

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

Farzaneh Karegar¹, Daniel Lindegren², Tobias Pulls¹, and Simone
Fischer-Hübner¹

¹ Department of Mathematics and Computer Science

² Department of Information Systems, Karlstad University (KaU)
{farzaneh.karegar,tobias.pulls,simone.fischer-huebner}@kau.se,
dlindegren93@gmail.com

Abstract. For enforcing transparency requirements pursuant to the EU Data Protection Directive and the General Data Protection Regulation we have implemented a Transparency Enhancing Tool called Data Track. The latest stand-alone version of the Data Track allows users to visualize personal data exports to data subjects. We have conducted usability tests in three iterations and observed that people were not aware of transparency, control functions and the consequences of disclosing information. They often did not have a clear understanding of data portability and its benefits either. Consequently, correct mental models should be evoked and awareness of consequences and availability of transparency and control functions should be addressed by future Data Track User Interface improvements.

Keywords: Transparency Enhancing Tools, Data Portability, Usability, 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 is collected, used, shared or accessed [10]. Consequently, transparency of personal data processing is enforced by most Western privacy laws, including the EU Data Protection Directive (DPD) 95/46/EC [1] and new General Data Protection Regulation (GDPR) [5] which will replace the DPD in 2018.

The GDPR grants enhanced data subject and transparency rights, including the right to access their data, receive an electronic copy of their data, the right to be informed about the logic involved in any automatic processing such as profiling and about significant envisaged consequences and the right to data portability. One way to exert the 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) [11].

The purpose of our work presented in this paper is to study an ex-post TET tool (which shows what data has been disclosed and how data has been processed), the Data Track (DT), developed at Karlstad University (KaU). The DT which started as a part of the European PRIME [2] and PrimeLife [3] projects and continued as part of the A4Cloud project [4], provides users with an overview of the data they have disclosed to service providers under an agreed-upon policy.

Besides functions to meet transparency requirements, we added a new functionality to the DT which helps visualize exports of personal big data to the data subjects. This function can provide users with an overview of the real location data they have disclosed to Google by exercising the right of data export which Google provides to its users via *myaccount.google.com*. The right of data portability pursuant to Art. 18 GDPR aiming at increasing user choices of on-line services, allows users to request all data from the controllers and provides users with the data in electronic form which can then be transmitted to any other controllers. Alternatively, the user can also request to transmit her data directly from a service provider to another one. Our general research objectives aim to discover whether the concept of data export from the service provider to the Data Track running at the user's machine by exercising the right of data portability is well-understood and to research how visualization of data exports can enhance usable ex-post transparency for end users. Accordingly, our study on the usability of the latest stand-alone DT version is based on the following sub-research questions:

1. Do users understand the concept of exporting data from a service provider (Google in this case) and importing it to their own machines?
2. Do users understand the differences between locally and remotely stored data? (i.e. data stored on their computers in the Data Track after being exported from a service provider vs. data stored at the service side)
3. Are users willing to take any actions after being exposed to the visualization of the data they disclosed implicitly or explicitly?
4. Are people aware of several transparency and control functions? (Especially the functions related to request to download data from Google)
5. Do users trust and use a tool such as DT using their personal data?
6. Does the interface convey that Google has more information about them other than the one that is sent explicitly?

In this paper, section 2 presents related work, section 3 briefly describes the Data Track and its different views, section 4 explains the methods used in our work and the test plan. Section 5 is devoted to analyzing the results. Finally, section 6 expounds conclusion and future work.

2 Related Work

A variety of Transparency-Enhancing Technologies has been presented during the past years with different properties and functionalities that make them suit-

able for various requirements. Besides TETs, some services like datacoup.org which try to return the control of personal data to the hand of users are growing. There are also some detail overviews of TETs and some discussions about challenges, risks and benefits of Personal Data Services (PSD) among literature [16, 17]. In [16], Janic et al. provide an overview and description of the sixteen available transparency enhancing tools at the time of writing (2013). As described in [16] the majority of TETs promote awareness. Identified tools were classified into the following categories according to the transparency functionality they offer [16]:

- Tools that provide insight into intended data collection, storage, processing and/or disclosure, based on website privacy policy
- Tools that provide insight in collected and/or stored data
- Tools that provide insight in third parties tracking the user
- Tools that provide insight in data collection, storage, processing and/or disclosure based on websites reputation
- Tools that provide insight into (possibly) unwanted user’s data disclosure (awareness promoting)

Data Track like Google Dashboard which is a publicly available TET [8] falls into the category of tools that provide insight into collected and/or stored data. Pulls describes a cryptographic scheme for storing all data disclosures tracked by the DT centrally in the cloud in a privacy-friendly way in [15] and more information about earlier UI versions of the DT is available in [10, 13, 14]. In accordance with the results of usability tests on the previous version of the DT [10], we extended the DT to visualize real exports of personal data from service providers. To the best of our knowledge, it is the first work focusing on usability of data portability and users’ perceptions of it which aims to help users exercise the right of data portability and visualize exports of data to have a better insight over implicit and explicit disclosed information.

3 Background

Different versions of the Data Track which is a user side ex-post transparency tool have been developed within the EU research projects PRIME [2], PrimeLife [3], and A4Cloud [4] with gradual enhancements.

In this paper, we focused on the latest version of the Data Track which is an open source and standalone program developed at the end of the A4Cloud project at KaU. It provides users with the visualization of data exported from the Google managing archive service. Among different Google products, 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 DT, participants have three views showing their personal data: timeline view, map view and trace view. Figure 1 shows the exact interface of the DT where users should select the desired view to visualize their personal data. The map view which is depicted in Figure 2 displays disclosed location pins to Google and the user sees

related information like activities in a small pop-up window by clicking on it. The map view also presents the movement patterns as described in the Google location history file. The timeline view which is shown in Figure 3 displays disclosed locations and related information in disclosure boxes chronologically. The activities shown in disclosure boxes or pop-up windows are derived by Google based on the location reported by mobile devices. Finally, the trace view which is depicted in Figure 4 displays different type of information disclosed to a service provider by drawing lines between them (the trace view is excluded from our test because we have just one service provider (i.e., Google) for now).

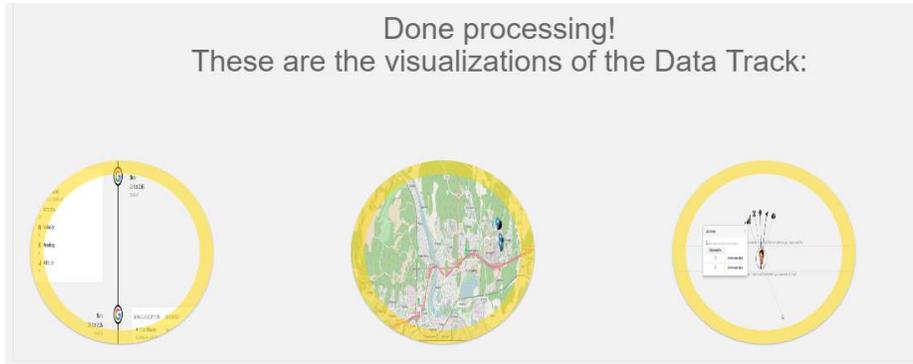


Fig. 1. Three different views shown to user after importing data to the DT

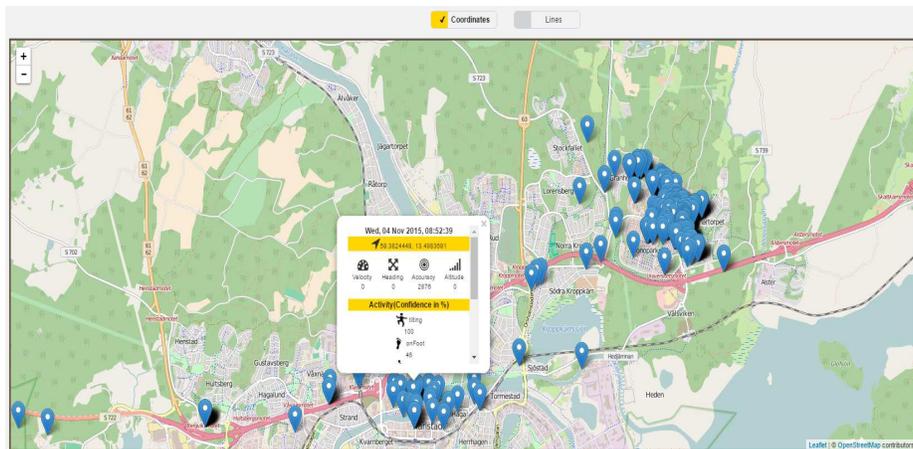


Fig. 2. Map view in the DT

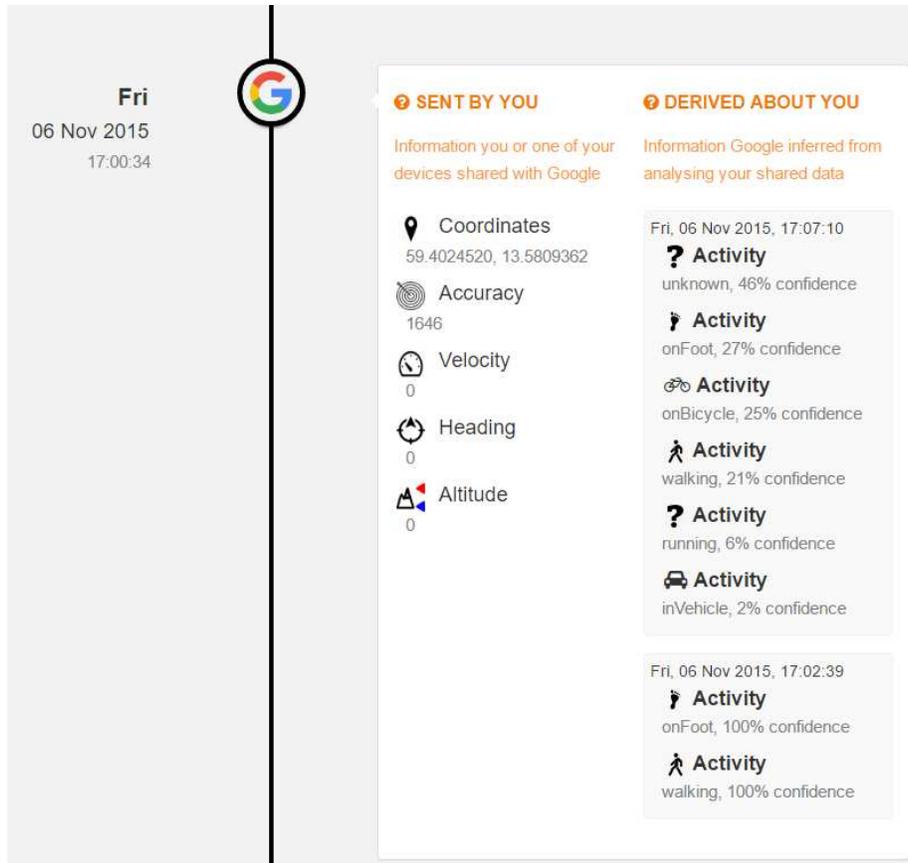


Fig. 3. Timeline view in the DT

4 Usability Tests and Methods

4.1 Participants

Participants were randomly recruited at KaU via recruitment emails sent to researchers' friends, family and coworkers and those who accepted the invitation were invited to participate in the tests. According to [12], it is important to retain the same conditions for sessions. Hence, all the tests were done in the same room, in a calm environment with an orientation script the test moderator used at KaU. As described in [6], the "best results come from testing no more than five participants and running as many small tests as you can afford". Rubin & Chisnell [12] also mention that most of the problems are found within the first four participants. As a result, we decided on running the tests for three user interface iterations with 16 participants in total (with six people in the first iteration to have the option to use counterbalancing [12]).

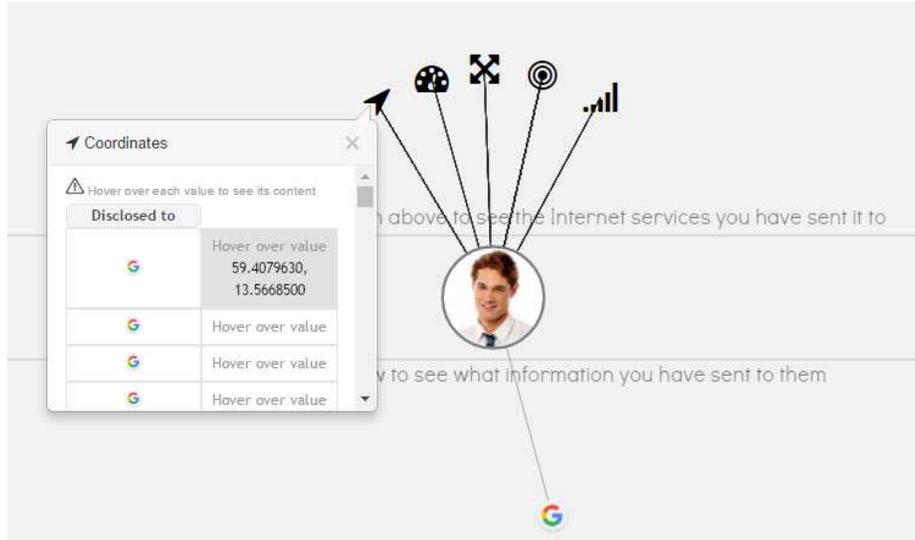


Fig. 4. Trace view in the DT

The order that participants experience the different views can influence the results by causing bias arisen from habituation or fatigue effects for example. Correspondingly, the order of experience the different views in the DT is balanced across the group in the first iteration so half of participants first experienced the map view and then the timeline view and the other half experienced the views in the reversed order. From the second iteration on, we excluded the map view from the tests because the results of the previous iteration showed that there was no difference between understanding the concepts based on the type of views. Furthermore, the pop-up window related to each location pin in the map view shows the exact same information with the same icons and terminologies as the disclosure box for each location in the timeline view. As a result, we focused just on the timeline view and prevented the repeated questions and answers for the map view for the second and third iterations.

Focusing on the users perceptions of different concepts in the DT, the goal of defining the subtask was not to time participants for efficiency in finding the disclosed location but to have the location as the starting point of discussing different elements shown to the user in each disclosure box. Consequently, the collection of verbal comments was prioritized in this research after the data was imported to the DT. Thus usability was measured objectively by the ability to export data successfully using the instructions and subjectively using questionnaires and answers to interview questions.

4.2 Tasks

We considered one main task and a subtask to be performed by users. The main task was to export data from Google and import it to the DT and the subtask was to find a specific location that was disclosed on a certain time in the timeline view. We used successful completion criteria to get a sense of how efficiently participants completed the main task and to measure success [12]. An internal test was conducted to setup the time for the task. Table 1 shows the main task description with the defined successful completion criteria.

Table 1. The main task with successful completion criteria

TASK COMPONENT	DESCRIPTION
Main Task	Follow the instructions and download your Location History-file and upload it to the Data Track.
State	Participant has arrived at the instruction page for downloading the Location History-file.
Successful completion criteria	Log into the Google account, download the Location History-file and manage to upload it to the Data Track by the help of the instructions.
Benchmark	From the point the participant is given the task to the point the data is uploaded to the Data Track, it should not take more than 4 minutes.

4.3 Study procedure

To begin with, a test plan was written to serve as a main communication vehicle as well as a blueprint for the tests. A test plan is a summary of all the containing documents needed for the usability tests [12]. Besides the order of experience the different views, the researcher bias could also be a problem in an experimental setting, thus procedures were standardized. Each participant received minimal instruction and followed the same blueprint. Apart from the standard instructions given to all participants, no further explanations were given if a participant asked for more clarification. In these cases, the moderator clarified in a friendly manner that we were testing the usability of the system and needed to see if people could use them without explanations. The participants' goals were to successfully export the location data from the Google and import it to DT and navigate through different views to see their previously disclosed information with its characteristics. To reach the goals, participants were given as much time they wanted.

Participants' personal data was not used in the tests. Instead they were given the role of a persona to play to visualize their data and exercise data portability right. Using a persona, participants feel secure that they are not compromising their personal details when taking part in the tests. 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 user name and password for a Google account. Before running our tests, we made a dedicated Google account for our tests and signed into the account on different mobile devices. Turning on the location on the mobile phones and enabling the Location History, allowed Google to create a private map of where we went with our signed-in devices. To create the map, Google regularly obtained location data from the devices we had enabled Location History, even when we were not using a Google product.

Data Collection. Multiple data collection methods were applied in the tests. The participants went through a consent-form, pre-test questionnaire to see if participant's characteristics had different impact on the understanding of DT, tasks which were annotated by the observer, semi-structured interview questions in which planned questions were asked and other questions emerged based on answers, and a post-test questionnaire to help us to better answer our research questions.

An observer sat with each participant throughout the session, recording observations, noting any difficulties, any obvious misconceptions in participants mental models of the Data Track and any comments made by participants.

Before beginning the tests, participants gave informed written consent to take part in the study. The informed consent imparted that participants agreed to have their screen recorded, alongside with their comments as they were interacting with the prototypes. Participants then answered short demographic information and at their own pace navigated through the DT. To improve observation, we used a screen-capturing software called OBS, which also recorded audio. The OBS recorder was then turned on and the semi-structured interview began. Sixteen semi-structured interviews were conducted, lasting between 16 and 36 minutes depending on how much the participants wanted to communicate.

Interviews consisted of some core questions, each with the candidate follow-up questions designed to encourage participants to give more information. Albeit the core questions pursued each other logically, it was not necessary to ask the core questions strictly in that order. If the participants took the discussion down a different route, the researcher could choose to ask a core question earlier or later than planned but in a way that all core questions were eventually covered. However, most interviews followed the structure listed in the test plan, and the first question was always the same: could you please tell me what does this box show? What do you think of different elements?

Captured screen videos were checked against notes taken in each interview. The recordings were transcribed and coded to extract participants' ideas and perceptions. In the view of the fact that the validation of the analysis had to be ensured, notes taken during the interviews were compared with corresponding screen recordings. Comparing the notes and videos reduced the observer bias and ensured the accuracy of data.

5 Results

5.1 Demographics

Demographic information extracted from pre-test questionnaires shows that 6 women and 10 men participated in our tests who were 27.25 years old in average. Five participants were quite familiar with security and privacy in computer science while seven of them were a bit familiar and four of them mentioned they had no knowledge in this area. Although half of our participants were from Computer Science and Information Systems Departments which might lead to have biased results, outcomes from usability tests showed even people with computing background experienced usability problems. So we assume that lay users would also have such problems.

5.2 Results from interviews and questionnaires

We describe the results derived from our data collection methods for each research question.

Data portability and control over data. *Research Questions 1 and 2:* Based on the data collected during and after the tests, it became obvious that it was hard for participants to think of a scenario in which they might need data portability. Some of the participants mentioned they felt being insecure and unsafe when they thought of transferring their personal data between controllers as one of them said: "It would feel safer just to type everything from scratch." It was not straightforward for participants to think separately of what was exported and what was imported, so most of them could not differentiate between what data was under their control and to what data controller still had access. Consequently, they thought any changes in one of the imported or exported data would affect the other. The type of data also had an impact on the users' willingness to transfer their personal information between controllers. They expressed a preference of downloading the data on their device before uploading it to the other controller over directly requesting to transfer their data between controllers as one of our participants mentioned: "I would rather download to my computer then upload rather than having something in between."

We asked participants what would happen on the Google side if you delete one of your disclosed location pin from the DT and vice versa. 4 out of 16 thought Google and the DT were synchronized. So they expected to see any changes in each of them reflected on the other. 6 participants answered that the DT and Google were not connected and if they wanted to see the new changes they had to download their location from Google again. Interestingly, 6 participants expressed the variations in a unidirectional way so that they thought any changes on Google might be reflected in DT but DT changes might not be mirrored on Google side and vice versa.

Although we provided participants with video and text/picture based instructions, it was hard for them to find the content control section in Google.

Just four out of sixteen were able to finish the main task in the normal time which was around four minutes. Although video or animations are useful to teach procedural skills [7] and the instruction video was helpful to show the connection between different steps for exporting Google data in the second iteration, placing instruction video before showing the text/picture based instructions increased the average time required to finish the task. We observed that participants played the instruction video several times, and it distracted them from following the text/picture based instructions. Also, because all different steps are not visible in the same frame when watching a video, they tried to move forward and backward to acquire what they needed. Therefore, for the last iteration, we changed the order of instructions. The instruction page of the DT is depicted in Figure 5.

Aiming to clarify different concepts the DT is conveying and to help people understand how their data is stored and transferred using the program, we added an introduction video at the homepage of the DT, which is presented in Figure 6. Since watching the video is not the primary task of users, it was usually skipped by individuals to continue the next steps of the program.

Finally, it is worth mentioning that some participants thought data portability was the same as *Sharing* options available in service providers. Using *Sharing* options, we can share the selected data from one service provider to another, and it is not usually a bidirectional option. Also in some cases, if we share something from service provider A to B and then delete it from A, it will also disappear from B.

Awareness of consequences. *Research Question 3:* In case of location data, 62.5% of participants said they had no idea whether they would take any actions to delete their location data or turn off the location on their phone after being exposed to what they had disclosed and what Google had derived about their data. Participants were not aware of the sensitivity of location data and what could be derived from it. Not being aware of consequences cause people to think location data is not as sensitive as for instance health data, we assume.

Awareness of transparency and control functions. *Research Question 4:* Results show that the participants were not aware of available transparency, control and data requesting functions of Google. None of the them were aware that they could request to download different type of data from Google.

Trust issues. *Research Question 5:* It was hard for participants to trust the DT in case of handling their data securely. Five participants thought they were the only ones who had access to their data, although three of them were not sure about it. 25% of participants directly mentioned Google had access to their data on the DT. There is an open question regarding trust situation when we extend the DT with exports from other service providers (such as Facebook): Is it harder for people to trust the DT if we extend it with different service providers?

We help you to import your data from Google
by text/picture based instructions and video

You can follow the instructions

1 First Step



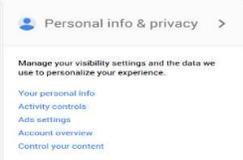
You have to log in to your Google account.

2 Second Step



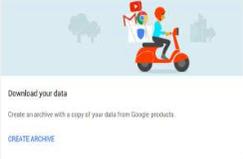
Select "My Account" setting at the top right corner.

3 Third Step



Among all, select "Personal Info & Privacy"

4 Fourth Step



Under "Control Your Content", press "Create Archive"

5 Fifth Step



Among all, just select "Location History".

6 Sixth Step



Press "Create Archive" button to create zip file.

Or you can watch the video



HOW TO
Download your location history from
Google

1:20

Are you done with importing data? Select your Google Takeout zip-file:

Choose File No file chosen

Fig. 5. Instruction page of the DT

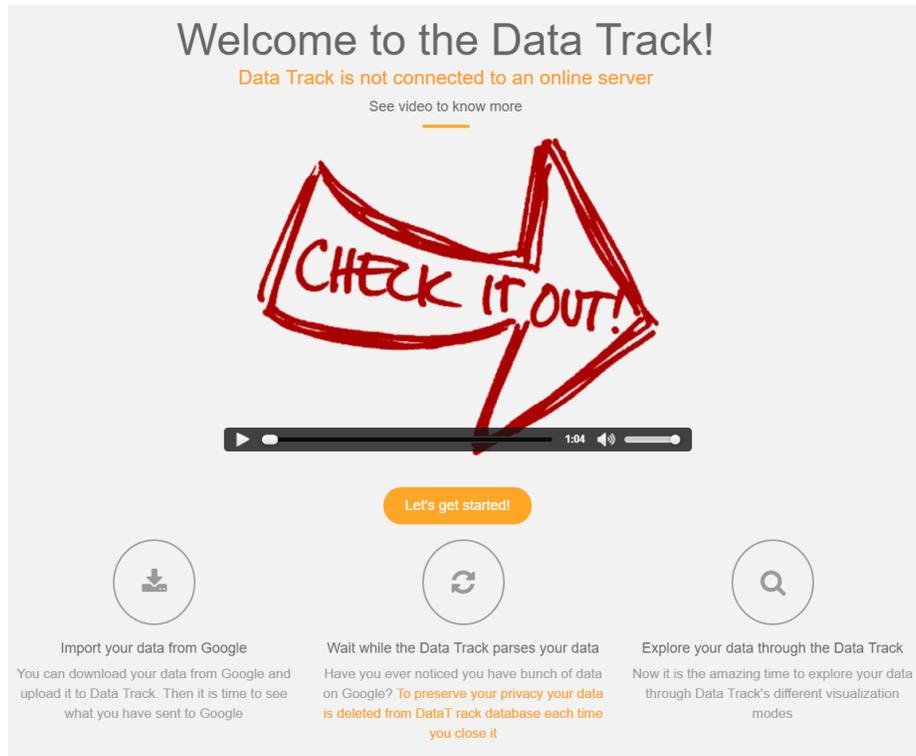


Fig. 6. Homepage of the DT

People may think available service providers have access to each other's data. Also the problem of trust issues would get worse if people want to exercise the right of data portability between different service providers. The results from our study showed that people did not have a clear understanding of how their imported and exported data is controlled and accessed. Accordingly, it may be difficult for users exercising their data portability right to understand how and where the original data and exported data is stored and accessed. Furthermore, current situation of some cloud services like Dropbox and Google Keep, which is a syncing notepad that connects to Google Drive, may influence the way people think their downloaded data is controlled and accessed. To answer what would happen if we change some of the location data on Google, one of the participants said: "The changes will be made automatically. It is the same with Google Keep."

Awareness of implicit and explicit disclosed data. *Research Question 6:* In the DT, contrasting colors, an explicit headline, adequate spacing and different brief description under each headline are used to differentiate between data that was explicitly submitted by the user from data that has been implicitly collected by the service provider or inferred from analysis of aggregated data

attributes. For example, Google infers user’s type of activity by analyzing the data it receives from user’s location. The Location History file which the user can request to download includes the activities like on foot, walking, in vehicle, tilling, still, on bicycle and unknown activity. By improving the design in three iterations and trying to clarify the terminologies used by giving an example, 4 out of 5 participants in the last iteration were all successful to understand the differences between derived and disclosed data. It is worth mentioning that one of our participants described the derived data as the data he did not allow the service provider to collect. Obviously it depicts the problem that some people do not pay attention to privacy policy and do not read them thoroughly. In addition, participants were generally aware that service providers could derive some extra information about them but the degree of awareness and concerns related to derived data are based on the context and type of data.

6 Conclusion and Future Work

As the results from usability tests show, people in general do not have a clear understanding of what data portability means and why they should use it. They are not aware of its benefits and cannot think of a scenario in which they may need to transfer their data to another service provider. Correct mental models for data portability need to be evoked and people should be informed about its benefits and cases in which data portability may help them have a better online experience.

Moreover, people are concerned about what a service provider potentially can send to the other parties when they request to transfer their personal data directly between controllers and thus would rather like to keep in control. In addition to concerns about the security and privacy of the personal information, exporting and importing data from and to different services and thinking about scenarios in which users can directly transfer information to other services confuse them about how and where their data is stored, how and by whom it is accessed and controlled. Now, many well-known service providers such as Google, Facebook or LinkedIn provide users with some transparency and control functions, but it seems people are not aware of their rights and they rarely exercise their rights by using the functions. Consequently, awareness of transparency, control, and data portability features besides awareness of consequences should be improved.

Using an introduction video at the home page of the DT did not help people understand the different concepts the program was trying to convey. Instead of a video which is usually skipped by individuals to continue the next steps of the program, we can use mandatory tutorial or policy notices with Drag and Drop (DaD) options like Drag and Drop Agreements (DADAs) introduced in [18]. By DaD people can play with different elements which show graphically what happens when their personal data is transferred from their computer to service providers or between service providers. We can help them understand the

data portability and differentiate between remote and local data when they are playing with graphical icons of their personal information and service providers.

We admit that our work presents some limitations and some items to future work. It is based on interviewing 16 people half of them from computer science department. In spite of the fact that the results show even people with computing background have some difficulties with understanding the program and data portability, we need more participants with various background to have better insight. Additionally, the novelty of concepts presented in DT such as exporting data from a controller and importing it to the other in order to visualize implicit and explicit disclosed data, may have artificial impacts on users' perceptions. The participants had only limited time to use the DT and reflect on the different questions asked. Therefore, we need to repeat similar interviews over time to divulge changes in perceptions as the data portability become more widely used.

Considering all the mentioned improvement above, we want to extend the DT also to be able to use real data from other service providers and help users to better understand and exercise their data portability rights between different services.

References

1. 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. Office Journal L. 281. 23.11.1995.
2. Privacy and Identity Management for Europe, 2010. [Online] Available at: <http://www.prime-project.eu>.
3. Privacy and Identity Management in Europe for Life, 2012. [Online] Available at: <http://primelife.ercim.eu/>.
4. Accountability for cloud and other future Internet services, 2013. [Online] Available at: <http://www.a4cloud.eu>.
5. European Commission, 2015. Proposal for a Regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Brussels, 15 December 2015.
6. Nielsen, J., Why you only need to test with 5 users, 2008. [Online] Available at: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [Accessed 1 March 2016].
7. Clark, R.C. and Lyons, C., 2010. Graphics for learning: Proven guidelines for planning, designing, and evaluating visuals in training materials. John Wiley and Sons.
8. Google. Google dashboard. <https://www.google.com/settings/dashboard>.
9. Fischer-Hübner, S., Angulo, J. and Pulls, T., 2013, June. How can cloud users be supported in deciding on, tracking and controlling how their data are used?. In IFIP PrimeLife International Summer School on Privacy and Identity Management for Life (pp. 77-92). Springer Berlin Heidelberg.
10. Angulo, J., Fischer-Hübner, S., Pulls, T. and Wstlund, E., 2015, April. 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). ACM.

11. Hildebrandt, M., 2009. D7. 12: Behavioural Biometric Profiling and Transparency Enhancing Tools. FIDIS WP7 deliverable, Available at: <http://www.fidis.net/>.
12. Rubin, J. and Chisnell, D., 2008. Handbook of usability testing: how to plan, design and conduct effective tests. John Wiley and Sons.
13. Pettersson, J.S., Fischer-Hbner, S. and Bergmann, M., 2007. Outlining Data Track: privacy-friendly data maintenance for end-users. In *Advances in Information Systems Development* (pp. 215-226). Springer US.
14. Fischer-Hbner, S., Hedbom, H. and Wstlund, E., 2011. Trust and assurance HCI. In *Privacy and Identity Management for Life* (pp. 245-260). Springer Berlin Heidelberg.
15. Pulls, T., 2012, October. Privacy-friendly cloud storage for the data track. In *Nordic Conference on Secure IT Systems* (pp. 231-246). Springer Berlin Heidelberg.
16. Janic, M., Wijbenga, J.P. and Veugen, T., 2013, June. Transparency enhancing tools (TETs): an overview. In *2013 Third Workshop on Socio-Technical Aspects in Security and Trust* (pp. 18-25). IEEE.
17. Acquisti, A., Krontiris, I., Langheinrich, M. and Sasse, M.A., 2013. 'My Life, Shared-Trust and Privacy in the Age of Ubiquitous Experience Sharing (Dagstuhl Seminar 13312). *Dagstuhl Reports*, 3(7), pp.74-107.
18. Pettersson, J.S., Fischer-Hbner, S., Danielsson, N., Nilsson, J., Bergmann, M., Clauss, S., Kriegelstein, T. and Krasemann, H., 2005, July. Making PRIME usable. In *Proceedings of the 2005 symposium on Usable privacy and security* (pp. 53-64). ACM.