How can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used?

Simone Fischer-Hübner, Julio Angulo, and Tobias Pulls

Department of Computer Science
Karlstad University, Sweden

Abstract. Transparency is a basic privacy principle and factor of social trust. However, the processing of personal data along a cloud chain is often rather intransparent to the data subjects concerned. Transparency Enhancing Tools (TETs) can help users in deciding on, tracking and controlling their data in the cloud. However, TETs for enhancing privacy also have to be designed to be both privacy-preserving and usable. In this paper, we provide requirements for usable TETs for the cloud. The requirements presented in this paper were derived in two ways; at a stakeholder workshop and through a legal analysis. Here we discuss design principles for usable privacy policies and give examples of TETs which enable end users to track their personal data. We are developing them using both privacy and usability as design criteria.

Keywords: Privacy, transparency, transparency-enhancing tools, usability.

1 Introduction

Transparency of personal data processing is an important principle for the privacy of individuals as well as for a democratic society. As for instance the German constitutional court declared in its Census Decision[1], a society, in which citizens could not know any longer who does, when, and in which situations know what about them, would be contradictory to the right of informational self-determination. For these reasons, transparency of personal data processing is enforced by most western privacy laws, including the EU Data Protection Directive 95/46/EC [9], by granting data subjects extensive information and access rights. Transparency is also an important factor of social trust, since trust in an application can be enhanced if procedures are clear, transparent and reversible, so that users feel in control of their personal data [2], [19]. However, particularly when data are processed in the cloud, multiple processors and subcontractors along a cloud chain can be involved that may belong to different legal entities and may be located in different jurisdictions. End users often lack transparency with regard to who is processing their data, under which conditions, and how they can exercise their data subject rights.

[1] German Constitutional Court, Census decision, 1983 (BVerfGE 65,1).

M. Hansen et al. (Eds.): Privacy and Identity 2014, IFIP AICT 421, pp. 77–92, 2014.
© IFIP International Federation for Information Processing 2014

The concept of transparency comprises both 'ex ante transparency', which enables the anticipation of consequences before data are actually disclosed (e.g., with the help of privacy policy statements), as well as 'ex post transparency', which offers information about any consequences if data already have been revealed (what data are processed by whom and whether the data processing is in conformance with negotiated or stated policies) [14]. The A4Cloud European Union (EU) Seventh Framework Programme (FP7) project[2] is creating ex ante and ex post transparency enhancing tools to support cloud users[3] in *deciding* on and *tracking* and *controlling* how their data are used by cloud service providers [23].

Transparency Enhancing Tools (TETs) that allow the tracking of the processing of personal data can, however, also endanger privacy, if personal data about a data subject or information about how data have been processed are made available to unauthorised parties (e.g., the information that a psychiatrist has accessed a patient record may reveal sensitive information that the patient may want to keep confidential and that according to Art.8 EU Data Protection Directive 95/46/EC require special protection). Hence, TETs for enhancing privacy should be designed in a privacy-respecting manner.

Moreover, as pointed out in Patrick & Kenny [22], legal privacy principles, such as the transparency principle, have Human Computer Interaction (HCI) implications as "they describe mental processes and behaviour that the data subjects must experience in order for a service to adhere to these principles". In particular, the transparency principle requires that data subjects *comprehend* the transparency and control options, are *aware* of when they can be used, and *are able* to use them. Therefore, another important design criterion for transparency enhancing tools is usability.

In this paper, we will discuss our work, mainly conducted within the scope of both the A4Cloud project and a Google Research Award project, on transparency enhancing tools that are both usable and privacy-preserving. The remainder of this paper is structured as follows. In section 2, we discuss and describe requirements that have HCI implications, which we have elicited through a stakeholder workshop and legal analysis. Section 3 discusses how parts of these requirements for ex ante transparency can be mapped to HCI requirements, principles and design proposals. In section 4, we present ex post TETs that we are developing at Karlstad University and discuss how they are designed to be both privacy-preserving and usable. Section 5 briefly presents related work. Section 6 concludes the paper by discussing some of the remaining challenges for usable and privacy-preserving TETs.

2 Problems and Requirements

Within the scope of the A4Cloud project, we follow a human centred design approach for analysing end user problems and eliciting and testing HCI requirements for TETs. These comprise different methodologies, including stakeholder workshops, user

[2] http://www.a4cloud.eu/

[3] In this paper, cloud users refer to both individual user as well as organisations that are outsourcing their data to the cloud. The focus of this paper is however on individual cloud users.

experiments, usability tests, legal analyses and literature studies. This section of the paper reports on the main results emerging from a stakeholder workshop and a legal analysis. Further results from usability tests are described in section 4.2 below.

2.1 End User Challenges - Results from a Stakeholder Workshop

Stakeholder workshops provide an opportunity for active face-to-face interactions between different influential actors who can express their opinions and needs for systems being developed. This workshop method is strongly encouraged during the initial design process, as a way of ensuring that the needs of those who might be impacted by the system are taken into account [20]. In February 2013, a stakeholder workshop was held at Karlstad University to elicit HCI-related requirements for A4Cloud tools including TETs for end users. Participants at the workshop included IT experts of service providers from the private and public sectors that are adopting or are planning to adopt cloud technologies as well consumer organisation representatives who are well aware of the problems that individuals face regarding cloud computing and who thus represent the stakeholder group of individual cloud users. In addition, a lawyer from the Swedish Data Protection Agency (Datainspektionen) was attending the workshop: through her work, she is familiar with the kinds of privacy concerns that data subjects have with regard to the handling of their personal data in the cloud. The results of the stakeholder workshop are reported in detail in Angulo et al. [4]. Most notably, the workshop revealed problems for individual end users with respect to the unclear responsibilities of cloud service providers. In particular, it is often not clear to end users who the data controller is and what liabilities data processors and service brokers have. It is also difficult for them to find out how to obtain redress if something goes wrong and what (national) laws apply. This is especially an issue if Swedish service brokers use services that reside in other countries or if a service provider appears to be located in Sweden (as it has a website in the Swedish language or with a Swedish domain/address/telephone number), but is in fact located in another country.

Furthermore, the shortcomings of trust seals and privacy policies were brought up in the workshop. Often individual end users do not make truly informed choices. It can be easy to deceive people because they often do neither read nor understand legal terms and agreements. There are no established trust seals for cloud services, and even if there were, how would the end users know what labels to trust? It was mentioned that individuals are often not interested in understanding all the details of trust seals, but would rather like to know in general whether their data are “secure”.

Another problem that workshop attendees pointed out is that there is usually insufficient support for service cancellation or data export. While registration for a service is usually made easy, it is often made difficult for end users to de-register or terminate a service contract, delete data, or transfer data to other service providers. It is not always clear to end users whether they "own" their own data, as they do not check the terms and conditions carefully.

In conclusion, the stakeholder workshop revealed several end user challenges with regard to privacy policies and the exercising of data subject rights that should be

addressed by ex ante and ex post TETs. We will therefore also address these issues in the following sections.

2.2 Legal Requirements for Transparency and User Control

This section discusses essential legal privacy principles for transparency and accountability for the cloud, for which HCI requirements and principles can be derived. Our legal analysis will mainly refer to the principles of the EU Data Protection Directive 95/46/EC, but we will also cite other legal requirements deriving for instance from the opinions of the Article 29 Data Protection Working Party. In view of the ongoing review of the European legal framework on data protection, our analysis will also take into account legal principles that are being proposed in the draft EU General Data Protection Regulation (GDPR) [10] and the compromise text of the proposed GDPR that was passed by the Committee on Civil Liberties, Justice and Home Affairs (LIBE) of the European Parliament on October 21, 2013 [11].

This section will place an emphasis on legal provisions for transparency and accountability for the cloud that have implications for HCI and that thus need to be addressed by the design of graphical user interfaces. These legal provisions mainly comprise transparency rights as well as detective and corrective control rights that data subjects have in regard of data controllers[4]. The proposed EU regulation also highlights the importance of usable transparency and user control by requiring that data controllers have *"transparent and easily accessible policies with the regard to processing of personal data and for the exercise of data subjects' rights"* (Art. 11 draft GDPR) – which according to the compromise text of the GDPR also need to be *concise* and *clear* [11].

Information Rights (Ex Ante Transparency)

Ex ante transparency is a condition for data subjects to be in control and to render a consent[5], which has to be informed, valid. Article 10 of the Data Protection Directive defines what information relating to the processing of their personal data needs to be given to data subjects when information about them are collected and processed. This includes at least the identity of the data controller, and the data processing purposes. Moreover, further information needs to be given for example on the recipients or categories of recipients of the data, on whether replies to questions are obligatory or voluntary and on information about the individual's rights in so far as such further information is necessary to guarantee fair data processing. Such information has to be provided to the data subjects not only when the information is collected from the data subjects, but also when the data have not been obtained from them (Art. 11 Data Protection Directive).

---

[4] According to EU Directive 95/46/EC, a data controller is defined as the entity that alone or jointly with others determines the purposes and means of personal data processing.

[5] The data subject's consent is defined by the Data Protection Directive as "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed".

The processing of personal data has to be based on one of the grounds that are mentioned in Art. 7 of the Data Protection Directive. The consent of the data subject (Art 7 (a) Data Protection Directive) can be taken as a legitimisation of personal data processing in the cloud. Information that needs to be given to data subjects for a valid (informed) consent should cover at least the elements of information required by Art. 10 Data Protection Directive.

The draft GDPR in Art.14 extends the information that should be provided to data subjects by information about data retention periods, the right to lodge a complaint to the supervisory authority, and – what is especially of relevance in the cloud context – information about the data protection level of a third country or international organisation to which the data controller intends to transfer data.

The compromise text of the GPDR includes a new Article 13a requiring that data controllers use *standardised information policies* for providing the data subject with the following particulars *before* providing information pursuant to Article 14: whether personal data are collected or retained beyond the minimum necessary for each specific purpose of the processing, whether personal data are processed for purposes other than the purposes for which they were collected, whether personal data are disseminated to commercial third parties or are sold or rented out, and whether personal data are retained in encrypted form [11].

Recently, the Art. 29 Working Party discussed in its Opinion 5/2012 on Cloud Computing [7] a lack of transparency in regard to the cloud services' processing operations. Privacy threats may arise from the controller not knowing or not informing the data subjects about the:

- Chain processing that involves multiple processors and subcontractors;
- Data being processed in different geographic locations within the European Economic Area (EEA);
- Data being transferred to third countries outside the EEA;
- Disclosure requests by law enforcement.

The last of these four threats is also important for the reason that, even if data are processed at a services side located in the EEA, data transfers to the United States of America (US) may take place on request by US American law enforcement services.

Furthermore, increased transparency over the chain of data processors and subcontractors is important as, in practice, the roles of data controllers or processors cannot always be clearly assigned to entities. The Art. 29 Working Party, in its "opinion 1/2010 on the concepts of 'controller' and 'processor'", argues that these roles should therefore be determined by "factual elements and circumstances" [6]. The proposed EU data protection regulation also recognises that data processors may, under certain circumstances, have increased control over the data processing and should be made directly accountable to the data subjects (cf. Art. 24, 26 IV).

As discussed above, our stakeholder workshop revealed another transparency problem, namely that data subjects are often not well informed about the applicable

consumer laws and rights, especially if cloud brokers or mediators[6] are involved in cross-border eCommerce transactions.

Hence, in a cloud setting, it may be argued that more policy information beyond the minimum that is required by Art. 10 of the Data Protection Directive should be provided to the data subjects, including:

- Contacts and obligations of all data processors along the cloud chain (as far as data processors can be determined ex ante);
- Geographic locations of all data centres along the cloud chain and, in the event that they are located outside the EEA, information about their data protection levels;
- How disclosure requests by law enforcement agencies are handled; and
- Consumer rights and applicable laws.

It will remain a challenge, however, how to inform users of these aspects both unobtrusively and, at the same time, in a way that they can really understand and are conscious of these aspects.

Right of Access (Ex Post Transparency) and Other Data Subject Rights

The EU Data Protection Directive provides data subjects with the right of access to their data. This comprises the right to information about the data being processed, data processing purposes, data recipients or categories of recipients, as well as information about the logic involved on any automatic processing (Art. 12 (a)). This data subject right, which provides ex post transparency, is also a prerequisite for exercising the data subject rights to correct, delete or block data that are not processed in compliance with the Directive (Art. 12 (b)).

The proposed EU Data Protection Regulation, with its Art. 15, extends the information to be provided by the controller to include also information about the data retention period, the right to lodge a complaint with the supervisory authority, and "*the significance and envisaged consequences*" of the data processing at least in the cases of profiling. The data subjects shall also have the right to obtain this information electronically if they have made their requests in an electronic format. Besides, the compromise text even states that "*where possible, the data controller may provide remote access to a secure system which would provide the data subject with direct access to their personal data*" [11]. In addition, the proposed GDPR extends the data subjects' rights by the right to be forgotten (Art. 17 - which is however replaced by a so-called right to erasure in the compromise amendment to GDPR (see [11])) and the right to data portability (Art. 18) and introduces the obligation of data breach notification of the controller to the supervisory authority (Art. 31) and data subject (Art. 32).

Furthermore specific ex post transparency rights are, for instance, provided by the Swedish Data Patient Act [28] to data subjects by requiring that health care providers have to inform patients upon request about who has accessed their medical information.

---

[6] A cloud broker or mediator is a third-party that acts as an intermediary between the customer of a cloud service and the seller of this service. It may for instance help to negotiate contracts with cloud providers on behalf of the customers.

3 HCI for Policy Display for Ex Ante TETs

Ex ante TETs include policy tools and languages, such as P3P [29] and the PrimeLife Policy Language PPL [25], which can help to make the core information of privacy policies and information on how far a service side's policy complies with a user's privacy preferences more transparent to an end user at the time that he or she is requested to consent to any form of data disclosure.

As pointed out in [22], legal privacy principles for transparency, consent and data subjects' rights "*have HCI implications as they describe mental processes and behaviour that the data subjects must experience in order for a service to adhere to these principles*". In particular, the principle of transparency requires that data subjects are aware of and comprehend all privacy policy information. Complex privacy notices are, however, usually neither read nor easily understood. This can be due to the limited cognitive capacity that people usually have, such as limited attention spans memory, as well as a restricted ability to process a large amount of complex information at any one time [22]. Hence, suitable HCI concepts have to be chosen to make the policies displayed by ex ante TETs easily noticeable and understandable.

3.1 Multi-layered Policy Notices

Comprehension of policy information can also be facilitated by a multi-layered structure of policy notices, as it was recommended by the Art. 29 Data Protection Working Party in its opinion on "More Harmonised Information Provisions" [5]. This recommendation takes the approach of structuring complex policies into different layers, where the top layer only provides a short privacy notice with the policy information that is at least required by Art. 10 EU Data Protection Directive (i.e., at least the identity of the controller and data processing purposes). Further detailed policy information can be obtained from the condensed and full privacy notices in other (lower or later) layers. Each layer should offer to the data subjects the information needed to understand the position and make decisions. Examples of user interface (UI) designs for short privacy notices of multi-layered policies are design proposals based on a privacy nutrition label metaphor [17] or PPL (PrimeLife Policy Language) UIs that were elaborated for more complex PPL policy presentations [3].

However, if data are processed in the cloud, it may be argued that more policy information beyond that which is required by Art. 10 should, depending on the circumstances, be displayed to the data subjects to provide transparency. Such information listed in 2.2, may also have to be displayed on the top layer in order to enable users to comprehend the implications of the specific policy.

3.2 Policy Icons

Furthermore, user interfaces, which use real-world metaphors, e.g. in the form of suitable icons, are easier to learn and understand (following Jakob Nielsen's usability heuristics of a "match between system and the real world" [21]). Privacy policy icons have been researched and developed for visualising policy elements in privacy

policies with the objective of making the content of legal policy statements easier to access and comprehend. Policy icons should preferably be standardised in the future and be usable across different cultures.

Within the scope of the PrimeLife EU project, a set of policy icons addressing the legal transparency requirements of the EU Data Protection Directive was developed. These icons can be used to illustrate core privacy policy statements in short privacy notices, namely statements about what types of data are collected/processed, for what purposes, and at what processing steps [15]. An intercultural comparison test of the policy icons conducted at Karlstad University with Swedish and Chinese students as test participants gave insights into which icons seem to be well understood by students of both cultures and which icons were understood differently by persons with different cultural backgrounds [12]. Those icons that were easily understood by both Swedish and Chinese students were, for instance, the ones shown in Fig. 1, displaying types of data (personal data, medical data, payment data), the purpose "shipping" and the processing steps (storage, retention).

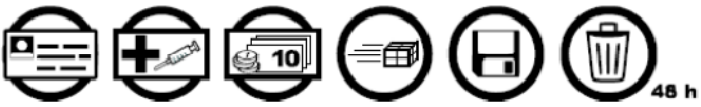

Fig. 1. Example of well understood PrimeLife policy icons

Other Creative Common-like privacy icons have been initiated by Aza Raskin in 2010 [27] and further developed by a Mozilla-led working group (which, however, stopped its work more than a year before this chapter was written). Interestingly, it includes special icons informing end users about how service providers are handling requests made by law enforcement (see Fig. 2 for examples of the alpha release of icons). As already pointed out, and as it also became apparent after the revelation of the existence of the PRISM program in the summer of 2013, whether and under which conditions data are given to law enforcement is an important aspect that is often not transparent to cloud users.

Fig. 2. Icon proposals (alpha version) by Aza Raskin on the handling of disclosure requests by law enforcement [27]

To meet the demand for higher transparency for data processing in the cloud, further policy icons could be helpful to inform end users about the geographic locations of all data centres along the cloud chain, and in particular whether they are located

inside the EEA. In the event that they are located outside the EEA, icons should also include information about their data protection levels.

The compromise amendment of the EU GDPR presents in an annex graphical policy icons to be used by standardised policies in yes/no icon-based tables along with textual descriptions for informing data subjects about policy particulars pursuant to the new Art.13a. While the approach of having standardised policy icons can facilitate an easier recognition and comparison of policy aspects, the icons of the compromise amendment, which were initially suggested and developed by the vice president of the European Parliament, Alexander Alvaro [1], do not seem to be very intuitive in their meaning, and they should definitely undergo further improvements and HCI testing.

4 Ex Post TETs

In this research, we have looked at the architecture and user interfaces of ex ante TETs. An example of such a tool has been named the Data Track. Its description and design is presented in the following sections, along with evaluations of an implemented prototype.

4.1 Data Track

The Data Track is a user side ex post transparency tool, which includes both a history function and online access functions. For each transaction, the history function stores in a secure manner, in which a user discloses personal data to a service, a record for the user on which personal data were disclosed to whom (i.e. the identity of the controller), for which purposes and under which agreed-on privacy policy. The Data Track's user interface version developed under the PrimeLife EU FP7 project provided search functions, which allow users to obtain easily an overview about who has received what data about them, as well as online access functions which allow end users to exercise their rights to access their data at the remote services' sides online and to correct or delete their data (as far as this is permitted by the services' sides). In this way, users can compare what data have been disclosed by them to a services' side and what data are still stored by the services' side, or what data have been implicitly been added (e.g., trust ratings of customers added by an eCommerce company) to the data records stored at the services' side. This allows users to check whether data have been changed, processed, added or deleted (and whether this was in accordance with the agreed-on privacy policy).

Online access is granted to a user if he or she can prove knowledge of a unique transaction ID (currently implemented as a 16-byte random number), which is shared between the user (stored in his or her Data Track) and the services' side for each transaction of personal data disclosure. In principle this also allows anonymous or pseudonymous users to access their data in the services' side.

Furthermore, a user function allowing users to excise the rights of data portability and the right to be forgotten/right to erasure (as proposed by the GDPD and to address precisely the issues pointed out in the context of the stakeholder workshop held by the research team) are developed.

4.2 Graphical User Interface for the Data Track

Complete descriptions of the Data Track proof-of-concept and user interfaces developed under the PrimeLife project can be found in Wästlund & Fischer-Hübner [30]. Usability tests of early design iterations of the PrimeLife's Data Track revealed that many test users had 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.

Therefore, in the A4Cloud project, we have tested alternative HCI concepts consisting of graphical UI illustrations of where data are stored and to which entities data have been distributed. Graphical illustrations of data storage and data flows have a potential to display data traces more naturally, like in real world networks, as discussed in the PRIME deliverable D06.1.f, Section 5.8.1 [17]. Besides, previous research studies suggest that network-like visualisations provide a simple way to understand the meaning behind some types of data ([13], [8] and other recent studies claim that users appreciate graphical representations of their personal data flows in forms of links and nodes [16], [18]).

Therefore, a new UI concept for visualising the users' information in the Data Track tool has been proposed and prototyped at Karlstad University, as shown in Fig. 3. This way of illustrating the tracking of the users' data has been called the "trace view", presenting an overview of which data (with data attributes) have been sent to service providers, as well as which service providers might have the users' data.

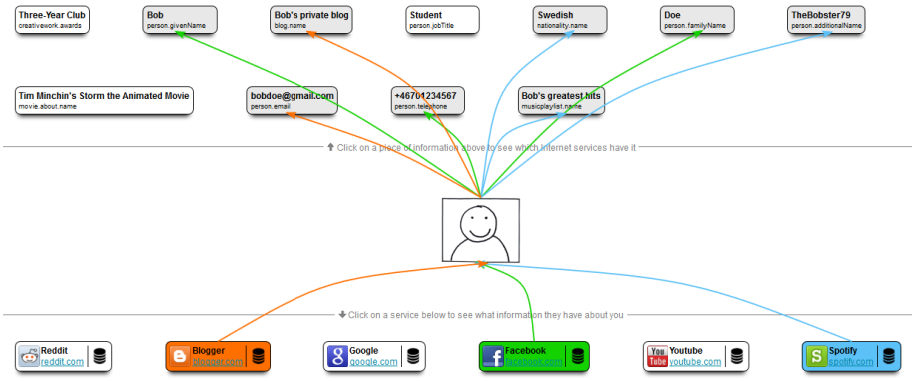

Fig. 3. The trace view user interface of the Data Track

The idea is that users should be able to see all the personal data items stored in the Data Track (displayed in the top of the UI) that they have submitted to services on the Internet (these Internet services are shown in the bottom panel of the interface). If users click on one or many of the Internet services, they will be shown arrows pointing to the information that those services have about them. In other words, they can see a *trace* of the data that the various services have about them. Similarly, if they select one or many data items (at the top of the figure), they will be shown arrows pointing to the Internet services that possess those data items.

Users can also access the data about them stored on the services' side by clicking on the corresponding icons, and are able to correct it or remove it, given that the respective service allows this.

4.3 Usability Evaluation

Usability evaluations of the Data Track's trace view have been carried out in order to test the extent to which users comprehend the ex post transparency features provided by the Data Track. An interactive mock-up of this tool was implemented providing a front-end for users to manipulate its different elements.

A total of 14 participants, aged between 19 and 40 years old, were recruited from different parts of the city of Karlstad in Sweden. They were asked for around 20 minutes of their time to evaluate the graphical interface of a computer program. From the total of 14 participants, 12 indicated that they were "experienced" or "very experienced" with computers; seven were employed at a company, six were students at the university, and one did not specify his occupation. As an introduction to the test, participants were told that they were about to evaluate a program that would let them see a history of the information that they had given to different online companies. They were also told that this program would let them verify that the information that they had released was the same as the information that was stored at the service, and that they could request a service to correct or delete their data if that service allowed it.

To start the test, participants were asked to pretend that they had already disclosed some information to other online companies and that, on this occasion, they were going to purchase a book. In order to complete the transaction, participants were asked to enter their personal information and submit it to this unknown online bookshop. However, they were given a fictitious credit card number to be used to complete the purchase and none of the personal information submitted was actually stored.

A cognitive walkthrough approach was adapted, in which participants were given a series of tasks to complete using the Data Track's user interface. A test moderator notated the answers and comments as the participants carried out the tasks. The order in which the tasks were presented was randomised in order to minimise possible biases. After completing the tasks, participants were asked to complete a post-questionnaire with the intention of capturing their subjective opinions on the interface.

Results revealed that all the participants clearly understood that elements in the bottom panel (cf. Fig. 3) represented different online services to which they had sent information, and the majority of participants (11 out of the total) clearly understood that the elements in the top panel of the interface represented their own information that was sent to online services. Also, it was intuitive for all participants to find out what data items they had sent to a particular service provider (by clicking on one of the services on the bottom panel). All the participants but one found it easy to discover which services had a particular data attribute.

On the one hand, these positive initial observations indicate that participants found the tracing feature of the interface easy to understand, intuitive and informative. On the other hand, participants had a harder time understanding that they could also access the data stored about them on the service's side (which was also a challenge in

earlier versions of the Data Track). The reasons for this might be due to the lack of users having mental models of transparency and control features on the services' side, or to the poor affordance and visibility properties of the UI elements that were supposed to allow users to access the services' side data. About half of the participants (eight out of the total), understood that the attributes displayed on the top panel were data that were under their control.

From the results of the post-questionnaires conducted, it is interesting to note that, when asked to rate "how much would you *trust* the Data Track program with the information you give to Internet companies" on a scale from 1 (would not trust at all) to 10 (would trust completely), 30.8% of participants gave a rating of 4, while 61.6% gave a rating of 6 to 8 (just one participant, i.e., 7.7%, gave a rating of 5, and one participant did not answer the question). Moreover, when asked how often they believed that they would use a transparency interface like the Data Track, 11 of the 14 participants indicated that they would use such a program a few times per month or more often (one participant suggested that she/he would use it almost never or never). Similarly, 12 participants responded that they would have the Data Track program turned on so as to track their Internet data releases 75% of the time or more, indicating the desire for this set of users at least to have such transparency tools.

Most test users found the graphical Data Track intuitive and useful. The local Data Track view was well understood by a majority of users. However, further improvements are needed to make users aware and to enable them to understand the control options allowing them to exercise their rights at a service provider side online.

While the current Data Track only allows the tracking of data disclosures to a primary services side, the future Data Track (when combined with transparency logging – see below) will also address more cloud-related scenarios, where users disclose many more data items to many different service providers, who may in turn forward the data to chains of cloud service providers. This consideration forces some redesign of the Data Track user interface, where users should be able to navigate through various elements without the interface being cluttered. **Fig. 4** shows an example illustration of how a more realistic scenario for the Data Track could look, depicting the flow of users' data through the chain of cloud providers.

4.4 Combining the Data Track with Privacy-Preserving Transparency Logging

While the Data Track provides an overview of data disclosures, and the ability to remotely access their data, one key missing aspect is a detailed record of *how* personal data have been processed. A privacy policy provides a description of intended data processor *before* a user discloses data. As data are being processed, *after* data disclosure, the data processor should log a detailed record of how it has processed personal data that are made available to data subjects: This is the goal of transparency logging.

Conceptually, with a detailed record of data processing available concerning personal data, a data subject could *verify* that the *actual* processing on personal data was in line with the processing for which the data subject gave consent to prior to

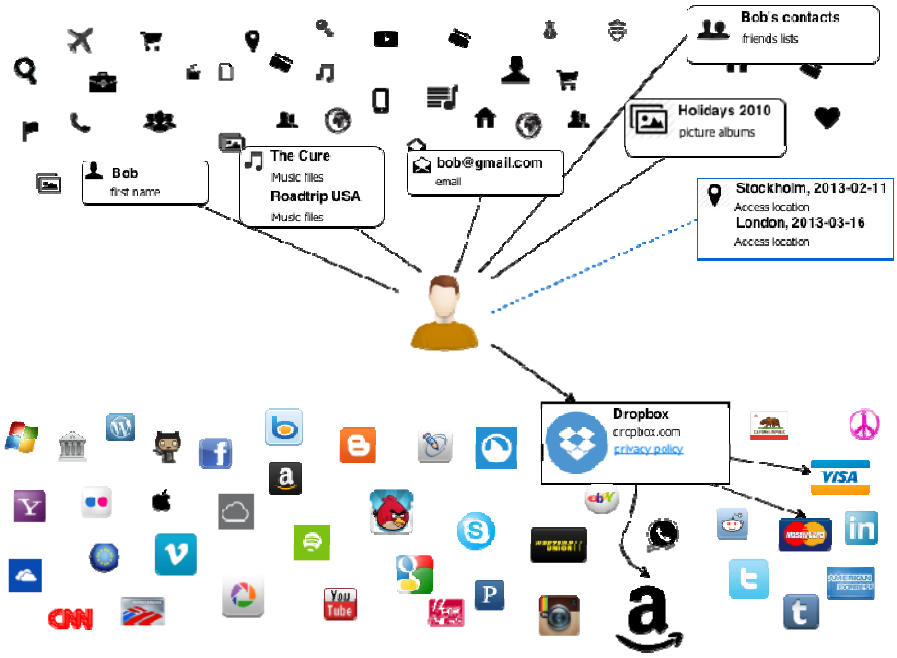

Fig. 4. Mock-up for a graphical Data Track for data a cloud

disclosing data, as stated in the privacy policy presented by the data processor. The detailed record of how personal data have been processed can be seen as the *provenance* of the personal data.

A transparency log contains a detailed record of how personal data have been processed, so the records are also personal data. For example, the records may reveal which doctor has read a patient's medical records in a hospital setting, or which type of car insurance a driver is qualified for based on previous accident history. This means that there is a need for preserving privacy when performing transparency logging. Some key considerations are:

1. Nobody should be able to make undetectable changes to recorded data.
2. Only the intended recipient user of recorded data (i.e., the data subject to whom the log entry refers) should be able to read the data.
3. It should be impossible to correlate data and users.

The first consideration ensures that, once a data processor has recorded data in the transparency log, no changes can be made, be it by the data processor who wishes to hide some processing or a malicious third-party. The second consideration captures the need for secrecy; only the data subject should be able to read this newly created personal data created to make processing of other personal data transparent. Last, but not least, the third consideration minimises the creation of new personal data in the form of metadata. For example, if it was possible to correlate the amount of encrypted data stored for a particular user especially over time, then such information might leak

everything from how often the data subject uses the processor's service to particular details about the user's personal data (how data change over time might serve as a signature of a particular event). In [26], the authors describe a cryptographic system for performing transparency logging for distributed systems[7] (e.g., a cloud-based system that provides these properties).

5 Related Work

To the best of our knowledge there is not much previous work on TETs for the cloud that have been designed to be both usable and privacy-preserving. Related work on usable policies, such as [16], [15], are not focused on the cloud context. Related data tracking and control tools for end users are, in contrast to the Data Track, usually restricted to specific applications and cannot be used directly to track data along cloud chains. One example of such related work is the Google Dashboard, which grants its users access to a summary of the data stored in a Google account, including account data and the users' search query history. In contrast to the Data Track, the Dashboard provides access only to authenticated (non-anonymous) users. Related to the Data Track are personal data vaults, such as [19], developed for participatory sensing applications. This includes a logging functionality which allows the display of transactions and transformations of users' data and enables users to track who precisely has accessed their data.

6 Concluding Remarks

Further work is needed to develop and enhance transparency-enhancing tools for the cloud that are privacy preserving and usable. Parts of this team's future work will focus on extending the Data Track to making data processing along the cloud chain more transparent. It will increase the usability of the Data Track's control functions, thereby allowing users to exercise their data subject rights including the right to data portability.

Additional relevant research questions that we would like to address include the following: How can policy interfaces better inform users about the consequences of data disclosures in an unbiased and unobtrusive fashion? How can privacy-preserving ex ante TETs be technically designed to allow users to track who has accessed their data, what logic has been involved in processing their data and what are the consequences of this, while not leaking trade secrets in regard to the data controller's business processes (cf. problem discussed in recital 51 GDPR)?

---

[7] As described above, this system targets at protecting the privacy of individuals whose data are processed. Protecting the privacy of employees processing personal data, whose activities are logged and who can thus be monitored at their working place, is another problem that is outside the scope of this solution. Also, the problem that there may be data referring to more than one data subject is not addressed yet.

Acknowledgements. Parts of this work has been conducted for the EU FP7 project A4Cloud, which received funding from the Seventh Framework Programme for Research of the European Community under grant agreement no. 317550. We especially want to thank our A4Cloud project colleagues, John Sören Pettersson, Erik Wästlund, Eleni Kosta, Maartje Niezen and Karin Bernsmed for cooperation and feedback on the related A4Cloud D:C-7.1 deliverable. Besides, we are very grateful to Marit Hansen and Diane Whitehouse who provided very helpful review comments for this paper to us. The development of the graphical user interfaces for the Data Track was funded by the Google Research Award Project on "Usable Privacy and Transparency II". We want to thank our contact partners at Google for their kind support and fruitful discussions.

References

1. Alvaro, A.: Life Cycle Data Protection Management – Ein Beitrag zur Anpassung der europäischen Datenschutzgesetzgebung an die Erfordernisse des 21. Jahrhunderts (January 30, 2013), http://www.alexander-alvaro.de/inhalte/lifecycle-data-protection-management-ein-beitrag-zur-anpassung-der-europaischen-datenschutzgesetzgebung-an-die-erfordernisse-des-21-jahrhunderts/
2. Andersson, C., Camenisch, J., Crane, S., Fischer-Hübner, S., Leenes, R., Pearson, S., Pettersson, J.S., Sommer, D.: Trust in PRIME. In: Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology. IEEE Xplore (2005)
3. Angulo, J., Fischer-Hübner, S., Wästlund, E., Pulls, T.: Towards usable privacy policy display and management. Information Management & Computer Security 20(1), 4–17 (2012)
4. Angulo, J., Fischer-Hübner, S., Pettersson, J.S.: General HCI principles and guidelines for accountability and transparency in the cloud. A4Cloud Deliverable D:C-7.1. A4Cloud Project (September 2013)
5. Art. 29 Data Protection Working Party (2004). Opinion 10/2004 on More Harmonised Information Provisions. European Commission (November 25, 2004)
6. Art. 29 Data Protection Working Party. Opinion 1/2010 on the concepts of "controller" and "processor". European Commission (February 16, 2010)
7. Art. 29 Data Protection Working Party. Opinion 5/2012 on Cloud Computing. European Commission (July 1, 2012)
8. Becker, R.A., Eick, S.G., Wilks, A.R.: Visualizing network data. IEEE Transactions on Visualization and Computer Graphics 1(1), 16–28 (1995)
9. European Commission. 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 (November 23, 1995)
10. European Commission. Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). COM (2012) 11 Final. Brussels (January 25, 2012)
11. European Commission. 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) (COM(2012)0011 – C7 0025/2012 – 2012/0011(COD)). Compromise amendments on Articles 1-29 (Passed October 21, 2013), http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_01-29/comp_am_art_01-29en.pdf

12. Fischer-Hübner, S., Zwingelberg, H.: UI Prototypes: Policy administration and presentation - Version 2. PrimeLife Project Deliverable D.4.3.2 (2010), http://primelife.ercim.eu/
13. Freeman, L.C.: Visualizing social networks. Journal of Social Structure 1(1), 4 (2000)
14. Hildebrandt, M.: Behavioural biometric profiling and transparency enhancing tools. FIDIS Deliverable, D7.12. FIDIS EU project (2009), http://www.fidis.net/
15. Holtz, L., Nocun, K., Hansen, M.: Displaying privacy information with icons. In: PrimeLife/IFIP Summer School 2010 Proceedings, Helsingborg, August 2-6 2010. Springer (2011)
16. Kani-Zabihi, E., Helmhout, M., Coles-Kemp, L.: Increasing Service Users' Privacy Awareness by Introducing On-line Interactive Privacy Features. In: IAAC Symposium 2011 (2012) (Online)
17. Kelley, P.G., Bresee, J., Cranor, L.F., Reeder, R.W.: A "Nutrition Label" for Privacy. In: Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS 2009). ACM, Mountain View (2009)
18. Kolter, J., Netter, M., Pernul, G.: Visualizing past personal data disclosures. In: ARES 2010 International Conference on Availability, Reliability, and Security. IEEE (2010)
19. Lacohée, H., Crane, S., Phippen, A.: Trustguide: Final Report (2006)
20. Maguire, M., Bevan, N.: User requirements analysis. In: Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M., Govindan, R. (eds.) Personal Data Vaults: a Locus of Control for Personal Data Streams, CoNEXT 2010, vol. 17. ACM Digital Library (2002)
21. Nielsen, J.: Usability inspection methods. In: Conference Companion on Human Factors in Computing Systems. ACM (1995)
22. Patrick, A.S., Kenny, S.: From privacy legislation to interface design: Implementing information privacy in human-computer interactions. In: Dingledine, R. (ed.) PET 2003. LNCS, vol. 2760, pp. 107–124. Springer, Heidelberg (2003)
23. Pearson, S., Tountopoulos, V., Catteddu, D., Sudholt, M., Molva, R., Reich, C., Fischer-Hübner, S., Millard, C., Lotz, V., Jaatun, M.G.: Accountability for cloud and other future Internet services. In: IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom). IEEE (2012)
24. Pettersson, J.S.: HCI Guidelines. PRIME Deliverable D06.1.f. Final Version. PRIME project (2008), https://www.prime-project.eu/
25. PrimeLife, Privacy and Identity Management in Europe for Life - Policy Languages, http://primelife.ercim.eu/results/primer/133-policy-languages
26. Pulls, T., Peeters, R., Wouters, K.: Distributed Privacy-Preserving Transparency Logging. In: Workshop on Privacy in the Electronic Society. ACM (2013)
27. Raskin, A.: Privacy Icons: Alpha Release (2010)
28. Svensk Författningssamling Riksdagen. Patientdatalag 355 (2008)
29. W3C, P3P – The Platform for Privacy Preferences 1.1 (P3P1.1) Specification, W3C Working Group Note (November 13, 2006), http://www.w3.org/P3P/
30. Wästlund, E., Fischer-Hübner, S.: End User Transparency Tools: UI Prototypes. PrimeLife Deliverable D.4.2.2; PrimeLife project (2010), http://primelife.ercim.eu/