Formalising Requirements for a Biobank Case Study Using a Logic for Consent and Revocation

Ioannis Agrafiotis, Sadie Creese, and Michael Goldsmith

Department of Computer Science
University of Oxford, Oxford, England
{ioannis.agrafiotis,sadie.creese,michael.goldsmith}@cs.ox.ac.uk

Abstract. In this paper we focus on formalising privacy requirements for the Oxford Radcliffe Biobank (ORB) case study that has emerged within the EnCoRe project. We express the requirements using a logic designed for reasoning about the dynamics of privacy and specifically for capturing the lifecycle of consent and revocation (C&R) controls that a user may invoke. We demonstrate how to tackle ambiguities uncovered in the formalisation and to bridge the gap between user requirements for personal data privacy and system level policy languages effectively.

1 Introduction

It is evident that there is an ever-growing amount of personal information shared by individuals over the Internet in order to obtain as users access to various products and services. However, the number of incidents of unexpected uses of such information is increasing and individuals have practically no control over their data. Innovative applications such as Web 2.0 and cloud computing raise the complexity of handling personal data effectively, and provide new challenges for privacy advocates.

This paper is inspired by work undertaken in the EnCoRe project[1]. The purpose of EnCoRe [9] is to build a system which will enable enterprises to collect and handle personal data while providing the individuals with the appropriate C&R mechanisms [1] to control the flow of their data, together with a way of ensuring that these mechanisms are effective.

In this paper we present and apply the novelties of the logic that occurred from addressing the ambiguities created when formal methods are applied for verification of the privacy properties of a system [2] and our aim is to verify that no more ambiguities are created when applying the logic on a different context. We formalise the requirements of the EnCoRe system operating in a Biobank environment, using a logic for consent and revocation [3]. The requirements of the system were elicited by analysing the system currently in place and by

[1] The EnCoRe project [9] is an interdisciplinary research project, a collaboration between UK industry and academia, partially funded by the UK Technology Strategy Board (TP/12/NS/P0501A), the UK Engineering and Physical Sciences Research Council and the UK Economic and Social Research Council (EP/G002541/1).

J. Camenisch et al. (Eds.): Privacy and Identity 2011, IFIP AICT 375, pp. 232–244, 2012.

conducting several focus groups to gain a better understanding of the system's environment. In the second section of this paper we provide a synopsis of the C&R logic, in the third we describe the case study and in the fourth we discuss the requirements in more detail and express them in a formal notation. There is a progressive increase in the level of complexity of the formal notations. We begin by presenting the options that the patients may choose from when they donate their samples to the Biobank, then we demonstrate how these options are captured by the policies that the Biobank has in place and how an administrator could create new policies and we conclude by formalising different use cases of possible actions that could occur in the Biobank's system, named Sapphire. Finally, we outline the conclusion and indicate some opportunities for future work.

2 Related Work

To set this work in context, the application of formal methods to privacy mainly focuses on translating privacy policies which are mostly written in natural language, into machine readable formats [12]. Languages like P3P [8] and EPAL [4] are examples of these. Barth et al [5] designed a logic for managing the dissemination of information based on Nissenbaum's theory of privacy as 'contextual integrity' [10]; this logic describes how different roles are allocated to people according to context and sets constraints on how people in these roles exchange data. They applied this logic to privacy policies such as Health Insurance Portability and Accountability Act (HIPAA) [5] and the Children's Online Privacy Protection Act [5]. Further research includes formal frameworks for privacy preferences such as SecPAL, that has been developed by Becker et al. [6]. However, none of these methods handle consent while they completely neglect the notion of revocation. Only the work presented in SecPAL could be complimentary to the logic presented in this paper. SecPAL handles the negotiation of policies and credentials between two parties, while in C&R logic we handle the lifecycle of users' C&R options, after a successful negotiation.

There is a general lack of work specifically addressing the processes of C&R in the context of personal data. While the concept of consent has been studied extensively in the social sciences [14], leading to work on the necessity, meaning and consequences of *informed consent* [13], few computer scientists have given the mechanics of such processes due attention.

3 Logic for C&R

We have formalised the C&R processes in terms of a simple Hoare Logic [3]. We define a set of rights for principals that pertain to specific data δ. We identify a set of actions requested by principals in the system. These actions can be performed only when the involved principals have the adequate rights, and when completed can either affect the rights of the principals or create obligations for principals in the system, or both.

The rules for the logic are given in the form of Hoare triples, as follows:

$$\{\textit{pre-condition}\}$$
$$\textbf{action}(a, b, \delta)$$
$$\{\textit{post-condition}\}$$

where a is the person that invokes the action and b is the person that will execute the action.

The precondition is a Boolean combination of (statements about) rights, which must be satisfied before the action can be performed. Every time the action is performed the state of the system changes from the one described in the pre-condition to the one described in the post-condition. The new state of the system is a combination of new rights and possibly also obligations. An obligation is a higher-order predicate that allows us to record the necessity of further actions consequent on the one just performed.

The state of the system is a specified set of principals (which could increase or decrease according to the performed actions) and a relation R that allocates rights to each principal of the system. In the initial state of the system, the only rights are those of the data subject who possesses full rights over the data; by definition, the data subject is the person whose data is being processed and inherently possesses all the rights that are defined in the logic. The data controller who is the person that will process the data has initially no rights at all on the data. The state of the system also records other variables such as time, notification constraints, etc.

In a previous publication we applied the simple version of the logic to an Employee case study [2]. That exercise resulted in an enriched version of the logic, expressive enough to address ambiguities created either by the complex notion of privacy or by the translation of natural into formal language [2]. In this paper we apply the extended logic that derived from the first formalisation, and our aim is to formalise a Biobank case study without any further ambiguities, than those identified and addressed in [2].

In the extended logic, the state additionally includes consent variables and the actions correspond to smaller steps than in the simple logic used in our first attempt. Moving to a refined model we are able to offer finer granularity by introducing consent variables into the actions and model the intermediate states of the actions by refining them, by discriminating on who is performing an action (the data subject or the data controller) and by inserting new actions to evolve a small-step semantics. Although we are exhaustive in capturing all the actions that may be triggered in the system, we do not claim to be exhaustive when capturing the consent variables. On the contrary, new variables could be added and defined to enrich the logic's expressiveness and to capture the environment where the logic is deployed. It is not the purpose of the paper to illustrate the full range of actions, rights and variables available in the logic; thus, only those actions, rights and variables that will be used for the formalisation of the case study will be explained in the relevant section.

4 Description of the Case Study

The Oxford Radcliffe Biobank (ORB) is a "resource of tissue and blood samples donated by patients for use in medical research" [11]. As a result, the ORB collects and stores samples in accordance with regulatory requirements and provides fair access to researchers in order to improve diagnosis and treatment, and ultimately patient care. The ORB case study offers interesting issues in terms of managing C&R controls in a context where sensitive information is handled, legislation imposes strict controls and patients' requests need to be addressed.

We identify a number of use cases that provide an overview of the environment where the system will be implemented and we have elicited from these a list of requirements to explore the implications of invoking C&R controls. The use cases that are formalised in this paper are:

- The EnCoRe IT administrator creates C&R options that will be presented to the patient both for the sample and the data.
- The EnCoRe IT administrator creates ORB privacy access control policies.
- The EnCoRe IT administrator creates ORB privacy obligation policies.
- The EnCoRe IT administrator sets CR default choices.
- The data subject (patient) or ORB technician makes CR choices for specific study/studies.
- The ORB technician is registering a sample and/or personal data in a spreadsheet.

5 Formalisation of the Use Cases

In this section we apply the logic to the use cases and we provide evidence that no more ambiguities emerge. In the formalisations below we use the letter a to denote the patients of the ORB, the letter b to denote the ORB itself and the letter c to denote a researcher.

We list in a hierarchy the data handled in this case study by creating three different domains in order to capture generic requirements (delete all data about my sample) by applying a single action to more that one datum, in a similar rationale to the one described in [7]. Each datum δ can only be allocated in one domain. With the letter δ we will generally refer to all data that may exist in this case study. We create three different domains and all δ data that may occur will belong only to one of these domains. With the letter δ_1 we denote the physical sample, with δ_2 the data derived from the sample and any measurements undertaken in the Biobank regarding the sample and with δ_3 we denote any personal data of the patient such as demographic data, name etc.

5.1 The EnCoRe IT Administrator Creates C&R Options That Will Be Presented to the Patient

In this use case the administrator writes the set of C&R options to be offered to the patients. The options derived from the current consent form that the patient

needs to sign before the donation of the sample to ORB, and from the results of the focus groups. We describe only the actions available that could be triggered in the system.

Based on the analysis of the focus groups, the options available to the patients concern the purpose for which the sample is given, the notification process and the ability to revoke and delete, the data created from the sample and the request for the distraction of the sample. Thus, the patient may give consent to the Biobank to store, process and share the data, constrain these choices by specifying the purpose of use and the parties that the data/sample will and will not be shared with. Furthermore, the patient could set notification preferences. It is not clear yet whether there will be an option allowing the patient to delegate consent to next of kin or update some of her/his data.

Define the Options for Sharing Data with Researchers. The first option that a patient may express is to allow the Biobank to share the sample to researchers and provide restrictions regarding the purpose of the research, the background of the researchers and the duration of consent. This is formalised as:

$$\mathbf{grant}^*(a, b, \delta, \Phi)$$

where $\Phi = \text{destination:}\Pi \wedge \neg\pi \wedge \text{purpose:}u \wedge \text{time duration:}t \wedge \text{times processed:}t^*$ and $\Pi \subseteq \{\text{Pharmaceutical, University}\}$, $\pi \subseteq \{\text{Insurance Companies}\}$and $p \subseteq \{\text{teaching, cancer research, DNA}\}$, $t \subseteq \{\text{One year - 40 years}\}$ and $t^* \subseteq \{\text{One time - 100 times}\}$. Notice that we use the letter δ for data, thus the options refer to all the available data. We could provide more options to the patient to choose from and apply these options only for a specific domain of data e.g the purpose for which the sample will be used for.

With this formalisation the patient may choose to share data with researchers working in university laboratories, with researchers working for pharmaceutical enterprises but not with researchers working for insurance companies. Furthermore, they could control the purpose of the research and choose to share their data with researchers for teaching purposes, for cancer research, and to allow or forbid DNA analysis. The variable Π includes the parties that the ORB is allowed to share data with, the variable π the parties that the ORB is not allowed to share data with and the variable u describes the purposes for which data may be shared. The ORB may also provide the option to the patient to choose the duration of consent and how many times the data/sample may be processed.

Define Revocation and Deletion Options. We distinguish four different classes of revocation symmetrical to what the patients give consent to.[2] These options could affect the Biobank, the researcher or both and could be enabled

[2] There are four different types of consent formalised in the logic, namely consent to store data, consent to process, consent to share one step further, consent to share transitively. Thus there are four types of revocation formalised in the logic symmetrical to those of consent, namely deletion of data, revocation of processing, revocation of sharing one step further, revocation of sharing transitively.

either with a prospective or a retrospective effect. The options for prospective revocation that a patient may ask for are presented below:

- Revoke permission to process sample /data from the Biobank after the Biobank has finished processing it.
- Revoke permission to share sample/data from the Biobank after the Biobank has finished processing it.
- Revoke permission to process sample/data both from the Biobank and the researchers after the completion of the research.
- Destroy the sample/ Delete data

Based on the same rationale, the different options of retrospective revocation offered are:

- Revoke permission to process sample /data from the Biobank before the Biobank has finished processing it.
- Revoke permission to share sample/data from the Biobank before the Biobank has finished processing it.
- Revoke permission to process sample/data both from the Biobank and the researchers before the completion of the research.
- Destroy the sample/ Delete data

The option of revoking consent and destroying/deleting the sample/data is formalised below. Whether the revocation is retrospective or prospective is defined by the value of the variable p. If the variable is true the revocation is retrospective, otherwise the revocation is prospective.

With the option below, the patient revokes the right to process the data/sample from ORB. If the Biobank is in the process of transferring the sample/data to other researchers they should withdraw it. Samples and data shared to researchers previous to that action are not influenced. The option is formalised as:

$$\mathbf{revoke}(a, b, \Phi, \delta)$$

where $\Phi =$ currently processed : p and $p = \{true\}$.

With the option below, the patient revokes from ORB the right to share data/sample. If the Biobank is in the process of transferring the sample/data to other researchers they should withdraw it. Samples and data shared to researchers previous to that action are not influenced.The option is formalised as:

$$\mathbf{revoke}^{*}(a, b, \Phi, \delta)$$

where $\Phi =$ currently processed : p and $p = \{true\}$.

With the option below, the patient revokes from ORB the right to process the data/sample and cascades these changes to the researcher. Samples and data shared to researchers prior to that option should not be processed further. The option is formalised as:

$$\mathbf{revoke}^{\dagger}(a, b, \Phi, \delta)$$

where $\Phi =$ currently processed : p and $p = \{true\} \implies$ obligation .

With this option the sample/data that is stored to ORB is deleted/destroyed. The researchers that have acquired samples/data prior to that action should delete that as well.

$$\mathbf{delete}^{\dagger}(a, b, \delta_2, \Phi^*)$$

where Φ^* = disposal:x/*wedge*currently processed : p and $x \subseteq$ {Destroy sample, delete data}and $p = \{true\} \implies$ obligation.

We describe a retrospective revocation as the patient wishes to revoke rights while the data is currently processed by inserting the variable p. We have introduced the variable x to provide the patient with the opportunity to decide either to destroy the sample or to delete the data that derived from that sample.

Some of the revocation actions create obligations and all of them require actions to happen in the future before the patient may invoke them. For example, a patient cannot revoke permission from researchers to process their sample unless the Biobank has initially shared their sample with them.

Change of Consent. The patient may also decide to change his initial consent. In this formalisation we capture the change of restrictions in the patient's initial consent. For example, changing the time allowed to process data to 5 months:

$$\mathbf{change}^{*}(a, b, \delta, \Phi)$$

where Φ = use by:t and $t = 150$.

Set Notification Options. There is also the option for a patient to be notified under certain conditions.

$$\mathbf{setnotify}(a, b, \delta, \Phi)$$

where Φ = notify-how:$n^* \wedge$notify-what:$n^{\dagger}$ and $n^* \subseteq$ {e-mail, general practitioner, ORB website} and $n^{\dagger} \subseteq$ {results of the research, implication for health, new research, sample dispatched to researcher, sample destroyed}.

In this formalisation the patient may choose to be notified either by e-mail, or via their General Practitioner (GP) or via the ORB's website. Furthermore, the patient may choose to be notified when the results of the research are published, if the researchers by examining the sample identified that there could be further implications to her/his health, when the sample is dispatched for research or when it is destroyed.

EnCoRe IT Administrator Creates ORB Privacy Access Control Policies. In order to create ORB privacy access control policies the administrator needs to: (1) define, or select one of the templates offered by EnCoRe, and (2) deploy them into the system. The privacy policies will be created from the allowed actions and the variables that will provide further constrains and information regarding the implementation of those actions. The suggested policies for the ORB case study are:

1. I {consent/revoke consent} for ORB to {collect/store/use} my personal data for { any research (provided it has been approved by ORB and met all ethical

standards of research); DNA specific research; selected clinical trials [list]; not at all} with access by {the research team that contacts me; pharmaceutical companies; others} (subject to time constraints/notification constraints).
2. I {consent/do not consent/revoke consent} for ORB to {collect/store/use} my {sample and associated digital representations} for {Specified purpose} (subject to time constraints/notification constraints)
3. I {consent/do not consent/revoke consent} for ORB to share my sample (or its digital representations) for {Specified purpose} to {direct contacts of the researcher, anyone} .
4. I {consent/do not consent/revoke consent} for ORB to share data for {any research (provided it has been approved by ORB and met all ethical standards of research); selected clinical trials [list]; only the research team that contacts me}.

In the logic all the actions and the variables create a policy. The policy describes how the system will cope with each action and each variable and the patients' choices will define the values of the variables.Thus, the option of

$$\mathbf{grant}^*(a, b, \delta, \Phi)$$

where $\Phi =$ destination:$\Pi \wedge \neg\pi \wedge$ purpose:$p \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Pharmaceutical, University} , $\pi \subseteq$ {Insurance companies}and $p \subseteq$ {teaching, cancer research, DNA}, $t \subseteq$ {One year - 40 years} and $t^* \subseteq$ {One time - 100 times} is converted into an EnCoRe policy as: I {consent} for ORB to {share} my sample for {teaching, cancer research, DNA} to {Pharmaceutical, University} and not to {Insurance Companies} for the next {100 years} or being processed {100} times.

Each of the actions described in the previous section will create a separate policy with the same rationale. Some of the options include obligations which will be formalised in the section below.

The EnCoRe IT Administrator Creates ORB Privacy Obligation Policies. To create ORB privacy obligation policies the administrator defines a set of obligation policy templates (and/or uses the templates offered by EnCoRe).

The privacy obligation policies are perceived in two different ways in the Hoare Logic:

- Those created to ensure that upon completion of an action, another action will be triggered in the future (notification requirements, delete data after 5 years).
- Those created by the obligation of a data controller to request further action from third parties with whom data has been shared, for the initial action to be completed (request to propagate C&R changes to all parties that process data).

The form of obligation policies for notification requirements is:

1. I {consent/do not consent/revoke consent} for ORB to contact me about my data or sample via {e-mail, phone, post, GP} when {my sample is shared, results of the research have gone public}.

The option for notification formalised as:

$$\mathbf{setnotify}(a, b, \delta, \Phi)$$

where Φ = notify-how:$n^* \wedge$notify-what:$n^\dagger$ and $n^* \subseteq$ {e-mail, general practitioner, ORB website} and $n^\dagger \subseteq$ {results of the research, implication for health, new research, sample dispatched to researcher, sample destroyed} will be converted into an obligation policy as: I {consent} for ORB to contact me about my data or sample via {e-mail, general practitioner, ORB website} when {results of the research are finalised, implication for health, new research, sample dispatched to researcher, sample destroyed}.

The EnCoRe IT Administrator Sets CR Default Choices. To set the default C&R choices the administrator accesses the EnCoRe "admin tool" box that propagates configuration changes to required components.

$$\{aO\delta_2 \wedge bR^\dagger \Phi \delta_2\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_2)$$
$$\{bL\delta_2 \wedge bP\delta_2 \wedge bS\delta_2 \wedge bR\Phi\delta_2\}$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$p \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Pharmaceutical, University} , $\pi \subseteq$ {Insurance Companies}and $p \subseteq$ {teaching, cancer research, DNA}, $t \subseteq$ {One year - 40 years} and $t^* \subseteq$ {One time - 100 times} and δ_3 is data regarding the patient's profile (registration number, further information).

5.2 The Data Subject (Patient) or ORB Technician Makes CR Choices for Specific Study/Studies)

For specific studies the data concern a sample only. We could also introduce a new variable to denote that the patient is participating in a study and he/she should acquire more options regarding for example the purpose of the sample.

$$\{aO\delta_1 \wedge bR^\dagger \Phi \delta_1\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_1)$$
$$\{bL\delta_1 \wedge bP\delta_1 \wedge bS\delta_1 \wedge bR\Phi\delta_1\}$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$p \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Team that consults me} , $\pi \subseteq$ {Anyone else}and $p \subseteq$ {DNA}, $t \subseteq$ {One year - 40 years} and $t^* \subseteq$ {One time} and δ_3 is data regarding the patient's profile (registration number, further information).

Patient Registration. When a new patient (data subject) registers, or more likely an authorised ORB employee is acting on his behalf by interacting with the ORB's system (Sapphire), it will assign a new patient ID, and then a new ID all linked to the new trial. This action is formalised as:

$$\{aO\delta_3 \wedge bR^{\dagger}\Phi\delta_3\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_3,)$$
$$\{bL\delta_3 \wedge bP\delta_3 \wedge bS\delta_3 \wedge bR\Phi\delta_3\}$$

where $\Phi = \text{destination:}\Pi \wedge \neg\pi \wedge \text{purpose:}p \wedge \text{time duration:}t \wedge \text{times processed:}t^*$ and $\Pi \subseteq \{\text{Pharmaceutical, University}\}$, $\pi \subseteq \{\text{Insurance Companies}\}$and $p \subseteq \{\text{teaching, cancer research, DNA}\}$, $t \subseteq \{\text{One year - 40 years}\}$ and $t^* \subseteq \{\text{One time - 100 times}\}$ and δ_3 is data regarding the patient's profile (registration number, further information).

During Tissue Sample Collection and Data Entry. This use case is very similar to the one described in the section above, with the difference that before locally storing (in Sapphire) the data subject's sample, the ORB technician is asked to define preferences for the new sample added. When an ORB technician collects a sample, he logs onto Sapphire, selects the Samples tab and clicks Add Sample. He then fills in the required Sample details on the Sample Data Entry Page. If the sample has already been received, the status of the sample is set to 'Received'.

$$\{aO\delta_1 \wedge bR^{\dagger}\Phi\delta_1\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_1)$$
$$\{bL\delta_1 \wedge bP\delta_1 \wedge bS\delta_1 \wedge bR\Phi\delta_1\}$$

where $\Phi = \text{destination:}\Pi \wedge \neg\pi \wedge \text{purpose:}p \wedge \text{time duration:}t \wedge \text{times processed:}t^*$ and $\Pi \subseteq \{\text{Pharmaceutical, University}\}$, $\pi \subseteq \{\text{Insurance Companies}\}$and $p \subseteq \{\text{teaching, cancer research, DNA}\}$, $t \subseteq \{\text{One year - 40 years}\}$ and $t^* \subseteq \{\text{One time - 100 times}\}$ Because it is a sample for trial we may want to introduce further variables to better describe the purpose of the research, such as DNA purposes. The patient may also choose to set notification requests.

5.3 The Data Subject (Patient) or ORB Technician Changes CR Choices for Specific Study/Studies

In this use case the ORB technician logs into Sapphire and changes on the patient's behalf her/his choices.

$$\{aO\delta_1 \wedge bR^{\dagger}\Phi'\delta_1 \wedge bR\Phi\delta_1\}$$
$$\mathbf{change}(a, b, \Phi, \Phi', \delta_1)$$
$$\{\neg bR\Phi\delta_1 \wedge bR\Phi'\delta_1\}$$

where $\Phi = \text{destination:}\Pi \wedge \neg\pi \wedge \text{purpose:}p \wedge \text{time duration:}t \wedge \text{times processed:}t^*$ and $\Pi \subseteq \{\text{University}\}$ and $p \subseteq \{\text{DNA}\}$, $t \subseteq \{\text{One year}\}$ and $t^* \subseteq \{\text{One time}\}$ $\Longrightarrow$ obligation.

With the action above the patient chooses to change her/his consent regarding the sample and to limit the parties having access to it, the duration of consent and the times that it is allowed to be processed. This action implies an obligation because when the existing consent is no longer valid the patient should choose to delete his/her sample or to change his consent.

5.4 Tissue Sample Collection, Data and C&R Choices Entry

When the sample is collected by the ORB, the patient expresses specific consent and revocation choices that should be enforced and respected by the ORB and any researcher that may acquire the sample.

$$\{aO\delta \wedge bR^{\dagger}\delta\Phi\}$$
$$\mathbf{grant}^{*}(a, b, \delta, \Phi)$$
$$\{bL\delta \wedge bP\delta \wedge bS\delta \wedge bR\delta\Phi\}$$

where Φ = destination : $\Pi \wedge$ purpose$p$ and $\Pi = \{\text{University}\}$ and $p = \{\text{cancer research, DNA}\} \implies$ obligation.

In the above formalisation the patient donates a sample to the ORB. The precondition declares that the donor of the sample is also the owner of the sample ($aO\delta$), and that the ORB must be willing to accept their C&R choices ($bR^{\dagger}\delta\Phi$, which will presumably be automatic, unless there are some forbidden combinations of tick-boxes on the form). The sample is then registered in the ORB. As a result in the post condition the ORB has stored the sample ($bL\delta$) and it may process it ($bP\delta$), share it ($bS\delta$) but must always respect the restrictions the patient has imposed. In this case, the sample may only be shared with the university laboratories, specifically for cancer research purposes and DNA analysis.

5.5 The ORB Technician Is Registering a Sample and/or Personal Data in a Spreadsheet

In this use case, the ORB is sharing data about a sample with a researcher in digital form. A researcher requests measurements of data from a sample and the Biobank provides the data in a spreadsheet. This is personal data, so all the controls that a patient may have imposed on the sample should be passed on.

$$\{bS\delta \wedge bR\delta\Phi \wedge c \in \cup\Phi.\text{destination} \wedge \Phi^{*} \leq \Phi \wedge cR^{\dagger}\delta\Phi^{*}\}$$
$$\mathbf{grant}(b, c, \delta, \Phi^{*})$$
$$\{cL\delta \wedge cP\delta \wedge cR\delta\Phi^{*} \wedge$$
$$\langle bNa\delta\rangle\ \mathbf{setnotify}(a, b, \delta, \Phi')\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \text{“}shared\text{”} \in \cup\Phi.\text{reason for notification}$$
$$\wedge \text{“}e-mail\text{”} \in \cup\Phi.\text{means of notification}\rangle$$
$$\langle \forall c.\mathbf{notify}(b, a, \delta, \text{“}shared\text{”}, \text{“}e-mail\text{”})\rangle true :\}$$

: where Φ^{*} = purpose : p and p=$\{$cancer research$\}$.

Note that we make use of a different action that only allows the researcher to process the data but not to share it. Also, all the controls that are appropriate for the right to process are cascaded to the researcher. Furthermore, as the patient had set notification requirements, an e-mail is sent to notify the patient that the data has been shared with the researcher.

The focus groups highlighted the importance of notification in such cases. If there are notification choices then there is an obligation triggered for the data controller to notify the patient. It is crucial to make a decision that will determine the strategy of notification. Possible options could be notification by e-mail, using a link on the Biobank's website pointing to the published papers of the researchers, or requesting the contribution of the patient's GP.

There will also be cases where the consent will not be specific enough for the Biobank to determine whether the sample should be shared or not. Therefore, the consent will be then implied. The patient could decide whether to be informed in such cases. Thus, depending on her/his choices she/he might choose to allow such sharing by default or to be asked for approval. However, the focus groups with researchers also pointed out that any notification may disturb the patients and their family.

6 Conclusion and Future Work

In a volatile environment, constantly changing due to innovations such as cloud computing, individuals disclose more personal data than ever. While these innovations allow them to acquire access to a number of services and products, the control they may have over their personal information is declining. The implications of such lack of control are profound when sensitive data and more specific medical data is handled by data controllers.

In this paper we have formalised requirements for a Biobank case study by applying a logic designed to handle C&R controls. We illustrate how diverse mechanisms of C&R may allow donors of the ORB to express their preferences and acquire control of their samples and data. Furthermore, with the formalisation of the requirements for such a system we effectively validated that patient choices are unambiguously described in order to be translated into privacy policies and enforced into the system.

Various avenues for future work could be followed. We intend to validate the C&R logic in a different real-world case study, namely the Identity Assurance Program organised by the UK cabinet office, in order to enhance our confidence in the general applicability of the logic. Subsequent research could apply the logic to translating privacy policies and form a strategy to resolve policy conflicts.

References

1. Agrafiotis, I., Creese, S., Goldsmith, M., Papanikolaou, N.: Reaching for Informed Revocation: Shutting Off the Tap on Personal Data. In: Bezzi, M., Duquenoy, P., Fischer-Hübner, S., Hansen, M., Zhang, G. (eds.) Privacy and Identity. IFIP AICT, vol. 320, pp. 246–258. Springer, Heidelberg (2010)

2. Agrafiotis, I., Creese, S., Goldsmith, M., Papanikolaou, N.: Applying Formal Methods to Detect and Resolve Ambiguities in Privacy Requirements. In: Fischer-Hübner, S., Duquenoy, P., Hansen, M., Leenes, R., Zhang, G. (eds.) Privacy and Identity Management for Life. IFIP AICT, vol. 352, pp. 271–282. Springer, Heidelberg (2011)
3. Agrafiotis, I., Creese, S., Goldsmith, M., Papanikolaou, N.: The logic of consent and revocation (2011) (in preparation)
4. Ashley, P., Hada, S., Karjoth, G., Powers, C., Schunter, M.: Enterprise privacy authorization language (epal). Research report, 3485 (2003)
5. Barth, A., Datta, A., Mitchell, J.C., Nissenbaum, H.: Privacy and contextual integrity: Framework and applications. In: 2006 IEEE Symposium on Security and Privacy, p. 15. IEEE (2006)
6. Becker, M.Y., Malkis, A., Bussard, L.: A framework for privacy preferences and data-handling policies. Technical report, Technical Report MSR-TR-2009-128, Microsoft Research (2009)
7. Bonatti, P.A., Damiani, E., De Capitani di Vemercati, S., Samarati, P.: A component-based architecture for secure data publication. In: Proceedings of 17th Annual Computer Security Applications Conference, ACSAC 2001, pp. 309–318. IEEE (2001)
8. Cranor, L.F.: Web privacy with P3P. O'Reilly Media (2002)
9. `http://www.encore-project.info`
10. Nissenbaum, H.: Privacy as contextual integrity. Wash. L. Rev. 79, 119 (2004)
11. `http://wyvern.ndcls.ox.ac.uk/orb/`
12. Tschantz, M., Wing, J.: Formal Methods for Privacy. In: Cavalcanti, A., Dams, D.R. (eds.) FM 2009. LNCS, vol. 5850, pp. 1–15. Springer, Heidelberg (2009)
13. Whitley, E.A.: Perceptions of government technology, surveillance and privacy: the UK identity cards scheme. In: New Directions in Surveillance and Privacy, p. 133 (2009)
14. Whitley, E.A.: Information privacy consent and the 'control' of personal data. Inform. Secur. Tech. Rep. (2009), doi:10.1016/j.istr.2009.10.001