PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2010) **368**, 4057–4072
doi:10.1098/rsta.2010.0169

Privacy compliance and enforcement on European healthgrids: an approach through ontology

By Hanene Boussi Rahmouni[1,*], Tony Solomonides[1], Marco Casassa Mont[2] and Simon Shiu[2]

[1]*Bristol Institute of Technology, UWE, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK*
[2]*HP Labs, Stoke Gifford, Bristol BS34 8QZ, UK*

The sharing of medical data between different healthcare organizations in Europe must comply with the legislation of the Member State where the data were originally collected. These legal requirements may differ from one state to another. Privacy requirements such as patient consent may be subject to conflicting conditions between different national frameworks as well as between different legal and ethical frameworks within a single Member State. These circumstances have made the compliance management process in European healthgrids very challenging. In this paper, we present an approach to tackle these issues by relying on several technologies in the semantic Web stack. Our work suggests a direct mapping from high-level legislation on privacy and data protection to operational-level privacy-aware controls. Additionally, we suggest an architecture for the enforcement of these controls on access control models adopted in healthgrid security infrastructures.

Keywords: privacy compliance; health; grid; ontologies; eXtensible Access Control Mark-up Language

1. Introduction

The protection of patient privacy in a pan-European healthgrid (Breton *et al.* 2005) infrastructure is challenging and requires combined solutions from legislation (European and national), technology (privacy-enhancing technologies), organizational (policies and protocols) and public frameworks (principles and ethics). A European open grid infrastructure is still a challenging goal to attain, but necessary for those nations wishing to collaborate to advance medical research and public health. The challenge arises primarily out of the lack of harmonization in the legal framework governing privacy and data protection in Europe, not least the European Data Protection Directive 95/46/EC (EU Directive 95/46/EC 1995). On top of this, there are significant conceptual and technical challenges in expressing, interpreting and deriving consequences of high-level policies. Finally, although a lot of attention is paid to security concerns such as infrastructure integrity and access control (typically authentication and authorization), these

*Author for correspondence (hanene2.rahmouni@uwe.ac.uk).

One contribution of 15 to a Theme Issue 'e-Science: past, present and future II'.

do not naturally extend to cover privacy concerns (often requiring context and purpose). As such, it is an ongoing challenge to search for ways to narrow the gaps between various aspects of the problem: legal, technical, social and organizational.

Our approach is an attempt to show that the use of semantic Web technology can allow the specification and enforcement of privacy requirements that traditional access control languages and mechanisms have failed to achieve. We start from the high-level regulations that govern privacy and data protection in Europe. Our main contribution is to integrate privacy constraints interpreted from high-level policies along with traditional security controls. Policy decisions cannot be deduced from data identifiers and access control conditions that are evaluated against their attributes' values. Privacy controls require more information about resources and hence the need to record metadata about the protected resources.

In previous works (Rahmouni *et al.* 2009*a*,*b*), we have described how semantic Web technologies have been used to classify the resources we intend to protect, described in terms of metadata within the ontology, and how different scenarios of data/resource sharing have been modelled within the same ontology. In this paper, we describe extensions to the previous model (necessary metadata added) and extend the data-sharing scenarios to include privacy policy contexts. We then show how this allows the specification and editing of privacy and access control policies in terms of existing concepts within the ontology. The paper is organized as follows: in §2, we present an overview of privacy requirements and harmonization issues in Europe and the impact of these on the transfer of medical data across European borders. Section 3 explains in detail the different steps we take to transform a high-level privacy policy into a system-level control through use of semantic Web modelling languages and formalisms. Section 4 specifies the architectural impacts driven by the aim of enforcing the newly designed privacy policies on existing access control architecture. A summary of related work and future plans concludes the paper.

2. Privacy requirements and law harmonization issues in Europe

The governance of personal data in Europe imposes certain obligations of regulatory compliance. By 'privacy requirements', we mean that all the obligations must be fulfilled by all parties involved in the process of sharing and processing sensitive patient data for medical purposes, including healthcare and medical research, to preserve informational aspects of the patient's privacy. This entails conceptual information about rights, obligations and consequent actions; among these are the obligation to obtain and to maintain patient consent; actions such as anonymization or pseudonymization and encryption as a surrogate for these; and rights such as those of the data subject to dissent or to be notified. This ontological variety leads us naturally to an ontology-based model. Our model must be sufficiently flexible to reflect any differences or, indeed, conflicts between EU Member States in the specification and the provision for these requirements. In the following paragraphs, we analyse a selection of these requirements.

(*a*) *Patient anonymity: de-personalizing the data*

The majority of data protection law in Europe is designed to govern and control the processing of personal data. At a first glance, the legal frameworks tend to ban the processing of personal data, but exemptions to the ban do exist to allow conditional lawful processing of the data. In cases in which the user is unable to meet the requirements for an exemption, then the only way for them to be able to process the data is by eliminating the personal aspect of the data. In our context, we are interested in those aspects of a patient's medical data that are classified as sensitive data, one category of personal, in data protection law. One means of de-personalizing the data is known as data anonymization, in which all data that may identify the data subject (the patient) are scrambled or eliminated. The degree of required anonymity may vary from one Member State to another as a consequence of differences in defining the term 'personal data' within the different legal frameworks (McCullagh 2006).

— Conflicts may arise on whether to consider anonymized or pseudonymized data as personal or not. In French law, if some anonymized data are to be disclosed to a third party, the data may be considered to be non-personal only if the sender does not have a further record or copy of the data which makes identification possible. Estonian anonymized data are not personal in all cases.
— Conflicts may arise in the handling of deceased persons' data, including in the way different Member States deal with human materials that are carriers of personal information (e.g. DNA samples or human tissue).

(*b*) *Patient consent*

We analyse certain features of consent in Member State legislation, with a view to a more general yet precise formulation. The situation is made more complex by the ambiguous use of the same terms. Some Member States, such as Poland, do not distinguish between consent and explicit consent. In other states, including Estonia, consent must always be an explicit informed consent. However, this does not mean that it has to be written. Other states, such as the Czech Republic, distinguish between consent to the processing of personal data and 'consent to (processing of) sensitive data' (confusingly termed *explicit consent*), which must always be written (Beyleveld *et al.* 2004). The UK law requires explicit consent when sensitive data are to be processed on condition of 'active communication between the parties', which need not be written (Information Commissioner's Office n.d.). The period of validity of consent might differ from one state to another; for example, in Estonia, once given, the consent for the processing of a deceased person's data could last up to 30 years after death (Beyleveld *et al.* 2004).

In the investigation of different European and national legal frameworks on consent for the processing of patient data, we identified certain general features of consent that are found in most of the examined frameworks. These features and consent requirements formed part of the issues investigated by policy-makers and ethicists across Europe, either in a data protection context or in a medical context, including cases of medical negligence and duty of care. Our analysis aims to be comprehensive rather than complete: the intention is not to have captured every nuance of every Member State law, but to have included all the principal

features that impinge on the processing of data. The practical cases in which these principles have been tested have invariably involved only a small number of Member States. These features and requirements are described as follows:

— the necessity of the consent for the processing of the data;
— the explicitness or expressiveness of patient consent (explicit and implicit);
— the specificity of the consent (for a specific processing task or as a general statement);
— the ways the consent may be collected (verbal and written);
— who can contact the data subject to get his/her consent;
— how consent should be documented (e.g. as paper record or in electronic form);
— the legal competence of the data subject;
— who can give consent instead of the data subject (next of kin, proxy or legal representative);
— consent validity lifetime;
— practicability of consent (practicable and impracticable); and
— miscellaneous others of less-wide applicability.

Most national frameworks, whether legal or ethical, have included the topics stated earlier within their vocabularies. However, harmonization of consent requirements is still far from complete. This is not only because some Member States have not included provisions for some of the requirements; rather, it is linked more to the diversity of definitions and interpretations from one Member State to another. Thus, although it appears that consent requirements across Europe may be classified under a standard taxonomy, the description or definition of different entities in the taxonomy would differ from one Member State to another. It is therefore important to harmonize on the *substance* of different laws and to map to appropriate local terms accordingly.

(*c*) *Purposes of processing*

According to Article 6(b) of the EU Directive 95/46/EC (1995) 'data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes'. Moreover, the same Article 6(b) states that 'further processing of data for [...] scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards'.

Therefore, there is an assumption of compatibility between the original (collection) purposes and further scientific purposes. However, according to Article 11.1 of the Directive, data subjects must be informed of the secondary use of their data: in particular, they should be informed about the identity of the controller and the purpose of the processing. This duty of information can be lifted only if provision of this information is impossible or would involve a disproportionate effort. In these cases, 'Member States shall provide appropriate safeguards' (Article 11.2).

The transfer or the disclosure of personal data to third parties (within the jurisdiction of one Member State or that of different Member States) is considered as a processing operation, and, as such, it is subject to legal requirements imposed on all processing. Thus, articulating the conditions, this transfer will only be allowed if:

The data subject has given his explicit consent to the processing of those data or when processing is necessary for a number of purposes that include:

— *to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent or*
— *the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.*

3. From high-level policies to operational-level access control

(*a*) *Step 1: ontology-based specification of access control policies using Web Ontology Language and Semantic Web Rule Language*

In this step, we start from a high-level policy extracted from a given European legal text on data protection, and we seek to simplify or abstract the policy by rewriting it in a way that allows easy formalization. To be more specific, we must rewrite the legal policy using its original natural language, but, by keeping only the minimal vocabulary representing the concepts and classifications necessary for preserving the meaning of the policy, we transform the policy to an if 'condition' then 'action' syntax. The vocabulary describing European data protection policies usually includes the following concepts.

(i) *User categories*

Data controller and data processor. According to Article 2(d) of the Directive, the controller is the natural or legal person (e.g. a company) who alone, or jointly with others, determines the purposes and the means of the processing of personal data. In every case, it is important to identify the data controller because 'he' is the one liable for the legality of the processing. He also has to fulfil obligations towards his national data protection authority and towards the data subjects. In contrast, according to Article 2(e) of the Directive, the processor is the natural or the legal person, public authority, agency or any other body that processes personal data on behalf of the controller. This will typically be a specialized third-party company entrusted by the controller to conduct technical aspects of the processing, such as the sorting or the combination of the personal data.

Data subject. The data subject is generally defined as the person to whom the personal data relate. Sometimes, when the data subject is not considered as legally competent owing to his age or mental status, it is possible to delegate consent to a surrogate or legal representative that is nominated by law, by the family or by other good practice depending on the Member State's national framework.

(ii) *Data categories*

According to Article 2(a) of the Directive, the term personal data relates to any information relating to an identified or identifiable natural person. According to this definition, personal data may refer to any information that can allow

direct identification of the data subject; in other terms, data that can be easily related to a data subject and reveal his identity. This is the case with data such as names, addresses, date of birth or even medical and genetic data, as well as physiognomy or other physical traits. These data when combined allow the identification of a data subject with a small margin of doubt. *Indirect identification* requires further steps to make a link between a specific person and the data being processed.

(iii) *Data-processing actions*

The concept of processing is very broad. It covers any operation or a set of operations that is performed upon personal data, whether or not by automatic means. In this frame, data processing is considered to be the collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction of personal data. Data protection legislation covers both automated processing and non-automated processing. Both types of processing operations need to form part of a filing system or to be intended to form part of a filing system, i.e. 'any structured set of personal data which are accessible according to specific criteria, whether centralized, decentralized or dispersed on a functional or geographical basis'.

(iv) *Purposes*

The European Directive on Data Protection and other national frameworks have stated several generic processing purposes for which the processing of personal data is allowed, as described in Article 8.2:

— *the processing of personal data is permitted for any purpose where the data subject has given his explicit consent to the processing of those data. Except where the laws of the Member State provide that the prohibition of processing may not be lifted by the data subject's giving his consent; or when processing is necessary for other purposes including*
— *if processing is necessary to protect the vital interests of the data subject or of another person when the data subject is physically or legally incapable of giving his consent; or*
— *if processing is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services, and when those data are processed by a health professional subject to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.*

The European Directive offers the opportunity to the Member States to add exemptions to those listed earlier, for reasons of substantial public interest. These exemptions should be subject to suitable safeguards. For instance, under Article 8.4 of the Directive, national exemptions might be adopted for scientific research.

(v) *Obligations*

Obligations are further steps that the processor of the data or the data controller has to take once the processing has been allowed. They are considered as further safeguards that the user of the data must agree to, such as secondary usage safeguards concerning the way the data would be handled after the processing.

Retention. This is to comply with the fifth principle of the directive indicating that 'Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes'.

Disposal. This is to explain the process the data would undergo to be deleted once the lifetime period has expired including careful examination and selection of which data should be disposed of and which may or must be kept.

Providing proper notice to the data subject. In cases in which the requirement of consent could be alleviated for reasons such as impracticality of the task of collection, the data subject continues to have the right to be notified of the collection and the processing of his data and should be given the choice to withdraw his consent.

(*b*) *Step 2: transforming high-level policies to simple semi-structured rules*

Similar to the work done in Powers *et al.* (2004) that simplifies policies dictated by the US Health Insurance Portability and Accountability Act, the policies specified by European data protection legislation could be expressed in simplified ways using the different concepts that constitute the vocabulary described earlier:

A [*user category*] *should be* [*allowed*] *the ability to perform* [*action*] *on* [*data category*] *for* [*purpose*] *under* [*conditions*] *yielding an obligation to* [*obligation*].

For example:

[*A processor*] *is allowed* [*to process*] [*sensitive personal data*] *for* [*any purpose*] *provided that* [*the processing is fair and lawful*], *and, in particular*, [*carried out with appropriate safeguards for the rights and freedoms of data subjects*].

Or in a slightly modified 'if then' syntax:

If [*a processor*] *is required to* [*process*] [*sensitive personal data*] *for* [*any purpose*] *provided that* [*the processing is fair and lawful*] *then* [*the processing*] *is* [*allowed*] *and should be* [*carried out with appropriate safeguards for the rights and freedoms of data subjects*].

Therefore, the abstracted high-level policies could be transformed to an 'if condition then consequence' rule as follows:

If

[*Condition on User*], [*Condition on data*], [*Condition on Purpose*], [*Condition on Other*]

Then

Allow [*action*] *and Ask for* [*Obligation*]

The first part of the rule (the condition part) aims to capture information about the context of sharing. This information would be described as a set of constraints applied on different entities involved in the context. All these conditions must be

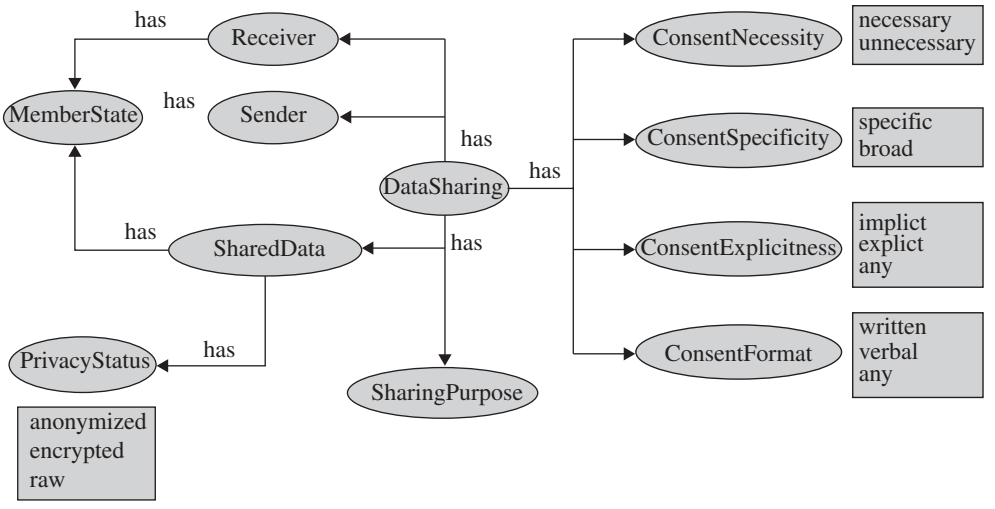

Figure 1. Ontology model of data-sharing contexts and requirements.

satisfied in order for the consequential permission and action specified by the rule to be achieved. Formalization of such rules requires the use of a Horn-like rule language that allows reasoning on context information stored in a model or ontology in order to infer tasks or actions to be fulfilled by some entities forming part of the context of sharing. In the next section, we will show how the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL) may be used to fulfil these requirements.

(*c*) *Step 3: modelling data disclosure contexts and the applicable privacy-aware access control policies*

The range of information that needs to be captured and the description of the context of sharing requires the use of a sufficiently complex model that allows contexts of data sharing from the real world to be specified through specialization of generic data-processing cases from high-level directives and regulations. The model must be built through the use of a modelling language able to abstract the generic concept of data-sharing context or scenario by specifying its components through the use of interrelated conditions or constraints. Semantic Web technologies and specifically the OWL (McGuinness & van Harmelen 2004) and the SWRL (Joint US/EU Ad hoc Agent Markup Language Committee 2004) have the ability, respectively, to specify and capture context information within a very complex domain using a highly expressive logic language and to specify the horn-like rules by reasoning on OWL axioms so as to infer an access control decision and privacy obligation that is specified as additional information of the context. As illustrated in figure 1, OWL allows us to model the conceptual domain of 'data sharing' or 'data disclosure' and its components as a hierarchy of classes and subclasses and a corresponding hierarchy of properties to represent the relationships between them. A class or concept is any physical or conceptual item that we can identify in the domain: for example, a class

might be *Sharing*, *SharingRequest*, *SharingParty*, *SharedData* and so on. Based on RDF (Brickley *et al.* 2004), an RDF-Schema (RDFS; Brickley *et al.* 2004) model can be represented as a list of statements of the form *class-property-class* or *class-property-literal*. Privacy requirements such as consent requirements could be modelled as RDFS classes and assigned to the *dataSharing* resource as object properties: for example, *dataSharing-hasConsentNecessity-consentNecessity*.

OWL, which is built on RDFS, has the ability to describe classes in more complex ways than RDFS alone, with additional class expressions (such as class restriction and enumeration). For example, consider a cardinality constraint, such as a data item must have one and only one place of origin. OWL is more appropriate for the modelling of the domain of interest than RDFS as it allows more complex relationships to be expressed between classes, including cardinality restrictions. Moreover, OWL provides additional features to allow overlapping models of a concept to be merged, even when different naming conventions have been used for the same resource; for example, 'Explicit Consent' in one model may be termed 'Express Consent' in another, but the two concepts have the same meaning.

The 'if then' rules specifying high-level policies described in the previous section could be translated to access control policy using SWRL, an extension of RuleML (Joint US/EU Ad hoc Agent Markup Language Committee 2004) with OWL features. The SWRL condition and consequence clauses of the rules are formulated using OWL axioms described in the ontology model of data-sharing contexts. The policy is applicable if the context described in the condition clause holds. One advantage of storing the rules along with the ontology model is that it allows an easy check on whether the conditions are satisfied. If the context is the case, the access control decision described by the OWL axioms in the consequence clause of the rule will be stored in the ontology as new inferred properties of the 'Data Sharing' instance. The following example is an SWRL representation of a high-level policy specifying that 'In the UK, if patient consent is required for the disclosure of their data and the patient has provided his informed consent for the purpose of sharing then the sharing is allowed'. Here we are assuming that another rule, similar to the privacy requirement rules presented in our previous work (Rahmouni *et al.* 2009*b*), will be fired up first in order to infer the necessity of acquiring patient consent for this type of sharing.

$$\begin{aligned}&\text{dataSharing}(?x)\\ &\wedge\ \text{hasPurpose}(?x, ?p)\\ &\wedge\ \text{concerning}(?x, ?data)\\ &\wedge\ \text{belongsTo}(?data, \text{UK})\\ &\wedge\ \text{isForPatient}(?data, ?pt)\\ &\wedge\ \text{hasConsentNecessity}(?x, \text{Necessary})\\ &\wedge\ \text{provided}(?pt, \text{InformedConsent})\\ &\wedge\ \text{consentedFor}(?pt,?p)\\ &\qquad\rightarrow\ \text{hasSharingDecision}(?x, \text{allow})\end{aligned}$$

In the next section, we describe, referring to a case study, how OWL ontology and the semantic rules we have created could be used to specify privacy-aware access control decisions while sharing medical data within a healthgrid context.

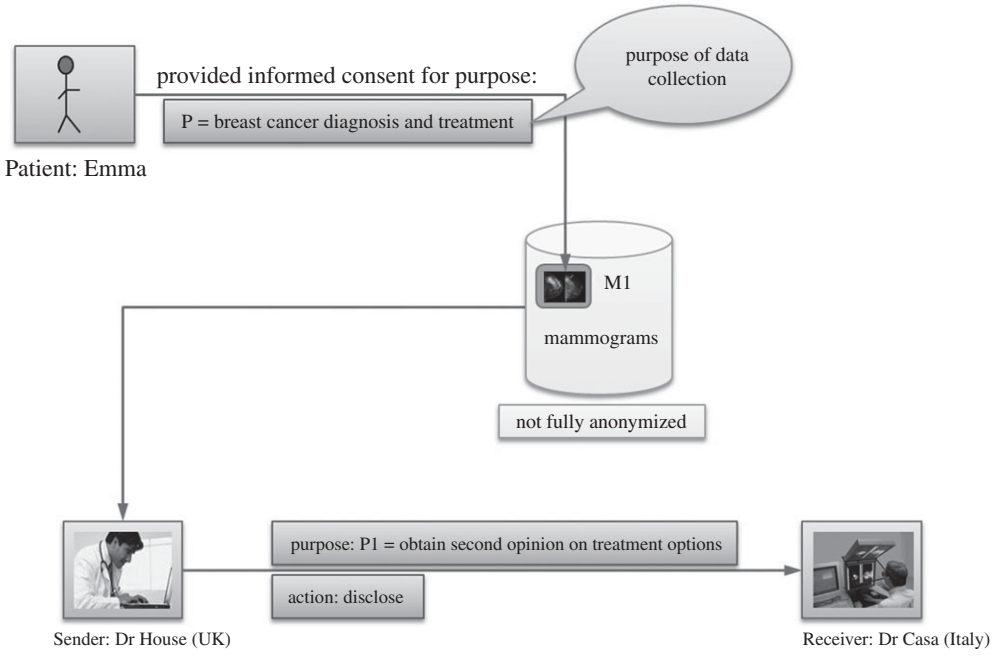

Figure 2. Data-sharing scenario: getting second opinion on patient treatment.

(i) *Case study—data-sharing scenario: getting second opinion on patient treatment*

The case study described in figure 2 presents a data-sharing scenario in which the data are shared between two nodes of a European healthgrid infrastructure. It involves the sharing of a patient's mammogram collected by the NHS Breast Screening Programme at a UK hospital from a patient, 'Emma', for the purposes of diagnosis and treatment. Emma provided her consent at the time of collection. 'Dr House', the radiologist who is taking responsibility for Emma's case, needs a second opinion from 'Dr Casa', a colleague at an Italian hospital. The mammogram cannot be fully anonymized before sending as some personal data about Emma are necessary in order to make an accurate judgement about the treatment plan. Analysing the legal context of this sharing scenario, we note that the mammogram is to be shared for a different but compatible purpose with the original purpose of collection. Recalling the statement in Article 6(b) of the European Data Protection Directive that 'Personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards', then, in order to comply with the purpose compatibility principle, we must check that the following conditions are satisfied before allowing the sharing:

— *if the mammogram was collected for a specified, explicit and legitimate purpose and*
— *if the secondary purpose is compatible with the original purpose of 'breast cancer diagnosis and treatment'.*

As a safeguard of the patient's right, Dr House may need to send the hospital 'secondary data usage policy' along with the mammogram in order to express and associate the UK's privacy requirements on how this mammogram may be used for further purposes. If Dr Casa wishes to use Emma's mammogram in the future, he will need to comply with UK policy. Indeed, this safeguard may be required of Dr House, e.g. through a disclaimer message at the time of allowing the sharing of the mammogram. An SWRL representation of the policy required for this case is presented in the following example (we use the suffixes p and p1 for space restriction purposes only):

dataSharing(sharing1)
∧ hasSender(sharing1, Dr House)
∧ hasReceiver(sharing1, Dr Casa)
∧ hasPurpose(sharing1, p:SecondOpinionOnTreatment)
∧ locatedIn(Dr House, UK)
∧ locatedIn(Dr Casa, Italy)
∧ concerning(sharing1, M1)
∧ belongsTo(M1, UK)
∧ isForPatient(M1, pt1)
∧ provided(pt1, InformedConsent)
∧ hasConsentPurpose(M1, p1:BreastCancerDiagnosisAndTreatment)
∧ compatibleWith(p, p1)
→ hasSharingDecision(sharing1, allow)
∧ hasObligation(sharing1, attachSecondaryUsePolicy)

(*d*) *Step 4: extending the ontology to enable the specification of enforceable privacy policies in order to ensure compliance*

Good governance of a European integrated grid for health (http://www.eu-share.org/) demands that legal and ethical requirements for privacy be enforced at the operational level as formal privacy policies to complement traditional security policies focusing on who can access what, based on access rights. In the previous steps, we have specified high-level privacy policies using a rule-based approach. Enforcement of these rules would require an adequate enforcement architecture and would drive major changes to existing access control architectures. The enforcement of semantic Web access control rules on existing access control architectures such as the eXtensible Access Control Mark-up Language (XACML) requires the transformation of our rules into policies. This requires an extension to our privacy ontology by adding necessary classes in order to represent the concept of a 'Policy' and to edit policy instances in a seamless way. In order to be easily enforced at the infrastructure level, we suggest that our policies should eventually be specified in a way that conforms to a widely adopted policy language or standard that has proven efficiency in the enforcement of privacy policies, hence our choice was XACML (OASIS 2005).

Privacy policies (PRIME 2007) are divided into two main categories according to their chronological order of enforcement compared with the traditional security and access control policy associated with them. The first category of policies comprises those that may affect a system *access control decision* or might cause

access permissions to roll back. For example, if patient consent is required for a specific context of sharing, the system will check for availability of such patient consent before allowing the user to access their data. The second category is *privacy obligations* (PRIME 2007), which do not affect access control decisions but are rather dealt with after access to control usage of data at a later stage, i.e. usage of data for a secondary purpose, disclosure of data to third parties, notifications, data deletion and retention. These may, respectively, be regarded as pre- and post-conditions on the access action.

Privacy policies in XACML are specified using some standard extensible mark-up language (XML) elements, including triples of (Policy, Target, ListOfRules), where the Target refers to the resource we are controlling access to and the Rules attached to the Policy are described in terms of other standard elements of XACML including Rule Effect (permit, deny, defer, refer, etc.), Rule Target (Subject or Requester, Resource, Requested Action) and Rule Conditions (OASIS 2005). The conditions attached to each rule are specific constraints on the subject or requestor, resource and others, depending on the context. XACML also allows users to add more user-defined components or elements to the traditional vocabulary (OASIS 2005).

Example of Purpose Compatibility Rule. In this example, we show how we model the privacy policy stating that:

A user may access a patient mammogram if patient has provided informed consent for a specific purpose of processing and the processing purpose is compatible with the purpose consented for.

First, we have rewritten the policy as an SWRL rule using the OWL classes and properties; specified in the ontology, the rule is as follows:

$$
\begin{aligned}
&\text{dataSharing(?x)}\\
&\wedge\ \text{hasSender(?x, ?s)}\\
&\wedge\ \text{hasReceiver(?x, ?r)}\\
&\wedge\ \text{hasPurpose(?x, ?p)}\\
&\wedge\ \text{locatedIn(?s, UK)}\\
&\wedge\ \text{locatedIn(?r, Italy)}\\
&\wedge\ \text{concerning(?x, ?m)}\\
&\wedge\ \text{belongsTo(?m, UK)}\\
&\wedge\ \text{isForPatient(?m, ?pt)}\\
&\wedge\ \text{provided(?pt, InformedConsent)}\\
&\wedge\ \text{hasCollectionPurpose(?m, CollectionPurpose)}\\
&\wedge\ \text{compatibleWith(?p, CollectionPurpose)}\\
&\qquad\rightarrow\ \text{hasSharingDecision(?x, allow)}\\
&\qquad\quad\wedge\ \text{hasObligation(?x, attachSecondaryUsePolicy)}
\end{aligned}
$$

For easy mapping to an XACML rule, the SWRL rule has to be specified in terms of attributes of only the generic entities that constitute an XACML Rule Target (discussed earlier) and other elements that are used to specify the general policy that the rule in question belongs to, i.e. the purpose of processing. The OWL property 'provided (Patient, Informed Consent)' is a property of the patient whose data are to be shared and indicates that the patient has provided informed consent. The patient or the data subject is not one of the XACML 'Rule Target'

components; therefore, we express the same condition in terms of property of the class 'O' (representing the data or the object required by the subject). The result is as follows:

$$\begin{aligned}
&\text{hasRuleContext}(?r, ?rc)\\
&\wedge\ \text{hasContextSubject}(?rc, ?s)\\
&\wedge\ \text{hasContextObject}(?rc, ?o)\\
&\wedge\ \text{hasContextAction}(?rc, \text{send})\\
&\wedge\ \text{hasContextPurpose}(?rc, ?p)\\
&\wedge\ \text{hasRole}(?s, \text{doctor})\\
&\wedge\ \text{consent}(?o, \text{true})\\
&\wedge\ \text{hasConsentType}(?o, \text{InformedConsent})\\
&\wedge\ \text{hasConsentPurpose}(?o, ?cp)\\
&\wedge\ \text{compatibleWith}(?p, ?cp)\\
&\wedge\ \text{sendingTo}(?s, ?rec)\\
&\wedge\ \text{locatedIn}(?s, \text{UK})\\
&\wedge\ \text{hasReceiverLocation}(?s, \text{Italy})\\
&\qquad\rightarrow\ \text{hasRuleEffect}(?r, \text{allow})\\
&\qquad\quad\wedge\ \text{hasObligation}(?r, \text{attachSecondaryUsePolicy})
\end{aligned}$$

4. Enforcement architecture

Our approach includes two enforcement modes, which depend on the number of data records involved. A *unique privacy context mode* is involved when only one record of data is subject to a sharing request; *a multi-privacy context mode* is involved when large amounts of data need to be shared. In this latter case, the contexts might differ in terms of their associated privacy constraints. For example, in the case of a request to share a large set of mammograms, some of them may have the consent attribute set to true and others have the consent attribute set to false. We have added a context handler component to the current XACML access control model. This access control model relies on a Policy Enforcement Point (PEP) to enforce policy decisions made by a Policy Decision Point (PDP; OASIS 2005). Our component will answer requests from the PEP in order to retrieve values of attributes specified within the access/sharing request. In both modes, the context handler will interact with the ontology model of data sharing in order to obtain the values of required attributes. In the first mode, as shown in figure 3, these values will be passed to the PDP in an attribute-value map. The PDP will work out the applicable policy to the request context and will return an access control decision and a related obligation (if any) to the PEP. The PEP will interact with specific services of the infrastructure to enforce the decision on the requested data and to deal with the obligation.

In the second enforcement mode, the context handler will identify the general context of sharing using the attributes sent by mapping the attributes sent by the PEP to those described in the ontology. The context handler can then identify different subcontexts of the general context passed by the PEP and return different instances of the original context to the PEP. The PEP will invoke a data

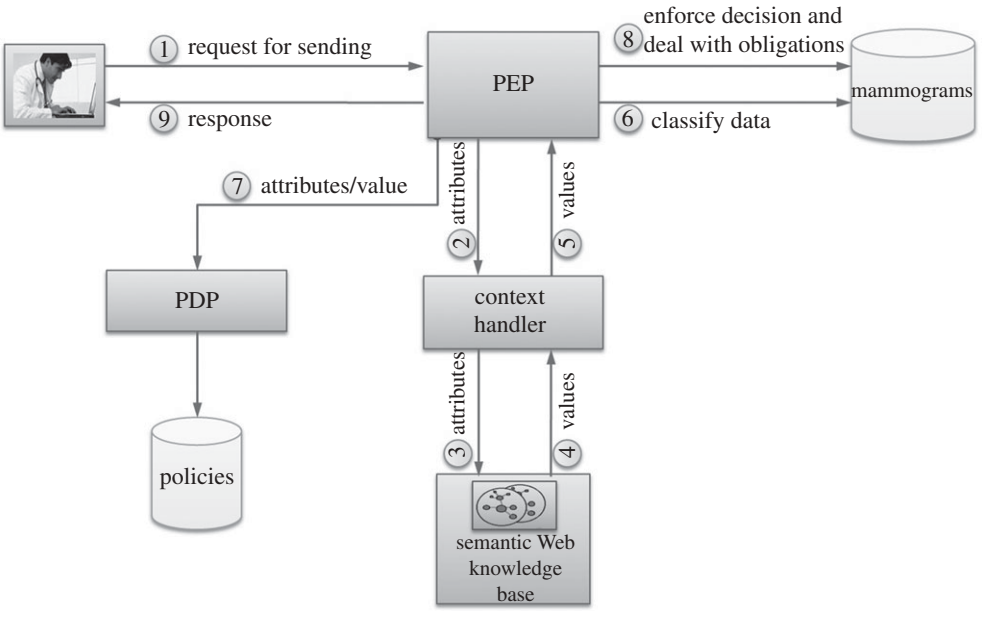

Figure 3. Extended XACML access control model/data flow.

classification service to divide the data into subsets, conforming each to the same properties and types of privacy constraints. The PEP will make several calls to the PDP in a sequential way to obtain authorization decision and obligation, if any, for each set of data. For example, consider three patient records, Rec1, Rec2, Rec3, and that only two of them, Rec1 and Rec3, have the value of the consent attribute set to true. At a primary assessment, the access request to the three records would be classified under the general context A:{subject: Dr House; Object: Mammogram1; Action: Send; BreastCancerResearch; Sendingto: Dr Casa}. Then, the context handler will identify two subcontexts as follows: A1:{subject: Dr House; Object: Mammogram1; Action: Send; BreastCancerResearch; Sendingto: Dr Casa; consent: true} and A2:{subject: Dr House; Object: Mammogram1; Action: Send; BreastCancerResearch; Sendingto: Dr Casa; consent: false}. The PEP will then classify Rec1 and Rec3 under A1 and Rec2 under A2. In this case, the PEP will call the PDP twice; first in order to work out the sharing decision for Rec1 and Rec3 and second in order to work out the decision for Rec2.The remaining steps of the data flow will continue as in the first mode of enforcement.

5. Conclusion and future work

The work reported here has been based in part on experience in current projects. Policy rules have been implemented in the Protégé 3.4 toolkit (http://protege.stanford.edu/) and have been tested by being passed to the Jess rule engine (http://www.jessrules.com/jess/) over an SWRL bridge. Details of this can be found in Rahmouni *et al.* (2009*c*). Although a full implementation

of the approach has not yet been incorporated into a working healthgrid system, there are specific examples that have been based on such a design. In the Health-e-Child project, access to UK hospital patient data has been filtered through a process that notifies a local 'data controller' of any request to access by a European partner. Permission is granted on a time-limited basis to minimize repeat requests compatible with the purpose of the access. In the neuGRID project, similar issues are under consideration but with the added requirement that imaging scientists' workflows may themselves be 'protected objects'. A full application and implementation of this method has been incorporated into the design of a proposed 'VPH e-Lab'.

Other research using semantic access control policies includes the work done for projects looking at enhanced handling of identifiable data and privacy enforcement on pan-European IT infrastructures, including the EU PRIME (PRIME 2007) and PRIMELife (http://www.primelife.eu/) projects and the UK TSB EnCoRe project (http://www.encore-project.info/). These projects have addressed several issues of privacy management and the specification and enforcement of privacy policies as privacy-aware access control and obligations. Our work has the special focus of explicitly addressing how to map high-level privacy policies to enforceable ones. Interesting work has also been published on the extensions to XACML policy language and architecture (Priebe *et al.* 2006; Demchenko *et al.* 2008). These approaches do not investigate the problem of enhancing privacy compliance in the design of security policies and did not look at an attempt at closing the gap between high-level legislation and system-level controls and post-condition obligations. Enterprise Privacy Authorization Language (Powers *et al.* 2004) was designed to enable the translation of privacy policies into an XML-based computer language. The resulting coded translation of human policy into IT policy allows complex description of necessary internal data-handling practices to enforce the privacy policy. In this work, only the specification of privacy obligations was provided but with no attempts to integrate conditions that ensure meeting privacy requirements and to enforce them as access controls.

In our research, we have managed to model high-level policies interpreted from European and national data protection law as privacy-aware access control policies. The use of semantic Web technologies such as OWL and SWRL allowed integration of privacy requirements highlighted in text law such as requirements of consent and other safeguards of patient rights as policy constraints. For ease of enforcement at the security architecture of highly distributed infrastructures such as the grid, we have used mapping templates to transform the semantic Web access control policies into a de facto and highly portable standard of access control, which is XACML.

The management of privacy and personal information within a multi-cultural domain such as European data grids and universal collaborative systems requires intrinsic compliance-checking and assurance modules in order to increase public and social trust and acceptance (http://www.eu-share.org/). The use of ontologies and semantic technologies can provide relatively easy interpretation of legislation at an operational level and can allow the recording or logging of data access events, and the way the system has managed these events, to be used as proof of compliance or liability for privacy breach accidents by auditors. These assumptions will be investigated in future work.

The first two authors wish to acknowledge the European Commission and the HealthGrid association for supporting this work at different stages. Colleagues in the SHARE, Health-e-Child and neuGRID projects have helped us refine our ideas through discussion of actual requirements in practical settings.

References

Beyleveld, D., Townend, D., Rouille-Mirza, S. & Wright, J. 2004 *Implementation of the Data Protection Directive in relation to medical research in Europe*. Farnham, UK: Ashgate.

Breton, V., Dean, K. & Solomonides, T. (eds) 2005 *The HealthGrid White paper*. See http://www.healthgrid.org/documents/pdf/HealthGrid-whitepaper-full.pdf.

Brickley, D., Guha, R. V. & McBride, B. (eds) 2004 *RDF Vocabulary Description Language 1.0*. RDF Schema, W3C Recommendation, 10 February. See REC-rdf-schema-20040210/ REC-rdf-schema-20040210/.

Demchenko, Y., de Laat, C., Koeroo, O. & Sagehaug, H. 2008 Extending XACML authorisation model to support policy obligations handling in distributed applications. In *Proc. 6th Int. Workshop on Middleware for Grid Computing* (*MGC 2008*), *Leuven, Belgium*, *1 December 2008*. ISBN: 978-1-60558-365-5. See http://portal.acm.org/citation.cfm?id=1462704.1462709.

EU Directive 95/46/EC. 1995 *The Data Protection Directive*. See http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf (last accessed 22 April 2010).

Information Commissioner's Office. n.d. *Data Protection Guide: the principles of the Data Protection Act in detail*. See http://www.ico.gov.uk/for_organisations/data_protection_guide.aspx.

Joint US/EU Ad hoc Agent Markup Language Committee. 2004 *SWRL: a semantic Web rule language combining OWL and RuleML*. See www.w3.org/Submission2004/SUBM-SWRL-20040521/ (last accessed 18 June 2009).

McCullagh, K. 2006 A study of data protection: harmonization or confusion? In *Proc. 21st BILETA Conf. Globalisation and Harmonisation in Technology Law, Malta, 6–7 April 2006*. See http://www.bileta.ac.uk/ pages/Conference%20Papers.aspx (last accessed 22 April 2010).

McGuinness, D. L. & van Harmelen, F. 2004 *OWL Web Ontology Language Overview2*, W3C Recommendation, 10 February. See www.w3.org/TR/owl-features/ (last accessed 18 June 2009).

OASIS 2005 *Privacy policy profile of XACML v2*. OASIS Standard. See http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-privacy_profile-spec-os.pdf (last accessed 18 June 2009).

Powers, C., Adler, S. & Wishart, B. 2004 *EPAL translation of the Freedom of Information and Protection of Privacy Act* [*version 1.1*], 11 March, IBM. See www.ipc.on.ca/images/Resources/up-EPAL_FI1.pdf (last accessed 22 April 2010).

Priebe, T., Dobmeier, W. & Kamprath, N. 2006 Supporting attribute-based access control with ontologies. In *Proc. 1st Int. Conf. on Availability, Reliability and Security, ARES 2006, Vienna, Austria, 20–22 April 2006*. Piscataway, NJ: IEEE. (doi:10.1109/ARES.2006.127)

PRIME: Privacy and Identity Management for Europe. 2007 *PRIME architecture—version 2*. See https://www.prime-project.eu/prime_products/reports/.

Rahmouni, H. B., Solomonides, T., Mont, M. C. & Shiu, S. 2009*a* Privacy compliance in European healthgrid domains: an ontology-based approach. In *Proc. 22nd IEEE Int. Symp. on Computer-Based Medical Systems, CBMS 2009, Albuquerque, NM, 3–4 August 2009*. (doi:10.1109/CBMS.2009.5255423)

Rahmouni, H. B., Solomonides, T., Mont, M. C. & Shiu, S. 2009*b* Modelling and enforcing privacy for medical data disclosure across Europe. In *Proc. MIE 2009, Medical Informatics in a United and Healthy Europe, XXIInd Int. Congr. European Federation for Medical Informatics*, *Studies in Health Technology and Informatics*, *Sarajevo*, vol. 150 (eds K.-P. Adlassnig, B. Blobel, J. Mantas & I. Masic), pp. 695–699. Amsterdam, The Netherlands: IOS Press.

Rahmouni, H. B., Solomonides, T., Mont, M. C. & Shiu, S. 2009*c* Ontology-based privacy compliance on European healthgrid domains. In *Proc. 11th Int. Protégé Conf., Amsterdam, The Netherlands, 23–26 June 2009*. See http://protege.stanford.edu/conference/2009/abstracts/S13P1Boussi.pdf for extended abstract (last accessed 22 April 2010).