Privacy Compliance in European Healthgrid Domains: An Ontology-Based Approach

Hanene Boussi Rahmouni[1a], Tony Solomonides [a],
Marco Casassa Mont [b], Simon Shiu [b]
[a] *Bristol Institute of Technology, UWE, Bristol/BS16 1QY*
[b] *HP Labs, Stoke Gifford, Bristol/BS34 8QZ*

Abstract. The integration of different European medical systems by means of grid technologies will continue to be challenging if technology does not intervene to enhance interoperability between national regulatory frameworks on data protection. Achieving compliance in European healthgrid domains is crucial but challenging because of the diversity and complexity of Member State legislation across Europe. Lack of automation and inconsistency of processes across health-care organizations increase the complexity of the compliance task. In the absence of automation, the compliance task entails human intervention. In this paper we present an approach to automate privacy requirements for the sharing of patient data between Member States across Europe in a healthgrid [1] domain and ensure its enforcement internally and within external domains where the data might travel. This approach is based on the semantic modelling of privacy obligations that are of legal, ethical or cultural nature. Our model reflects both similarities and conflicts, if any, between the different Member States. This will allow us to reason on the safeguards a data controller should demand from an organization belonging to another Member State before disclosing medical data to them. The system will also generate the relevant set of policies to be enforced at the process level of the grid to ensure privacy compliance before allowing access to the data.

Keywords. Privacy, Healthcare, EU, Grid, OWL, Rules, Biomedical Research

Introduction

When sharing medical data between different healthcare and biomedical research organisations in Europe, it is important that the different parties involved in the sharing handle the data in the same way indicated by the legislation of the member state where the data was originally collected as the requirements might 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 of Member States. While most EU Member States are now governed by similar personal data protection rules, harmonization remains more

[1] Corresponding author: Hanene2.Rahmouni@uwe.ac.uk

978-1-4244-4878-4/09/$25.00 ©2009 IEEE

apparent than real. This is due first to the fact that subject to the provision of suitable safeguards the directive leaves some space for Member States to lay down simplifications and exemptions to some of the obligations that are dictated [1], e.g. the obligation to notify the data subject of the processing of their data. Also for reasons of substantial public interest, Member States may lay down exemptions to the ban on processing of sensitive personal data in addition to those laid down in the directive, either by national law or by decision of the supervisory authority [2]. Second, as noted by some studies [3], the definitions used do not lead to a uniform understanding of the key concepts underpinning the directive. Focusing on the concept of "Personal Data", many Member States find it difficult to interpret. The UK found that in some cases data is not easily classified as personal or non personal. And this classification could be relative according to the circumstances. Some data that is normally considered as non personal has been shown to be capable of being personal data in special circumstances, like shoe size, details of death, business data and encrypted data. Overlaps in the interpretation of "Personal Data" have also resulted in different ways of governing anonymized and pseudonymized data [3] [6].

Consequently, the frameworks in some Member States such as the UK [4] tend to be less favorable to the processing of personal data for medical research compared to the Italian data protection framework. They impose more constraints on medical researchers, specifically in requirements such as necessity and specificity of patient consent. Despite complaints from researchers, no simplification was provided, unlike, e.g., in Italy. This was found to be an obstacle to the participation of the UK in some European and international research projects. In contrast, the Italian data protection law seems to grant more privileges to medical researchers by allowing consent for the processing of medical data across different healthcare organizations to be given in a single, one-off statement [5]. This raises the potential of ethical objection that *informed* consent should only be given for specific research tasks that are known to the data controller at the time the consent is collected and should be later required for any other processing that may be performed in future. [6]

These issues explain the diversity, complexity and dynamicity of the rules governing privacy protection. Privacy requirements could not be generalized to cover all cases of sharing. It is rather dependent on different aspects of the data and the context. This includes the type of data and its level of sensitivity, the entities sharing the data and the purpose of sharing. Many issues of privacy compliance in the healthcare domain are due to the gap between legislation and its technical implementation within healthcare and medical research organizations. We believe modelling could simplify and abstract the complexity of rules from the real world to allow their automation and enforcement at the organizations' process level. Figure1 illustrates our vision of the smooth shift from the complexity of the real world to operational controls for privacy protection.

For this paper our ideas will be structured as follows: in section one, we present our technical solution to the modelling and automation of privacy requirements. Section two presents a proof of usability of the model for building decision support applications to help the grid's medical users to share medical data while complying with privacy obligations. In section three we extend our ontology to allow the specification of privacy policies and the mapping of this specification to a standard privacy policy language such as the Extensible Access Control Markup Language (XACML). Finally we conclude and hint to future tasks that look automatic generation of enforceable privacy policies.

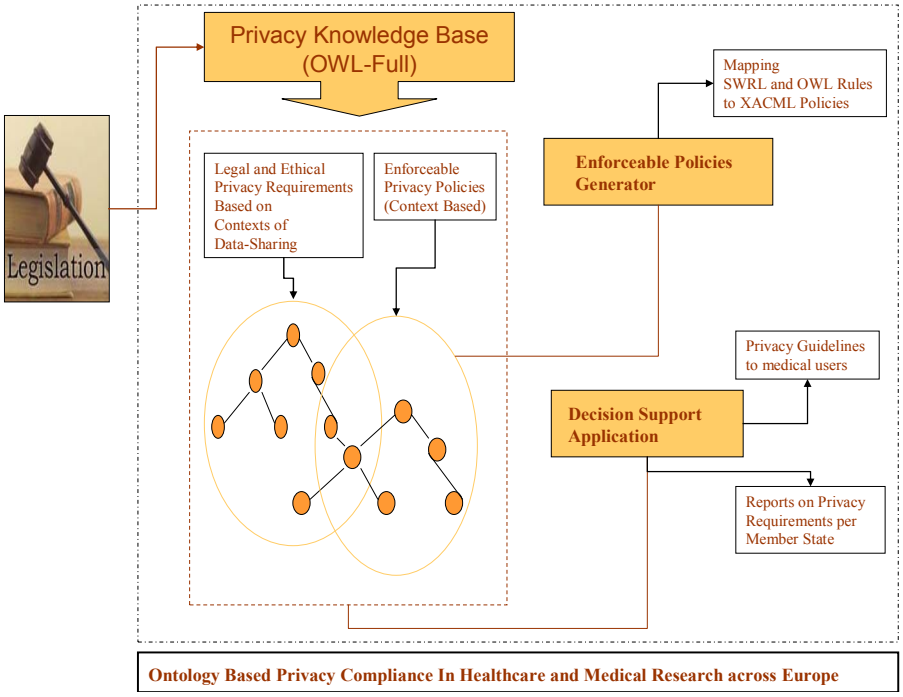

Figure1: Ontology Based Privacy Compliance for Medical Data Disclosure across Europe

1. Modeling Privacy requirements: OWL and SWRL

The diversity, complexity and dynamicity of the rules governing privacy protection in Europe explains the need for a modelling approach that is able to abstract this complexity and facilitate its automation and enforcement at the process level. We mean by privacy requirements all the obligations that 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 the patient privacy. This includes patient consent, data anonymity, and the rights of the data subject including their right to dissent and to be notified [6]. Our model should reflect any conflicts between the EU Member States in the specification and the provision of these requirements. In the following paragraphs we are presenting our attempt to model and automate privacy requirements in the context of medical data disclosure in Europe.

Our approach uses W3C Web Ontology Language (OWL) [7] to represent privacy obligations in medical data disclosure. OWL allowed us to model the conceptual domain of "data sharing" or "data disclosure" and its components as a hierarchy of classes and a hierarchy of properties to represent the relationships between them. Privacy requirements such as consent requirements could be modelled as OWL classes and assigned to the "dataSharing" resource as object properties.

Moreover, OWL allows overlapping models of a concept to be merged, even when different naming have been used for the same resource; e.g., Explicit Consent might be named Express Consent in another model but both concepts have the same meaning.

With complex legal domains, we need to model relationships that cannot be expressed in OWL because the logic for describing porperties is not rich enough. Legal rules are usually expressed in the form of *if–then*-like rules. For example, we may want to model a rule stating that if the data belongs to the UK then patient consent is necessary for the processing. Expressing this kind of rule requires the use of a semantic web rule language to allow building sets of rules in terms of the different concepts of the sharing process already described in the ontology and their properties. This will allow us to reason on the relevant set of rules and ontology classes in order to infer privacy requirements for different instances of sharing contexts.

As a rule language we have relied on a promising approach based on the Rule Markup Language (RuleML) that is the Semantic Web Rule Language (SWRL) [8]. The following example is a SWRL representation to the rule stating that the patient consent is necessary for the sharing of a UK medical data item that is anonymized. E.g.

$$\text{dataSharing}(?x) \wedge \text{hasSender}(?x, ?s) \wedge \text{hasReceiver}(?x, \text{any}) \wedge \text{locatedIn}(?s, \text{UK}) \wedge \text{hasStatus}(?d, \text{Anonymised}) \rightarrow \text{hasConsentNecessity}(?x, \text{Necessary})$$

2. Decision support for clinicians and medical technicians to enhance compliance with privacy regulations

Our system should reason on the model described in the previous section to generate protocols for medical users to guide them through the different processing tasks. For this purpose we developed a semantic web application that allows users to specify details of the different entities that constitute a sharing process and invoke a Jess rule engine [9] to fire up the relevant SWRL rules from our model. The result is a set of new inferred axioms that are added to the model as attributes of the instance of the "Data Sharing" class in question. These are returned to the user as the set of privacy requirements to allow the sharing of the data. (See Figure 2 below.)

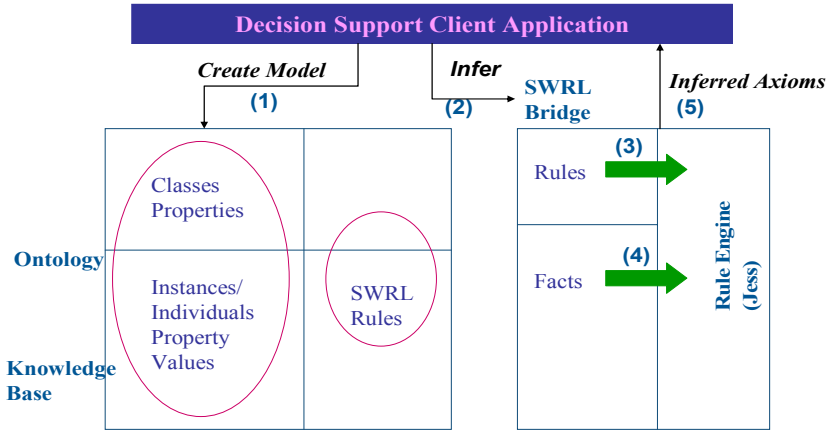

Figure2: Architecture of the privacy Decision support Application

In this section we explain how privacy requirements are integrated within some real world workflows of medical data sharing. We focus on a real word grid scenario, the MammoGrid project [11], which was an EU-funded collaboration between hospitals and research centres from UK, Italy and Switzerland. The project aimed to standardize scanned mammograms for use in epidemiological studies, quality control for breast cancer screening, comparative diagnosis and validation of computer aided detection algorithms for mammographic images. For this case study we focus mainly on the requirement of patient consent for two critical phases of the data lifecycle: 1- Uploading the data from local resources to the grid and 2- sharing the data on the grid. For the sake of highlighting more possible conflicts between the different Member states we have chosen to suppose that France is a grid party as well.

The data that is subject to processing for this project are patient breast mammograms along with other data revealing the age and some body metrics of the patient. Data Anonymization was not a preferred option for protecting patient identity as the data that should be hidden forms important clinical variables for comparative diagnosis. The justification to the processing of patient data was patient consent and/or ethical approval. For the UK, patient consent is considered necessary even when ethical approval was granted. However ethical approval is a sufficient condition for Italy. When a technician at one of the grid nodes tries to upload some local data to the grid shared data base, our system will automatically generate a set of privacy guidelines to assist them through their data uploading task. For example the following rule will be inferred in order to indicate to a radiologist at a French hospital that more that an express and specific patient consent is required in order to share data on the grid:

$$\begin{aligned} &\text{dataSharing}(?x) \wedge \text{hasSender}(?x, ?s) \wedge \text{locatedIn}(?s, \text{France}) \\ &\quad \wedge \text{concerns}(?x, ?d) \wedge \text{belongsTo}(?d, \text{France}) \\ &\qquad \rightarrow \text{consentNecessity}(?x, \text{Necessary}) \wedge \text{con\_Explicitness}(?x, \text{Explicit}) \\ &\qquad\quad \wedge \text{consentSpecificity}(?x, \text{SpecificConsent}) \end{aligned}$$

Similarly indicating to an Italian technician that patient consent is not necessary for uploading medical data to the grid will be based on firing up the following rule:

$$\begin{aligned} &\text{dataSharing}(?x) \wedge \text{hasSender}(?x, ?s) \wedge \text{locatedIn}(?s, \text{Italy}) \\ &\quad \wedge \text{concerns}(?x, ?d) \wedge \text{belongsTo}(?d, \text{UK}) \\ &\qquad \rightarrow \text{consentNecessity}(?x, \text{Necessary}) \wedge \text{con\_Explicitness}(?x, \text{Any}) \\ &\qquad\quad \wedge \text{consentSpecificity}(?x, \text{SpecificConsent}) \end{aligned}$$

The data now is uploaded to the grid and ready for processing for the specific medical purposes the MammoGrid projects was aiming to achieve. It is very likely that patients' mammograms would be shared with clinicians across European borders. In many cases researchers will require the data to be downloaded to their personal storage devices. At this stage we are more concerned with the usage of data for future purposes. A British organization might insist that when their data is to be processed by an Italian grid user, either the new processing purpose should be compatible with the purpose the patient has consented to or patient consent must be collected for the new purpose. The following rules determine who can contact the patient in order to collect consent, first for the UK:

dataSharing(?x) ∧ concerns(?x, ?data) ∧ belongsto(?data, UK)
∧ about(?data, ?patient) ∧ hasPurpose(?x, ?p) ∧ isa(?p, SecondaryPurpose)
∧ generalPractitioner(?gp, ?patient)
→ consentPointofContact(?gp)

and for Italy:

dataSharing(?x) ∧ concerning(?x, ?data) ∧ belongsto(?data, Italy)
∧ about(?data, ?patient) ∧ hasPurpose(?x, ?p) ∧ isa(?p, SecondaryPurpose)
∧ hasRequestor(?x,?r)
→ ConsentPointofContact(?r)

3. Extending the ontology to enable the specification of enforceable privacy policies to insure compliance

For better governance of European integrated health systems, legal and ethical requirements for privacy must be enforced at operational level as formal privacy policies. Through the use of OWL [7] we were able to represent the concept of a "Policy" and edit policy instances in a seamless way.

Privacy policies [12] are divided into two main categories according to their order of enforcement compared to the access control policy associated with them. The first category of policies is the ones that might affect a system access control decision. For example if patient consent is required for a specific context of sharing, the system will check for availability of patient consent before allowing the user to access their data. The second category are Privacy Obligations [12] that do not affect access control decisions but are rather dealt with after a decision is made in order to control usage of data at a later stage i.e. usage of data for a secondary purpose, disclosure of data to third parties, data deletion and retention.

In order to be easily enforced at the system level we suggest that our policies should 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. Our choice was the extendable access control markup language (XACML) [10]. Privacy policies in XACML are specified using some standard extendable markup language (XML) elements including (Policy, Target, and List of rules) 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…), Rule Target (Subject or Requester, Resource, Requested Action) and Rule Conditions [10]. The conditions attached to each rule are specific constraints on (the subject or requestor, resource, and others (depends on the context). XACML also allows users to add more user defined components or elements to the traditional vocabulary [13].

Our model captured all the components that constitute the XACML privacy policy specification and extends the Rule target component to allow setting constraints according to the purpose of data processing and the member state to which the data belongs.

Example: Purpose Compatibility Rule:
In this example we show how we modelled the 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 a SWRL rule using the OWL classes and properties specified in the ontology; the rule is as follows:

$$\text{provided}(?\text{patient}, \text{InformedConsent}) \wedge \text{isa}(?\text{consentPurpose}, \text{SpecificPurpose}) \wedge \text{compatibleWith}(\text{Purpose}, \text{ConsentPurpose}) \rightarrow \text{allow}(?\text{requester}, \text{Access})$$

For easy mapping to an XACML rule, the SWRL rule needs to be specified in terms of attributes of only the generic entities that constitute an XACML Rule Target (described above) and other elements that are used to specify the general policy that the rule in question is belonging to i.e. the purpose of processing. The OWL property "Provided (Patient, InformedConsent)" is a property of the patient whose data is to be shared and it is indicating that the patient has provided an informed consent. The patient or the data subject is not one of the XACML "Rule Target" components therefore we decided to express the same condition in terms of property of the class "Subject" (the requestor of a required action on a resource or object). The result is as follows:

$$\text{obtained}(?\text{subject}, \text{InformedConsent}) \wedge \text{hasCollectionPurpose}(?\text{object}, ?\text{collnPurpose}) \wedge \text{CompatibleWith}(?\text{collnPurpose}, ?\text{consentPurpose}) \rightarrow \text{allow}(?\text{subject}, \text{Access})$$

This rule could be easily mapped to the following XACML rule:

```
<Rule RuleId = "1" Effect= "Permit">
    <Target>
        <Subjects>< AnySubject/> </Subjects>
        <Resources> <AnyResources/> </Resources>
        <Action> <Any Action/> </Action>
    </Target>
    <Condition FunctionId = "string-equal">
        <Apply FunctionId –"string-one-and-only">
        <PurposeAttributeDesignator attributeId= "Disclosure-Purpose" DataType = "string"/>
        </Apply>
        <Apply FunctionId –"string-one-and-only">
        <ResourceAttributeDesignator attributeId = "Consent-Purpose" DataType = "string"/>
        </Apply>
    </Condition>
</Rule>
<Purpose>
    <Attribute AttributeId= "purpose-id" DataType="String">
        <AttributeValue> Disclosure-Purpose-1</AttributeValue>
    </Attribute>
    <Attribute AttributeId= "compatibleWith" DataType="String">
        <AttributeValue> Consent-Purpose-1</AttributeVAlue>
    </Attribute>
</Purpose>
```

Conclusion and Future Work

Throughout our research we have managed to model high level policies interpreted from European and national data protection law as privacy requirements for data disclosure. We have been able to capture similarity and possible conflict between the different frameworks across Europe through the use of SWRL rules, JESS and the protégé API. We have also specified by means of an ontology the concept of "Enforceable Privacy Policy" conforming to the XACML Standard. Privacy policies could therefore be created as instances of the Policy class and could be assigned to an equivalent privacy requirement. Linking between privacy requirements that are generated by the data disclosure decision support application and the policies that are enforced at the operation level is the basis for privacy compliance assurance on integrated medical systems. For future work we are looking at developing a Semantic Web application to construct XACML privacy policies and obligations from the policies specified in our OWL ontology model of enforceable privacy policy.

References

[1] Breton, V. *et al. The HealthGrid White Paper*, in "From Grid to Healthgrid" Proceedings of the Third HealthGrid Conference 2005, IOS Press. (Published online at http://initiative.healthgrid.org/the-initiative/healthgrids-concept/white-paper.html in 2004, last accessed on 21/06/09.)

[2] EU Directive 95/46/EC *The Data Protection Directive* last accessed on 18/06/09 at: http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf.

[3] McCullagh, K. *A study of data protection: harmonization or confusion?* 21st BILETA Conference: "Globalisation and Harmonisation in Technology Law", 2006.

[4] Iversen, A. *et al. Consent, confidentiality, and the Data Protection Act*, British Medical Journal 332(7534):165–169, 2006.

[5] Italian Personal Data Protection Code Legislative Decree no. 196, 30 June 2003

[6] Beyleveld, D. *et al. Implementation of the data protection directive in relation to medical research in Europe,* Ashgate, 2004.

[7] McGuinness, D.L., van Harmelen, F. *OWL Web Ontology Language Overview2*, W3C Recommendation,10 February 2004, last accessed 18/06/09 at www.w3.org/TR/owl-features/.

[8] Joint US/EU *ad hoc* Agent Markup Language Committee *SWRL: A Semantic Web rule language combining OWL and RuleML*, 2004, last accessed on 18/06/09 at www.w3.org/Submission/2004/SUBM-SWRL-20040521/.

[9] Friedman-Hill, E. *Jess in Action: Java Rule-Based Systems.* Manning Publications, Greenwich, 2003.

[10] OASIS, *Privacy policy profile of XACML v2. OASIS Standard*, 2005. last accessed on 18/06/09 at http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-privacy_profile-spec-os.pdf

[11] R. Warren, A.E. Solomonides, *et al*. *MammoGrid — a prototype distributed mammographic database for Europe* Clinical Radiology 62 (11): 1044-1051 (2007)

[12] PRIME: Privacy and Identity Management for Europe. *PRIME Architecture - Version 2*, 2007 https://www.prime-project.eu/prime_products/reports/