Privacy Protection in Data Sharing: Towards Feedback Based Solutions

Mortaza S. Bargh[1,2], Ronald Meijer[1,2]

[1]Creating 010
Rotterdam University of Applied Sciences
Rotterdam, Netherlands
m.shoae.bargh@hr.nl

[2]Research and Documentation Centre
Ministry of Security and Justice
The Hague, Netherlands
{m.shoae.bargh, r.f.meijer}@minvenj.nl

Sunil Choenni[2,3], and Peter Conradie[4]

[3]Creating 010
Rotterdam University of Applied Sciences
Rotterdam, Netherlands
r.choenni@hr.nl

[4]Industrial System & Product Design
Faculty of Engineering and Architecture, Ghent University
Ghent, Belgium
peter.conradie@ugent.be

ABSTRACT

Sharing data is gaining importance in recent years due to proliferation of social media and a growing tendency of governments to gain citizens' trust through being transparent. Data dissemination, however, increases chance of compromising privacy sensitive data, which undermines trust of data subjects (e.g., users and citizens). Data disseminators are morally, ethically, and legally responsible for any misuse of the disseminated data. Therefore, privacy enhancement techniques are often used to prevent unsavory disclosure of personal data. Data recipients, nevertheless, are sometimes able to derive (part of) privacy sensitive information by, for example, fusing the shared data with other data. This can be considered as a sort of data misuse. In this contribution, we investigate how having a feedback from data recipients to data disseminators is instrumental for detecting such data misuses (i.e., privacy breaches). We also elaborate on using feedback for defining and deriving context-dependent privacy-preferences of data disseminators. In this case, feedback acts as a means of privacy prevention. We provide a categorization of existing feedback based solutions and, in addition, describe our implementation of a feedback-based data dissemination solution in an eGovernment setting. Finally, we elaborate on the importance of real-time partial feedback mechanisms, as a rising and promising solution direction for preserving privacy.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues – *ethics, privacy, regulation*

General Terms

Design, Economics, Human Factors, Legal Aspects, Management, Reliability, Security

Keywords

eGovernment; Data Fusion; Data Sharing; Feedback; Privacy; Trust

ICEGOV2014, October 27 - 30 2014, Guimaraes, Portugal
Copyright 2014 ACM 978-1-60558-611-3/14/10...$15.00
http://dx.doi.org/10.1145/2691195.2691279

1. INTRODUCTION

Governments, governmental organizations and scientific community have pursued more openness and transparency with their (research) data. This transparency contributes to trust of stakeholders such as citizens, organizations and enterprises in governmental and scientific institutes [37]. Considering the rising popularity of social networks, we also witness that individuals increasingly share their personal data in social networks to gain friendship, support, recognition, knowledge, etc. [3]. Data sharing, on the other hand, increases the chance of compromising privacy sensitive data items such as names, email and postal addresses, dates of birth, geo-locations, bank account numbers, photos and political/personal opinions.

Potential violations of privacy and solutions to prevent these violations have been received a lot of attention in the literature. These technical solutions focus on supporting principles such as 'privacy by design', 'access control' and 'need to know' in order to prevent disclosure of privacy sensitive information, as seen for example in [6] [15] [20] and [34]. In [34] a framework is proposed to protect the privacy of citizens. The authors of [20] and [2] propose comprehensive architectures to minimize violations of privacy laws and regulations. A so-called ambient law is proposed in [15] that articulates fundamental legal protections such as those for privacy preservations within socio-technical infrastructures. There are also many papers that address privacy issues within specific application domains [6]. Despite having abundant amount of literature on privacy protection, it remains a big challenge to prevent privacy violations in practice. The difficulty stems from the complex nature of the notion of privacy in being subjective and context dependent. In a certain context, privacy pertains to the identity of a person; while in another context where the identity of a person is known, it pertains to the actions of the person. For example, in the healthcare domain the identity of a patient is known by a hospital but unnecessary tracking of movements and actions of the patient might be regarded as a privacy violation.

In this paper we consider privacy protection in data sharing settings, where data about a phenomenon is scattered among and can be found at several entities. In terms of the Data Protection Act [13] there are three entities involved in disseminating personal data, namely: data subject, data controller, and data processor, whose relations are shown schematically in Figure 1.

- Data subject is the one about whom the (personal) data is.

- Data controller is the entity (e.g., individual or organization) who determines the purpose and the way of data processing.
- Data processor is any entity who processes the personal data on behalf of the data controller. Processing functionality includes obtaining, recording, holding, and doing operations (such as adaptation, retrieval, disclosure by transmission or dissemination, alignment, combination) on the data.

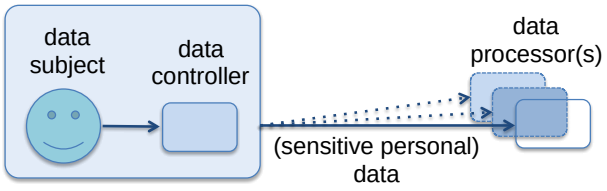

Figure 1. An illustration of our data-sharing model with three entities involved.

Inspired by the Data Protection Act model, we use the terms 'data controller' and 'data processor' to denote the 'data disseminator' and 'data recipient', respectively, throughout this paper. According to this notation, a data controller shares (sensitive personal) data with one or more data processors as shown in Figure 2. Note that each of these data processors may act as a data controller and, in turn, share the (processed) data with other data processors. Further, note that we will not show data subjects in our illustrations solely for simplification purposes. It is implicitly assumed that individuals hold both data subject and data controller roles in the first hop of any data sharing process. As an example of our model, Google can be regarded as a data processor, while the homepage of an individual person (i.e., the data subject) as a data controller.

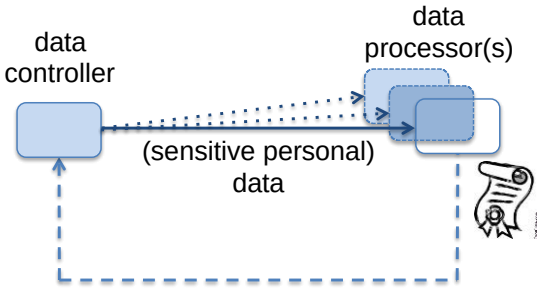

Figure 2. An illustration of feedback in a data dissemination setting.

Any careless and improper combination and use of data from multiple sources by data processors may give a wrong and/or unjustifiable view of a real-world phenomenon. Such views can be due to establishing inadequate relationships or exposing privacy sensitive information, respectively. For example, appearing on a so-called banga list – i.e., a list of girls which unscrupulous boys publish in social media, often groundlessly, accusing the girls of sexual promiscuity – may have serious consequences for young girls. Any data processor, therefore, must access data and use it in authorized ways according to data use and access rules, agreements and policies. Following [5], we recognize two characteristics of a trustful data processing, namely: whether the data access is authorized or not and whether the data use is authorized or not. These categories lead us to the following four outcomes:

- Data processors may access and use data in an unauthorized way. For example, a culprit hacks to a company's database to access confidential customer data.
- Data processors may access data in an unauthorized way but use the data in an authorized way. For example, law enforcement forces access to some personal data in order to prosecute criminals.
- Data processors may access data in an authorized way but use the data in an unauthorized way. For example, a tax office employee skims through personal data of celebrities or neighbors solely out of curiosity.
- Data processors may access and use data in an authorized way but this may still lead to data misinterpretation and privacy breaches. For example, through information fusion privacy sensitive information or wrong conclusions may be derived.

For unauthorized data access and use, classical security solutions such as access control (including authentication and authorization) and privacy enhancement techniques (including data anonymization, aggregation, and confidentiality) are often used. Using data for a legitimate purpose while the access is obtained from illegitimate ways (i.e., the second category above) requires solutions of a more procedural nature such as having clear legislations and policies in place to define the conditions under which an access to data becomes authorized. For the third and fourth categories, which are concerned with authorized access (where data is used in an unauthorized and authorized way, respectively), monitoring solutions are mainly used. In such solutions the way that the data is processed is monitored and when a rule or policy is violated an alert event or message is fed back to data controllers as illustrated in Figure 1.

In this contribution we study various feedback mechanisms that can in a way be exploited to protect privacy and to prevent establishment of inadequate relationships. Within a feedback mechanism, as illustrated in Figure 2, a data controller passes data to a data processor who aggregates the data, integrates the data with other data, shares the data with others, etc. We explore the ways that the data processor can provide the data controller with feedback about the processing steps and results. Based on the feedback received, a data controller may change its policy for data sharing (e.g., decide not to share some data at all or to change the aggregation level of the data passed to data processors). Note that this paper focuses on 'detection' of privacy breaches or specifying of privacy preferences via feedback. How precisely a data controller deals with feedback information is out of our scope.

Exploiting feedback to protect privacy is a rather unexplored area to the best of our knowledge. In this paper we revisit and analyze the feedback mechanisms used in data sharing systems for protecting privacy. Subsequently we categorize the ways that these feedback mechanisms are used. To this end, we are mainly inspired by the feedback mechanisms used in telecommunication systems. Further, we present our realization of two (technical and procedural) feedback mechanisms for disseminating some judicial datasets of our organization. As a solution direction, further we advocate using partial and real-time feedback mechanisms when a breach of privacy occurs (for example, a user gets a notification as soon as her/his data, which is stored in the database of a service provider, is looked up by an employee or fetched by a third party in an unordinary way). This type of feedback has not been widely adopted in current data dissemination systems yet.

The remainder of the paper is organized as follows. Section 2 provides an overview of feedback usages in the area of privacy preservation. Section 3 gives some background information on using feedback in telecommunication and data sharing settings. Section 4 presents our taxonomy of feedback mechanisms and our framework on how to use feedback mechanisms to protect privacy in information sharing settings. Section 5 provides an overview of our realization of feedback mechanisms within a judicial data sharing setting. Section 6 discusses our results and Section 7 presents our conclusions and describes future research directions.

2. RELATED WORK

Feedback can serve as a trust enhancement mechanism by giving a good feeling directly to the data controller, encouraging her/him to share her/his data. In the context of location information sharing, for example, Tsai et al. [32] investigate the effect of feedback on location information requests for Locyoution system (a location sharing system for mobile devices). The system keeps logs of who viewed a user's location. Those users who had access to logs of their location information requests reported greater comfort levels in using the system and a reduced privacy concerns, compared to those who received no feedback of their location information requests [32]. Jedrzejczyk et al. [19] similarly present Buddy Tracker, where location information requests are shown in real time to the users. In this way the Buddy Tracker system provides real time feedback and the authors concede that such notifications could get annoying and that the timing of feedback is important. According to [19], providing such feedbacks contributes to social translucency of users, whereby users use group-based systems more efficiently.

Feedback as a trust enhancement mechanism is also considered in [21] and [22], which aim at enforcing accountability in cloud computing. To this end, the papers propose to use reporting tools for generating summaries of, for example, audit trails, file access history, file lifecycle and suspected irregularities to end users. Similarly [28] mentions one of the contractual agreements that can be considered between cloud providers and users is to require providing "immediate notification by specified means (e.g., via telephone with written follow-up), for any suspected data breach". These papers do not, however, elaborate any further on generating these reports and do not consider them as feedback for privacy protection as we do here. Their focus is merely on the accountability aspects that encompass a wider range of issues such as accountability phases and functions. Examples of accountability phases are: policy planning, sensing and tracing, logging, log-data storing, reporting and replaying (N.B.: this aspect is related to feedback), auditing, and optimizing. The accountability functions include data collection, automated auditing; and policy, law/regulation management. As such these works cover a more generic scope than ours in this contribution.

In [26] and [27], the authors focus on 'consent' management in the context of data usage. The authors devise building blocks such as a personal consent/revocation assistant, a privacy-aware policy enforcement manager, and a disclosure/notification manager. The latter tracks and intercepts personal data flows between organizations, and informs end-users about these flows through the personal consent/revocation assistant (thus it provides feedback). Hereby end-users can consent or dissent to their personal data flows passing across organizations.

In [16] and [18] a so-called obligation framework is presented to define precise data protection policies. The framework allows a way to impose contextual conditions in controlling access to data in the future. The authors argue that yes or no access is not adequate when requesting data. To overcome this, they suggest granting access to data only when certain conditions are met. For checking such obligations (i.e., future conditions) one can define consent type mechanisms.

Nowadays we also witness a surge of tools in the market to enable organizations to monitor how their information resources are used. Example tools are Security Operations Center (SOC) systems for large organizations and the VDSS (Vita Data Security Systems) system [35] for small and medium size enterprises. Such tools provide (real-time) feedback for detecting security and privacy irregularities in data sharing systems.

Although all works and tools considered in this subsection somehow rely on using feedback to enhance trust (and protect privacy) of data controllers, to best of our knowledge, there is no work that systematically and explicitly bases its privacy protection approach around the concept of feedback as we do in this contribution. Furthermore, our implementation of feedback is similar to the implementations mentioned in [24] and [26] for sharing user-context and user-identity attributes with third parties, respectively. In all these the feedback is used to further specify data access and privacy policies per data sharing instances. While the feedback mechanisms presented in [24] and [26] are of technical nature, our feedback solution encompasses both technical and procedural aspects and, as such, offers a cross-organizational solution for preserving privacy.

3. BACKGROUND

Feedback has widely been used for controlling the behavior of mechanical, physical, biological, cognitive, and social systems for many years. In 1948 Norbert Wiener introduced the so-called Cybernetics theory based on the idea that the information transmitted in a system is an effort to control the surrounding environment, pp.15 [36]. Hereto cybernetic systems sense feedback from the environment and adapt their behavior accordingly to influence some aspects of the environment.

One can track the influence of feedback to various engineering disciplines such as mechanical, electrical and software engineering; and to various scientific disciplines such as social science, climate science, biology, economic and finance. An interesting application area of feedback is developed in Shannon's information theory for transmission of messages, expressed in bits and bytes, from a source to a destination. Both Wiener's theory and Shannon's theory are concerned with transmission of information; nevertheless their scopes differ. While the former is concerned with achieving a (desired) change in the environment, the latter is specifically concerned with delivering data from its source to destination in an errorless way.

3.1 Feedback for Privacy Preservation

This paper explores the role of feedback in disseminating or sharing of data in a privacy preserving way. In other words, we are interested in transmitting data to its destination correctly, in the sense of being privacy preserving, rather than exercising a control over the environment through (the content of) the disseminated data. Therefore, we define feedback as having a communication medium or channel in the backward, i.e., from the destination (as we denoted it as data processor) to the source (as we denoted it as data controller) direction in order to facilitate the data transmission/dissemination process in the forward (i.e., from the source to the destination) direction. The facilitation of data dissemination in the context of our paper means preserving

privacy. Figure 3 illustrates the concept of data dissemination and feedback.

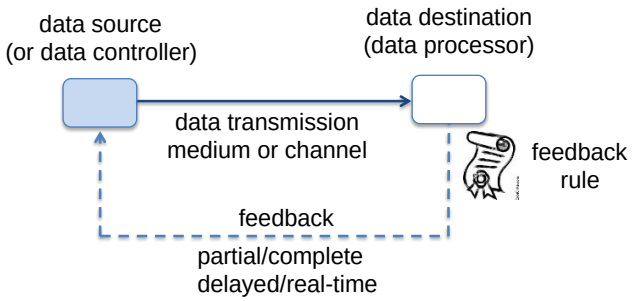

Figure 3. An illustration of the concept of feedback in data transmission.

Our focus lies more on detecting privacy breaches in an automatic way rather than on how individuals experience privacy in a given context. As such, our scope comes closer to the scope of Shannon theory than to that of Wiener theory. Therefore, in the following subsection we will review some properties of feedback from Shannon Theory mainly. This review will help us to derive a framework to categorize the feedback mechanisms used for privacy preservation in data dissemination settings.

3.2 Feedback Properties and Characteristics

Using feedback for data transmission has advantages such as reduced complexity of data transmission [8]. The price paid for these benefits is having an extra communication channel in the reverse direction for feedback. A feedback channel is solely meant for facilitating the data transmission in the forward direction. Communication with feedback, in other words, is not meant for two-way communication where the destination tries to send its own message, i.e., which is other than the source's message, in the backward direction. In telecommunication the feedback channel is often assumed to be noiseless in the sense that it does not introduce uncertainty to the feedback data. A noiseless feedback truly conveys what the data destination wants to share with, i.e., to feed back to, the data source.

One can distinguish between complete feedback and partial feedback. In case of complete feedback the destination sends back exactly whatever it has received, which is the transmitted data plus the noise introduced in the forward channel. In case of partial feedback, the destination sends back a condensed and processed version of the received data. The procedure of data condensation and processing in the latter case is done according to a rule that is well known at both destination and source (this rule is indicated as "feedback rule" in Figure 3). Hereby, the source practically gets a transformed image of what is received at the destination. The Automatic Repeat reQuest (ARQ) data transmission model is an example of partial feedback. In ARQ model the destination should have a baseline rule to know when the data is received wrongly and inform the source about the occurrence of the error using a short message (i.e., partial feedback). Subsequently the source retransmits the corresponding message.

4. FEEDBACK FOR PRIVACY ENHANCEMENT

In this section we start with a categorization of feedback methods used for preserving privacy in various data dissemination settings. Subsequently we provide a number of design principles in order to use feedback for preserving privacy.

4.1 Feedback Taxonomy

Within the context of privacy protection, feedback can be an enabler for putting data controllers in charge of revealing their data to other parties. Comparing the data transmission settings described in Subsection 3.2 and our privacy-preserving data-dissemination settings, we assume that the data source corresponds to the data controller and the data destination corresponds to the data processor, which receives privacy sensitive data. This correspondence is also indicated in Figure 3.

The complete feedback in our settings is characterized as reporting to the data controller all data processing actions (to be) done by a data processor. The trigger for feedback can originate from anywhere, ranging from these actions at the data processor (i.e., pushed to the controller model) to a request of the data controller (i.e., pulled by the controller model). As an example of the latter, Google is the well-known data processor that collects data about everything including individuals. If someone enquiries Google about herself/himself, she/he can get feedback about what Google has gathered about she/he from different data sources. This is a full feedback as one is directed to all sites wherein the relevant data resides. The partial feedback, on the other hand, reports only some specific information about (aspects of) data processing. For example, when a privacy policy violation occurs at the data processing node, an alert is sent to the data controller in the VDSS [35] system.

In terms of its timing, feedback can be categorized as real-time or delayed, depending on whether the feedback is sent immediately or with a (certain) delay. We define real-time feedback when it is given as soon as the data is processed or as soon as a policy rule is triggered. Otherwise, we regard it as delayed feedback. Note that in case of Google provisioning its information about a user based on her/his request, one is concerned with delayed feedback normally. A real-time feedback, however, would be the case where Google informs individuals over any privacy sensitive information it derives, immediately and proactively (assuming that the privacy policy of individuals is known at Google).

The two defined feedback types specify four feedback categories as illustrated in Figure 4. For each category we provide here a number of examples from Section 2 in the following.

- An example of 'complete & delayed' feedback is given in [32] where data controllers can look at data access logs at data processors.
- Examples of 'complete & real-time' feedback are [19] (for allowing data controllers to monitor the way that data is processed) and [26] (for posing consent questions whenever a data request arrives – thus before data release – from a third party, i.e., the data requester, at the data processor to transfer the data to the data requester).
- Examples of 'partial & delayed' feedback are given in [21] and [22], where the policy violations detected by analyzing data access logs are reported to the data controllers.
- An example of 'partial & real-time' feedback is the 'immediate notification' feature proposed in [28] that can be included in the contractual agreements between cloud providers and users. Another example of 'partial & real-time' feedback is embedded into a new Dutch law proposal. This proposal requires those parties that collect personal information to immediately report to individuals when these

individuals' personal information is breached and if these are qualified as serious-impact data-breaches [29].

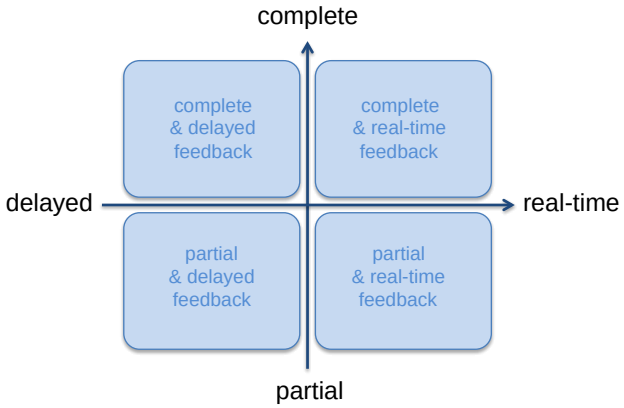

Figure 4. An illustration of feedback categories.

4.2 Guiding Principles

Compared to partial feedback, complete feedback may cause information overload at data controller side and inflict a high data transmission cost on the backward channel. Therefore if the data controller does not have enough processing power or if the feedback channel does not have enough bandwidth, then partial feedback is preferred to complete feedback.

An important issue in partial feedback is to determine when feedback should be triggered. Such a trigger can, for example, be based on occurrence of privacy breaches. Hereto the data processor may rely on privacy policies that define the rules and conditions of privacy violations. These privacy policies can be domain-specific (i.e. defined by the organization that has governance over the data processors) or data specific (i.e. defined by data controllers). The technical realization of data specific policies can be based on, for example, sticky policies [4] [25]. In such systems one may attach attributes or sticky policies to the data objects when disseminating them from data controllers to data processors [23].

Timing of feedback depends on the degree of urgency for notifying data breaches. Domain administrators or data subjects can define the timing of feedback in their privacy policies based on their preferences and the risks associated with possible data breaches. Nowadays we witness a trend towards more real-time feedback to contain risks of privacy breaches on time.

Often in data sharing settings the privacy policy cannot be defined beforehand in details due to, for example, complexity of and/or not knowing the context of data usage. Here feedback can be used to refine the data-access or privacy-policy on the scene and based on the context of data usage. To this end, feedback is used as a means of detecting privacy preferences rather than of detecting privacy breaches. In other words, the feedback enables us hereby to realize a self-learning system for specifying (context dependent) privacy policies. To this end, feedback works as a means of privacy breach prevention than privacy breach detection. Such cases typically occur in need-to-know scenarios (e.g., when asking for consents), where 'complete & real-time' feedback is used. Such a feedback is considered complete because of being initiated per every data request/sharing and it is real-time because of being initiated at time of data need. In the following we describe our implementation of such a feedback-based solution that enables the data-controller to define data sharing policy per data request.

5. REALIZATIONS

Our organization, i.e., the Research and Documentation Center (abbreviated as WODC in Dutch), is the research center of the Dutch Ministry of Security and Justice. The center systematically collects, stores and enhances Dutch criminal-justice information to define, address and assess the ministry's future research agenda, policy-related questions and the possible implications of standing policies, respectively. The research center also aims at sharing its research data of the completed research projects and its statistical information in order to allow scrutinizing and validating its data. As the research center works with confidential judicial data, issues such as confidentiality and privacy-sensitivity are thoroughly taken into account before, during and after sharing its data.

In the following we elaborate on the ways that feedback is used to detect privacy issues and preferences as a first step towards enhanced privacy protection. Through this feedback it is possible to check the adherence to privacy laws and regulations such as the Data Protection Directive of the European Union and the Dutch Privacy Protection Act (DPPA). These regulations include finality (related to the purpose of collecting data), legitimacy (related to the process of data collection), proportionality (related to the means of data collection), subsidiarity (related to using other alternative means), transparency (related to data controllers knowing about the processing of their data) and data subjects' rights. The necessity of receiving such feedback stems from the fact that the data controller (i.e., our research center) is morally, ethically and legally responsible for any misuse of the disseminated data.

5.1 Feedback Procedure

To share our research center's datasets with scientists, we use the servers of the Data Archiving Networked Services (DANS) organization [11], i.e., data processor, in accordance with Dutch government guidelines. Datasets of completed research may be considered for dissemination if being in compliance with some criteria such as not being confidential, not being reused by us in our monitoring or longitudinal research, not being insufficiently representative, and not being unreliable/invalid. The steps of sharing our research data, as indicated in the message sequence diagram of Figure 5.

- Step 0: Our anonymized data and its metadata are uploaded to the DANS servers, using a file transfer protocol.
- Step 1: A data requester, e.g., scientific researcher, looks up the DANS site to find about our center's interesting datasets using the metadata on the DANS servers. The researcher fills in web form at the DANS's website, using the https protocol.
- Step 2: DANS sends a Data Request (DR) derived from the filled Web form to the WODC via email to inform about the data processing (i.e., the requested data transfer). The DR, which acts as feedback (similar to a consent type), includes an elaborated research design and is transferred automatically to the Data Request Service (DRS) mailbox of our research center.
- Step 3: The data request goes through a rigorous procedure aimed at protecting privacy through deriving a case-specific data dissemination policy. There are three procedural steps for data sharing: The first step concerns an examination of

the request by an experienced data manager to see which variables are necessary and whether they could be delivered from the data at DANS. Subsequently, an advice document is sent to a DPPA workgroup to examine the legal conditions of the request. Hereto it is also possible to contact the data requester for further information via a traditional means such as email, telephone or face-to-face meeting. Finally a board of directors judges the request to grant or deny access to the requested data.

- Step 4: The decision of the board of directors is sent to DANS via email. Granting/denying access to a specific data requester, considering the situation and context, can be seen as specifying the privacy policy (for that requester) as shown schematically at the end of this step in Figure 5.
- Step 5: If the access is granted, DANS delivers the data to the data requester via email. Note that the data requester typically signs a standard agreement and specific conditions of reuse before obtaining the data.

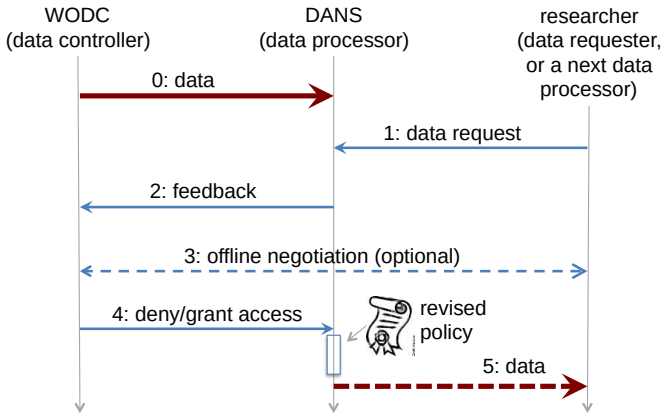

Figure 5. Illustration of the steps of our data sharing procedure.

5.2 Implicit Feedback

To address privacy, misuse and misinterpretation issues we have allowed so far only a limited sharing of our data with scientists and only for research purposes. Such data sharing has another inherit feedback mechanism, namely: the review process of the outcomes of scientific exercises, as shown with Step 6 in **Erro! A origem da referência não foi encontrada.**Figure 6. Because these outcomes are generally peer-reviewed by fellow scientists or by practitioners at our research center, there is a guarantee, up to an acceptable level, that the disseminated data is used correctly and according to privacy requirements and professional ethics. We call this as implicit feedback because (1) the data controller receives it indirectly from (subsequent) data processors, i.e., research institutes, through the published papers/reports, and (b) it is based on good practice principles (i.e., it is not enforced by contracts).

Note that we regard implicit feedback as a specific case of partial feedback where a data controller obtains incomplete feedback via third parties instead of directly via data processors. We used to share minimum or no data due to privacy concerns and considerations practically. The peer review process, however, enabled us to *dare* sharing some information with scientists, thereby we moved up a step towards being transparent. In other words, implicit feedback enabled us to trust in the data dissemination process and to dare to disseminate data as suggested in [32], [19] and [13] (due to, for example, being a means that contributes to "social translucency" [19]).

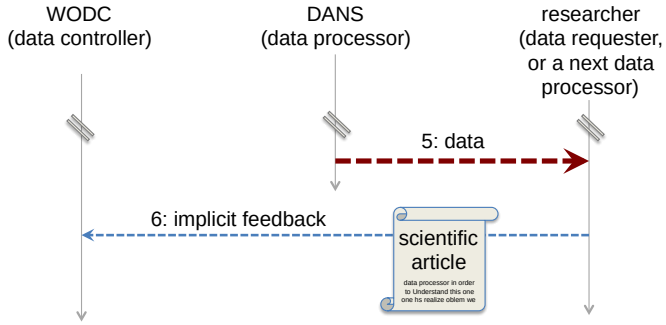

Figure 6. Using implicit feedback to facilitate (and enhance trust in) data sharing.

6. DISCUSSION

We realized a combination of software system and organizational procedures with 'complete & real-time' feedback in order to share information in an organizational setting. The resulting system places the data of data subjects at a data processor in order to be, for example, accessed by others, shared with others, and fused with other information. Here the feedback is used to ask permission of the data controller to access data and the data controller may authorize the access based on her/his context-dependent privacy preferences. In this way, consequently, the data controller can refine the data access policy on spot and based on data usage context through, for example, negotiating the purpose and terms of use.

In principle, the approach adopted for realizing this system is similar to the approaches of the systems in [24] and [26] designed for sharing user-identity attributes and user-context information, respectively, with third parties through using consent questions (i.e., feedback). Our contribution applies similar feedback principle in order to tailor privacy policies to the setting of the cross-organizational data sharing. While feedback mechanisms presented in [24] and [26] are of technical nature, our feedback solution combines both technical and procedural aspects and offers a feedback-based solution for eGovernment and cross-organizational privacy preservation. Within our solution the organizations involved can negotiate terms of use and access at a procedural level using also traditional means like postal correspondences, telephone calls and face-to-face meetings. This compatibility with traditional mechanisms makes, therefore, the solution suitable for cross-organizational settings where processes are not (or cannot be) fully automated.

In cases of consent questions or our data sharing permission requests the feedback is used at every data access instance (i.e., before a data access is granted). We categorized these cases as 'complete & real-time' feedback. On the other hand, 'partial & real-time' feedback is given whenever for example a privacy violation occurs. Such a partial & real-time feedback can be used to refine data dissemination policies (one should note that such a policy refinement influences only the future releases of personal data and cannot remedy the impacts of the already disseminated personal data). To the best of our knowledge 'partial & real-time' feedback is not used in privacy preservation settings widely. An example that comes close is the warnings that users get when their services are accessed from an unexpected context (e.g., one accesses his/her Gmail for the first time from a different country

than the usual one). These warnings can be considered as 'partial & real-time' feedback that is triggered based on the rule of suspicious account accesses. Bargh and Choenni [1] elaborate on the necessity of the generic case of informing data controllers as soon as their privacy is breached at data processors. Another application of 'partial & real-time' feedback would be in social networks like Facebook to somehow issue a warning to you when a friend publishes your picture on his/her page against your privacy preferences. In this direction, it is for our future work to design a database monitoring system that provides a partial feedback as soon as a privacy policy is violated when accessing or using the corresponding dataset.

For a technical realization of feedback to report privacy breaches one should closely follow the lessons learnt from various initiatives like [29] and [12] that aim at notifying individuals whose privacy is seriously breached. A challenge in (real-time) notification of individuals in case of privacy breaches is to determine when issuing a notification is appropriate, especially in low risk data breaches [12]. The feedback mechanisms in eGovernment settings can have both technical and non-technical natures (thus asking for developing new laws, policies and/or technologies). As an example of the latter type, we have already mentioned an implementation of feedback in the Dutch law proposal that requires those parties that collect personal information to immediately report to individuals when these individuals' personal information is breached and if these are qualified as serious-impact data-breaches.

Feedback mechanisms can play an important role in monitoring and enforcement of data usages by data controllers. Currently one observes a surge of interest in policy and eGovernment research and practice to investigate and realize mechanisms for monitoring and enforcement of authorized data usage as complementary to those of authorized data access. Involving third party service providers for information management and starting open data / data transparency initiatives are examples of such trends in eGovernment settings. Steadily public institutes and governmental organizations rely more on third party service providers for managing their datasets and information. National and international cloud service providers and organizations like DANS [11] (for sharing information of various Dutch ministries) and SURFNET [30] (for providing, among others, identity management services for/across Dutch universities) are examples of such (independent) third party service providers. These third parties (often by their constituency) not only share and mediate information with/among parties such as citizens, user-groups and institutes; but also (are capable to) deduce and collect (privacy sensitive) information about users and citizens through fusing information from various sources. For example, the myIdP service is proposed in [24] as an extension to the Swiss eID infrastructure to handle personal attributes (like address, telephone number, email) that are not collected by SuisseID identity providers (due to legal restrictions). The myIdP service can collect and reuse the data that users share in the Internet when they fill in Web application forms. The myIdP service collects these attributes for reuse in future transactions for users' comfort as well as for providing Internet service providers with trustful attributes about users. Such a service can easily cause privacy breaches by sharing the derived user attributes with unknown and untrusted service providers or can collect wrong/polluted user attributes. To deal with these threats, the myIdP service acquires users' consent before sending their attributes to a requesting web application and collects user attributes selectively from those trusted service providers with whom the user had had interactions in the past. According to [24], "the myIdP service is evaluated in a scenario of prefilling e-forms in an eGovernment application". Similarly, other eGovernment services can be foreseen in the near future that use such complete or partial feedback mechanisms in real-time to manage user privacy sensitive information in those scenarios that arise in unpredictable situations and circumstances.

In recent years governments have started open data initiatives by releasing public sector data to citizens as a measure of government transparency [9][10]. Such initiatives motivate using data from citizens, which can be collected directly or indirectly, and combining it with data from other sources in order to deliver added value services. An example of direct data collection occurs when a user provides his/her information to government agencies in order to use a public service. An example of indirect data collection occurs when methods like crowdsourcing [14] [31] are used to collect data using for instance smart mobile devices of citizens. In such data collection cases, however, there are potential risks of privacy breaches when the (self-provided) data of users is combined with other user data retrieved from elsewhere [1][5][33]. In direct data collection cases citizens are often aware of and they consent for their personal data to be collected. In crowdsourcing, however, this consent may not be present as citizens are often unaware of (the extent of) their contribution to the data collected and the extent of using their data by third parties. As such citizens may or may not consent to the data processor that makes use of their personal data or shares it with other organizations. That is why feedback – particularly when the data processor initiates the feedback as a result of, for example, processing the data for any purpose other than it was collected for – can be instrumental in preserving citizens' privacy.

Governments have to protect both national security and privacy. Hereto difficult choices should be made, which may stress the delicate balance between security/privacy interests and other interests [7]. These choices have to be made in a political and legal context; which makes it a great challenge because developing new laws often lags behind the rising threats (like privacy breaches in this information age). Current privacy laws and regulations do not always provide sufficient tools to cope with new developments. For example, it is not possible to monitor communications on a large scale in order to prevent terrorist attacks. Feedback in our opinion is essential to transparency and accountability (by its definition, even after the facts) as it shall create the openness that is necessary to maintain and, if necessary, to restore the public trust in government.

7. CONCLUSIONS

Feedback has been used as a trust enhancement mechanism by giving a good feeling to data subjects directly and encouraging them to share their data with others, as seen in the cases of social translucency and accountability. Exploiting feedback to protect privacy is a rather unexplored area to the best of our knowledge. In this paper we revisited and analyzed the feedback mechanisms used in data sharing systems. We used two criteria to categorize feedback mechanisms, namely: completeness and timing of feedback. Consequently we arrived at four feedback types: 'complete & delayed', 'complete & real-time', 'partial & delayed', and 'partial & real-time'. Depending on the capacity available for processing and transmission of feedback information and on the risk associated with privacy breaches, one can choose one of these feedback types. Based on the insight gained we foresee the usefulness and necessity of using 'partial & real-time' feedback to inform data controller as soon as a breach of privacy policies occurs due to any data processing.

In addition to being a means of detecting privacy breaches, feedback can be used as a means of detecting privacy preferences within specific contexts of data usage. When the privacy policy cannot be defined beforehand in details, due to for example not knowing the data usage context, feedback can be used to refine the data privacy policy on the scene before or after data access. In the latter case, feedback becomes instrumental to prevent privacy breaches. In this direction we described our realization of a software system combined with organizational procedures that relies on 'complete & real-time' feedback in order to share information in an organizational setting for privacy preservation. Having the possibility of feedback to monitor the results of future data processing activities was a source of encouragement for us to share our data.

It is for our future research to investigate the ways that feedback information can be used for enforcing and achieving a particular privacy objective, i.e., to define what to do after receiving feedback. Further, feedback can be exploited for other objectives than preserving privacy. To this end, dealing with misinterpretation of disseminated information can be considered as another topic for future research. We are currently looking forward to investigating/implementing 'partial & real-time' feedback mechanisms in integrating various databases within our research center.

8. REFERENCES

[1] Bargh, M.S., Choenni, S. 2013. On preserving privacy whilst integrating data in connected information systems. *In Proceedings of the International Conference on Cloud Security Management (ICCSM'13)*, Seattle, US, 17-18 October.

[2] Bertino, E. and Sandhu, R. 2005. Database security, concepts, approaches and challenges. *In IEEE Trans. On Dependable and Secure Computing*, vol. 2, no. 1, IEEE Press.

[3] Boyd, D. 2010. Privacy and publicity in the context of big data. Opening keynote at WWW'10, Raleigh, North Carolina, [Online]. Available: http://www.danah.org/papers/talks/2010/WWW2010.html.

[4] Chadwick, D.W. and Lievens, S.F. 2008. Enforcing sticky security policies throughout a distributed application. *In Proc. of the workshop on Middleware Security*, ACM.

[5] Choenni, S., Bargh, M.S., Roepan, C., and Meijer, R. 2014. Privacy and security in data collection by citizens. A chapter of *Smarter as the New Urban Agenda: a Comprehensive View of the 21st Century City*, edited by Gil-Garcia, J.R., Pardo, T.A. and Nam, T., Springer (in press).

[6] Choenni, R. and Leertouwer, E. 2010. Public safety mashups to support policy makers. In *Proc. of Int. Conf. on Electronic Government and the Information Systems Perspective (EGOVIS)*, Bilbao, Spain, August 30- September 3, LNCS 6267, Springer, pp. 234-248.

[7] Clarke, R.A., Morell, M. J., Stone, G.R., Sunstein, C.R., and Swire, P. 2013. Liberty and security in a changing world. Available: http://www.whitehouse.gov/sites/default/files/docs/2013-12-12_rg_final_report.pdf

[8] Cover, T.M. and Joy A.T. 2012. *Elements of Information Theory*, John Wiley & Sons Inc.

[9] Dawes, S. S. 2010. Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, 27, pp. 377-383.

[10] Dawes, S. S. 2010. Information Policy Meta-Principles: Stewardship and Usefulness. In Proceedings of the 43rd IEEE Hawaii International Conference on System Sciences, pp. 1-10.

[11] Data Archiving Networked Services (DANS) website, http://www.dans.knaw.nl/en/content/about-dans

[12] Dennett, J. 2013. Inquiry into privacy amendment (privacy alerts) bill 2013. Available: http://www.oaic.gov.au/news-and-events/submissions/privacy-submissions/inquiry-into-privacy-amendment-privacy-alerts-bill-2013#ftn21.

[13] Erickson, T. and Kellogg, W. A. 2000. Social translucence: an approach to designing systems that support social processes. In *Trans. on Computer-Human Interaction (TOCHI)*, volume 7, no. 1, ACM Press, pp. 59-83.

[14] Ganesan, D. and Corner, M. (2011). Crowd sourcing for data collection [Online]. Available http://sensorlab.cs.dartmouth.edu/NSFPervasiveComputingAtScale/pdf/15693928 97.pdf

[15] Hildebrandt, M. and Koops, B.-J. 2010. The challenges of ambient law and legal protection in the profiling era. *Modern Law Review*, Vol. 73, No. 3, pp. 428–460.

[16] Hilty, M., Basin, D. and Pretschner, A. 2005. On obligations. In *Proc. of the 10th European Symposium on Research in Computer Security (ESORICS)*, pp. 1–20.

[17] Information Commissioner's Office, 22 April 2013. Key definitions of the data protection act. Available: http://ico.org.uk/for_organisations/data_protection/the_guide/key_definitions.

[18] Jajodia, S., Kudo, M. and Subrahmanian, V.. 2001. Provisional authorizations. *E-commerce Security and Privacy*, Springer, pp. 133-159.

[19] Jedrzejczyk, L., Price, B.A., Bandara, A.K. and Nuseibeh, B. 2010. On the impact of real-time feedback on users' behavior in mobile location-sharing applications. In *Proc. of the 6th Symposium on Usable Privacy and Security (SOUPS)*, ACM Press, New York, pp. 1.

[20] Kalidien, S., Choenni, S. and Meijer, R. 2010. Crime statistics online: potentials and challenges. In *Proc. of the 11th International Conference on Digital Government Research*, DG.O, Puebla, Mexico, May 18-21, New York.

[21] Ko, R. K., Lee, B. S. and Pearson, S. 2011. Towards achieving accountability, auditability and trust in cloud computing. In *Advances in Computing and Communications*, Springer, pp. 432-444.

[22] Ko, R. K., Jagadpramana, P., Mowbray, M., Pearson, S., Kirchberg, M., Liang, Q. and Lee, B. S. 2011. TrustCloud: a framework for accountability and trust in cloud computing. In *Proc. of World Congress on Services (SERVICES)*, IEEE Press, pp. 584-588.

[23] Krishnan, R., Sandhu, R., Niu, J. and Winsborough, W.H. 2009. A conceptual framework for group-centric secure information sharing. In *Proc. of the 4th International Symp. on Information, Computer, and Communications Security*.

[24] Laube, A. and Hauser, S., 2013. myIdP - the personal attribute hub. In *Proc. of the 5th International Conferences on Advanced Service Computing (SERVICE COMPUTATION).*

[25] Mont, M. C., Pearson, S. and Bramhall, P. 2003. Towards accountable management of identity and privacy: sticky policies and enforceable tracing services. In *Proc. of the 14th International Workshop on Database and Expert Systems Applications*, pp. 377-382.

[26] Mont, M. C., Pearson, S., Creese, S., Goldsmith, M. and Papanikolaou, N. 2011. A conceptual model for privacy policies with consent and revocation requirements. In Fischer-Hübner, S., Duquenoy, P., Hansen, M., Leenes, R. and Zhang, G. (Eds.), *IFIP Advances in Information and Communication Technology*, V. 352, Springer, pp. 258–270.

[27] Mont, M. C., Pearson, S., Kounga, G., Shen, Y. and Bramhall, P. 2009. On the management of consent and revocation in enterprises: setting the context. *Technical Report HPL-2009-49*, HP Labs, Bristol.

[28] Pearson, S. and Charlesworth, A. 2009. Accountability as a way forward for privacy protection in the cloud. In *Cloud Computing*, Springer, pp. 131-144.

[29] Reijerman, D 17 april 2014. Kabinet zwakt wetsvoorstel meldplicht datalekken af (translation from Dutch: Government weakened the bill of hailing data-leaks). From Tweakers hompage, Available: http://tweakers.net/nieuws/95477/kabinet-zwakt-wetsvoorstel-meldplicht-datalekken-af.html#r_6889054.

[30] SURFnet website, http://www.surf.nl/en/about-surf/subsidiaries/surfnet

[31] Taylor, J. 2010. Citizens as public sensors, [Online]. Available (retrieved on 26 February 2014) here http://radar.oreilly.com/2010/04/crowdsourcing-the-dpw.html

[32] Tsai, J. Y., Kelley, P., Drielsma, P., Cranor, L. F., Hong, J. and Sadeh, N. 2009. Who's viewed you? The impact of feedback in a mobile location-sharing application. In *Proc. of Computer Human Interaction (CHI)*, ACM Press.

[33] Wang, Y., Huang, Y., and Louis, C. 2013. Towards A Framework for Privacy-Aware Mobile Crowdsourcing. In Proceedings of IEEE International Conference on Social Computing (SocialCom), pp. 454-459.

[34] Warner, J. and Chun, S.A. 2008. A citizen privacy protection model for e-government mashup services. In *Proc. of the 9th Annual International Digital Government Research Conf*erence.

[35] Website VDSS (Vita Data Security Systems). Available: http://www.vdss.nl/vdss-in-your-network.ashx.

[36] Wiener, N., 1954. Cybernetics in history. In *The Human Use of Human Beings: Cybernetics and Society*, Free Association Books in Great Britain in 1989, pp.15-27.

[37] Zuiderwijk, A., Janssen, M., Meijer, R., Choenni, R., Charalabidis, Y. and Jeffery, K. 2012. Issues and guiding principles for opening governmental judicial research data. In *Proc. Of EGOV*, LNCS 7443, Springer, pp. 90–101.