A Survey on Transparency Tools for Enhancing Privacy*

Hans Hedbom

Dept. of Computer Science Karlstad University Karlstad, Sweden
Hans.Hedbom@kau.se

Abstract. This paper provides a short survey on transparency tools for privacy purposes. It defines the term transparency tools, argues why they are important and gives examples for transparency tools. A classification of transparency tools is suggested and some example tools are analyzed with the help of the classification.

1 Introduction

At our department we are involved in EU research projects (among them FIDIS [9], PRIME [6] and PrimeLife [12]) aiming at understanding the consequences to privacy for a user[1] in a networked world and at constructing concepts and tools that can help a user to regain control over her personal sphere. One goal of these projects is to increase the possibilities that a person has to know what really happens with her personal data, i.e. what data about her are collected and how they are further processed, by whom, and for what purposes. This is important in order to judge if the data are processed in a legal manner and whether they are correct. The concept usually used to describe these properties is the notion of transparency. Consequently, one of our goals is to develop tools and concepts for increased transparency. As a first step to reach this goal and to get an idea of what has been done in the area and the current state of the art, a short survey of transparency tools for privacy purposes has been conducted. In this process we have also tried to find a way of categorizing these tools. Even though we realize the great importance legal, social and economical tools, frameworks and

* Part of the research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n 216483. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The above referenced consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

[1] We imply that "user" and "end user" throughout this paper are also data subjects in the system.

V. Matyáš et al. (Eds.): The Future of Identity, IFIP AICT 298, pp. 67–82, 2009.

sanctions play in the transparency area[2], the focus of the survey has been on technical tools. This paper describes the results of the survey.

2 Why Transparency Tools?

In todays Internet environment, information on individuals tend to get collected and revealed to a number of different actors. The distributed nature of the World Wide Web and services like e-shopping, e-health, on line community services and e-government makes it hard for a user to keep track on where information about her is stored, to whom it is handed out and for what purposes it is used. This situation will be even worse with the advent of so called intelligent environments or AmI environments which are highly distributed networks of sensors and computers gathering information on their environments and possibly trying to adopt the environments to a users preferences. Some authors have argued [16,13] that in these environment the traditional privacy paradigm of concealment (i.e. controlling the access to (or even the existence) and distribution of personal data) does no longer hold or is impossible to maintain. Instead they claim that the main focus must be on controlling the proper use of the data. In order to do this a user must be able to get information on how her personal data is used and possibly from which sources it originated. To achieve this type of control transparency tools play an important role. Transparency is a legal privacy principle, which also can be derived from the EU Data Protection Directive 95/46/EC [10]. When a data controller is requesting personal data from a data subject, the data controller must inform the data subject about her identity (name and address), the purposes of data processing, the recipients of the data and all other information required to ensure the processing is fair ([10] Art. 10) The data subject has the right to access all data processed about her, to demand the rectification, deletion or blocking of data that is incorrect or is not being processed in compliance with the data protection rules ([10]Art. 12). The users right to access also includes the right to obtain knowledge of the logic involved in any automatic processing of data concerning her. Even though there is no legal requirement that users can exercise their rights on line, we believe that such a state of affairs would be beneficial for all parts involved and could also make the process more administratively efficient.

3 The Scope of the Survey

In order to define the scope of the survey and to understand what we are examining we need to define what we mean by a transparency tool for privacy purposes. First of all, transparency as such can be required for more than privacy purposes e.g. different types of audit and control to make sure that company finances are in order or that procedures and processes are managed and used in an appropriate

[2] Even though there exists technical tools for transparency we believe many of them require additional legal tools or technologies such as reputation systems and black lists in order to be fully effective. This is because there is limited use in getting the information if the person involved cannot act against the service if the promises are broken or her personal data is misused in some way.

manner and of course there exists tools to aid in those cases. In this survey we have limited ourselves to consider tools that have the objective to help the user to enhance her privacy. Thus, we focus on transparency tools for enhancing privacy. So, what is a transparency tool for privacy purposes then? FIDIS [9] has in its deliverable D7.12 [5] defined a concept called Transparency Enhancing Technologies (TETs). Their provisional definition of TETS is literally [5]:

"Type A: legal and technological instruments that provide (a right of) access to data processing, implying a transfer of knowledge from data controller to data subjects, and/or

Type B: legal and technological instruments that (provide a right to) counter profile the smart environment to 'guess' how one's data match relevant group profiles that may affect one's risks and opportunities, implying that the observable and machine readable behaviour of one's environment provides enough information to anticipate the implications of one's behaviour."

However, their vision on TETs is for tools that make it possible for individuals to assess how profiles will be used on them and to be able to judge how different actions will influence the outcome of this profiling. In our view this definition is too narrow considering the implications of the word transparency. Further, since we do not consider legal tools the definition is too wide in that sense.

In [7], Hansen defines transparency tools as follows: "When dealing with personal data and privacy, transparency tools are tools which can provide to the individual concerned clear visibility of aspects relevant to these data and the individuals privacy." This definition is, we believe, too narrow since it only takes into account the end user and not entities that act on behalf of user's or in the interest of the user to increase the user's privacy such as data protection officers.

Based on the definitions above and on the classification on privacy mechanisms given in [13] we would like to give the following definition on transparency tools for privacy purposes (please note that by a proxy acting on behalf of the user we also include organizations authorized by other entities than the user to protect the privacy interests of the user) : A transparency tool for privacy purposes is a technological tool that has one or more of the following characteristics:

- gives information on intended collection, storage and/or data processing to the data subject, or a proxy acting on the behalf of the data subject, in order to enhance the data subject's privacy;.
- provides the data subject, or a proxy acting on the behalf of the data subject, with access to stored data and/or to logic of data processing in order to enhance the data subject's privacy;.
- provides counter profiling capabilities for a data subject, or a proxy acting on behalf of the data subject, in order to 'guess' how her data match relevant group profiles that may affect her risks and opportunities, implying that the observable and machine readable behavior of her environment provides enough information to anticipate the implications of her behavior.

To lessen the scope further we have excluded technologies that we deem as enabler technologies such as policy languages, obligation management and transfer protocols from the survey. Thus, technologies like P3P [15] and EPAL [4] are not

considered in the survey even if they would be considered as a tool by themselves. However, tools that use these technologies and make them more accessible for the user are included.

4 The Privacy Risks of Transparency Tools

Transparency tools as such cannot only help the user to increase her privacy but could also if improperly designed actually be a severe privacy threat to the user. The reason for this is that since some of them give a lot of information to the user about her personal data and in some cases limited control over this data people masquerading as the user also will get this information and this control. Some of the transparency tools that provide information on how data has been processed, such as TAMI, also require the services side to keep extensive logs on user data and processing, which as such might be privacy-sensitive. Thus, systems that use these types of transparency tools need to have good access control mechanisms and routines to not turn the tool into a privacy threat. This in turn implies that there must be mechanisms in place to guarantee that data is only handed out and controlled by either the user concerned or somebody authorized by the data controller. In a networked environment with a lot of users this can be a complex and costly system to implement and manage.

5 An Attempt for a Classification

For designing privacy enhanced systems it is helpful to have a classification that can be used in order to compare different tools or choose the right tool for the system. Since we want to compare and analyze the tools in this survey we need some characteristics as parameters that can help us in this process. Because of this we will, in this section, describe a brief classification of transparency tools based on a number of characteristics that we believe are important to take into consideration. Please note that since we are in reality talking about two types of activities, i.e. data storage and processing, some tools might fall under different category regarding the data stored and regarding access to processes.

5.1 Possibilities of Control and Verification

One of the more interesting aspects of a transparency tool is how much control and verification on the process of gathering and processing personal information is given to the user. This gives an indication on how much the user/proxy can learn about the actual processing and also get a view on what is really stored about her. Roughly this parameter is divided into three categories.

1. **Promises:** In this case the user gets information on what the data controller promises to do or not to do with the data in the future. This category encompasses the tools that will present, or in other ways give access to, the privacy policy or other types of commitments from the gathering side in a more or less user-friendly manner, but give no on line or automatic way for the user/proxy to verify these claims.

2. **Read only:** In this case the user or her proxy can get access to information on what processing the data actually has gone through up to a specific point in time and/or to the stored personal data itself in a read only manner. This category could be combined with the "Promises" category either in the tool itself or by using another tool to retrieve/store the privacy policy to be fully effective. This is because we believe that the privacy policy negotiated with the data controller (in combination with applicable laws) is needed in order to make a sound judgment on whether a privacy violation has occurred or not.
3. **Interactive:** In the interactive category the tools, in addition to the properties in 2), have the ability to let the user or her proxy actively influence the stored data and/or the processing of the data in some way according to legal requirements or agreed on policies. This category could also be subdivided into "Fully Interactive" and "Partly Interactive" depending on whether the user can use the tool to manipulate all stored data and/or processing or just parts of it.

5.2 Target Audiences

Transparency tools can also be categorized according to their expected audience (i.e the users of these tools). In essence these could be divided into professionals and non-professionals (i.e. people that professionally do audits for privacy protection and the data subjects whose personal data is processed and stored). In the following we will call these categories Auditors/Proxies respective Data Subjects. Of course users and professionals come "in many shapes and sizes" and could probably be divided into further levels, e.g. beginners, intermediate, experts and so on and tools made for Auditors/Proxies could certainly be used by Data Subjects and vice versa. However, in this classification we will concentrate on the high level differences that one might expect between tools for these two target audiences and the properties that one would expect to find in a tool for the specific audience.

1. **Tools for Data Subjects:** Tools for data subjects are expected to have a high level of "user friendliness". In the transparency case this will generally mean that the information is presented in an easy to understand manner and that the privacy implications of different choices and actions are explained so that the data subject understands what she is doing and what consequences her action will have. In order to achieve this, these tools usually limit the information that is presented by using predefined choices and filters with limited customization and try to find alternative ways (e.g. icons or graphs) of presenting complex information properties. Tools for users are also expected to have a high degree of automatisation when it comes to interpreting policies or finding privacy violations. Finally, one would expect these tools to give advice on how to proceed or who to contact in case of privacy violations or questions.
2. **Tools for Auditors/Proxies:** Tools targeted towards Auditors/Proxies do not necessarily produce output that is presented or explained in a way that is supposed to be read or understood by non-professionals. One would also

expect these types of tools (especially if they are used for audits) to be transparent in themselves (i.e. to produce their own logs and audit trails) in order to get an understanding in how they have been used and on what data decisions are based on. Finally, these types of tools might give direct access to data or processes that are outside of what a Data Subject would be allowed to access or expected to handle or understand.

5.3 Scope

Another categorisation parameter is the amount of privacy information that the tool can make accessible to the user. From a privacy perspective this will give an indication on what level of transparency the tool offers and what a user can expect to gain by using the tool and from a security perspective it will help to judge the amount of information that could be revealed or compromised if the tool is compromised. We have chosen to call this aspect "the scope" of the tool and divided it into four levels. These levels are constructed from a user view and are based on what answer the tool is expected to give on the hypothetical question "Please give me all info that x has on me and how x has handled that information"

1. **Service Scope:** The tool will give transparency to information stored and processed by a single service.
2. **Organizational Scope:** The tool will give transparency to information stored and processed by a single organization.
3. **Conglomerate Scope:** The tool will give transparency to information stored and processed by a conglomerate of organizations (e.g. multiple governmental offices or big corporations).

5.4 Trust Requirements

Many solutions have requirements in order to achieve trust on the data controller side. With trust in this case we mean the level of assurance that a user can have that the data controller behaves as expected and that she does not try to cheat or deceive the user (e.g. by not following the negotiated privacy policy). These trust requirements can be either directly expressed in the solution or implicitly presented due to assumptions on the operation environment or the technology used. The levels we have chosen for the trust requirements are strictly speaking not a classification but rather consist of a number of high level trust components. This means that a solution can require more than one of the components described below. The ideal situation is when the user does not need to place any trust in the data controller at all in order to protect her privacy, thus the less trust needed the better it is. In the list of components below we do not discuss the technical, legal, social or economical means in order to implement the component since there are a number of ways of solving this. The high level trust components are the following:

1. **Trusted Server:** The server environment used in the solution is assumed to behave in a trusted manner. This generally means that enough mechanisms

to prevent or deter the server from cheating are implemented on the server side. Thus, we have to expect that the server behaves as expected and in a fair manner. Note that this notion also works in a p2p environment since a p2p connection at any point in time and for any transport direction can be divided into a client/ server relationship: In essence this means that both sides are a potential server and thus both fall under the same assumptions that are put on the server side.

2. **Trusted Third Party:** In this case the solution requires that parts of the responsibilities and functions in the solution are taken over by an impartial third party component. This component guarantees that even if one of the parties tries to cheat or violate negotiated policies, one can trust that the solution as a whole will continue to behave in a fair and trusted manner.
3. **Trusted Client:** The client environment used in the solution is assumed to behave in a trusted manner. This generally means that enough mechanisms to assure to a certain level that the client is not compromised or under the control of an attacker are present and that the client does not release data in an uncontrolled manner. Generally, one could infer that a solution that does not itself try to protect the data it uses on the client side is assuming a trusted client environment.
4. **No trust needed:** The solution itself is designed in such a manner that it, in some way, prevents (or makes it exceedingly hard for) the server and the client from cheating or misbehaving. This is achieved without the use of an external trusted third party to guarantee the trustworthiness of the solution.

5.5 Information Presented

In a sense transparency is all about achieving a balance of information. Because of this, it is valuable to know what type of information can be gathered and presented by the tool. Information of interest is not only personal data stored and processed by the data controller and the logic of the processing, but also information about the data controller herself (or rather the service provider or organization she represents). Such information need not necessarily be acquired from the data controller but can be harvested from other information sources. We have chosen to classify the type of information into three categories. Note that these categories are not orthogonal but rather complementary:

1. **Required information:** The tool gathers and presents information that a service provider has to provide according to the Law (in a EU context this would e.g. be national laws based on the EU Data Protection Directive 95/46/EC Art. 10 [10] (type of data processed, identity of the controller, for what purposes,)).
2. **Extended information:** The tool gathers and presents information given or harvested from the service provider that is not legally required but that increases the transparency for the user in a privacy context.
3. **Third party information:** The tool gathers and presents information given or harvested from other sources than the service provider that increases the

transparency for the user in a privacy context. This might e.g. be privacy seals, whether the service provider is blacklisted, reputation systems or security breach reporting systems.

5.6 Other Aspects

There are other aspects that could be interesting when comparing solutions but more from a designer/implementer perspective than from a user/provider perspective. In this section we will mention them briefly.

Technologies Used. The technology used in and by the tool probably plays an important part in both making the tool economically feasible and getting a widespread use. Standard protocols, languages and frameworks tend to mean that less work is needed to integrate the tool in new as well as legacy systems. Concerning technologies used we would like to differ between three high level aspects: Communication, Information retrieval and Infrastructure Requirements. Regarding communication the interesting aspects from our view point are the standards used and since many Internet services today are based on different web standards we would like to suggest a classification into three categories: predominately Web Standards, predominately non-web Standards and predominately proprietary solutions. The question of information retrieval is more or less a question of sophistication of the tool i.e. does it just retrieve the my stored data or can I get more information and will directly reflect in the possibilities of control and verification properties and scope properties that can be achieved by the tool. This aspect is not so much a categorization but rather a list of possible capabilities or technologies used e.g. data mining, transaction logging and direct storage access. Finally the Infrastructure Requirements are usually reflected in the trust requirements of the tool and might result in specialized hardware, software and architectural components being needed for e.g. trusted computing.

Security Requirements. As mentioned before transparency tools might impose security risks. Exactly how severe this risk is or what security requirements are needed is probably hard to judge looking at the tool itself since this also depends on the data being processed and the implementation of the tool. We will not elaborate these issues much further but rather list some aspects that will influence the security requirements of the solution. The list is not meant as an exhaustive list, but rather mentions the more important aspects that, from a transparency perspective (in a privacy context), influence the requirements. Normal server (and client) security practices and tools should be evaluated and used to secure the tool as one would do for any other application.

1. **Sensitivity of data:** The more sensitive the data is the higher the requirements on how they can be handled and who can get access to it. The consequences of a privacy violation can also be considered to be more severe for sensitive data. In many cases it can be hard to judge if the data is sensitive or not since the sensitivity is dependent not only on the data itself but

also on the data processing purposes and the context in which it is processed and stored.
2. **Concentration of data:** The higher the concentration of data, i.e. the larger the amount of identifiable information about an individual the tool has access to, the higher the privacy impacts are if the data is compromised.
3. **Ease of access:** The easier it is to get access to a transparency service and the more well known it is the better it is from a usability perspective. However, one would also expect well known and easy accessible services to be more prone to attacks especially if they contain information of value for potential attackers.

6 Examples of Transparency Tools

In this section we will give an overview of different types of transparency tools that are either available, under development or suggested in research papers. We also elaborate on the differences and commonalities of the example tools and classify them according to the classification given in section 5. Please note that the amount of space given to any specific solution is not meant in any way to reflect the importance of that tool.

6.1 The TAMI Project

TAMI [16] is a project at MIT/CSAIL laboratory aimed at creating a Transparent Accountable Data Mining (TAMI) system. The idea is to use technology present in (or developed in connection with) the Semantic WEB efforts. In connection with this it is part of a bigger project aimed towards making the WEB policy aware. The current descriptions of TAMI is highly geared towards law enforcement agencies and other governmental agencies using data mining to find evidence or other information about persons.

In [16] Weitzner et al. identify three distinct classes of rule violations that could occur in connection with data mining.

Adverse actions premised on factual incorrect antecedents.

Impermissible sharing of data beyond the collecting organization.

Adverse actions premised on interference from data where the data, while factually correct and properly in the possession of the user, is used for an impermissible purpose.

The TAMI system is designed to detect these types of violations and consists of a set of general-purpose interference components:

I. The Inferencing Engine: Used to analyze available data and to assess compliance with relevant rules.

II. The Truth Maintenance System: A persistent store fed by 1 and used to assess reliability of inferred results and to record justifications and proof antecedents.

III. Proof Generator: Used to construct proofs that adverse actions and critical transactions are justified by facts and permissible under applicable rules.

Using these components it is possible to construct an audit trail that can be used to trace the sources of a decision and also see if the data has been used and handled in a correct manner.

The TAMI system is still under development and does in the state described by [16] use XML and RDF in N3 format for data sources and transaction logs and N3 logic to express rules and policies. As far as we know there is no practical implementation of the TAMI system.

Looking at TAMI first we can easily infer that it is not primarily meant to be what we classified as a tool for data subjects but rather is meant as an Auditor/Proxy tool. Since it is based on information mined from different data sources without sending the usage policy to the data subject or inform the data subject on what data is gathered it can be considered as a pure "read only" system regarding the data and a "combination of read only and promises system" regarding the processing of the data since the proof-engine will give information on what processing rules that were used. The scope of the tool is hard to judge since it currently is only a research system and in itself has the potential to fall into any of the scope categories dependent on which sources it takes its data feeds from. Regarding the trust requirement one can derive that it does not really trust its clients (data sources) since it stores where the information comes from and where this source got it from i.e. the origin of the data and based on this it judges the trustworthiness of the data. However, as far as we can judge, the system as a whole requires a trusted server since there are still ample opportunities for the server to cheat regarding policies and data that is feed into the proof engine. Since the tool is aimed primarily as an audit tool one might argue that the auditor might act as a trusted third party and thus prevents the server from cheating. There is also the fact that the tool as it is currently described is meant for law enforcement and one might argue that these types of organizations are assumed to be trusted to play by the rules as default (at least in a democracy). Finally, since the primary purpose of TAMI is to act as an audit-trail for law enforcement and to be used in court we would argue that it presents legally required information and in some sense extended information since it presents the originating sources of the information.

6.2 Privacy Bird

Privacy Bird [1] is a browser plug-in that helps people to decide if the web pages they enter are compliant with their own privacy preferences. At the heart of the plug-in is a P3P policy interpreter and tools for constructing P3P privacy preferences in a user friendly fashion. When installed it will manifest itself as a bird icon in the browser that have different colors depending on how well the web servers P3P policy compares to the users preferences. If the policies match the users preferences the bird will be green, if they do not match, it will be red and if the web server does not have a policy it will be yellow. Different sounds are associated with the different states of the bird and can be used to further enhance the awareness of the user. It is also possible to get more information on the policy of the web server by using menus that turn up when the bird is

clicked. This is information on what in the server policy that did not match the user policy, a summary of the server policy in human readable form, contact information to the web page owner and links to the full privacy policy of the web server.

Regarding privacy bird, one can deduce that it is definitely a "Promises" tool aimed at users. It has limited functionality and does not store transactions or promises and thus it is not usable as a Auditor/Proxy tool. Based on the discussion above, one could also infer that the tool really needs a trusted server if it is to be considered as a transparency tool. Regarding the scope it depends on the policy described but generally we would consider this as having a service scope. The information presented by privacy bird is strictly legally required since it only presents the privacy policy of the service.

6.3 The PRIME Project

PRIME [6] has been a European project that aimed at developing tools and concepts for privacy enhanced identity management systems. Within the project a proof of concept prototype was developed. This PRIME prototype consists of a PRIME-enabled server side that communicates with the PRIME enabled user side components. For PRIME-enabled web applications, a plug-in has been developed that will give access to the different tools developed by PRIME. Among those tools, four are interesting from a transparency perspective: The "Send Personal Data?" Dialog, the concept of PrifPrefs (privacy preferences), the Assurance Control Function (ACF) and the DataTrack. Below we will discuss each of these tools. The ACF has the main purpose of assuring the trustworthiness and integrity of the PRIME server. It performs this duty by using sub components to check whether the service provider is blacklisted or has a privacy seal and to verify the integrity of the hardware and the prime code. Since the tools are currently prototype tools and still further developed within the scope of the PrimeLife project we will describe their intended functionality and not the functionality actually implemented at this point.

The "Send Personal Data?" Dialog is in essence a policy aware automatic form filler that issued to obtain informed consent from the user for the disclosure of her personal data to a services side. The "Send Personal Data?" Dialog is following the approach of multi-layered privacy notices as suggested by the Art.29 Working Party [11]. When data needs to be sent to the server it will pop up and present the privacy policy of the web server and also help the user decide what privacy implications the data will have. The policy is presented to the user on a purpose by purpose manner acting as an interactive form filler wizard. It will start by asking the user which PrifPref she wants to use in this specific case. PrifPrefs are privacy preferences stored at the user side describing basically what data or types of data the user is willing to communicate and for what purposes those data may be collected and used. These privacy preferences can be bound to a web service (recipient), a pseudonym or a combination of these or they could be generally applicable based on a desired level of privacy. There are predefined PrifPrefs for anonymous usage and minimal data release. Based on the PrifPref the "Send

Personal Data?" Dialog will present the information the server wants purpose by purpose indicating if the data asked for and the purpose specified conforms with the stated PrifPref. The user can get more information on why and how the requested data violates her chosen PrifPref and possible consequences if the user decides to send the data anyway. If the actual data to use is stored in the PrifPref it is automatically filled in the form otherwise the user is asked to provide the information. If new information or new purposes are added to the selected PrifPref in this process the user can save this as a new PriPref for later use.

The Data Track is a view handler towards a history data base. The purpose of the tool is to let the user keep track of what data she has disclosed and to whom. The data is basically presented in two different ways. One view is a table with the different receivers of the data, how many times data has been sent out to this receiver and the dates of the different receiver sessions. By double clicking on a row in the table the receiver can get a more detailed view on exactly what data was sent during this session and the privacy policy that was agreed on when the transfer was performed. The other view is based on a card metaphor where the data are presented as a deck of cards that can be browsed through. The cards basically contain the same information as a table row with the addition of three buttons. These buttons are used for communication with the web server that the cards relate to and are used to either interactively (if the server has the ability) or in an offline manner request, the deletion of data, correction of data or access to the data that the server currently has stored about the user. The idea here is to make it easy for the user to exercise her legal rights towards the data controller. When double clicked, the card view will display the same detailed information as mentioned for the table view above. The Data Track also includes search functionality so that the user more easily can find answers to questions such as in what sessions certain information was given or what information a specific receiver has on the user.

The PRIME project tools in their current state of implementation are also to be considered as a "Promises" tool. However, the storage capabilities and the tracking of transactions makes it possible to verify and to some extent prove privacy violations if data or logs are apprehended. By implementing the transparency capabilities planned (service data access and secure logging on the server side) the data track would end up as a "interactive" tool both regarding the processing and storage of data. The PrifPrefs by themselves are just a tool for constructing privacy preferences and cannot be seen as a transparency tool. However, in connection with the "Send Personal Data?" Dialog and the local Data Track database it could be used to inform the user about what the collected data is used for and whether the services side really requests only the minimal amount of data from the user for the purposes of a requested service. As with TAMI the scope of the tool is dependent on the data sources used which in the data track case depends on the search capabilities of the data track and the data stored there. However in its current prototype implementation we would argue that it has a service scope.

Concerning the trust requirement the PRIME solution in its ideal implementation does not require a trusted server for the transparency services. However, the current prototype does require a trusted server and there are other parts of the PRIME solution that require trusted third parties (e.g. identity providers, black list providers and privacy seal granting authorities). Regarding the information presented the tool will present legally required information through the data track and the "Send data dialog" and third party information through the ACF.

6.4 Privacy Evidence

In a couple of articles (e.g [13]) Sackmann et al. discuss an approach based on what they call privacy evidence. The key components in this system is a secure logging facility and an automated privacy audit component to give the user information on how well a system fulfills the promised (or user provided) privacy policy. The general work-flow of the system is the following:

1. The data subject delivers her privacy policy to the system.
2. The data subject interacts with the system in some way and every action of the system is logged to a secure log file.
3. The data subject can inspect the logs with a specific low view tool that will provide the record that belongs to the respective data subject.
4. The log view created by the tool can be sent to an automatic audit facility that compares the log view with the provided privacy policy and construct privacy evidence. These give the user an indication of whether there has been any violation against the policy.

Central to this setup are, besides the policy language, three components: the secure log, the log view and the automated audit facility. The secure log used is a file of encrypted log entries where hash chains are used to make sure that the logs integrity is not tampered with and for key generation to insure forward security. Further some user identification information is used to create the keys for the encrypted entries, so that only entries related to a specific data subject are readable by that data subject (further details are given in [Sackman06]). The log view is constructed by traversing this file entry by entry and it constructs the view based on the identifier of the data subject. Finally, the automated audit is performed by constructing a violation set (i.e the set of rules describing violations of the rules described in the policy). This violation set is then compared with the log view and any match in this comparison process constitutes a violation of a policy rule.

Classifying Privacy Evidence we can first conclude that it is a "read only" tool regarding processing and as far as we can judge a "promises" tool regarding the stored data. This is however dependent on how extensive the logging is and what goes into the log's. Given the right log instructions, it might be a "read only" tool in the stored data area as well. It is also developed as a tool for data subjects. Concerning the scope it is dependent on how it is deployed, but because of the intensive logging needed it is hard to see that it would scale well

to anything but a service scope. Concerning the trust classification the solution in its current state requires both a trusted server and a trusted third party. The information given is also dependent on the implementation but will, dependent on the legal context, be either required or extended.

6.5 The Amazon Book Recommendation Service

It is debatable whether this tool really falls under the scope of the survey. However we have chosen to include it since it is an example of customer influenced profiling.

Zwick et al. [17] discuss the Amazon book suggestion service as an example of a service where the customers can directly influence their user profile. As an Amazon customer it is possible to subscribe to a book recommendation service. This service will recommend different books to you based on your previous purchase. By clicking a link in the recommendation a window will appear. This window tells you which of your previous purchases were used to generate the recommendation. The user can then choose whether she wants to remove any of the "input" purchases from her profile so that it is not used as a base for recommendations any more.

The Book Recommendation Service is a tool for data subjects and a partially interactive tool regarding processing, but gives almost no information regarding stored data. Despite this it will give a very limited insights into the processes or the profiles used. Thus, we believe that the user has minimal capabilities as a customer to influence the result. She will only know that a specific input generated a specific result, but not why and how or even how the different input parameters relates to each other when multiple purchases are used to generate a result. Nevertheless, it is a good start since it makes part of the profile visible to the user. Regarding scope it has a service scope. Given the trust requirements there are implicit trusted server requirements to guarantee fairness since the data subject currently has no choice but to trust that Amazon actually behaves. The information given will, dependent on the legal context, be either required or required and extended.

6.6 Other Solutions

Of course there exist other transparency solutions than the once described in the example. However, due to space limitations in the paper we have chosen to just briefly mention some of them in this section. Concerning web services the Norwegian government gives its citizens the ability to see data stored on them by connected governmental offices through the "minside" web portal [14] similar portals are discussed or planed in other European countries. Regarding keeping track of transaction (i.e. similar responsibilities as part of the PRIME data track) "iJournal"[2], a part of Mozilla Privacy Enhancement Technologies (MozPETs) and "iManager"[8] for use with PDAs and mobiles should be mentioned. Microsoft CardSpace [3] also have some transaction tracking capabilities.

7 Conclusions

In the paper we have given an overview on transparency tools for enhancing privacy. We have given a definition on what we consider such a transparency tool and discussed and suggested a number of parameters that can be used to classify and compare implemented and suggested solutions for transparency tools. Finally, we have given a short analysis and comparison on some example solutions. The conclusion that can be drawn from this survey, taking both the example solutions and the referred solutions are the following:

On the control and verification side there are very few tools that can be classified as "interactive" (i.e. have the ability to let the user/proxy actively influence the stored data and/or the processing). This could be due to the fact that many companies see the information as a big asset and that it is necessary to have a very well developed identity management system and a good access control in order to not turn that type of functionality into both a privacy and a security risk. And thus they are reluctant to provide this type of service on line, but rather stick to manually based and analog methods for the service.

Regarding trust, all the actually implemented solutions and some of the suggested solutions do require (or assumes) a trusted server and some of them also require (or assume) some form of trusted third party. The reason for this might be the problems of practically implementing and maintaining a trusted computing environment and the lack of standards and requirements regarding privacy and privacy auditing. However, according to our experience, many companies and service providers behave in a responsible and fair manner since they are usually dependent on a good reputation in order to be profitable.

Most of the presented or referred solutions, as far as there is a possibility to make a judgment, have a service scope. However, there is one notable exception to this in the referred solution. This exception is the "minside" web service that in our opinion has a "Conglomerate Scope" or at least,as far as we know, has the intention to have this when it is fully implemented.

Acknowledgment

This work has partially been done within the scope of the European network of excellence FIDIS (www.fidis.net), PRIME (www.prime-project.eu) and PrimeLife (www.primelife.eu). The author would also like to thank Simone Fischer-Hbner, Stefanie Poetzsch, Marit Hansen and the participants of the FIDIS/IFIP summer school 2008 for helpful comments and discussions on this paper. Further I would like to thank the participants in the upcoming FIDIS D 7.12 deliverable working group for interesting and helpful discussions on transparency tools. FIDIS receives research funding from the European Unions Sixth Framework Program and the Swiss Federal Office for Education and Science. PRIME has received has received funding from the European Unions Sixth Framework Program and PrimeLife receives funding from the European Community's Seventh Framework Program.

References

1. Privacy Bird, http://www.privacybird.org
2. Brckner, L., Voss, M.: Mozpets – a privacy enhanced web browser. In: Proceedings of the Third Annual Conference on Privacy and Trust (PST 2005), Canada (2005)
3. Chappell, D.: Introducing windows cardspace. Technical report, Windows Vista Technical Articles (2006)
4. World Wide Web Consortium. Enterprise privacy authorization language (epal 1.2). W3C Member Submission (November 2003)
5. Hildebrant, M. (ed.): D 7.12: Biometric behavioural profiling and transparency enhancing tools. FIDIS Deliverable (work in progress)
6. Fischer-Hbner, S., Hedbom, H. (eds.): Deliverable d14.1.c framework v3. PRIME Project Deliverable (March 2008)
7. Hansen, M.: Marrying transparency tools with user-controlled identity management. In: Proceedings of Third International Summer School organized by IFIP WG 9.2, 9.6/11.7, 11.6 in cooperation with FIDIS Network of Excellence and HumanIT, Karlstad, Sweden (2007)
8. Jendricke, U., Gerd tom Markotten, D.: Usability meets security– the identity-manager as your personal security assistant for the internet. In: Proceedings of the 16th Annual Computer Security Application Conference (2000)
9. FIDIS (Future of Identity in the Information Society), http://www.fidis.net
10. European Parliament. 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. Official Journal, L 281/31– L 281/39 39 (October 1995)
11. Article 29 Data Protection Working Party. Opinion on more harmonised information provisions. 11987/04/EN WP 100 (November 2004)
12. PrimeLife, http://www.primelife.eu/
13. Sackmann, S., Strker, J., Accorsi, R.: Personalization in privacy-aware highly dynamic systems. Communications of the ACM 49(9) (September 2006)
14. Min Side, http://www.norge.no/minside
15. W3C. The platform for privacy preferences 1.0 (p3p1.0) specification (April 2002)
16. Weitzner, J., Abelson, H., Berners-Lee, T., Hanson, C., Hendler, J., Kagal, L., McGuiness, D.L., Sussman, G.J., Waterman, K.: Transparent acountable data mining: New strategies for privacy protection. Computer Science and Artificial Intelligence Laboratory Technical Report MIT-CSAIL-TR-2006-007, Massachusetts Institute of Technology, Cambridge, MA, USA (2006)
17. Zwick, D., Dholakia, N.: Whose identity is it anyway? consumer representation in the age of database marketing. Journal of Macromarketing 24, 31 (2004)