# Visualizing Exports of Personal Data by Exercising the Right of Data Portability in the Data Track - Are People Ready for This?

Farzaneh Karegar, Tobias Pulls, Simone Fischer-Hübner

# Visualizing Exports of Personal Data by Exercising the Right of Data Portability in the Data Track - Are People Ready for This?

Farzaneh Karegar, Tobias Pulls, and Simone Fischer-Hübner

Department of Mathematics and Computer Science, Karlstad University (KaU)
{farzaneh.karegar,tobias.pulls,simone.fischer-huebner}@kau.se

**Abstract.** A transparency enhancing tool called Data Track has been developed at Karlstad University. The latest stand-alone version of the tool allows users to visualize their data exports. For analysing the users' perceptions of the Data Track in regard to transparency features and the concepts of data export and data portability, we have conducted a qualitative user study. We observed that although users had rather little interest in the visualization of derived data activities revealed in the Google location file, they were interested in other kinds of derived data like usage patterns for different service providers. Also, as earlier user studies revealed, we again confirmed that it is confusing for users to differentiate between locally and remotely stored and controlled data. Finally, in spite of being concerned about the security of the data exported to their machines, for exercising data portability rights pursuant to the General Data Protection Regulation, most participants would prefer to first export and edit the data before uploading it to another service provider and would appreciate using a tool such as the Data Track for helping them in this context.

**Keywords:** Transparency Enhancing Tools, Data Portability, Visualisation, Data Track.

## 1 Introduction

Transparency of personal data processing is an important principle for the privacy of individuals as well as for a democratic society [9]. People rarely have a clear understanding about how their personal data are collected, used, shared or accessed [1]. Consequently, transparency of personal data processing is enforced by most Western privacy laws, including the EU Data Protection Directive (DPD) 95/46/EC [6] and new General Data Protection Regulation (GDPR) [7] which will replace the DPD in 2018. The GDPR grants enhanced data subject rights for transparency and intervenability, such as the right of access by the data subject including the right to receive a data copy of her personal data undergoing processing in a commonly used electronic format (Art. 15), the right

to rectification and erasure (Art. 16, 17), and the right to data portability (Art. 20). The right to data portability is aiming at increasing user choices of online services and allows users to request all their data from a data controller that in turn has to provide the users with the data in a structured, commonly-used machine readable format which can then be transmitted to any other controllers. Alternatively, the users can also request to transmit their data directly from a service provider to another one, if technically feasible. One way to exert these rights pursuant to GDPR is using technologies which enhance transparency and provide user control. These kinds of technologies are commonly referred to as Transparency Enhancing Tools (TETs) [12].

The Data Track (DT) developed at Karlstad University (KaU) is an example of a TET that shows users what data they have disclosed to which service providers under what agreed-upon policies and how their data have been processed. The Data Track development started as a part of the European PRIME[1] and PrimeLife [2] projects and continued as part of the A4Cloud project[3]. This paper reports about a user study on the perception of a new function for visualising exports of personal big data to the data subjects, which we added recently to the Data Track tool. Already today, many service providers, such as Google and Facebook, provide users with data export functions for downloading their personal data. The newly added functionality to the Data Track for visualizing personal data exports from a service provider, which is also available in the form of a stand-alone open source Data Track version[4], can for instance provide users with an overview of the location data they (or more precisely their devices) have disclosed to Google by first exporting their data from *myaccount.google.com* as a file and then importing the data to the Data Track for visualisation. At least when the GDPR will apply in May 2018, users could export their personal data from all types of service providers (beyond those providing export functions already today) by exercising their right to receive an electronic copy (Art. 15) or their data portability right (Art. 20) and could import them to the Data Track for visualising their disclosed data to different services providers.

In this paper, we present a qualitative user study which has the objective to analyse the users' perceptions of the stand-alone Data Track in regard to the transparency features that it is providing and in regard to the concept of data export from a service provider to the Data Track running at the user's machine by exercising the rights to access and of data portability.

More precisely, we have been addressing the following two research questions and related sub-questions:

1. **What are the users' perceptions of transparency with the stand-alone Data Track?:** Does the interface convey that Google has more information about the users other than what they have sent explicitly or implicitly? What kind of transparency options are the users interested in and

[1] EU FP6 project PRIME, http://www.prime-project.eu
[2] EU FP7 project PrimeLife, http://primelife.ercim.eu/
[3] EU FP7 project A4Cloud, http://www.a4cloud.eu
[4] https://github.com/pylls/datatrack

would they like the Data Track to provide more transparency information related to their data? Do users have any concerns in regard to using the Data Track?

2. **What are users' perceptions of data export and portability with the stand-alone Data Tack?:** Do users understand and value the idea and the concept of exporting data from a service provider (Google in this case) and importing it to a tool running on their own machines or to another service provider? Consequently, do users understand the differences between locally stored (and thus user controlled) and remotely stored data? (i.e. data stored on their computers in the Data Track under their control after being exported from a service provider vs. data stored at the service's side)?

In the remainder of this paper, Section 2 briefly presents background and related work in regard to TETs and related user studies, Section 3 explains the methods used in our work and the test plan. Section 4 is devoted to analyzing the results. Finally, Section 5 discusses our conclusion and future work.

## 2 Background and Related Work

In this section, we first explain the different kinds of TETs. Then we elaborate more on Data Track versions and finally we describe the related user studies.

### 2.1 Transparency Enhancing Tools

There is a variety of TETs that have been developed and evaluated with different types of user tests in the past. TETs can in general be divided into ex-ante TETs—which enable the anticipation of consequences before data are actually disclosed (e.g., with the help of privacy policy statements)—and ex-post TETs which inform about consequences if data already have been revealed (cf. [12]).

TETs can be further categorised, in dependence on where the transparency information is stored and controlled, into services side TETs, user side TETs and Third Party TETs. Services side TETs run at the service provider's side and allow authenticated users to receive information about collected, processed or forwarded data at those sides. Examples of services side TETs are the Google Dashboard[5] or PrivacyInsight [4]. A Third Party TET requires the user to entrust a third party with the user's personal data for providing transparency services. An example for a Third Party TET is the DataBait tool by the EU project USEMP [20], which derives guesses and predictions about the user's personality by analyzing the user's social media and browser data with machine learning software. User side (or user controlled) TETs store the user's personal information to be made transparent locally on the user's device under the user's control. While user side TETs require the user's device to keep the data safe and may be more demanding to set up and get running from a usability perspective,

[5] Google dashboard. https://www.google.com/settings/dashboard.

they are in principle the more privacy-friendly solution, as the users retain control over their data. Examples of user side TETs are Mozilla's Lightbeam [15] or personal data vaults, such as Mun's et al. work [16], and the different versions of the Data Track that are briefly presented in the next section.

### 2.2 Data Track Versions

The first version of the Data Track was developed within the PRIME project [18] and included a "history function" for each transaction, in which the user's disclosed personal data to a service, a record describing to whom the personal data was disclosed (i.e. the identity of the controller), for which purposes and, more precisely, under which agreed-upon privacy policy, as well as a unique transaction pseudonym are stored in a secure manner. It was later complemented in the PrimeLife project with online access functions, which allow users (authenticated as data subjects of data held by a service provider via those transaction pseudonyms) exercising data subject rights to access, correct, rectify or erase those data at the service provider's side [10].

For the next Data Track versions developed in the A4Cloud project [9], [1], we have mainly improved the user interfaces (UIs) and interaction concepts, replacing the tabular presentations of the PrimeLife Data Track with the graphical UI illustrations, as previous research studies suggest that network-like visualizations provide a simple way to understand the meaning behind some types of data [2], [11]. Therefore, for the A4Cloud Data Track (also called "GenomSynlig"), we developed the so-called "trace view" (see Figure 1), presenting an overview of the data items sent to service providers, as well as the data items service providers received about the user. In addition to this "local view" of the trace view, which is graphically displaying the information that is stored locally in the Data Track about what data has been disclosed to whom, a user can also execute online access functions for exercising her data subject rights by clicking on the cloud icon next to the service provider's logo thereby switching to a "remote view" of what (disclosed or derived) data about the user are stored at the service provider side. In addition, an alternative timeline view has been developed for the A4Cloud Data Track, which lists the information about data disclosures in the Data Track records in chronological order for selected time intervals.

The latest version of the Data Track which is, as mentioned earlier, an open source and stand-alone program developed at the end of the A4Cloud project at KaU, is subject of this paper. It provides users with the visualization of data exported from the Google managing archive service. For our first version, we focused on the Google location history to be included in our archive. After successfully exporting the location data from Google and importing it to the Data Track, in addition to the trace and the timeline views, participants have a newly developed map view that allows to visualize location, activity and movement patterns as described in the location history provided by Google (see Figure 2). Activities are data derived by Google based on the locations reported by their devices (i.e., activities are derived by Google and not by the user's device).

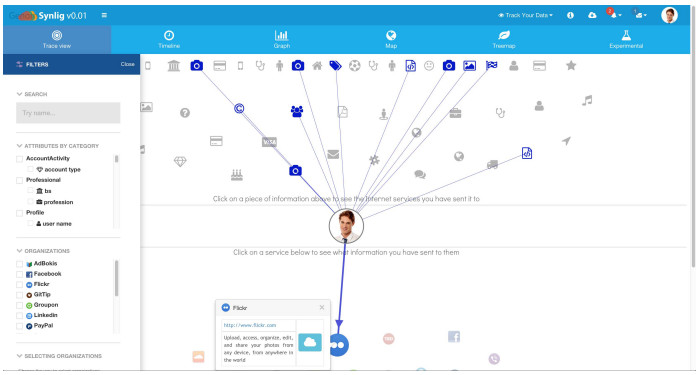

**Fig. 1.** Trace view of the A4Cloud Data Track

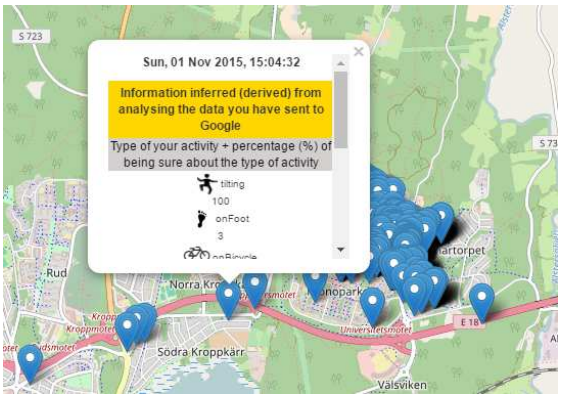


**Fig. 2.** Location data and activities on the map view of the stand-alone version of the Data Track

### 2.3 Related User Studies

User studies have been conducted for various ex-post TETs, which are mainly demonstrating the usability of graphical network-like presentations to illustrate data flows. For instance, Bier et al. present recent usability studies of the user interfaces of the PrivacyInsight tool in comparison to GenomSynlig and their different network-like data flow representations [4]. Moreover, Kane-Zabihi et al. present usability tests of an "interactive social translucence map" [13]. Other user evaluations studies or TETs using both network-like presentations and chronological presentations of data disclosure events comprise user tests of previous Data Track versions (with its trace and timeline views) and a user test by Kolter et al. of a tool for visualising transaction logs [14].

Prototypes of the trace and timeline views of the A4Cloud Data Track have been evaluated with usability tests and two focus group workshops. These user evaluations revealed that while test participants mostly valued the transparency

functionality of the Data Track and could successfully use it for tracking data disclosures, many test users had however problems to understand whether data records were stored in the Data Track client on the users' side (under the users' control) or on the remote service provider's side [8][3]. The specific feature of the Data Track trace view that allows to easily switch from a local view of Data Track records to a remote view providing users with online access to the data stored at the service provider's side, might have contributed to this confusion. As previous A4Cloud Data Track user tests and usability tests conducted in the PRIME project [19] showed the user's confusion of discerning between the locally and remotely data control and access, our user study also analysed the user's understanding of locally and user-controlled data (at the user's side) vs. remotely stored and controlled data (at the service provider's side) for the stand-alone Data Track with its new map view.

In contrast to those previous user studies of the Data Track, this paper evaluated the new stand-alone version of the Data Track with its newly added map view. This paper presents the first evaluation of the perception of a transparency tool based on exports of personal data from a service provider (Google in this case) and of its perceived value when exercising the right to receive an electronic copy of the personal data or the data portability right pursuant to the upcoming GDPR.

## 3 User Study and Methods

We conducted a user study with ten participants with the objective to receive insights on the users' perceptions of transparency functions and of data export and portability with the latest stand-alone version of the Data Track. Our study is primarily a qualitative study using semi-structured interviews, which are allowing to follow and explore new directions as they come up in the interview process, and a grounded theory based approach [5] to surface key themes that arise in our interviews.

Before the user study, we conducted an incremental and iterative pilot study with 16 participants. The reason to conduct the pilot user study was twofold: 1) To test, fine-tune the task and adopt the timing. 2) To tailor and manipulate the questions we ask during the study for better answering our main research questions.

Based on the feedback that we received during the pilot study, we manipulated our interview questions on the grounds that some of them were not suitable enough to answer our two main research questions. Ultimately, we also concluded to guide users during the interview through how they can download the location file from Google. In the pilot test, the participants were supposed to follow the instructions or watch a video on the Data Track to detect where on the Google they can order to export their location data. The results showed that it was really time-consuming and sometimes irritating for the users and due to the fact that we do not intend to test the usability and clarity of Google settings, we

decided not to time the users but to guide them. In the following, we present the recruitment, study procedure, and demographic information of the participants.

### 3.1 Recruitment of Participants

We strived to get an unbiased sample of participants by recruiting arbitrary people in Karlstad city center (P1-P4, P10), via a Facebook group related to Karlstad (P5, P9) and participants of an innovation seminar in Örebro (P6-P8). Those who accepted our invitation were compensated with a 100 SEK gift card. All interviews were conducted in English in October 2016.

### 3.2 Study Procedure

To begin with, a study plan was written to serve as the main communication vehicle as well as a blueprint for the study. A study plan is a summary of all the containing documents needed for the user studies [21]. To avoid an active researcher (study moderator/interviewer) bias which includes mannerisms and statements made by the researcher that provide the participants with information about the researcher's preferences [17], the procedure was standardized. The leading questions were avoided in the interviews and before conducting the user study and in the recruitment advertisements, we told participants that Data Track was implemented at KaU and that we were just responsible for conducting the user study of it.

During the study, each participant received the same instructions and followed the same blueprint. The study took 30-60 minutes based on how much each individual participant wanted to communicate and consisted of four parts: 1) a welcome session in which we thanked them, briefly talked about what they were expected to do and we obtained informed consent from all of participants. The informed consent imparted that participants agreed to have their screen and audio recorded, alongside with their answers. Consent to the recording was not required, though all participants agreed to be recorded, 2) a pre-task questionnaire for collecting demographics, 3) a role-playing task with a fake Google account to download the location data from Google, upload the same file to the Data Track and view the location data in the map view, and 4) the semi-structured interview during which participants answered to the questions while they were still allowed to use and navigate through the tool. Two researchers, one as an interviewer (moderator) and one as a note keeper, participated in the studies.

Participants' own Google accounts were not used in the study. Instead, they were given the role of a persona to play to visualize their data. Using a persona, participants feel secure that they are not compromising their personal details when taking part in the study. Moreover, it allows full control of what each participant encounters, avouching a standard experience that can be compared between participants. The persona details in each case included a username and password of a Google account.

#### 3.2.1 Task

After filling in the pre-task questionnaire, the participants had to conduct a task in which they export (download) the location data from Google, import (upload) it to the Data Track and view the visualization of the data and its characteristics in the map view.

Focusing on the users' perceptions of different concepts in the Data Track, the goal of defining the task was not to measure the participants' efficiency in finding how they should download the data. However, we aimed to have the location data imported to the Data Track as the starting point of discussing the interview questions, providing all participants with a common ground for enabling them to have a better insight into what downloading data from a service provider and uploading the same data to another party mean.

#### 3.2.2 Semi-structured Interview

For learning more about the users' perceptions of transparency and of data export (via their right of access) and data portability, after the task, the participants were asked to answer different questions in semi-structured interviews, in which planned questions were asked and other questions emerged based on answers which were annotated by the note keeper (observer).

The interviews consisted of some core questions, each with the candidate follow-up questions designed to encourage participants to give more information. All interviews followed the structure listed in the study plan including but not limited to the questions below:

- How do participants understand and perceive the concept of derived and disclosed information visualized in the Data Track? What other type of information about their data are they interested in? What do they value in the Data Track and what do they suggest to improve and what are their concerns regarding using the tool?
- Who has access to their data in the Data Track and where is the data that they uploaded to the Data Track stored? Will any changes of the data in the Google account affect the uploaded data to the Data Track and the other way around ? In what circumstances would they like to download and upload their data from/to service providers? How can the Data Track help them if it provides users with the option to edit/filter data and it saves the changes? What is the preferable way for them to transfer the data between service providers (directly or via the Data Track)?

Captured screen videos were checked against notes taken in each interview. The recordings were transcribed and coded to extract participants' ideas and perceptions. Notes taken during the interviews were compared with corresponding screen recordings to reduce the observer bias and ensure the accuracy of data.

### 3.3 Demographic Information

Demographic information extracted from the pre-study questionnaires and summarized in Table 1 shows that six women and four men participated in our study with different age ranges. All the participants but one have Google accounts, all of them work with computers and use the Internet daily or almost every day, and all of them possess smartphones. As discussed above, we tried to get an unbiased sample by mostly inviting arbitrary people from the city center or the Facebook group for citizens of Karlstad. Nonetheless, the table shows that many of the test participants have an academic background—probably because people with a higher education are more interested to participate in a research study by Karlstad University in English and also the fact that Karlstad is a university town with a high percentage of academics and students probably contributes to this effect.

**Table 1.** Summary of participants' information

| Demographic Information | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ID | Age range | Gender | Educational background | Google account | Smartphone | Computer usage | Internet usage | Knowledge of computer security and privacy |
| P1 | 21-25 | Female | Bachelor: Law | Yes | Yes | Almost everyday | Everyday | A bit familiar |
| P2 | 26-30 | Female | Master: Psychology | Yes | Yes | Almost everyday | Everyday | A bit familiar |
| P3 | 61-65 | Male | PhD: Natural sciences | No | Yes | Everyday | Almost everyday | No knowledge |
| P4 | 51-55 | Male | High school | Yes | Yes | Everyday | Everyday | A bit familiar |
| P5 | 31-35 | Female | High school | Yes | Yes | Everyday | Everyday | Quite familiar |
| P6 | 36-40 | Male | Bachelor: Physiotherapy | Yes | Yes | Everyday | Everyday | A bit familiar |
| P7 | 46-50 | Female | PhD: Business administrator | Yes | Yes | Everyday | Everyday | A bit familiar |
| P8 | 51-55 | Female | Bachelor: Social sciences | Yes | Yes | Everyday | Everyday | A bit familiar |
| P9 | 46-50 | Female | Bachelor: Psychology | Yes | Yes | Everyday | Everyday | A bit familiar |
| P10 | 61-65 | Male | Human resource management | Yes | Yes | Everyday | Everyday | Professional |

## 4 User Study Results

We analysed the answers that we received during the semi-structured interviews, categorized them using the grounded theory method [5] and identified common themes. In this section, we will provide the results of our user study in relation to our two main research questions.

### 4.1 Users' Perceptions of Transparency Functions

The results of our questions aiming to identify users' perceptions about transparency can be categorized in three main domains: 1) the users' understanding of data derived by service providers vs. explicitly or implicitly disclosed data to service providers, 2) users' attitudes about sensitivity and importance of derived data, and 3) desired transparency functions from the users' point of view.

#### 4.1.1 Derived vs. Disclosed Data

We intended to analyse whether the users were aware of or surprised about derived data and whether they were interested in having it visualized. The exact terms used by Google for derived activities from location data displayed in the Data Track (such as "tilting" or "in vehicle") were not meaningful for the participants and most of them were not much interested in this type of derived activity data that Google reveals in the exported file. However, all of the participants expressed that they were aware that some services providers have more data about them comparing to what they disclose directly. They mostly referred to Facebook and they stated their interests in other kinds of derived data like what Facebook learns about them by analysing the keywords and pages they search for. They were also more interested in derived data related to their long-term behavioral patterns (see Section 4.1.3).

We conclude that visualising data exports using the exact wording for the derived data categories that service providers such as Google provide, may thus not always be perceived as useful and more meaningful information about this type of data may be more appropriate to be shown by TETs.

#### 4.1.2 Sensitivity and Importance of Derived Data

Some participants expressed that they do not consider the derived activity data visualized in the Data Track as sensitive. To justify, they mentioned that they are not generally concerned about their privacy but they know that on the other hand other people do care about it. In addition, some other participants explained that it depends on different factors like what service providers can learn from analyzing the information and the combination of information and how they intend to use it. Correspondingly, P6 mentioned about activities users would like to hide and said: "**if we are on the highway and speeding in our car or if we are doing something illegal, it is really sensitive**."

Interestingly, just two of the participants said it is an important and sensitive kind of information. However, one of them mentioned she herself is not interested to know about it because too much information would only scare her.

Whether to have control or not and whether to know about the risks and benefits of derived data were the other parameters participants mentioned that contribute to the fact of importance of visualising derived data.

It is worth mentioning that the context of derived data was location data and the kind of information that people think could be derived from. However, asking them to think of other contexts than location data, their attitudes differed from previous ones. One Participant (P2) mentioned Facebook advertisements and said: "*based on the recipe you look for, they know if you are vegan or based on your membership in different groups they know if you are depressed and it is really sensitive.*" and the other (P4) expressed his concerns regarding user profiling rather than derived activities from his locations and said: "*political view is sensitive and wrong analysis could be harmful for my reputation.*"

#### 4.1.3 Transparency Functions

To clarify more about users' requirements regarding transparency features of the Data Track, we asked them about what they value in the Data Track, what are their concerns regarding using it and what they think should be improved. Moreover, to have a better insight into those transparency functions in which participants could be more interested and not to limit ourselves and our participants to the location context, we told them that they can think more generally about their most used service providers and what they would like to see visualized. In the following, we explain our observations and findings.

*Values and concerns related to the Data Track.* Regarding what participants value in a transparency tool like Data Track we avoided asking and reporting positive comments about whether they like the tool but we focused on the particular comments that explain how they may use such a tool. Some participants mentioned they cannot see the point of visualizing their data and they think it is not needed. Some other mentioned they prefer the visualization of the information of which they are not usually aware like behavioral patterns and statistical data about them.

Six participants expressed their concerns regarding privacy and security of their data if they want to use the tool. Also, two referred to trust and symbols making them feel less concerned. Surprisingly, one of the participants (P7) said: "*It is certified by the Google* (pointing out to the Google logo on the timeline view while navigating through the tool) *so I am not concerned. Google has a good reputation in my opinion*" and the other (P6) said: "*as it is developed at KaU I am less concerned*". In any case, branding may play a role in lowering concerns.

*What to be visualized.* 1) Behavioral patterns: Most of the participants were interested in knowing more about their movement and travel patterns, usage patterns for different service providers, some statistical data about their behaviors and some information about to whom their data is sold, how it is exchanged and how they receive related advertisements. "*One does not know how much data an application really gets, it is interesting to know about it.*", said participant one (P1). P3 also mentioned about the speed and movement profile: "*by knowing the speed of movement when walking they know that I am not young. it is good for me to know my speed.*"

2) Have more control: Three participants explained explicitly that they would like to have some functions in the Data Track to exercise more control over their data: "*Now that I am informed about the data, what can I do about it? I need to react on it.*"

3) Knowing about benefits and usage: One participant (P8) clarified that she cannot see the point of using the tool and she needs more information about the advantages of being aware of all her disclosed data.

Some suggestions about more practical information based on the exact locations and more representative icons for location pins were also among the comments.

### 4.2 Users' Perceptions of Data Export and Portability

For user-side TETs like the Data Track which visualizes data exports, it is important that users comprehend the visualized data in the tool is under their control. However, as mentioned earlier, previous Data Track usability tests and related tests conducted in the PRIME project revealed the users' problems to differentiate between user and services sides and confusions about where their data are stored. So first in this section, we report on users' understanding of locally and remotely stored data and we represent the results of people's opinions about who has access to their data uploaded to the Data Track.

Furthermore, adding the new functionality for visualizing personal data exports from a service provider to the latest stand-alone version of Data Track has the objective to allow to export and visualize personal data from users' accounts. In the future, it could also help people first to visualize and then edit (in future versions) the data exports according to their requirements before sending it to other service providers while they are exercising the right of data portability. Thus we aim to learn more about people's attitudes about Data Track as an intermediary TET when they want to transfer the data from one service provider to another. Since data portability is a new concept for users and it is not offered yet by service providers, we first investigate users' perceptions about downloading their data from a service provider, uploading it to another one and whether they can consider a scenario in which they may need or use these features. Then we investigate about their preferable way when they want to transfer their data from a service provider to another. Finally, we ask about their attitudes towards Data Track as a tool with which they can visualize exports of data, change and then send the edited version to the desired service provider. The results are reported in Section 4.2.2.

#### 4.2.1 Locally vs Remotely Stored Data and Access to the Uploaded Data to the Data Track

When we asked participants about who has access to the data that they uploaded to the Data Track, we received different answers that, with participants' justifications about their statements, are summarized in Table 2. Also in Table 3 we represent participants' opinions about where the data uploaded to the Data Track are stored. Although some participants (P1,P10) correctly answered who has access to the data, they were confused about where the data are stored. On the contrary, P4 who recognized that the data are stored on his machine thought that Google had access to his data uploaded to the Data Track.

Data Track is running within the browser. Also, the term *upload* is usually used for transferring files from the user side to online websites. We assume both of these facts suggest people that the Data Track is a web application connecting to some servers and they usually forget about the file they download to their machines and upload to the Data Track. Also, we assume because they can upload the data which they downloaded previously from Google to the Data Track, they think the Data Track is somehow connected to Google and synchronised with it. It needs more investigations and is the subject of future works.

**Table 2.** Participants' attitudes about who has access to the uploaded data to the Data Track

| Who has access to your uploaded (imported) data to the Data Track? | | | | |
|---|---|---|---|---|
| ID | P1-P10 | P2-P5-P9 | P4-P7 | P6-P8 |
| Answer | Just me | Some companies - probably all the world - my phone, my account | Google | Tool developers |
| Justification | If hackers cannot, it is just me - I downloaded the file and gave it to Data Track | Google sells the data to others - everything online can be accessed by others - you know, big brother! | It is synchronized with Google - functioning within Google | It is developed at Kau and I uploaded my data to it |

**Table 3.** Users' attitudes about where the data are stored

| Where the uploaded data to the Data Track are stored? | | | | |
|---|---|---|---|---|
| ID | P1-P2-P3-P7-P8 | P6 | P4 | P5-P10 |
| Answer | Somewhere on the Cloud | Hopefully in a server at Kau | On my computer | Google and my computer |

To further investigate about people's perceptions of locally and remotely stored data, we asked participants whether some changes in Google data would affect the visualized data in the Data Track or the other way around and we observed different answers that with participants' justifications about their statements are summarized in Table 4 and 5.

Comparing Tables 4 and 5, it reveals that although some participants correctly recognized that to see the changes made on the Google data they should again download and upload it to Data Track, they thought that modifying the uploaded data to the Data Track would affect the data on the Google side (P9 and P1). Moreover, although some participants correctly said that changes on the data in Data Track would not affect the Google data, the justifications about their answers were not the real reasons. One said if it removes the Google data it is a "bad software" (P3). In addition, some participants (P2,P3,P5,P6) understood that changes in the uploaded data to the Data Track or Google would not affect the other one (until we download and upload the data again to see the changes we made on Google data). However, interestingly, they could not recognize where the uploaded data to the Data Track were stored and who had access to it.

#### 4.2.2 Users' Attitudes of Data portability, Preferable ways and Usefulness of Data Track

About the usefulness of download and potential upload features in service providers and ability to think of a scenario in which people may use it, we observed different opinions. Some participants explicitly said that they do not need these options because they cannot think of a scenario in which they will use it or they think

**Table 4.** Users' attitudes about changes in Google data and its effect on the uploaded data to the Data Track

| Does editing the Google location data affect the uploaded data to the Data Track? | | |
|---|---|---|
| ID | P1-P2-P3-P5-P6-P9 | P4-P7-P8-P10 |
| Answer | No | Yes |
| Justification | I should download and upload first to see the changes, not simply refresh the page | It is synchronized with Google - It should be updated automatically - I hope so, I do not know really |

**Table 5.** Users' attitudes about changes in uploaded data to the Data Track and its effect on Google data

| Does editing the uploaded data to the Data Track affect the Google location data? | | |
|---|---|---|
| ID | P2-P3-P5-P6-P10 | P1-P4-P7-P8-P9 |
| Answer | No | Yes |
| Justification | They do not work together, they are not connected - It would be a bad tool otherwise - Google does not allow, it is downloaded | If you refresh Google, it will fetch new changes - They are connected - It is a service from Google |

the risks would outweigh benefits. Some others mentioned they think download and upload options are useful. They explained a scenario like downloading all of their Instagram pictures to have them on their machines or downloading all of their WhatsApp groups and messages to save them somewhere. Impressively, five out of ten people expressed their concerns regarding the risks in regard to security and two out of ten asked about the benefits:

"*If I can see the benefits and usage of it I will think of these options*" (P8) and the other said: "*I am not willing at all to download my data on my machine. It is fine on the cloud. It is too risky to have it on my machine because I am responsible for its security and if something happens it is my fault*" (P7).

We asked participants to consider a scenario in which they want to transfer their data (Facebook data including all advertisements they have clicked or the people they have searched about in Facebook) from Facebook to a new social network (because their friends moved to the new social network or according to the news the new one has better features) and tell us about their preferable way. They assumed they had two options: 1) download the data from Facebook, change/filter information and then upload it to the new website. 2) use a button on Facebook that directly sends the data to the new website. Different opinions and preferences are summarized in Table 6.

Interestingly, one of the people (P9) who preferred the button emphasized that she needs some kind of information or messages showing what is happening before she sends the data by clicking the button like the information she receives when she uses Facebook login button. The one (P5) who was not sure about her preferred option also mentioned about social login buttons and expressed her privacy concerns about what is transferred when she uses these buttons.

**Table 6.** Participants' preferable way to transfer data from one Service Providers (SP) to another

| Do you prefer to use a button directly or download the data and then upload it to a new SP? | | | |
|---|---|---|---|
| ID | P1-P2-P3-P4-P6-P9 | P7-P8-P10 | P5 |
| Preferable way | Download and then upload | The button | I am not sure |
| Justification | I feel safer - I want to have control over my data - By using button, I do not know what is going on, I want to see what is transferred | less time-consuming - no need to think - easy and convenient | I remembered the Facebook login button. It transfers the data but I am not sure what Facebook sends to the others. |

Finally, we asked people to consider the same scenario of transferring their Facebook data to another social network and we told them to imagine that Data Track would provide edit functionality and would be able to save the new version of data exports. Then we asked about participants' attitudes of the usefulness of Data Track in this scenario. All participants said that they can see the usefulness of the Data Track to have control over their data, change the parts they do not want to be included and it will be really helpful to adjust the data and visualize what will be transferred. However, some participants also again mentioned that they have privacy and security concerns (see Section 4.1.3).

## 5 Conclusion and Future Work

We conducted a user study with ten participants using semi-structured interviews aiming to understand the users' perceptions of data export, data portability and transparency functions in the latest stand-alone version of the Data Track. Albeit most of the participants showed rather little interest in the visualization of derived activity data revealed in the Google exported location data file, they stated their interests in other kind of derived data (e.g. by Facebook or online marketing services), like movement and travel patterns, usage patterns for different service providers, statistical data based on their behaviors and information about to whom their data are sold, how it is exchanged and how they receive related advertisements. In addition to the kind of transparency functions of their interests, some of them also stated that they would like to exercise more control over their data via this tool, e.g. they would like to have added functionality allowing them to delete or correct data (such control functions are actually offered by the previous (non stand-alone) A4Cloud Data Track version).

Analyzing users' perceptions of data export and portability, as we experienced in previous user studies, again confirmed that it is for many users confusing and difficult to differentiate between locally and remotely stored and controlled data. Several test participants were thinking that the data in the Data Track were synchronised with the data in the Google account. Several users were also concerned about the security of their data when they think of downloading the data on their own machines and being responsible for its security. Nonetheless,

most participants stated that for the purpose of exercising their right to data portability, they would prefer to first export their data, inspect and filter out some information before uploading it to another service provider, and would appreciate to use a tool such as the Data Track for helping them visualising and filtering data in this context. They thereby clearly would like to be in control when exercising the right of data portability over the easier option of having their personal data transmitted directly from one controller to another one.

We want to note that while we used fake location data of a persona in the first task of our user study, it will be interesting to conduct future user studies for our transparency tools with real data of test participants to analyse how they are reacting if they are confronted their own data traces. This will, however, require further careful preparation for recruiting suitable volunteers, for setting up suitable data protection and ethical procedures and for getting ethical approval by the university's ethics review board.

We are currently extending the stand-alone Data Track for allowing also to visualise data exports of other service providers like YouTube search history or Facebook data, which will let us make other types of disclosed and derived data transparent that participants in our user study showed interest in. Moreover, it will allow the users to compare what data different service providers know about them and what data the service providers have in common (a feature that the previous non-stand alone A4Cloud Data Track version already provided with its trace view).

Moreover, users should be provided with helpful instructions on how they can subsequently exercise their rights to erase or rectify data electronically by logging into the service provider's side (if these control functions are made available by the respective service provider—which is the case at least partially with Google today).

Besides, given the interest by participants to use the Data Track as a visualisation and filtering tool when porting data from one service provider to another one, we intend to expand the functionality of the Data Track supporting users in all data portability steps, i.e., supporting them to export the data from one service provider, to visualise and filter data and to import the altered data set to the new service provider.

## Acknowledgment

The authors gratefully acknowledge Daniel Lindegren for his contribution in conducting pilot study and John Sören Pettersson for his reviews that helped to improve the pilot study design. Furthermore, the authors also thank anonymous reviewers and participants of IFIP Summer School (2016), whose comments and suggestions greatly contributed to enhance and clarify our work.

## References

1. Angulo, J., Fischer-Hübner, S., Pulls, T., Wästlund, E.: Usable transparency with the Data Track: a tool for visualizing data disclosures. In: Proceedings of the 33rd

Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1803–1808. CHI EA '15, ACM, New York, NY, USA (2015), http://doi.acm.org/10.1145/2702613.2732701
2. Becker, R.A., Eick, S.G., Wilks, A.R.: Visualizing network data. IEEE Transactions on visualization and computer graphics 1(1), 16–28 (1995)
3. Bernsmed, K., Fischer-Hübner, S.: A4Cloud deliverable D.D-5.4 user interface prototypes (2015)
4. Bier, C., Kühne, K., Beyerer, J.: PrivacyInsight: The next generation privacy dashboard. In: Annual Privacy Forum. pp. 135–152. Springer (2016)
5. Charmaz, K.: Constructing grounded theory. Sage (2014)
6. European Commission: European Commission, 1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. (1995)
7. European Commission: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016)
8. Fischer-Hübner, S., Angulo, J., Karegar, F., Pulls, T.: Transparency, privacy and trust–technology for tracking and controlling my data disclosures: Does this work? In: IFIP International Conference on Trust Management. pp. 3–14. Springer (2016)
9. Fischer-Hübner, S., Angulo, J., Pulls, T.: How can cloud users be supported in deciding on, tracking and controlling how their data are used?, pp. 77–92. Springer Berlin Heidelberg, Berlin, Heidelberg (2014), http://dx.doi.org/10.1007/978-3-642-55137-6_6
10. Fischer-Hübner, S., Hedbom, H., Wästlund, E.: Trust and assurance HCI. In: Privacy and Identity Management for Life, pp. 245–260. Springer (2011)
11. Freeman, L.C.: Visualizing social networks. Journal of social structure 1(1), 4 (2000)
12. Hildebrandt, M.: Behavioural biometric profiling and transparency enhancing tools. FIDIS WP7 deliverable. http://www.fidis.net/
13. Kani-Zabihi, E., Helmhout, M.: Increasing service users privacy awareness by introducing on-line interactive privacy features. In: Nordic Conference on Secure IT Systems. pp. 131–148. Springer (2011)
14. Kolter, J., Netter, M., Pernul, G.: Visualizing past personal data disclosures. In: Availability, Reliability, and Security, 2010. ARES'10 International Conference on. pp. 131–139. IEEE (2010)
15. Mozilla: Lightbeam add-on for Firefox. https://addons.mozilla.org/en-US/firefox/addon/lightbeam/
16. Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M., Govindan, R.: Personal data vaults: A locus of control for personal data streams. In: Proceedings of the 6th International Conference. pp. 17:1–17:12. Co-NEXT '10, ACM, New York, NY, USA (2010), http://doi.acm.org/10.1145/1921168.1921191
17. Onwuegbuzie, A.J., Leech, N.L.: Validity and qualitative research: An oxymoron? Quality & Quantity 41(2), 233–249 (2007), http://dx.doi.org/10.1007/s11135-006-9000-3
18. Pettersson, J.S., Fischer-Hübner, S., Bergmann, M.: Outlining "Data Track": privacy-friendly data maintenance for end-users, pp. 215–226. Springer US, Boston, MA (2007), http://dx.doi.org/10.1007/978-0-387-70802-7_18

19. Pettersson, J.S., Fischer-Hübner, S., Danielsson, N., Nilsson, J., Bergmann, M., Clauss, S., Kriegelstein, T., Krasemann, H.: Making PRIME Usable. In: Proceedings of the 2005 Symposium on Usable Privacy and Security. pp. 53–64. SOUPS '05, ACM, New York, NY, USA (2005), http://doi.acm.org/10.1145/1073001.1073007
20. Popescu, A., Hildebrandt, M., Breuer, J., Claeys, L., Papadopoulos, S., Petkos, G., Michalareas, T., Lund, D., Heyman, R., van der Graaf, S., et al.: Increasing transparency and privacy for online social network users–usemp value model, scoring framework and legal. In: Annual Privacy Forum. pp. 38–59. Springer (2015)
21. Rubin, J., Chisnell, D.: Handbook of usability testing: how to plan, design and conduct effective tests. John Wiley and Sons (2008)