual presentations, were developed in an iterative way.

First, the explanations were written out according to the different test conditions. The BASELINE version includes the three steps of *(1) input source, (2) run script* and *(3) output source*, without any further explanation of what is happening at each step (similarly to the steps in Figure 3, but without the explanation in smaller font). The ABSTRACT explanation also consists of three steps, but includes textual content for each step to explain to the user why the data is anonymized. The CONCRETE explanation details how the anonymization of the location data actually is performed, and consists of the three steps plus sub-steps which give concrete examples of anonymization from the DDP.

Research shows that users do not read when visiting websites, but rather scan them [29]. Therefore the texts were further developed to include icons showing a visual step-by-step procedure of the anonymization to the user. This was first implemented in the data donation UI just as an image, but through user testing this was revealed to not be a good user experience as the text was too small (Figure 3).

The following iteration built on the image showing the pipeline, and implemented this as a stepper element in the web interface. The idea being that an interactive element such as a progress stepper allows users to click through the steps of the explanation. This leads to more engagement than presenting a plain text or image to the user.

In the final iteration the explanation of the data pipeline is presented to the user in an infographic video (see Figure 4). The infographic videos are included in step 3 of the data donation procedure and auto-plays when the user opens this step. The infographic videos are linked in Appendix B.

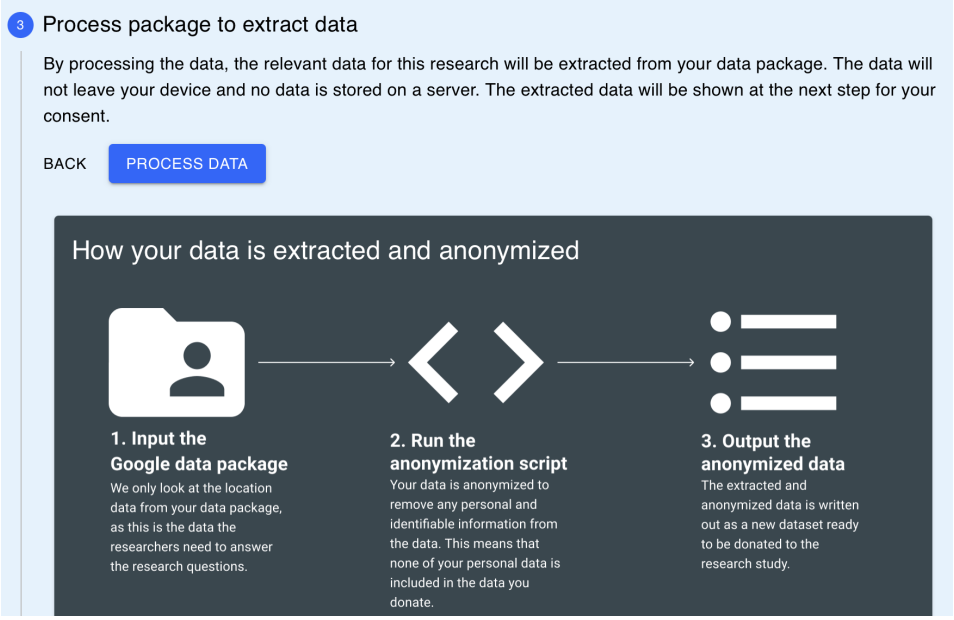

Figure 3: An early version of communicating the data pipeline in the ABSTRACT condition.

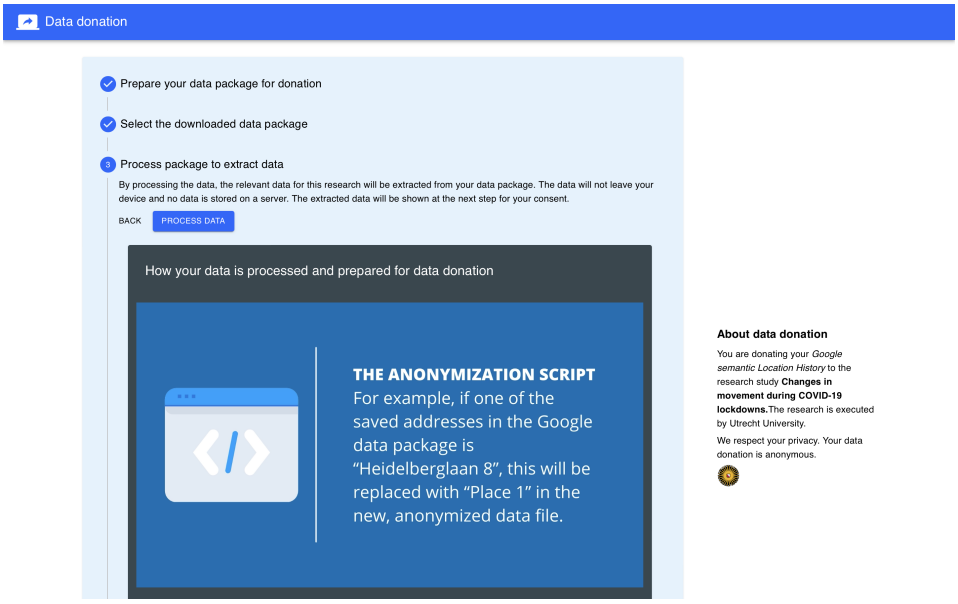

Figure 4: The final implementation of the explanation of the data pipeline in the CONCRETE condition.

3.4 Data donation interface

In the following section the design and development of the user interface design for the data donation website will be explained.

Based on user-centred design principles the data donation interface was developed through an iterative process. From the proposed donation framework and discussions with the researchers, a set of requirements were found. Based on the list of requirements the first sketches and wireframes were developed. These were again discussed and evaluated along with the researchers, before the low fidelity and subsequently high fidelity prototypes were made in design tool software Figma.

To test the user interface in the research experiment, a web application had to be developed, as the high fidelity prototype created in Figma lacked the possibility to give users an interactive experience when uploading a file for data donation. The web application was created using React, an open source JavaScript library for developing user interfaces.

3.4.1 Requirements

This thesis project and the development of the data donation user interface is based on the workflow on digital trace data collection using DDPs presented by Boeschoten et al. [5]. The following design requirements were established based on stakeholder interviews, previous research presented in the literature section, as well as design principles for UI design. When talking about the user in the requirements stated below, we are referring to participants who are donating data.

The researchers from the Human Data Science group were established as the stakeholders of this project and were interviewed about the needs and essential parts of a data donation system. The data donation system follows the five steps of the data donation procedure seen in Figure 1. On the donation page of the user interface, information should be included to guide the user through each of these steps and be transparent about the data donation procedure. Furthermore, the system should be easy to use. According to the researchers, the intent of the user interface is that if participants are able to interact with social media sites (i.e. they have an account), they should be able to use the data donation system with ease.

From stakeholder interviews the following seven requirements were established:

- The user should be able to donate data through the website without having to install a program on their device
- The user should be able to use the interface without having to log in to the system
- The user should be explained each step of the data donation procedure
- A dashboard should display the available research projects users can donate to
- The user should only be able to select a .zip file from their local device to donate
- The user should be able to **not** give consent to data donation and still finalize the data donation procedure
- The anonymization script should be available to the user for inspection and transparency

Requirements were also defined based on previous research on data donation and willingness to share data. Research by Diethei et al. [11], Maus et al. [26], and Struminskaya et al. [35] indicates the importance of transparency and communication to foster trust and increase willingness. This includes giving an overview of the transformed data that is to be donated from the uploaded DDP-file. Furthermore, including information about the institution will likely positively influence the user's willingness to donate data. Diethei et al. [11] suggests that information about the research project should be included on the data donation website, so users do not have to visit the research project's own website when seeking information.

The following four requirements were based on previous research:

- The user should be shown an example of a feedback report of the transformed data
- The user should be given the option to accept or reject to give informed consent
- The anonymization should be communicated to users
- Information about the institution should be visible on each page
- Information about the purpose of the research project should be visible each page

Finally, some requirements for the data donation system were determined from design principles. Following the design principle of visibility, all the steps in the data donation procedure are clearly visible in the user interface. This ensures transparency of the procedure before the user starts the data donation procedure, but also allows the user to see which steps have already been completed. When the user clicks through to the next step in the procedure, previous steps are marked with a check mark or greyed out to provide feedback to the user whether the action is completed or not.

The following four requirements were established from design principles:

- The design should make the data donation procedure easy to complete
- The UI should display previous, current and upcoming tasks to the user
- The user should be able to move back and forth in the data donation procedure
- It should be visible to the user which tasks were and were not completed

3.4.2 Low fidelity prototype

To start the design process initial sketches were drawn using pen and paper. These sketches were based on the established requirements to conceptualize the idea of a data donation UI. These were created using various design methods such as the *Crazy Eights* method, where the designer has to sketch eight ideas in eight minutes. These kinds of design thinking methods allow for brainstorming various ideas and concepts before continuing with the prototyping.

From the sketches, two different versions of low fidelity wireframes were created in Figma. The wireframes display the essential elements and frames that are to be included in the data donation UI, as well as the relation between the frames. The first wireframe design presented the data donation procedure to the user as a sort of slider, where only the current frame

is displayed to the user and the user clicks back and forth between the various steps like a slideshow. However, a second version was created to make the whole process visible. This version includes a row of sections, which expand and collapse. By doing so, the data donation procedure is transparent and visible to the user, which is more in line with the established design requirements for the data donation UI. Figure 5 shows the second wireframe design.

Initially a couple of ideas on how to implement ephemeral design elements in the data donation UI were sketched up and discussed with the researchers. One idea was to present the steps of the data donation procedure using circles instead of more traditional box elements. Another concept for implementing ephemerality in the interface was to imitate an old-fashioned screen being turned off and fading away after a participant completes the data donation procedure. Both of these ideas were supposed to indicate to the user the ephemerality of the process, and that none of the user data was stored besides the donated data file. Ultimately none of these ideas of ephemeral design were proceeded with in the later steps of the design process and in the development of the data donation system.

All of the sketches and wireframes designs, including the overview of the interaction steps, are included and linked in Appendix B.

Figure 5: A wireframe design of the data donation UI.

3.4.3 High fidelity prototype

After having created the sketches and wireframes, mock-ups were made to show the high fidelity prototype of the data donation UI. The prototype was also created using Figma, together with a Material-UI design kit (MUI) based on Google's Material Design. Material Design is a design system created by Google following common design principles [25]. Design

systems are a collection of standards to be used when designing user interfaces, usually including reusable components [15]. Using a design system when designing and developing an artefact makes it possible to focus on more complex issues, as the design system offers components and guidelines that are proven to work. In addition, MUI is available as a framework for React making it easy to implement the prototype design made in Figma when designing the web application.

According to research it would be beneficial to include information about the institution behind the research study [26, 35]. Therefore a dashboard page was created showing currently active and inactive research studies along with accompanying descriptions. The active research study is the research study participants engage with during the experiment, while the inactive dummy studies are included to fill up the dashboard page to make it more realistic. Information about the institution and the fictional research project which participants donate data to are also included on the data donation page.

A summary report of the data donated is included in step 4 of the data donation procedure, the final step before participants are asked to consent, as this was seen as a positive addition by users in previous research on data donation [26, 35]. The summary report is visualized with a table, presenting the user with an example of the transformed and anonymized data included in the data donation file. The data included in this table is just an example from a simulated data package from Google and not the actual data donated from participants. A screenshot of the high fidelity prototype showing the feedback table can be seen in Figure 6.

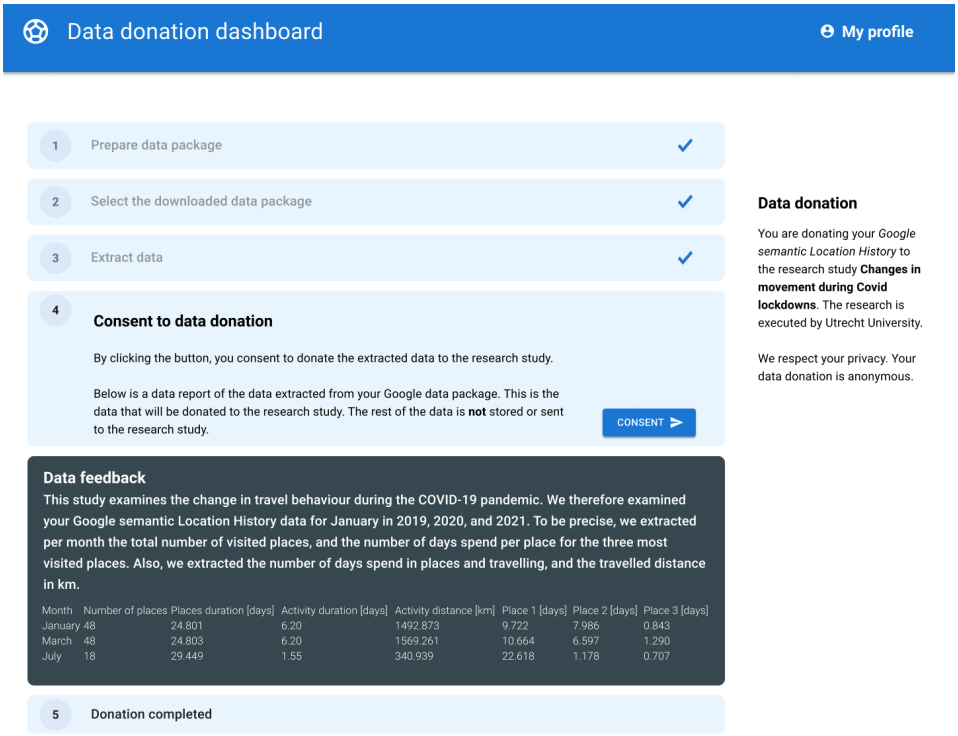

Figure 6: Figma prototype of the data donation UI.

3.4.4 React web application

The final working prototype of the data donation UI was developed in JavaScript along with the React library and MUI (Material-UI) framework. React is a free and open-source library for building interactive interfaces, while MUI is a framework for React which implements the design principles of Google's Material Design [16, 28]. These libraries provide a vast amount of components when developing an application.

The web application consists of two main pages, the dashboard page showcasing the research studies available for data donation, and the page for the data donation procedure (see screenshot Figure 7). The pages are responsive in case participants use smaller devices, although participants are advised to use a desktop or laptop during the research experiment.

The application was built and deployed using GitHub pages. GitHub pages can only host static pages, and to avoid making it obvious to computer proficient participants that the web application was a static page and not actually collecting data files, the web application was hosted on the researcher's own domain.

Various versions of the interface were created for the three different test conditions, and for the three experiment rounds to link the correct surveys for each condition and round. The full source code for the web application can be found in the GitHub repository linked in Appendix B.

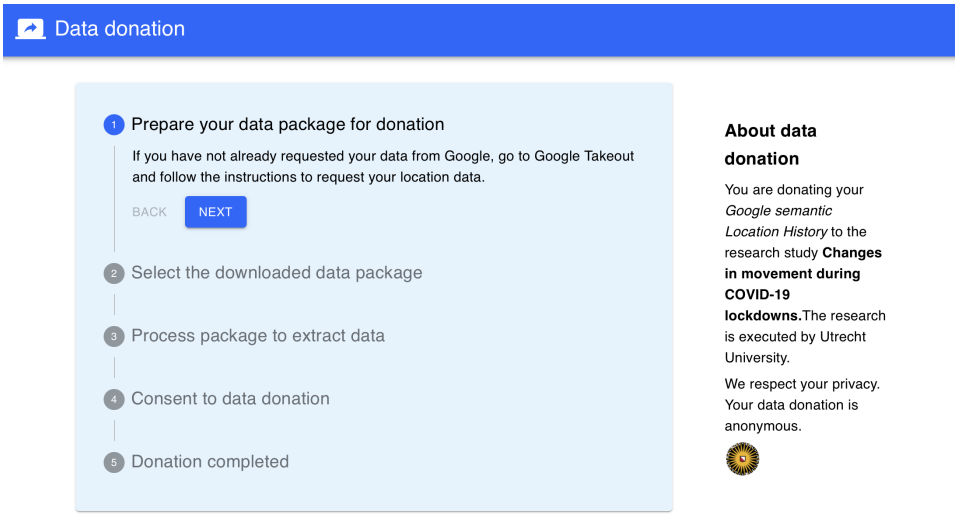

Figure 7: The final version of the data donation web application.

4 Method

For this thesis project we set up a within-subjects experiment where participants were asked to donate data files through a data donation interface, for a fictional research study. The communication of the data pipeline and anonymization to the participant was manipulated based on construal level theory by Trope and Liberman [37]. Three test conditions were created for the experiment: a BASELINE condition with no explanation, an ABSTRACT condition with abstract explanation, and a CONCRETE condition with concrete explanation of the data anonymization.

In the following section the research methods and study design will be explained in detail. The procedure and measurements of the experiment, as well as sampling are also included in this section.

4.1 Pilot study

A smaller-sized pilot study was conducted to evaluate the experiment setup and uncover potential for improvements in the efficiency and quality of the research experiment. The pilot study was conducted with four volunteer subjects (2 female, 2 male). The participants were provided with the links to the three rounds of data donation procedures to walk through at their own preferred time. The results of the pilot study indicated that the experiment design was feasible, but some minor changes were made.

As a result of the pilot study some changes were made to the data donation UI. One of the subjects indicated that seeing a long paragraph with the anonymization script was more confusing to them than beneficial, because they did not understand it anyway. On the other hand, another subject claimed the opposite, namely that showing the anonymization script showed transparency about the research and data handling. Therefore, instead of presenting a long paragraph of the anonymization script at step three, we implemented an expanding and collapsing paragraph with the script. By implementing this the script is still available to participants by expanding the pane, and if they are interested they can look through the code.

The explanations of the data anonymization technique were also piloted. The researcher interviewed the subjects to check their understanding and investigate possible misunderstandings. As a result of this, the illustrations showing the data handling steps were changed from a static image to an interactive stepper that the user could interact with themselves. This was later changed into an infographic video showing the steps of the anonymizations to the participants to offer a more engaging experience to the user.

4.2 Experimental study design

The research experiment was set up as a within-subject study, meaning that each participant tested all three different conditions in the experiment. An advantage of this experiment setup is that fewer participants are needed for the research. This was preferable in terms of recruitment and scheduling of the experiment. The individual differences between participants are minimized as the same participants test all conditions in the experiment. However, a drawback of the within-subject design is that the participants have to go through the data donation multiple times, which likely would not happen in a real-world scenario where a data donation website is used by a research study.

To account for confounding variables and order effects the experiment was randomized according to the Latin square principle [24]. To avoid carry-over effects and due to the odd number of conditions, a balanced Latin square design was implemented, using all possible combinations of the three test conditions. The setup was fully balanced, as all test conditions followed each other the same number of times. In the pre-experiment surveys, participants were given randomized identification numbers consisting of five digits, and appointed to one of six groups to distribute the experiment evenly. The randomization was set up and executed in Qualtrics survey tool as a part of the survey flow.

The independent variable in the experiment is the level of construal used when communicating the data anonymization to the user in the user interface. There are three test conditions: no explanation of data anonymization steps, explanation of the data anonymization steps in a higher level of construal and explanation of the data anonymization steps in a lower level of construal. The dependent variables measured in the experiments are users' propensity to trust, trusting stance, willingness, preference, and confidence in donation.

4.2.1 Conditions

The three experimental test conditions are presented below. The anonymization script is included in all of the test conditions, but the explanation of the anonymization steps is manipulated by applying construal level theory.

- **BASELINE condition:** No explanation of the data anonymization technique besides presenting the steps in the data pipeline
- **ABSTRACT condition:** An abstract explanation of the data anonymization technique (higher level of construal)
- **CONCRETE condition:** A concrete explanation of the data anonymization technique (lower level of construal)

4.3 Measurements

To answer the three hypotheses stated in the related work section, the following measurements were used. To measure users' trust in the data donation UI, McKnight's measures of *'Trust in Technology'* were used [27]. This includes measuring users' propensity to trust technology, which consists of two constructs *'Faith in General Technology'* and *'Trusting Stance'*. The two constructs consist of 4 and 3 item statements, respectively, where participants score each statement on a 7-point Likert scale from *Strongly disagree* to *Strongly agree*. The measurements are included in the pre-experiment survey, and are tried and tested questionnaires to evaluate users' beliefs about attributes in technology.

In the experiment rounds, McKnight's measures on *Trusting belief for Specific Technology* were adapted to measure users' trust in a specific system for each test condition. Three of the *Reliability* item constructs were used in the questionnaire (item number 3 is not included as it was deemed not to be relevant for the system), while all three of the *Functionality* item constructs were used. *Helpfulness* was not included in the survey, as the system does not include a help function in any of the three test conditions. By doing so, the workload of participants is also minimized, as they already answer multiple of these statements after going through the data donation procedure.

Based on previous research on willingness to donate data by Keusch et al. [22] and Struminskaya et al. [35] measurements on willingness were also used to investigate willingness

in the various test conditions. In the experiment rounds, participants were therefore asked if they consented to data donation, and if so, how willingly they consented to donate data on a scale from 1 to 9. We assume that if participants understand how the data is anonymized they will be more willing to donate data. Furthermore, in the post-experiment questionnaire a final question on willingness to donate data again in the future was included to investigate potential willingness to donate again through a data donation system.

In the post-experiment survey participants are also asked how confidently they donated data for each condition, on a scale from 1 to 10. The assumption is that the clearer a data anonymization is explained to participants, the higher their confidence in giving consent to donate data will be. Additionally, participants are asked about their version preference in regards to confidence in giving consent to data donation.

The complete survey texts, including informed consent and explanations given to participants, are included in Appendix C.

4.4 Procedure

The experiment was run over the course of a few weeks. To recruit participants for the main experiment, potential participants were asked to fill out a short pre-experiment survey in Qualtrics survey tool. In the pre-experiment survey participants were informed about the experiment and asked to consent to take part in the experiment. The survey included questions about demographics, as well as the propensity to trust questionnaires consisting of questions on users' trust in general technology and trusting stance. At the end of the pre-experiment survey, participants were asked for their contact information, specifically their email address, to contact and invite them to participate in the rest of the experiment. At the end of the pre-experiment survey participants were also instructed on how to request their Google location data through the Google Takeout tool. This was done to ensure that the DDP from Google was readily available on participants' devices for the data donation rounds as requesting the data can take a few days to be processed and completed.

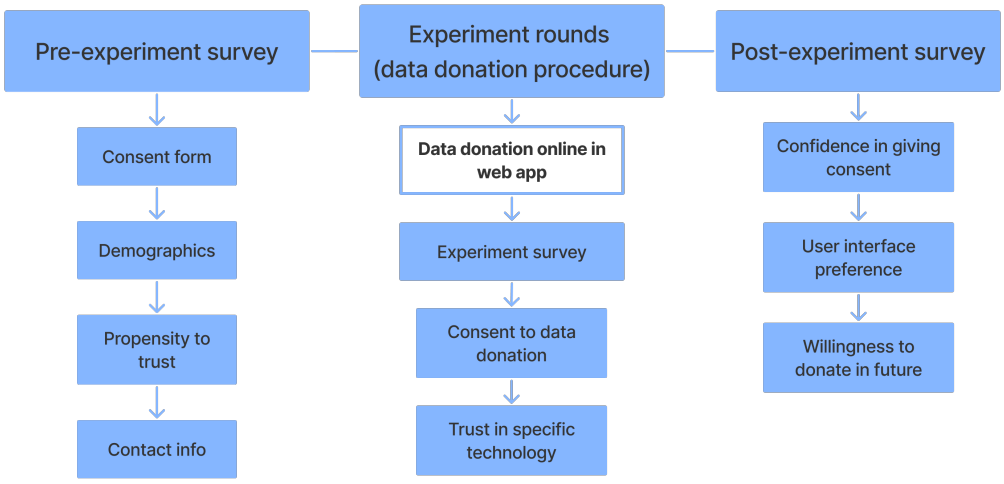

Figure 8: Protocol for the experimental study design.

Next, the participants were invited to the main experiment where they were asked to donate data in the three different test conditions. In the invitation email, the experiment instructions were included along with a link to the data donation website. In the final data donation step a button with a link to the accompanying questionnaire was included, where participants were asked to fill out the survey for the experiment round. In the survey participants were asked whether or not they consented to donate data, and if so how willingly they consented to donate data. Participants were also asked about trust in the specific system in this survey. When enough participants were recruited through the pre-experiment survey the first round of data donation started. The second and third data donation rounds were scheduled six days after the previous date of data donation. For the third data donation round, the survey also included post-experiment questions on users' confidence in donating data, version preference and future willingness to donate data. Open-ended questions were also included to offer more qualitative insights into the topic and the process of data donation in research. To increase the completion rate, reminders were sent to all participants two and four days after the initial invite for the data donation round. An overview of the study protocol can be seen in Figure 8.

4.5 Participants

A total of 61 participants *(N = 61)* were recruited to participate in the experiment. The participants were recruited via means of convenience sampling. Three recruitment criteria were established; participants needed to have a Google Account, their location settings turned on, and have access to a computer to perform the data donation rounds of the research study.

Of the initial sample size, 52 of the recruited participants consented to take part in the research project, filled out the pre-experiment questionnaire completely and provided an email address to be contacted for the subsequent parts of the study. Five participants were excluded due to not meeting the criterion of having location settings turned on and therefore did not have any location data to donate. One participant withdrew during the first round of data donation due to not wanting to share their location data with the research, while 2 participants withdrew by opting out from the mailing list distributed through Qualtrics.

4.5.1 Experiment sample

Twenty-four *(N = 24)* of the invited participants took part in at least one of the experiment rounds. Ten participants were male, while 14 of the participants were female (see Figure 9). The mean age of the participants was 33.04, *(M = 33.04, SD = 11.24)*. The plot in Figure 10 shows the distribution of participants' reported education level. Thirteen of the participants listed a Bachelor's degree as their highest level of finished education, nine participants listed a Master's degree, and two participants a PhD.

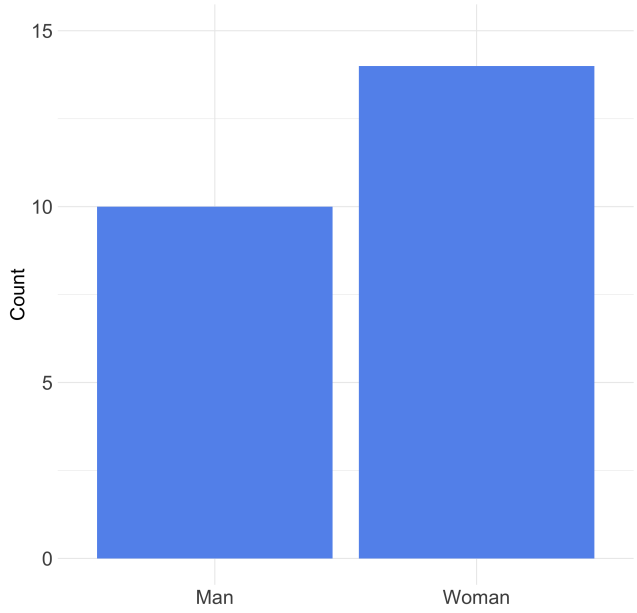

Figure 9: Gender distribution of the particpants in the experiment sample.

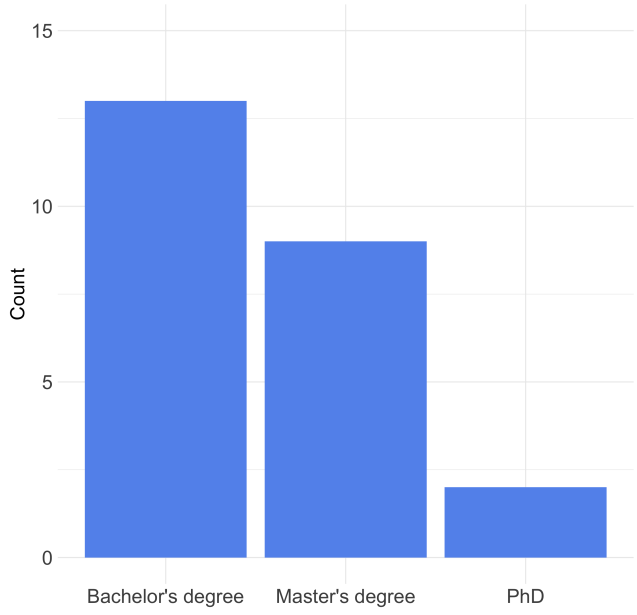

Figure 10: Highest level of education among participants in experiment sample.

5 Results

This section will present the results of the quantitative analysis and the qualitative analysis of the research experiment. First, the statistical analysis will be presented, followed by the results from the remaining quantitative measures. Then, the qualitative analysis and results will be presented.

5.1 Quantitative analysis

The following section presents the procedure for the statistical analysis, as well as the results of the hypothesis testing for the three hypotheses presented in the literature section of this thesis. Furthermore, the results from the quantitative measures of users' preference and willingness to donate again will be given.

Before conducting the analysis on the collected data, responses from excluded participants, non-responses and partial responses were removed from the dataset. Exploratory data analysis was also conducted to uncover initial findings before conducting the statistical analysis. The code for the statistical analysis is linked in Appendix B.

5.1.1 H1: Trust in data donation system

Before conducting the statistical tests for the hypothesis testing of **H1**, Levene's test and normality tests were performed on the data. Both the assumption of variance among the groups and normality were met. Normality tests were performed using both Q-Q plots and the Shapiro-Wilk test.

Condition	TiST M	TiST SD
BASELINE	5.64	0.94
ABSTRACT	5.38	0.83
CONCRETE	5.32	0.77

Table 1: Trust in specific technology (TiST) for the three test conditions.

For the BASELINE condition the mean score for users' trust in specific technology was 5.64, for the ABSTRACT condition the mean score was 5.38, and for the CONCRETE condition the mean was 5.32. The results can be seen in Table 1, as well as in the boxplot in Figure 11. A one-way ANOVA was used to test **H1**, comparing the effect of construal level in communicating data anonymization technique on users' trust in a specific system. Results of the ANOVA showed no significant difference on *Trust in specific technology* between the three test conditions, $F(2) = 0.69$, $p = 0.51$.

In addition to trust in specific technology, users' propensity to trust technology was also measured. The mean score for users' *Propensity to trust* was 5.1.

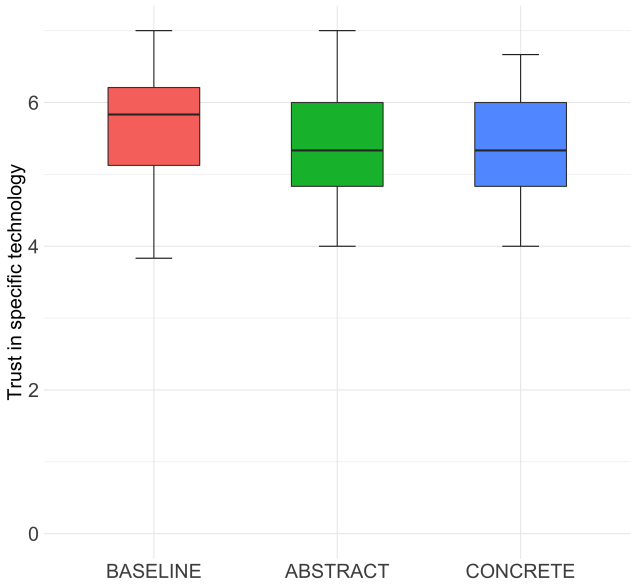

Figure 11: Boxplot of users' trust in specific technology per test condition. Likert scale from 1–7.

5.1.2 H2: Willingness to donate data

For hypothesis testing of **H2**, users' willingness to consent to donate data, assumptions of variance and normality in the data were examined. Levene's test showed that the assumption of equality of variances among the groups was met. Normality tests showed that the data was not normally distributed. A one-way ANOVA with aligned rank transform (ART) was performed to test the effect of explaining the anonymization in different construal levels on users' willingness to donate data through the data donation interface.

Condition	Willingness M	Willingness SD
BASELINE	8.05	1.10
ABSTRACT	7.59	1.53
CONCRETE	7.62	1.36

Table 2: Users' willingness to donate for the three test conditions.

The mean scores and standard deviations for each test condition can be seen in Table 2, and the boxplot in Figure 12. The highest mean score for users' willingness was for the BASELINE condition (8.05), while the mean score was 7.59 for the ABSTRACT condition and 7.62 for CONCRETE. We did not observe a significant effect of construal level theory in communicating data anonymization on users' willingness to donate data, $F(2) = 1.50$, $p = 0.24$.

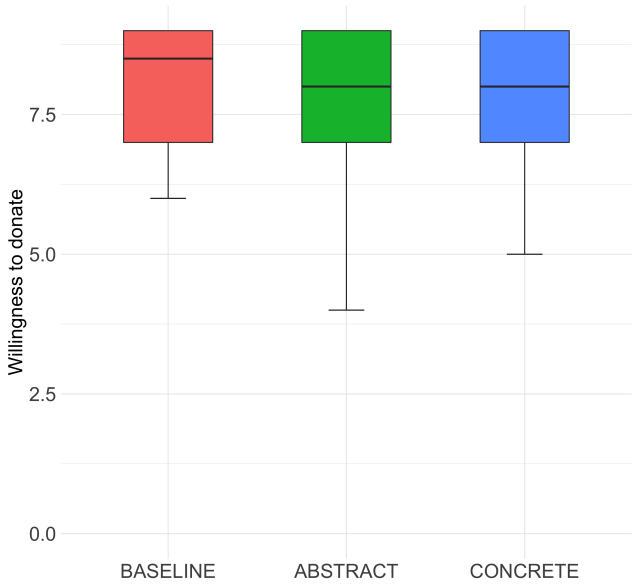

Figure 12: Boxplot of users' willingness to donate per test condition. Scale from 1–9.

5.1.3 H3: Confidence in donating data

Before conducting the statistical tests for **H3**, extreme outliers in the results were detected and removed using boxplots. Six outliers were in total removed before continuing with the statistical analysis. Levene's test showed that assumption of equal variance among the samples was not met. Both the Shapiro-Wilk test and Q-Q plots showed that the assumption of normality in the sample was met.

Condition	Confidence M	Confidence SD
BASELINE	6.75	2.12
ABSTRACT	7.79	1.44
CONCRETE	7.20	0.68

Table 3: Users' reported confidence when donating data.

The mean confidence score was lowest for the BASELINE condition with 6.75, while ABSTRACT condition had the highest mean, with a mean confidence score of 7.79. The CONCRETE condition had a mean confidence score of 7.65. The results can be seen in Table 3, and Figure 13 shows a boxplot of the results for the three test conditions. To test **H3**, the effect of construal level in anonymization on users' confidence in donating data, a one-way ANOVA with ART applied was conducted. We do not observe a significant effect of applying construal level theory when communicating data anonymization on users' confidence in donating data for the various test conditions, $F(2) = 3.20$, $p = 0.05$.

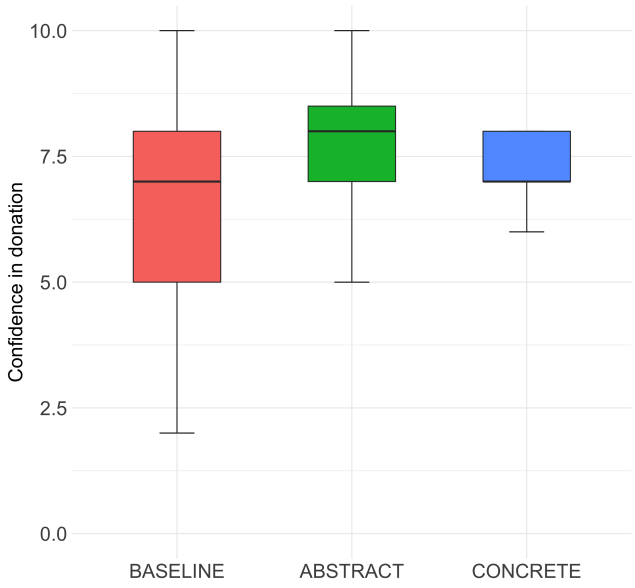

Figure 13: Boxplot of confidence in donation per test condition. Scale from 1 - 10.

5.1.4 Version preference

In addition to reporting their confidence when donating data on a scale from 1 to 10 for each of the test conditions, participants also reported for which of the three conditions they felt the most confident when giving consent to donate data. From the results of this question, it is clear that the majority of the participants felt the most confident when donating data in the CONCRETE condition, with 35% preferring this version over the others (Figure 14). For the BASELINE condition 20% of participants reported being most confident, and another 20% reported that they preferred the ABSTRACT condition. Furthermore, 25% of the participants had no preference, and answered that they did not know for which of the test conditions they felt the most confident donating data to.

5.1.5 Willingness to donate again

Participants were also asked whether they would be willing to donate data to a research study again in the future, using a data donation website. The results show that 90% of the participants would be willing to donate data again in the future (*'Yes, I would definitely be willing to.'* and *'Yes, I would probably be willing to.'*). The remaining 10% of the participants answered that they would probably not be willing to donate data again in the future (*'No, I would probably not be willing to.'*). None of the participants selected the option that they would definitely not be willing to donate data again in the future (*'No, I would definitely not be willing to.'*). The results of users' willingness to donate data again in the future can be seen in Figure 15.

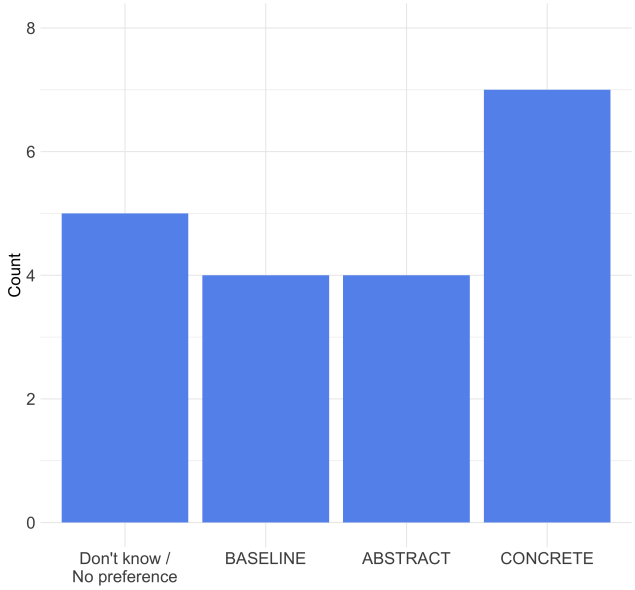

Figure 14: Reported preference for the different test conditions.

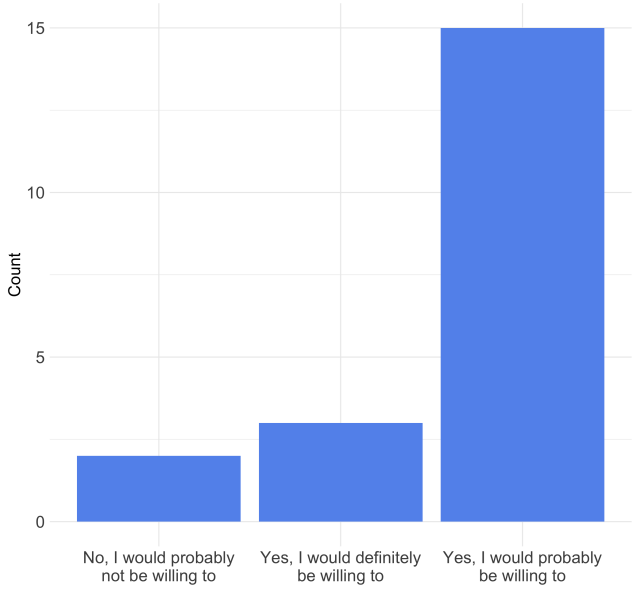

Figure 15: The reported willingness to donate data again in the future.

5.2 Qualitative analysis

In the following section the results from the qualitative responses of the research experiment will be presented. The qualitative analysis was performed using Thematic Analysis, an iterative coding process consisting of six steps: familiarize with the data, initial coding, searching for themes, reviewing themes, defining themes, and writing the results [6]. This method allows for flexibility while coding, and provides the researcher with a deeper understanding of the data through the iterative coding process.

First the qualitative responses to why participants gave consent to donate data will be presented, then the results of the version preference participants reported, and lastly the qualitative responses regarding users' willingness to donate data again in the future are shown.

5.2.1 Reasons to consent to data donation

After going through the data donation procedure, participants were asked to explain why they did or did not consent to donate their data through the data donation website. The 63 responses were coded and analyzed by following the Thematic Analysis steps explained in the introduction to the section. From these 63 responses, 14 initial codes were established. Some of the initial codes were: *'help research', 'interest in topic', 'why not', 'asked by study organizer', 'anonymous data', 'trust it is safe'* and *'usage was explained'*. These codes were then grouped into seven overarching themes, where similar codes were sorted together. The qualitative responses were then inspected again. In the end, five final themes were found: *'contribution to research'*, *'trust'*, *'anonymity'*, *'relationship to researcher'* and, *'compliance'*. The themes represent different explanations for why the participants gave consent to data donation. For all of the three test conditions, participants mentioned *contribution to research* as a reason as to why they gave consent to donate data.

'Because I know it's for a science project. And I think science is important.' (P5, ABSTRACT).

Multiple participants mentioned that *trust* and *anonymity* were reasons why they gave consent to data donation. Interestingly, participants mentioned the feedback table showing the transformed data in the fourth step of the data donation procedure, as being helpful to understand what data from the initial DDP would be shared with the research project.

'Trusting that the data will be handled somewhat anonymously and only the data shown to me in the last step is shared. I'm fine with that but it's based on trust.' (P20, CONCRETE).

'At first I was sceptical. I did not know what was in the file to be honest, and felt like I was giving away all my location data of the past X years. But then a small clear table was shown with the data that was needed (and thus handed in), which made it pretty clear they weren't 'ripping me off' of all my data. So in the end, I was pretty willing to consent. And I trust the fact that not all data was send to the research, just that small part they needed.' (P16, CONCRETE).

'[...] I looked at the data myself which was to be shared, and it's safe, not personal, just a few numbers.' (P23, CONCRETE).

'Compliance' was also uncovered as one of the themes in the qualitative responses. Participants mentioned that they wanted to give consent, that they did it for the purpose of the study, or that they saw no reason not to donate their data. Furthermore, the *'relation to researcher'* was another theme in the responses. Being asked by the study organizer or to help the researcher was mentioned by some participants.

5.2.2 Version preference

Participants who indicated in the post-experiment survey that they preferred a particular version of the video explaining the data anonymization, were asked to clarify why they preferred this version over the other two. As there were only 15 responses for all three conditions to analyze, the responses were not coded into themes. Instead the explanations for each test conditions are presented below.

For the BASELINE version participants reasoned that the simple explanation was clear and gave them a better overview of the information regarding the anonymization of the data:

'Simpler interface, makes it easier to feel confident. No twisted words, no tons of information to understand.' (P6, BASELINE)

One participant voiced a concern that providing the user with too much information could have an adverse effect and potentially make users more sceptical about the data donation procedure:

'Because it's the most simple version. When there is too much info, I think: should I be worried about it?' (P4, BASELINE)

On the other hand, for the ABSTRACT version participants felt that the amount of information used when communicating the data anonymization was on a appropriate level without becoming overbearing:

'It was not overloaded with information, yet ensuring and clear [...]' (P24, ABSTRACT)

'It answered the most of my questions and was not confusing.' (P1, ABSTRACT).

'It gives a bit more information, but not too much. I feel like 3 (CONCRETE) is just too much to read, and I'm not even sure if I read that in the first to be honest [...] and 1 (BASELINE) does not give any info [...]' (P16, ABSTRACT)

One participant who preferred the ABSTRACT version admitted that the confidence and trust when donating data was high regardless of the condition:

'The level of confidence was pretty much the same for all of them. I didn't even notice that the amount of text/detail increased [...] the second version seems like an appropriate amount of info for instructions. I'd be good with version 1 [...] since I know it's a thesis study (high initial trust level).' (P17, ABSTRACT)

Participants with a preference for the CONCRETE version preferred it due to providing them with more information and details about the data pipeline and the anonymization of the data:

'More written information on how the data is made anonymous' (P8, CONCRETE).

'Most explanations given, feels detailed and secure' (P23, CONCRETE).

5.2.3 Willingness to donate again

As stated in the quantitative analysis, most participants answered that they would be willing to donate data again in the future. Participants were then asked to explain why they would or would not be willing to donate data again in the future. In total 20 qualitative responses from participants were analyzed based on the principles of Thematic Analysis. Eight initial codes were defined: *'research is important', 'social importance', 'for science', 'interest in research', 'depends on research project', 'feels secure', 'good experience'* and *'convenience'*. After creating these codes the qualitative responses were again inspected and grouped into the codes. The responses and codes were then grouped into themes based on underlying themes. Finally, three themes were established: *'convenience'*, *'interest'* and *'privacy'*.

Participants mentioned that *'privacy'* and feeling secure about the data handling was important, as well as the *'convenience'* of donating data like this. P23 mentions that they would donate again in the future because *'it seems safe and easy'*. Other participants mentioned the data anonymization as a factor as well:

'If the research seems important I would donate if I feel secure that the data is handled safely and is anonymized.' (P8)

'As long as there is a good explanation of how it is anonymized (or at least a guarantee from a reputable party)' (P20)

The majority of the participants mentioned the importance of supporting science and research as the main factor for them being willing to donate data again in the future. Numerous of the participants emphasized that they would only be willing to donate data as long as they could be sure that it was for non-commercial use:

'If it is for science or it has a social importance (for example: data research about public health or local traffic) I don't think I have a problem with it. When it is for more commercial use, I would be more hesitated.' (P5)

'I think providing data for the sake of research is a good thing to do (If it doesn't lead to ad optimization [...])' (P17)

Two of the participants answered that they would probably not be willing. P18 stated that they would not be interested, while P21 answered *'I don't trust donating my data in any case.'*

6 Discussion

The goal of this research was to design and develop a data donation user interface, as well as to gain insights on how to communicate anonymization to regular users by measuring trust and willingness to donate data through such a system. Based on the proposed framework by Boeschoten et al. [5] to let participants donate partial DDPs and previous research on participants' willingness to donate data, we investigated how this applies in a practical research experiment where participants donate DDPs through a data donation website. The results will be discussed in relation to the research questions below and the related hypotheses:

> ***RQ1:*** *How can we foster trust in participants when designing a data donation user interface?*
>
> ***RQ2:*** *How do we communicate the anonymization of data to a regular user?*

In the following section the results of the quantitative and qualitative analysis will be discussed in regards to previous research , and the research questions. The limitations of the research project will also be examined, in addition to future work in relation to data donation within HCI.

6.1 Data donation systems

The work of this thesis project is to be regarded as exploratory research on how to implement the framework of digital trace data collection and use DDPs for research, as this has not been done before. Although no significant results were found for the three hypotheses that were tested, the findings of this research provide us with valuable insights on data donation systems, and on communication and transparency of a data pipeline within such a system. Overall there was a high trust in the data donation system and a high willingness among participants to consent to data donation, indicating that the framework and data donation interface can be implemented in future research projects using DDPs for data collection.

6.1.1 User interface design

The first research question is concerned with how to design a data donation user interface that fosters trust in participants, in an effort to maximize the consent level of participants donating data. As the trust in the data donation website was high, we can assume that users found the system to be both functional and reliable.

No significant effect was found for trust in a specific system for the three different test conditions, and therefore we cannot reject or accept the related hypothesis, **H1**. As there was overall high trust in the systems, this is partial proof that how the anonymization is communicated to users is not related to the level of trust users have in the interface. Previous research by Acquisti et al. [1] suggests that security is not a primary concern for users, and that until given a reason not to trust a system, users tend to not worry about privacy and security. It is possible that the data donation interface being novel to users and the limited interaction with the system affects the trust relationship positively. However, from the qualitative results it is clear that both anonymity and privacy are important for participants in relation to giving consent to donate data. In addition, privacy was a factor for participants' willingness to donate data again through a data donation website. Based on

these results, we propose that future data donation systems include some sort of explanation of the data pipeline, even though it is not clear whether an abstract or concrete explanation is most beneficial in regards to participants' trust.

The results indicate that most participants are willing to donate data again through a data donation website. This is in line with the previous research on participants' willingness to share data by social scientists Keusch et al. [22] and Struminskaya et al. [35]. The qualitative analysis revealed that contributing to research was important to a majority of the participants. Therefore, information about the institution should always be easily available to participants in a donation system. This is also supported by previous research from Maus et al. [26] and Struminskaya et al. [35]. As most participants stated an interest in research studies, providing information about the specific aims of the research study should also be included on the website. Preferably with links pointing to a university page with more information about the research study or the researchers. We suggest implementing this information in a sidebar element next to the main element displaying the donation procedure.

Based on the results of the statistical analysis we cannot reject or accept **H2**, but the results indicate a high willingness for participants to donate data. Applying construal level theory to communicate data anonymization does not appear to be a better option than presenting an overview of the steps in the data pipeline without further explanations, in relation to participants' willingness to donate data. Even though participants had the option to not give consent to donate data, all of the participants agreed to give consent. Research by Habib et al. [19] on privacy controls and opting out have shown that users tend to agree to default settings presented to them. Similarly, it could be that users agree to give consent in a data donation system when consenting is the more obvious action presented to them. On the other hand, it is likely that participants in this research study were already willing to donate data before starting the experiment, as they had already requested their DDP from Google and consented to take part in the experiment. Despite the fact that all of the participants reported that they gave consent, a data donation system should still include a button for participants to not agree to give consent as it can reveal potential biases in the research.

To sum up, the results indicate that the design of the data donation system is experienced as usable and trustworthy by participants. Future research planning to build a data donation website should communicate to the participants how the data is anonymized and processed. Furthermore, information about both the research institution and the research study should be available on the data donation website. Although all of the participants in this research experiment gave consent to donate data, implementing a button for users to not provide informed consent would be beneficial for research projects with more participants in an effort to uncover potential biases.

6.1.2 Communicating the data pipeline

How to explain the anonymization to the regular participant is the focus of the second research question. Similar to **H1** and **H2**, no significant effect was found for **H3**. It is therefore not possible to accept or reject the hypothesis that the degree of explaining data anonymization has an effect on users' confidence when donating data. However, there are some trends that can be noted from the results, but these results should be interpreted with caution due to limited sample size. It is possible that a larger sample size would have revealed significant results for an effect on users' confidence to donate data. The highest

mean score is measured for the ABSTRACT condition, indicating that explaining to users why the data is anonymized is appreciated more than focusing on how it is done, or not providing any explanation beyond the steps in the data pipeline. The BASELINE condition had more variety in the distribution of the data, than the ABSTRACT and CONCRETE conditions. Providing no further explanation of the anonymization technique in the data pipeline presented to the user appears to be more conflicting in terms of making users feel confident when donating data to scientific research.

In terms of preference, a majority of the participants opted for the CONCRETE version. That this version presents the user with the most detailed information is mentioned by nearly all of the participants who choose this version. The BASELINE and ABSTRACT versions were each preferred by the same percentage of participants as the ideal way to communicate data anonymization. It is also apparent that many participants do not have a preference between the versions, as this option is the second most selected option. The qualitative results show that participants who preferred the BASELINE condition felt that this version was ideal, as they became more sceptical when presented with a lot of information in the other conditions. On the other hand, participants who preferred the way the anonymization was explained in the ABSTRACT condition stated that the amount of information was clear without becoming confusing. We cannot draw any absolute conclusions from these results, but the overall results suggest that some form of communicating the data anonymization is appreciated by the participants.

It was noted by multiple participants that the table with the data feedback made them realize how the transformed data file was actually constructed. We therefore propose to include an example of the transformed data when communicating the anonymization to the user. Showing the result of analyzing their data could potentially also interest participants as an incentive to participate in the research study. Furthermore, it would be beneficial to explain the anonymization and processing of the data to the user at an earlier stage in the data donation procedure, such as in the infographic video explaining the anonymization. Presenting this as the first step in the data donation procedure and in the initial invitations to participate, might be helpful when recruiting participants to donate data in the first place.

For this particular use case where participants were asked to donate location data, it is somewhat logical how the data is processed and calculated from the original data. However, it is likely that for processing more complex DDPs, for example sentiment analysis of text or analysis of facial expressions in images, the actual processing of the data would be less obvious to the regular user. For research projects asking participants to donate these data types, communicating the anonymization is might be even more essential for participants' willingness and trust, but further research is needed.

6.2 Limitations

The following limitations have possibly impacted the result of this research project. First, due to time constraints the sample size used in the research experiment is limited, and therefore results cannot be generalized for the population. Future work should replicate this study with a larger sample size.

Second, participants were recruited through the means of convenience sampling as this allowed for easy recruitment via online communities and on campus recruitment. Due to this, it is possible that participants knowing the researcher might have had an impact on the trust they had in the research and their willingness to donate data. Future research

looking to replicate this study should keep this limitation in mind, and look into recruiting participants via other means of sampling.

Third, there is a non-response bias as individuals concerned with privacy are likely to be unwilling to donate data in any case and not participate in the research. Furthermore, these individuals might not even have a Google account due to their privacy concerns. Thus, they would already be excluded by not meeting the recruitment criterion and not have any data to donate. Although future research should keep this limitation in mind, it is difficult to avoid this bias for research concerned with data donation.

Finally, this thesis project is an exploratory research project where participants were led to believe they donated data through the data donation system. However, participants did not in fact donate any actual data to a research study as there was no need to collect this data for the research project. Moreover, there was limited time and resources available to build a website and system where the uploaded DDPs were processed. Future work should try to replicate this study and build a data donation system with a working back-end to process the participants' data.

6.3 Future work

Future research should replicate the study with the previous limitations in mind, with a larger sample size and a varied distribution of participants. Future research should also implement the anonymization script to process the data and let participants donate data through the data donation interface. In addition, it should be an option for participants to sign up for upcoming research before a research study is starting up, to allow participants who do not have data to donate at that time to still be able to generate and donate data when data donation is open.

This thesis project is an exploratory research project where quantitative data was gathered through questionnaires filled out by participants after interacting with the data donation system. In future research with more resources, it would be interesting to build a dynamic website for the data donation website instead of a static website. By doing so, it gives researchers the possibility to log users and collect various quantitative data using more objective measurements than questionnaires. Based on the results of this thesis, it would be interesting to measure the amount of time a user spends on the website for each of the various test conditions. This would make it possible to investigate if the time users spend interacting with the system changes accordingly with the length of the infographic videos explaining the data anonymization. Furthermore, a dynamic website would make it possible to log visitor data and clicking behaviour, to uncover the percentage of participants actually consented to donate data. Lastly, it could be interesting to use A/B testing to investigate at what it is most beneficial to explain the data anonymization to participants.

Future work should also expand on this exploratory research by investigating other aspects of the data donation framework. Outside of the participants' scope researched in this thesis, the interface could be expanded to allow researchers to upload research projects and anonymization scripts for DDPs. In doing so, the website can become a data donation platform where researchers can set up a research study using DDPs more easily. By doing so, the framework of using digital trace data collection and the data donation platform would encourage open data and sharing of anonymization scripts among researchers.

7 Conclusion

This exploratory research presented the impact of explaining data anonymization to users, on users' trust and consent when donating partial data download packages (DDP) through a data donation website. Based on requirements from previous research, stakeholder interviews and design principles, a high fidelity prototype for a data donation user interface was created. Then, a working prototype for a data donation website was developed using React and the user interface library MUI. Based on construal level theory by Trope and Liberman [37] three different test conditions were implemented on the data donation website, differing on how the data processing and anonymization of the DDPs was communicated and explained to the participants.

While the quantitative results of the research were not significant, the exploratory research indicates that previous research on users' willingness to donate data, also can be transferred to a practical research study where participants share data through a data donation website. Participants generally have a high trust in the system regardless of how the data anonymization is explained to the users. Anonymization and privacy were revealed as important for participants' willingness and confidence to donate data, meaning that some form of explaining the anonymization to the participants should be included in a data donation system. Further research is needed to explore how best to implement this in terms of abstraction or concretion of information. In addition to explaining the data anonymization, a data donation system should present the participants with a visual representation of what information is included in the transformed data file that is to be donated. A table showing the participants' processed data should be displayed before participants decide to give consent to donate their data. Furthermore, showing a table with an example of how the processed data will look along with an explanation of the anonymization would probably be beneficial for participants' understanding of the data handling. With this in mind, our findings can be useful for researchers when designing data donation systems, but also for executing research projects using DDPs.

Due to the limited sample size in this research, future research should replicate the research study with a larger sample in an effort to find significant results. Furthermore, future research should set up the data donation system as a dynamic website to collect useful insights from quantitative data on participants' interaction with the system, such as time spent on the website, reading time or clicking behaviour.

Bibliography

[1] Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., & Wilson, S. (2017). Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys*, *50*(3), 1–41. https://doi.org/10.1145/3054926

[2] Arbuckle, L. E., & Emam, K. E. (2020). *Building an anonymization pipeline: Creating safe data* (1st ed.). O'Reilly Media.

[3] Bannihatti Kumar, V., Iyengar, R., Nisal, N., Feng, Y., Habib, H., Story, P., Cherivirala, S., Hagan, M., Cranor, L., Wilson, S., Schaub, F., & Sadeh, N. (2020). Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text. *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, 1943–1954. https://doi.org/10.1145/3366423.3380262

[4] Bietz, M., Patrick, K., & Bloss, C. (2019). Data Donation as a Model for Citizen Science Health Research. *Citizen Science: Theory and Practice*, *4*(1), 1–11. https://doi.org/10.5334/cstp.178

[5] Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). A framework for digital trace data collection through data donation, 1–26. https://arxiv.org/pdf/2011.09851

[6] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[7] Chicaiza, J., Cabrera-Loayza, M. C., Elizalde, R., & Piedra, N. (2020). Application of data anonymization in Learning Analytics. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3378184.3378229

[8] Cranor, L. F. (2012). Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal on Telecommunications and High Technology Law*, *10*, 273–307. https://ieeexplore-ieee-org.proxy.library.uu.nl/document/7927931

[9] Dahlinger, A., Wortmann, F., Ryder, B., & Gahr, B. (2018). The impact of abstract vs. concrete feedback design on behavior - Insights from a large eco-driving field experiment. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*, 1–11. https://doi.org/10.1145/3173574.3173953

[10] Demir, L., Kumar, A., Cunche, M., Lauradoux, C., Demir, L., Kumar, A., Cunche, M., Lauradoux, C., & Pitfalls, T. (2019). The Pitfalls of Hashing for Privacy To cite this version : HAL Id : hal-01589210.

[11] Diethei, D., Niess, J., Stellmacher, C., Stefanidi, E., & Schöning, J. (2021). Sharing Heartbeats: Motivations of Citizen Scientists in Times of Crises. *2020*(April), 1–15. https://doi.org/10.1145/3411764.3445665

[12] Döring, T., Sylvester, A., & Schmidt, A. (2013). A design space for ephemeral user interfaces. *TEI 2013 - Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, 75–82. https://doi.org/10.1145/2460625.2460637

[13] Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E., Kyba, C. C. M., Bowser, A., Cooper, C. B., Sforzi, A., Metcalfe, A. N., Harris, E. S., Thiel, M., Haklay, M., Ponciano, L., Roche, J., Ceccaroni, L., Shilling, F. M., Dörler, D., … Jiang, Q. (2017). Citizen Science Terminology Matters: Exploring Key Terms. *Citizen Science: Theory and Practice*, *2*(1), 1. https://doi.org/10.5334/cstp.96

[14] European Parliament, C. o. t. E. U. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Pro-

tection Regulation). *Official Journal, L 119/1*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679#d1e2513-1-1

[15] Fessenden/, T. (2021, April 11). *Design systems 101*. Retrieved December 1, 2021, from https://www.nngroup.com/articles/design-systems-101/

[16] *Getting started – react*. (n.d.). Retrieved December 2, 2021, from https://reactjs.org/docs/getting-started.html

[17] Google. (n.d.[a]). *Google Takeout*. Retrieved September 1, 2021, from https://takeout.google.com

[18] Google. (n.d.[b]). *Manage your Location History*. Retrieved November 25, 2021, from https://support.google.com/accounts/answer/3118687?hl=en

[19] Habib, H., Pearman, S., Wang, J., Zou, Y., Acquisti, A., Cranor, L. F., Sadeh, N., & Schaub, F. (2020). "It's a scavenger hunt": Usability of Websites' Opt-Out and Data Deletion Choices. *Conference on Human Factors in Computing Systems - Proceedings*, 1–12. https://doi.org/10.1145/3313831.3376511

[20] Habib, H., Zou, Y., Yao, Y., Acquisti, A., Cranor, L., Reidenberg, J., Sadeh, N., & Schaub, F. (2021). Toggles, dollar signs, and triangles: How to (in)efectively convey privacy choices with icons and link texts. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–25. https://doi.org/10.1145/3411764.3445387

[21] Janic, M., Wijbenga, J. P., & Veugen, T. (2013). Transparency enhancing tools (TETs): An overview. *Workshop on Socio-Technical Aspects in Security and Trust, STAST*, 18–25. https://doi.org/10.1109/STAST.2013.11

[22] Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to Participate in Passive Mobile Data Collection. *Public Opinion Quarterly*, *83*, 210–235. https://doi.org/10.1093/poq/nfz007

[23] Lankton, N. K., Harrison Mcknight, D., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

[24] MacKenzie, I. S. (2013). *Human-computer interaction: An empirical research perspective*. Morgan Kaufmann.

[25] *Material design*. (n.d.). Retrieved December 2, 2021, from https://material.io

[26] Maus, B., Salvi, D., & Olsson, C. M. (2020). Enhancing citizens trust in technologies for data donation in clinical research: Validation of a design prototype. *10th International Conference on the Internet of Things Companion*. https://doi.org/10.1145/3423423.3423430

[27] Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2). https://doi.org/10.1145/1985347.1985353

[28] *Mui*. (n.d.). Retrieved December 2, 2021, from https://mui.com

[29] Nielsen, J. (2013, June 24). *Website reading: It (sometimes) does happen*. Retrieved November 27, 2021, from https://www.nngroup.com/articles/website-reading/

[30] Prainsack, B. (2019). *Data Donation: How to Resist the iLeviathan*. https://doi.org/10.1007/978-3-030-04363-6_2

[31] Preece, J. (2016). Citizen Science: New Research Challenges for Human–Computer Interaction. *International Journal of Human-Computer Interaction*, *32*(8), 585–612. https://doi.org/10.1080/10447318.2016.1194153

[32] Preece, J., Rogers, Y., & Sharp, H. (2019). *Interaction design: Beyond human-computer interaction* (5th ed.). John Wiley & Sons.

[33] Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

[34] Spotify. (n.d.). *Data rights and privacy settings*. Retrieved September 1, 2021, from https://support.spotify.com/us/article/data-rights-and-privacy-settings/

[35] Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (2020). Understanding Willingness to Share Smartphone-Sensor Data. *Public Opinion Quarterly*. https://doi.org/10.1093/poq/nfaa044

[36] Thornsby, J. (2020, November 18). *How to download your personal data on google, apple, facebook and twitter*. Retrieved November 22, 2021, from https://www.maketecheasier.com/how-to-download-your-personal-data-on-google-apple-facebook-and-twitter/

[37] Trope, Y., & Liberman, N. (2010). Construal-Level Theory of Psychological Distance. *Psychol Rev*, *3*(2), 3–23. https://doi.org/10.1037/a0018963.Construal-Level

[38] Wiggins, A., & Wilbanks, J. (2019). The Rise of Citizen Science in Health and Biomedical Research. *American Journal of Bioethics*, *19*(8), 3–14. https://doi.org/10.1080/15265161.2019.1619859

A Anonymization scripts

In this appendix the two versions of the anonymization scripts are included. First the Python code for the initial first draft of the script is shown, then the Python code for the final anonymization script is presented.

A.1 First draft anonymization script

```
# From https://medium.com/@ggonzalezzabala/graph-your-own-google-location-history-in-tableau-e362d1d8f18d
# github: https://github.com/gabrielgz92/location_history_data/blob/master/key_value_parsing.py

import pandas as pd
import numpy as np
import datetime as dt
import json
import csv
import re

## User input of home address
address_input = input("Please enter your home address:\n")
print(f'You entered home address: "{address_input}"')

#Creates list from placeVisit data
def placeVisit(placeVisit_dict):
    address = placeVisit_dict["location"]["address"].replace("\n",", ")
    start_time = placeVisit_dict["duration"]["startTimestampMs"]
    end_time = placeVisit_dict["duration"]["endTimestampMs"]

#Formatting variables
start_time_dt = timeStampToDate(int(start_time))
end_time_dt = timeStampToDate(int(end_time))
home_address = bool(re.match(address_input, address))
place_visit = [address, start_time, end_time, start_time_dt, end_time_dt, home_address]
return place_visit

#Convert milliseconds timestamp into a readable date
def timeStampToDate(milliseconds):
    date = dt.datetime.fromtimestamp(milliseconds/1000.0)
    date = date.strftime('%Y-%m-%d %H:%M:%S')
    return date

#Method to run all the scripts
def parse_data(data):
    for data_unit in data["timelineObjects"]:
      if "placeVisit" in data_unit.keys():
        write_places_csv(placeVisit(data_unit["placeVisit"]))

#Write to CSV
def write_places_csv(place_data_list):
  with open('FULL_places.csv', 'a', newline='') as file:
```

```
    writer = csv.writer(file, delimiter=',')
    writer.writerow(place_data_list)
print("\nList exported to csv file")

files = ["2019/2019_JANUARY.json", "2020/2020_JANUARY.json"]

for file in files:
  with open(f"SemanticLocationHistory/{file}") as f:
    data = json.load(f)
  parse_data(data)

full_places = pd.read_csv("FULL_places.csv", header=None,
    names=['address', 'start_time', 'end_time', 'start_time_dt', 'end_time_dt', 'home_address'])

# #### Duration
# Change to datetime to calculate duration
full_places["start_time_dt"] = list(map(
        lambda x: dt.datetime.strptime(x, '%Y-%m-%d %H:%M:%S'), full_places["start_time_dt"]))
full_places["end_time_dt"] = list(map(
        lambda x: dt.datetime.strptime(x, '%Y-%m-%d %H:%M:%S'), full_places["end_time_dt"]))

#duration in days and hours
full_places["duration"] = full_places["end_time_dt"] - full_places["start_time_dt"]
#duration in total seconds
full_places["duration_sec"] = list(map(lambda x: x.total_seconds()/60, full_places["duration"]))

# #### Home address and year
## Select rows where home address is true
df_home = full_places.loc[full_places['home_address'] == True]

## Add column with year for extraction
df_home['year'] = df_home['start_time_dt'].dt.year
## Calculate length of month
start_month = pd.Timestamp('20200101000000')
end_month = pd.Timestamp('20200131235959')
jan_len = end_month - start_month

## Calculate the time duration at home in months for 2019 and 2020
month2019 = df_home.loc[df_home['year'] == 2019]['duration'].sum()
month2019_per = (month2019/jan_len*100)

month2020 = df_home.loc[df_home['year'] == 2020]['duration'].sum()
month2020_per = (month2020/jan_len*100)

# Contingency table to export
hometimes = pd.DataFrame([[month2019_per, month2020_per]], columns=['Month 2019', 'Month 2020'],
                         index=['Percentage at home'])
print(hometimes)
```

A.2 Final anonymization script

```
import json
import itertools
import re
import zipfile
import pandas as pd

# years and months to extract data for
YEARS = [2019, 2020, 2021]
MONTHS = ["JANUARY"]
NPLACES = 3
TEXT = "This study examines the change in travel behaviour during \
the COVID-19 pandemic. We therefore examined your Google semantic \
Location History data for January in 2019, 2020, and 2021. To be \
precise, we extracted per month the total number of visited places, \
and the number of days spend per place for the three most visited \
places. Also, we extracted the number of days spend in places and \
travelling, and the travelled distance in km."


def __visit_duration(data):
    """Get duration per visited place
    Args:
        data (dict): Google Semantic Location History data
    Returns:
        dict: duration per visited place, sorted in descending order
    """
    placevisit_duration = []
    for data_unit in data["timelineObjects"]:
        if "placeVisit" in data_unit:
            address = data_unit["placeVisit"]["location"]["placeId"]
            start_time = data_unit["placeVisit"]["duration"]["startTimestampMs"]
            end_time = data_unit["placeVisit"]["duration"]["endTimestampMs"]
            placevisit_duration.append(
                {address: (int(end_time) - int(start_time))/(1e3*24*60*60)})

    # list of places visited
    address_list = {next(iter(duration)) for duration in placevisit_duration}

    # dict of time spend per place
    places = {}
    for address in address_list:
        places[address] = round(sum(
            [duration[address] for duration in placevisit_duration
                if address == list(duration.keys())[0]]), 3)
    # Sort places to amount of time spend
    places = dict(sorted(places.items(), key=lambda kv: kv[1], reverse=True))

    return places

```

```
50
51  def __activity_duration(data):
52      """Get total duration of activities
53      Args:
54          data (dict): Google Semantic Location History data
55      Returns:
56          float: duration of activities in days
57      """
58      activity_duration = 0.0
59      for data_unit in data["timelineObjects"]:
60          if "activitySegment" in data_unit.keys():
61              start_time = data_unit["activitySegment"]["duration"]["startTimestampMs"]
62              end_time = data_unit["activitySegment"]["duration"]["endTimestampMs"]
63              activity_duration += (int(end_time) - int(start_time))/(1e3*24*60*60)
64      return activity_duration
65
66
67  def __activity_distance(data):
68      """Get total distance of activities
69      Args:
70          data (dict): Google Semantic Location History data
71      Returns:
72          float: duration of activities in days
73      """
74      activity_distance = 0.0
75      for data_unit in data["timelineObjects"]:
76          if "activitySegment" in data_unit.keys():
77              activity_distance += int(data_unit["activitySegment"]["distance"])/1000.0
78
79      return activity_distance
80
81
82  def process(file_data):
83      """Return relevant data from zipfile for years and months
84      Args:
85          file_data: zip file or object
86      Returns:
87          dict: dict with summary and DataFrame with extracted data
88      """
89      results = []
90      filenames = []
91
92      # Extract info from selected years and months
93      with zipfile.ZipFile(file_data) as zfile:
94          file_list = zfile.namelist()
95          for year in YEARS:
96              for month in MONTHS:
97                  for name in file_list:
98                      monthfile = f"{year}_{month}.json"
99                      if re.search(monthfile, name) is not None:
```

```
100                         filenames.append(monthfile)
101                         data = json.loads(zfile.read(name).decode("utf8"))
102                         places = __visit_duration(data)
103                         results.append({
104                             "Year": year,
105                             "Month": month,
106                             "Top Places": dict(itertools.islice(places.items(), NPLACES)),
107                             "Number of Places": len(places),
108                             "Places Duration [days]": round(
109                                 sum(value for value in places.values()), 3),
110                             "Activity Duration [days]": round(__activity_duration(data), 3),
111                             "Activity Distance [km]": round(__activity_distance(data), 3)
112                         })
113                         break
114
115     # Put results in DataFrame
116     data_frame = pd.json_normalize(results)
117
118     # Anonymize by replace PlaceIds with numbers
119     number = 0
120     for column in data_frame.columns:
121         if column.split(".")[0] == "Top Places":
122             number += 1
123             data_frame.rename(columns={column: f"Place {number} [days]"}, inplace=True)
124
125     return {
126         "summary": TEXT,
127         "data_frames": [
128             data_frame.fillna(0)
129         ]
130     }
```

B Digital artefacts

In this appendix sketches and prototypes from the design process of the data donation user interface are included. In addition links to Figma projects of the various prototypes are included, as well as links to the web application. The GitHub repositories with code for the data analysis and the code for the web application, including the infographic videos are also included.

B.1 Design process

B.1.1 Sketches

In this section various sketches from the beginning of the design process are included.

Basic UI ideas To start the design process various sketches were made for the initial design of the user interface. Crazy Eights design exercise was also performed during this stage.

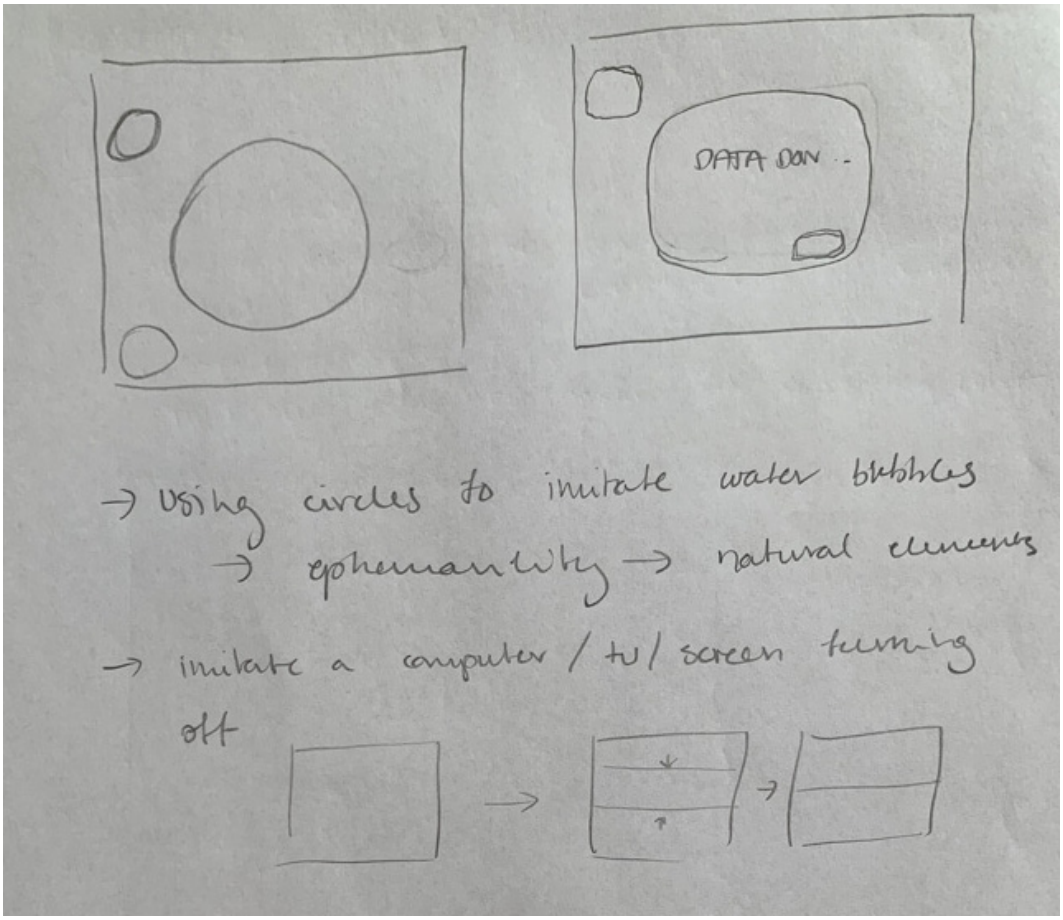

Figure 16: Sketches for ephemeral design elements.

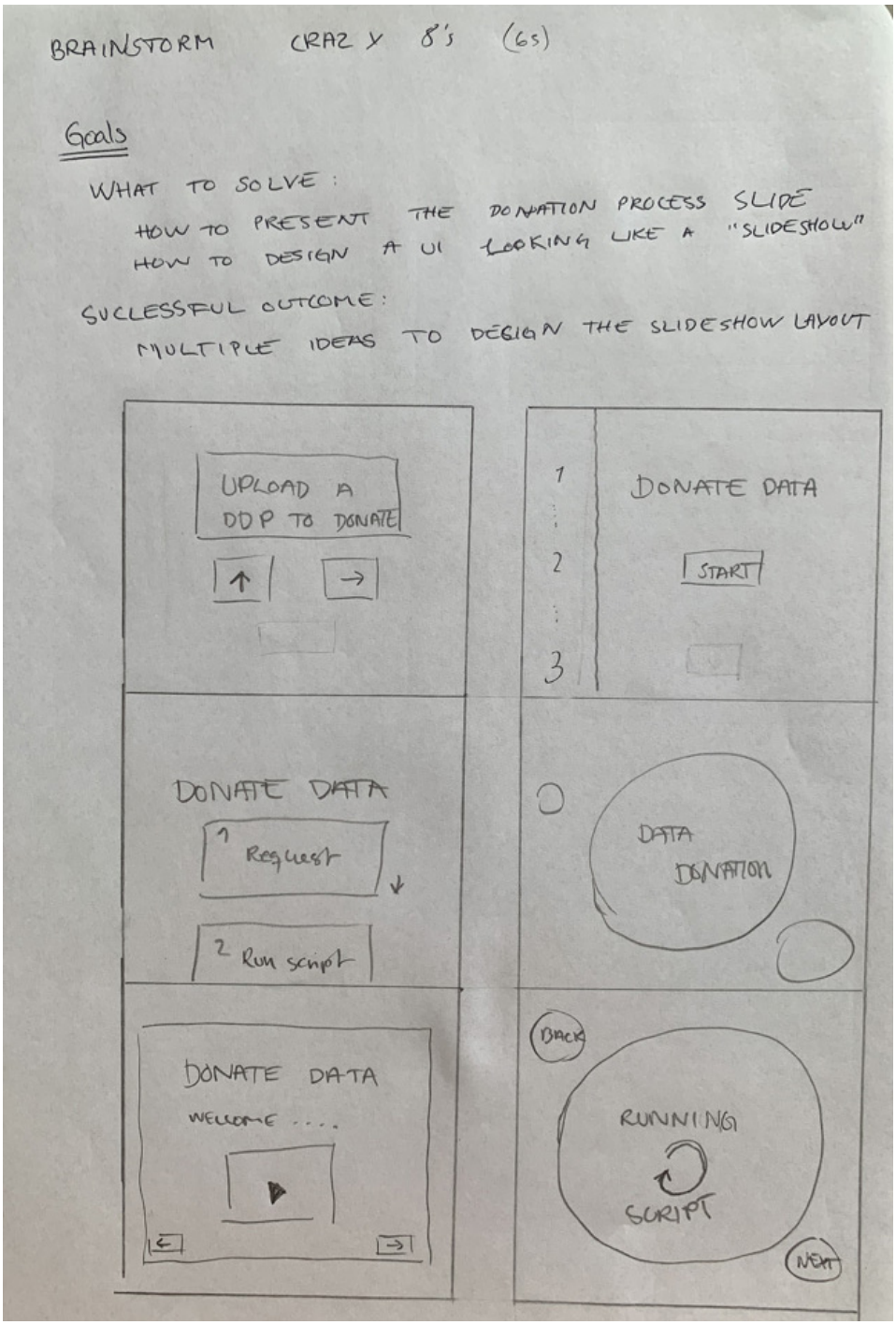

Figure 17: Some of the sketches from the Crazy Eights exercise.

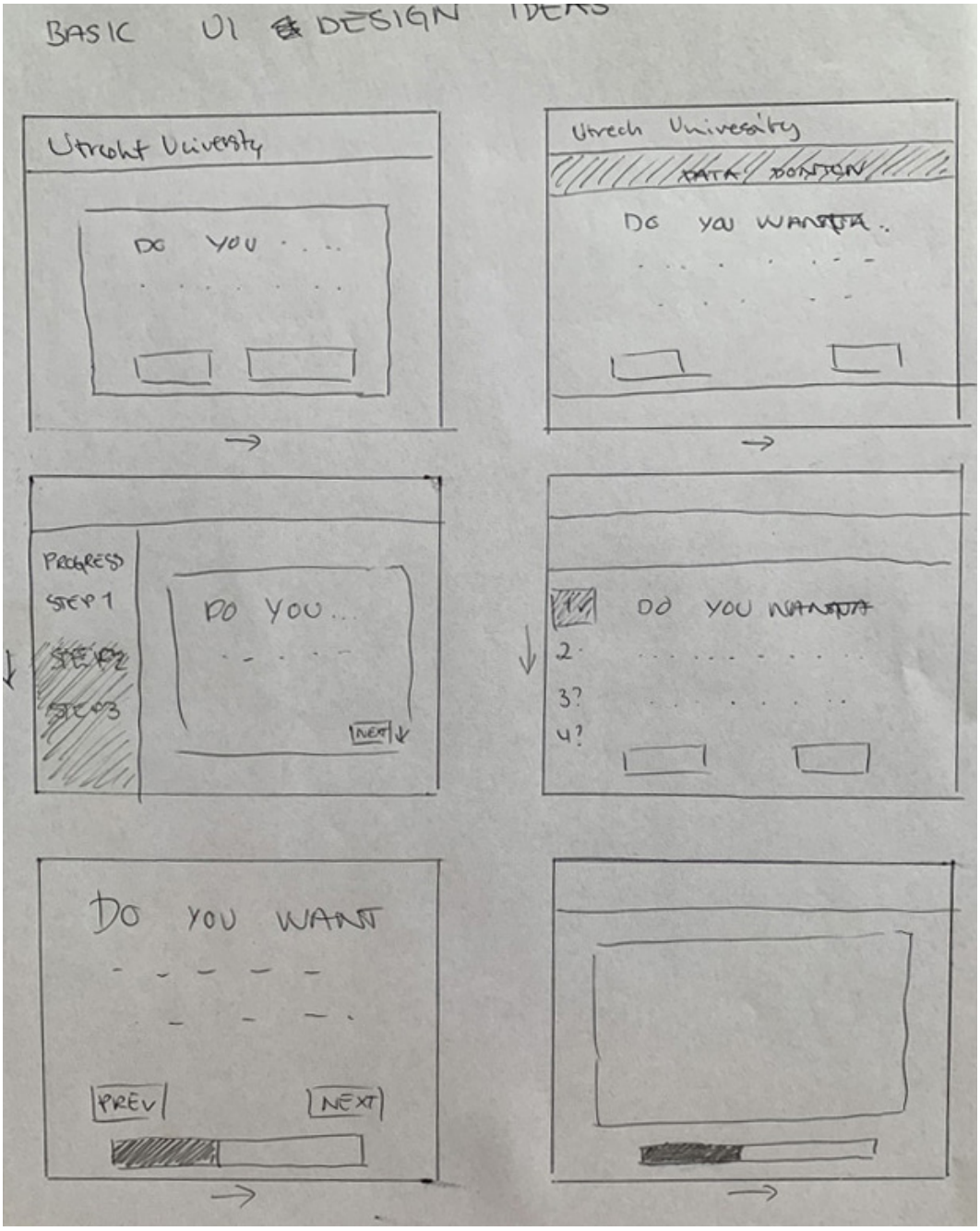

Figure 18: Sketches of various ideas for the UI design.

Sketch Sketches showing the two different concept ideas for a UI, one version with a slideshow UI, and the second version displaying a progress stepper with expanding and collapsing sections.

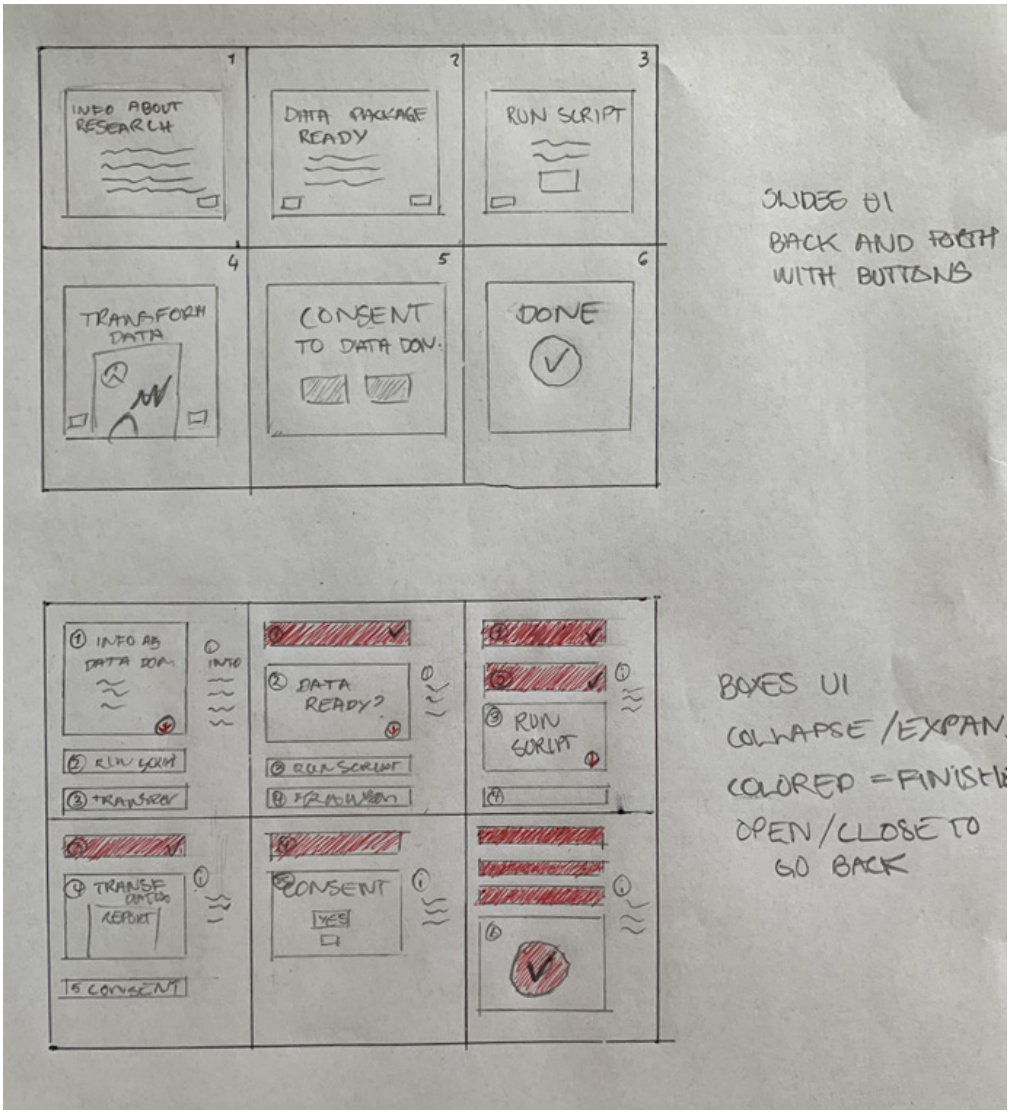

Figure 19: Sketches of two types of UI design.

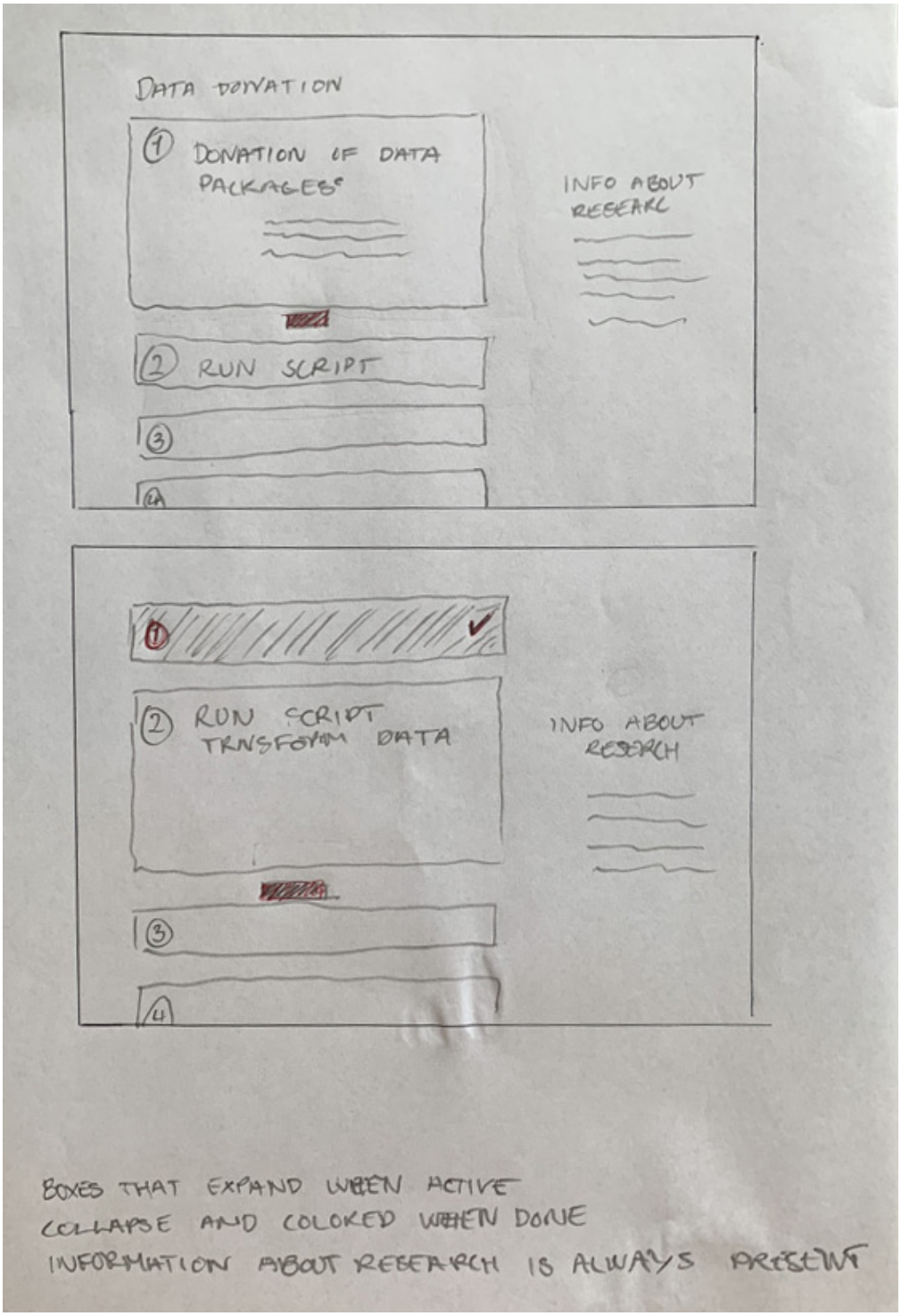

Figure 20: Sketches of the idea for the final UI design.

B.1.2 Wireframes

Initial version Two wireframe designs were created using Figma. The frames and the interaction between the steps in the first version is shown below. Link to interactive presentation: https://www.figma.com/file/y5CnooIfS5dgJZMf8vaqFL/Data-donation-UI-wireframes-v1?node-id=8%3A4

Figure 21: Version 1 of wireframe design of the data donation UI.

Final version The final wireframe is shown below, which is the basis for the continued design iterations. Link to final version: https://www.figma.com/file/KZUTltk93Iom7fQihgh2qA/Data-donation-UI-wireframes-v2?node-id=8%3A4

Figure 22: Wireframe design of the data donation UI.

B.1.3 High fidelity prototype

Figma prototype Below are screenshots of all the frames of the high fidelity prototype of the data donation system, including the dashboard landing page, and the frames of the five steps in the data donation procedure.

Link to the high fidelity prototype created in Figma: https://www.figma.com/file/MKkcf4dcCE3EaqE0JCWdvY/Data-donation-UI-mockup-version-1?node-id=1%3A3

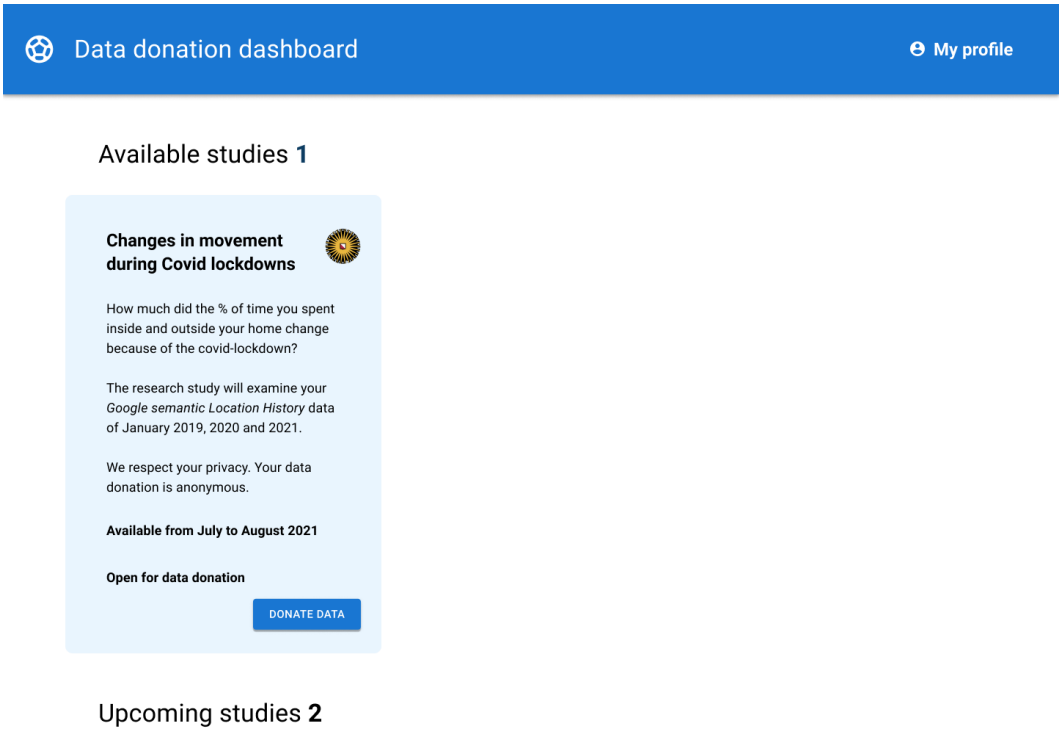

Figure 23: Dashboard homepage in the high fidelity prototype.

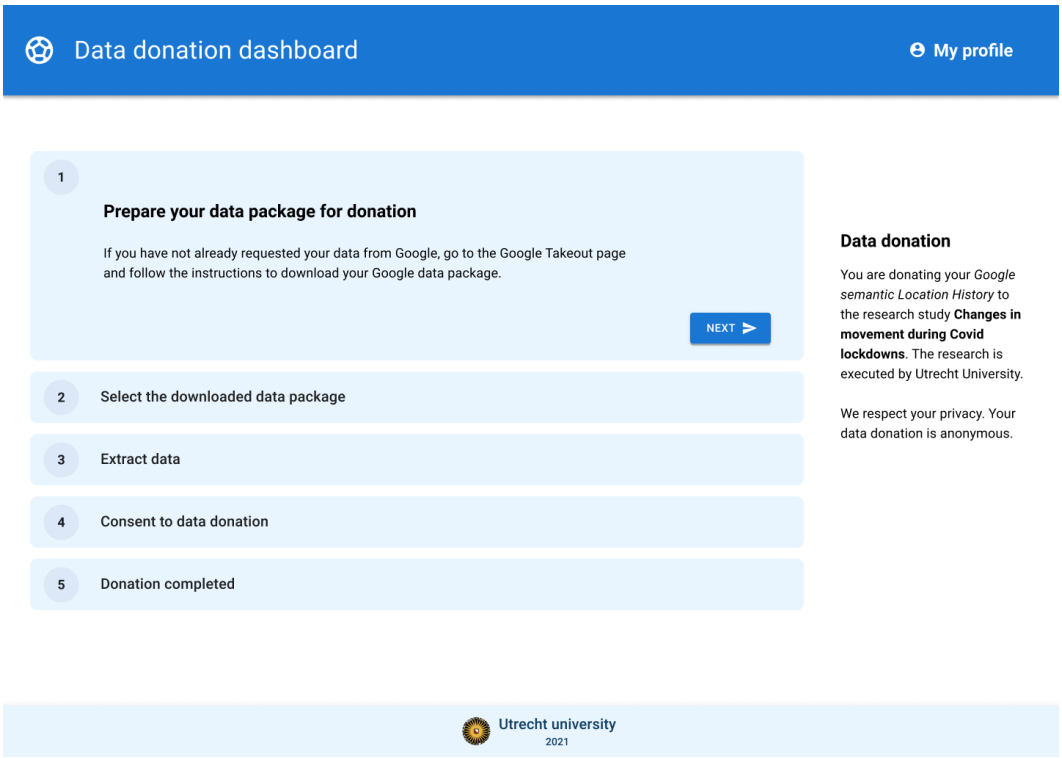

Figure 24: Step 1 of the data donation procedure in the high fidelity prototype.

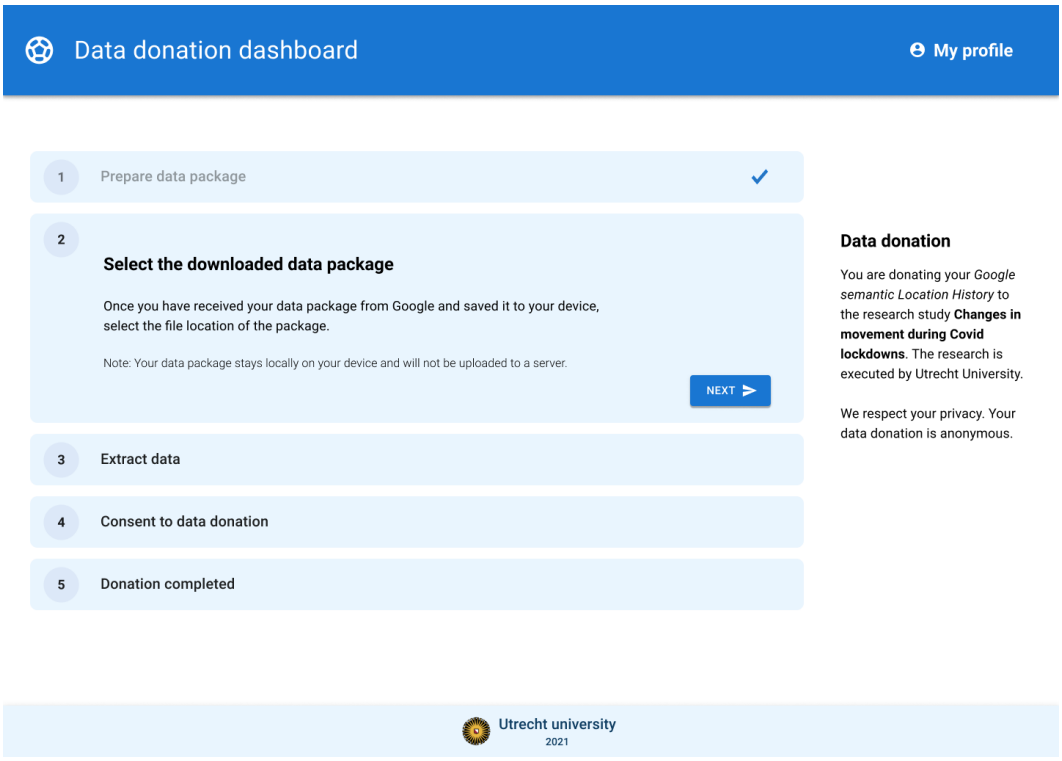

Figure 25: Step 2 of the data donation procedure in the high fidelity prototype.

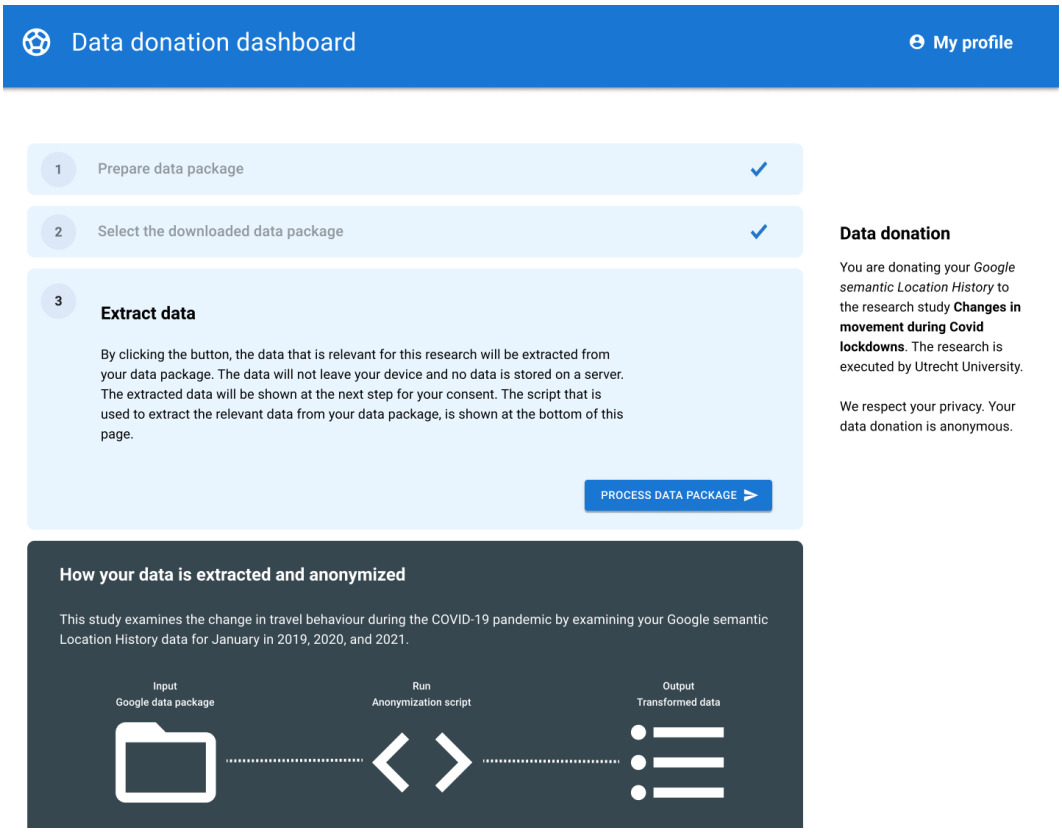

Figure 26: Step 3 of the data donation procedure in the high fidelity prototype.

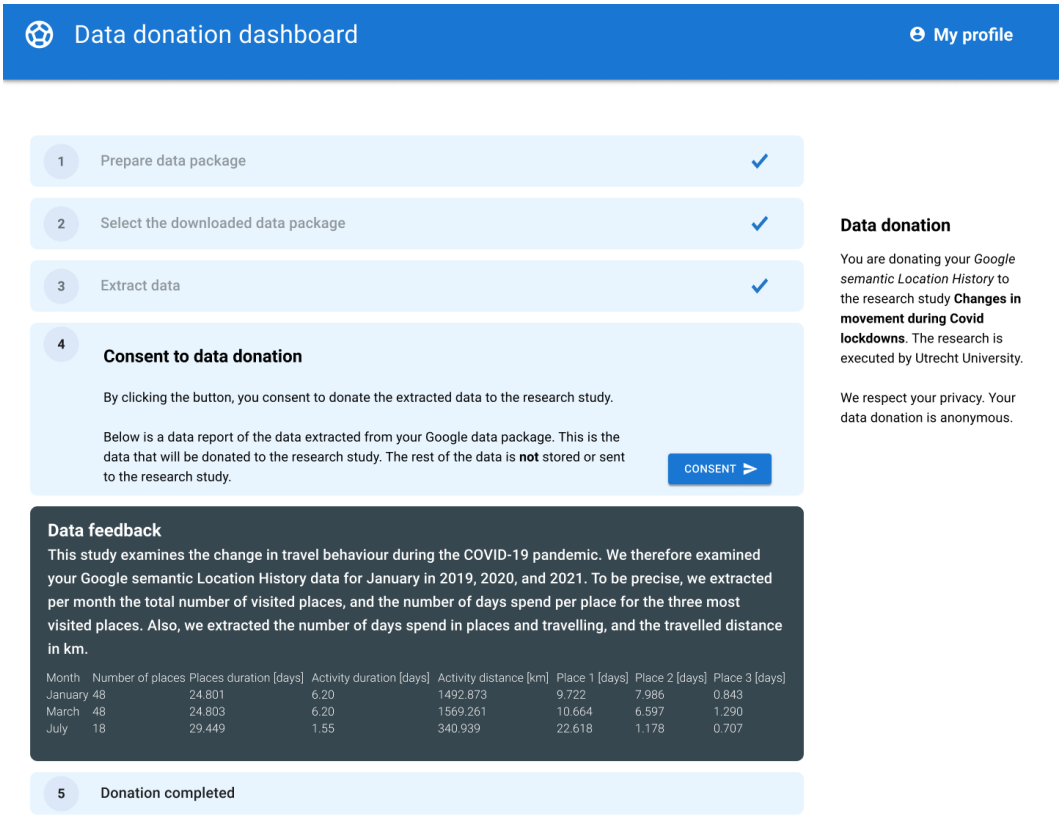

Figure 27: Step 4 of the data donation procedure in the high fidelity prototype.

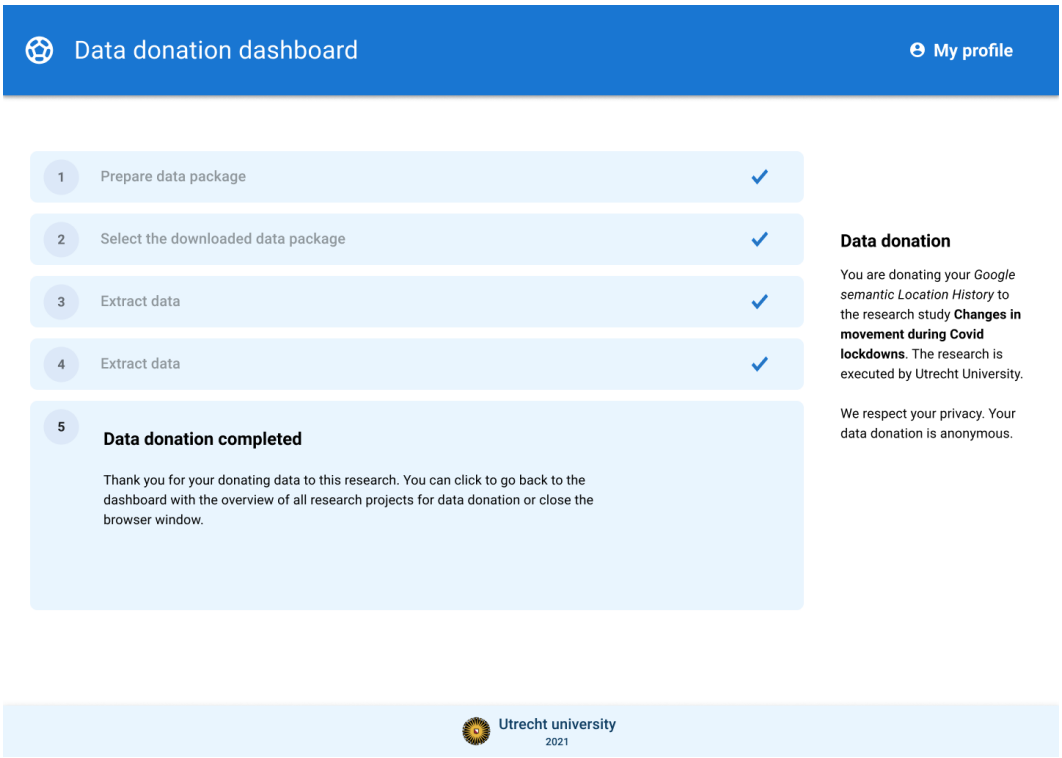

Figure 28: Step 5 of the data donation procedure in the high fidelity prototype.

B.2 Web application

Links to the web application showing the data donation interface are included in the list below. Note that at the time of the research experiment the web application was hosted on the researcher's personal domain https://katrinekuiper.com instead of a github project page as shown in the links.

There are also dashboards and donation versions for the two following rounds of each test condition (just replace A1, B1, C1 with A2, B2, C2, A3, B3 or C3), but due to the length of the list these are not included. These various versions were only made for the counterbalancing of the experiment, and to have separate links for each round. Besides this, the second and third round versions are identical for each test condition.

Dashboard homepage (not active) https://knkuiper.github.io/datadonation/#/home

Dashboard A1 https://knkuiper.github.io/datadonation/#/dashboarda1

Donation A1 https://knkuiper.github.io/datadonation/#/donationa1

Dashboard B1 https://knkuiper.github.io/datadonation/#/dashboardb1

Donation B1 https://knkuiper.github.io/datadonation/#/donationb1

Dashboard C1 https://knkuiper.github.io/datadonation/#/dashboardc1

Donation C1 https://knkuiper.github.io/datadonation/#/donationc1

GitHub repository All of the code for the web application can be found at the GitHub repository *datadonation*: https://github.com/knkuiper/datadonation.

B.3 Infographics videos

Link to folder in GitHub repository with the infographic videos for each condition: https://github.com/knkuiper/datadonation/tree/master/src/assets/video. Alpha is the video for the BASELINE condition, Beta for the ABSTRACT condition and Gamma for the CONCRETE condition.

B.4 Data analysis

Link to GitHub repository with the experiment data and code for the statistical analysis: https://github.com/knkuiper/Data-donation-experiment-analysis.

C Questionnaires

The following section includes the questionnaire setup for the various surveys used in this research. First the pre-experiment survey is presented, which was used to recruit participants. Then the survey for the experiment rounds and post-experiment survey are presented. Finally, any correspondence with participants is included.

C.1 Pre-experiment survey

The pre-experiment survey contains the informed consent, survey and experiment instructions. The survey consists of 12 questions, including questions about demographics, participants' propensity to trust and trust in technology, as well as asking participants about their contact information.

C.1.1 Consent

Q1.1 Welcome to this research study on data donation.

We are interested in understanding the use of data donation in scientific research. For this study, you will be asked to donate a data package to research through a data donation website.

Specifically, you are asked to donate Google location to a research study. You will be asked to donate data three times over the course of a few weeks. Afterwards, you will be asked to answer some questions about it. Your data and responses will be kept completely confidential.

To participate in the research study, you must have a Google account, and location settings turned on, on your devices. To donate your data, you will first have to request your Google data from Google Takeout. When the data export is finished you will be sent an email on how to download the .zip-file to your device (note it may take a few days before the export is done). We also ask you to please use a laptop or computer when participating in this research study.

Next we ask you to fill out a short questionnaire. We ask you to please provide your email address so we can send you the instructions on how to participate and complete the data donation procedure in the upcoming weeks.

Your participation in this research is voluntary. You have the right to withdraw at any point during the study. If you have any questions about the study, please contact Katrine Nordeide Kuiper at k.n.kuiper@uu.nl.

By clicking the button below, you acknowledge:

Your participation in the study is voluntary. You are 18 years of age or above. You are aware that you may choose to terminate your participation at any time for any reason.

- I consent, begin the study.
- I do not consent, I do not wish to participate.

C.1.2 Information

Q2.1 Your participant ID is $e://Field/ParticipantID.

C.1.3 Demographics

Q3.1 What is your gender?

- Woman
- Man
- Non-binary
- Prefer not to disclose
- Prefer to self-describe *(with text entry box.)*

This question allows for multiple answers according to HCI Guidelines for Gender Equity and Inclusivity. https:// www.morgan-klaus.com/ gender-guidelines.html

Q3.2 What is your age?
(*Text entry box allowing numbers from 18 to 99.*)

Q3.3 What is your highest level of finished education?

- Less than high school
- High school
- Bachelor's degree
- Master's degree
- PhD

C.1.4 Propensity to trust - Faith in General Technology

Measures of propensity to trust, *Faith in General Technology* from Mcknight et al. [27].

Q4.1 Please answer the following statements.
(*Statements Q4.2–Q4.5 are answered on a 7-point Likert scale.*)

- Strongly disagree
- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

Q4.2 I believe that most technologies are effective at what they are designed to do.

Q4.3 A large majority of technologies are excellent.

Q4.4 Most technologies have the features needed for their domain.

Q4.5 I think most technologies enable me to do what I need to do.

C.1.5 Propensity to trust - Trusting Stance - General Technology

Measures of propensity to trust, *Trusting Stance - General Technology* from Mcknight et al. [27].

Q5.1 Please answer the following statements.
(*Statements Q5.2–Q5.4 are answered on a 7-point Likert scale.*)

- Strongly disagree
- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

Q5.2 My typical approach is to trust new technologies until they prove to me that I shouldn't trust them.

Q5.3 I usually trust a technology until it gives me a reason not to trust it.

Q5.4 I generally give a technology the benefit of the doubt when I first use it.

C.1.6 Contact information

6.1 Please fill in your email address in order to continue participation in this research study. *(With text entry box.)*

This information will only be used to send you email reminders about the next parts of the research study. Once the research study is completed your email information will be deleted.

C.1.7 Google takeout information

7.1 To donate data in this research study you have to request your location history data from Google. The video shows you how to request your data.

Please go to Google Takeout tool to request your data.

Video showing how to go to Google takeout and request the location data.

C.1.8 End of survey message

We thank you for your time spent taking this survey. Your response has been recorded.

If you did consent to participate in the research study you will receive an email in the next few days about the first data donation procedure.

If you have any questions about the research please contact Katrine Nordeide Kuiper at k.n.kuiper@uu.nl.

C.2 Experiment data donation rounds surveys

The questionnaire for the experiment rounds include 10 questions. Participants answer whether or not they consented to donate data through the data donation website, their willingness to on a scale 1–9, and their trust in the data donation system.

C.2.1 Participant ID

Q1.1 We will now ask you a few questions about the data donation procedure. This will only take a few minutes.

Please fill in your participant-ID provided in the mail. (*With text entry box.*)

C.2.2 Data donation

Q2.1 Did you donate data through the data donation website?

- Yes
- No

Q2.2 How willingly did you donate your data?
(*Slider going from 1 to 9.*)

Q2.3 Why did you donate data?
(*Text entry box. Displayed if the participant answered 'Yes' to Q2.1.*)

Q2.4 Why did you not donate data?
(*Text entry box. Displayed if the participant answered 'No' to Q2.1.*)

C.2.3 Trust in technology measures

Questions adapted from Mcknight et al. [27], *Trusting Belief - Specific Technology*: *Reliability* (Q3.2–Q3.4) and *Functionality* (Q3.5–Q3.7).

Q3.1 Please answer the following statements.
(*Statements Q3.2–Q3.6 are answered on a 7-point Likert scale.*)

- Strongly disagree
- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

Q3.2 The data donation interface is a very reliable piece of software.

Q3.3 The data donation interface does not fail me.

Q3.4 The data donation interface does not malfunction for me.

Q3.5 The data donation interface has the functionality I need.

Q3.6 The data donation interface has the features required for my tasks.

Q3.7 The data donation interface has the ability to do what I want it to do.

C.2.4 End of survey message

We thank you for your time spent taking this survey. Your response has been recorded.

If you have any questions about the research please contact Katrine Nordeide Kuiper at k.n.kuiper@uu.nl.

C.3 Post-experiment survey

The following questions are added to the final round of data donation, to the experiment survey. The participants are asked to answer 12 questions in the post-experiment questionnaire. Participants are asked about their confidence when donating data in each test condition, version preference and about willingness to donate again in the future. Questions Q5.2–Q5.8 include screenshots giving an overview of the infographic videos to reference the different test versions.

C.3.1 Post survey information

Q4.1 Before you are completely done, we ask you to fill out some additional information that might be of interest to us.

C.3.2 User interface version confidence

Q4.2 Please answer the following questions on a scale 1–10.

Q4.3 How confident did you feel about giving consent to donate data for version 1?
Slider going from 1 to 10.

Q4.4 How confident did you feel about giving consent to donate data for version 2?
Slider going from 1 to 10.

Q4.5 How confident did you feel about giving consent to donate data for version 3?
Slider going from 1 to 10.

Q4.6 Which version made you feel the most confident about giving consent to donate data to the research study?

- Version 1
- Version 2
- Version 3
- I don't know

Q4.7 Why did you feel the most confident giving consent to donate data for version 1?
Text entry box. Displayed if the participant answered 'Version 1' to Q4.6.

Q4.8 Why did you feel the most confident giving consent to donate data for version 2?
Text entry box. Displayed if the participant answered 'Version 2' to Q4.6.

Q4.9 Why did you feel the most confident giving consent to donate data for version 3?
Text entry box. Displayed if the participant answered 'Version 3' to Q4.6.

C.3.3 Willingness to donate data

Q6.1 Would you be willing to donate data to a research study through a data donation website in the future?

- Yes, I would definitely be willing to
- Yes, I would probably be willing to
- No, I would probably not be willing to
- No, I would definitely not be willing to

Q6.2 Why?
(*Text entry box. Displayed if the participant answered either 'Yes, I would definitely be willing to' or 'Yes, I would probably be willing to.' to Q6.1.*)

Q6.3 Why not?
(*Text entry box. Displayed if the participant answered either 'No, I would probably not be willing to' or 'No, I would definitely not be willing to' to Q6.1.*)

C.3.4 End of survey message

We thank you for your time spent taking this survey. Your response has been recorded.

We would like to inform you that we have not actually collected and saved any of the data from the data donation procedure. We are interested in researching trust and willingness in regards to data donation in research, and have only stored the answers from the survey questionnaires.

If you have any questions about the research please contact Katrine Nordeide Kuiper at k.n.kuiper@uu.nl.

C.4 Correspondence with participants

In this section the various types of correspondence with participants are listed, as well as the content of these messages, such as invite and reminder emails for the experiment rounds and the thank-you email sent to participants after the research experiment was finished.

C.4.1 Pre-survey recruitment on social media

Participate in master's research on data donation!

Hi!

For my master's in Human Computer Interaction, I'm researching the use of data donation in research, specifically the potential of using data from f.ex. social media in scientific research. I would appreciate it a lot if you would help me out and participate in my research experiment.

You will be asked to donate a Google data file. To participate you need to have a Google Account and your location settings turned on. The experiment runs over two weeks and consists of three rounds of data donation. Each round will only take around 10-15 minutes to complete.

If you are interested, please go to the survey linked below.

Thanks a lot!

C.4.2 Invite emails for the experiment rounds

Dear participant,

Thank you for participating in this research study on data donation. This is a reminder to take part in the *first/second/final* round of data donation.

Please go to the data donation website (or copy and paste this URL in your webbrowser: https://katrinekuiper.com/ *(with link to the dashboard for the specific round and test condition)* and choose the available study *'Changes in movement during COVID-19 lockdowns'*. Follow the instructions there to donate your data. It is important that you perform the experiment using a webbrowser on a **laptop** or **desktop**.

After completing the data donation procedure, please fill in the survey. You will be asked to fill in your participant-ID.

Your participant-id is: $m://FirstName

In case you accidentally close the window after the data donation procedure is completed, but before finishing the survey the link is provided below.

$l://SurveyURL

If you still need to request your data from Google Takeout please follow the instructions in the video below. You can either request all of your data from Google, or select only 'Location History'('Locatiegeschiedenis', 'Posisjonslogg').

(Video showing the Google takeout tool instructions.)

In a few days you will receive an email to complete the *second/third* part of the research study. If you have any questions please contact k.n.kuiper@uu.nl.

Thank you for taking part in this research study, your contribution is appreciated a lot!

Kind regards,
Katrine Nordeide Kuiper (k.n.kuiper@uu.nl)
HCI master's student at Utrecht University.

Follow the link to opt out of future emails:
$l://OptOutLink?d=Click here to unsubscribe

C.4.3 First reminder emails for the experiment rounds

Dear participant,

This is a reminder to take part in the *first/second/final* round of data donation. If you have already completed this round of data donation, please disregard this reminder mail.

Please go to the data donation website (or copy and paste this URL in your webbrowser: https://katrinekuiper.com/ *(with link to the dashboard for the specific round and test condition)* and choose the available study *'Changes in movement during COVID-19 lockdowns'*. Follow the instructions there to donate your data. It is important that you perform the experiment using a webbrowser on a **laptop** or **desktop**.

After completing the data donation procedure, please fill in the survey. You will be asked to fill in your participant-ID.

Your participant-id is: $m://FirstName

In case you accidentally close the window after the data donation procedure is completed, but before finishing the survey the link is provided below.

$l://SurveyURL

If you still need to request your data from Google please go to Google takeout tool. You can either request all of your data from Google, or select only 'Location History' ('Locatiegeschiedenis', 'Posisjonslogg').

In a few days you will receive an email to complete the *second/third* part of the research study. If you have any questions please contact k.n.kuiper@uu.nl.

Thank you for taking part in this research study, your contribution is appreciated a lot!

Kind regards,
Katrine Nordeide Kuiper (k.n.kuiper@uu.nl)
HCI master's student at Utrecht University.

Follow the link to opt out of future emails:
$l://OptOutLink?d=Click here to unsubscribe

C.4.4 Final reminder emails for the experiment rounds

Dear participant,

This is a final reminder to take part in the *first/second/final* round of data donation. As the next round will start soon, this *first/second/final* round will close tomorrow evening at 23:59. If you have already completed the *first/second/final* round of data donation, please disregard this reminder mail.

Please go to the data donation website (or copy and paste this URL in your webbrowser: https://katrinekuiper.com/ *(with link to the dashboard for the specific round and test condition)* and choose the available study *'Changes in movement during COVID-19 lockdowns'*. Follow the instructions there to donate your data. It is important that you perform the experiment using a webbrowser on a **laptop** or **desktop**.

After completing the data donation procedure, please fill in the survey. You will be asked to fill in your participant-ID.

Your participant-id is: $m://FirstName

In case you accidentally close the window after the data donation procedure is completed, but before finishing the survey the link is provided below.

$l://SurveyURL

If you still need to request your data from Google please go to Google takeout tool. You can either request all of your data from Google, or select only 'Location History' ('Locatiegeschiedenis', 'Posisjonslogg').

In a few days you will receive an email to complete the *second/third* part of the research study. If you have any questions please contact k.n.kuiper@uu.nl.

Thank you for taking part in this research study, your contribution is appreciated a lot!

Kind regards,
Katrine Nordeide Kuiper (k.n.kuiper@uu.nl)
HCI master's student at Utrecht University.

Follow the link to opt out of future emails:
$l://OptOutLink?d=Click here to unsubscribe

C.4.5 Thank-you emails

Dear participant,

Thank you for taking part in this research study on the use of data donation in scientific research. Your contribution is much appreciated!

We would like to inform you that **no data** from the Google data file donated through the data donation website was actually collected or saved. We are interested in researching trust and willingness in regards to data donation in research, and have only stored the answers from the survey questionnaires.

If you would like to know more about the research please contact Katrine Nordeide Kuiper at k.n.kuiper@uu.nl.

We will delete your contact information and you will no longer receive any emails from us.

Kind regards,
Katrine Nordeide Kuiper (k.n.kuiper@uu.nl),
HCI master's student at Utrecht University.