Privacy Compliance Decision Support for Medical data Sharing in Europe: a rule–based approach

Hanene BOUSSI RAHMOUNI [a,1], Tony SOLOMONIDES [a],

Marco CASASSA MONT [b], Simon SHIU [b]

[a] *Bristol Institute of Technology, University of the West of England, Bristol, UK*

[b] *Trusted Security Lab, HP Labs, Bristol, UK*

Main Author: Hanene Boussi Rahmouni

Address: 66 Gayner Road, Bristol BS7 0SW, UK

E-mail: Hanene2.rahmouni@uwe.ac.uk

[1] *Corresponding Author: Bristol Institute of Technology, University of the West of England, Bristol, UK, BS16 1QY; E-mail: Hanene2.Rahmouni@uwe.ac.uk.*

Summary. The harmonization of data protection legislation in Europe has been theoretically achieved by means of the EU directive on data protection. In practice the harmonization is not absolute and conflicts and inconsistencies continue to exist in the way Member States are implementing the directive. 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. In this paper we present an approach to model and automate privacy requirements for the sharing of patient data across within a semantic knowledge base. Then we approach the usage of the model for the purpose of providing automated decision support mechanism which would help medical professional complying with legal privacy requirements. Our methods starts with the capturing and the semantic modelling of privacy obligations that are of legal, ethical or cultural nature. These requirements are for the sharing of personal data between different European Member States. Our model reflects both similarities and conflicts, between the different Member States. We then use the resulting model in order to allow the reasoning on the safeguards a data controller should ask from an organization belonging to another Member State before disclosing medical data to them. This work shows that it is feasible; through the use of ontologies and semantic web technologies; to minimize unintentional breaches of privacy and data protection principles while sharing personal data on European healthgrid domains.

Keywords. privacy, EU data protection directive, health-grid, Semantic Web technologies

1. Introduction

When sharing medical data between different health organizations in Europe, it is important that the different parties involved in the sharing handle the data in the way indicated by the legislation of the Member State where the data was originally collected from. Privacy requirements, such as patient consent, may be subject to conflicting conditions between different national frameworks. Conflict also arises between different legal and ethical frameworks of the single Member State. Whilst most EU Member States are now governed by similar personal data protection rules, harmonization remains more apparent than real. This is due first to the fact that subject to the provision of suitable safeguards the European data protection directive [1] leaves some space for Member States to lay down exemptions to some of the obligations [1]. For example 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 additional exemptions to the ban of the processing of sensitive personal data [1]. Second, as specified by some studies [2], the definitions used do not lead to a uniform understanding of the key concepts underpinning the directive. Focusing on the concept of "Personal Data" for example, 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. Overlaps in the interpretation of "Personal Data" have also resulted in different ways of governing anonymised and pseudonymised data [2]. Consequently, the frameworks in some Member States such as the UK [3] tend to be less favourable to the processing of personal data for medical research compared to other frameworks. This includes the Italian data protection framework. The latter seems to grant more privileges to medical researchers in allowing consent to be given in a single, one-off statement [4]. This raises ethical concerns on handling secondary usage of the data [5].

These issues explain the diversity, complexity and dynamicity of the rules governing privacy protection. 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 as a way of privacy compliance management. In previous work presented in [16] we have showed the usefulness of our privacy requirement knowledge-base for closing the gap between high level policies and operational access controls by suggesting a privacy aware access control model and architecture. In this work we use our knowledge-base for providing an automated privacy guidelines and advices to medical users which would help assisting them with their every day duties of medical data disclosure. For this paper our ideas will be structured as follows: in section two, we analyze a selection of privacy requirements and issues associated with the sharing of patient sensitive data across European borders. In section three, we present our technical solution to the modelling and automation of privacy requirements. Section four presents a proof of usability of the model for building decision support applications to help the healthgrid's [6] medical users to share medical data while complying with privacy obligations. Finally we conclude and hint to related work and future tasks.

2. Materials and Methods

The principal problem addressed in this work is, how to encode privacy legislation and related regulatory frameworks (e.g. institutional rules on ethics) in such a way that they are amenable (a) to provision of decision support for the non-expert user, (b) to automation of compliance at an operational level, and (c) to documentary support for compliance audit. This paper reports only on (a) decision support; [16] reports on (b) and a planned paper will cover audit.

Our case study carries the additional complexity of a supra-national "directive" which has been variously interpreted as national legislation. Indeed, our thesis is that this additional layer of complexity allows us to demonstrate the power of our method better than would be the case under a single regulatory regime. As is the case with EU directives in general, the European Data Protection Directive 95/46/EC has been "transposed", as the official jargon has it, into national legislation in the Member States. These national laws are not necessarily in complete agreement with the Directive, nor are they necessarily entirely compatible with each other. (Examples of this will be discussed below.) In any case, it is generally accepted that text law, i.e. the statutes themselves, are not ordinarily well understood and acted upon by non-experts, so that between the law and any potentially questionable action stands an interpreter of the law, a "lawyer", who provides expert opinion or professional guidance. In our case, the need to interpret data protection legislation in the various Member States of the EU is of such importance to business that there are many immediate sources of guidance, such as, in the UK, the Information Commissioner's Office website guidelines [17]. In our work, we have sought to codify such guidelines, at least in cases where they are not controversial, rather than attempt the legal text itself. A standard reference work for research in this field is that published by the *Privireal* project [5] and we have largely relied on this.

A relatively recent approach to harmonizing fields in which different languages or data structures are applied to a common domain is through so-called "ontologies". An ontology is a standard method of organizing the concepts in a domain in such a way that it can map to the various linguistic or informatic practices that may occur in that domain. Inter-relationships between concepts, such as equivalence, subsumption, specialization and generalization, and so on, are also mapped. A commonly used language for ontology description is the Web Ontology Language (OWL, after the character in AA Milne's children's story) which forms the basis of the tool *Protégé* from Stanford University [18]. It is possible to reason with OWL concepts using the Semantic Web Rule Language (SWRL) and we have adopted both. Through the use of these technologies we have captured the legal requirements and modelled them as a semantic web knowledge base. This may be interpreted by a 'reasoner' or rule engine to work out the applicable privacy requirements for a given case of medical data sharing

Through the use of the *Protégé* application programming interface (API) and the *Protégé OWL* API we develop a semantic web application that allows a professional user to specify facts describing a specific case of proposed data sharing in order to get as output a list of privacy requirements that sender and recipient must comply with. Also users can choose to generate a report of privacy requirements per Member State. In our practical examples, we have used the rule-based system environment *Jess* [19] to demonstrate such reasoning in particular use-cases. Last but not least among our tools is the eXtensible Access Control Markup Language (XACML) which allows us to

interpret our high level policy rules into actionable permissions and obligations; this is important, but figures somewhat less in the work reported here than it does in subsequent work.

3. A Selection of Privacy Requirements

The governance of personal data in Europe imposes certain obligations of regulatory compliance. By 'privacy requirements' we mean those obligations that must be fulfilled by all parties involved in the process of sharing or processing sensitive patient data, whether for healthcare or medical research, to preserve informational aspects of the patient's privacy. This entails understanding of conceptual information about rights, obligations and consequent actions; among these is the obligation to obtain and 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 and, indeed, conflicts between EU Member States in the specification of and provision for these requirements. In the following paragraphs we analyze a selection of requirements specifically taking into consideration the degree of challenge faced when trying to comply with them. A fuller analysis of such challenges has been published in joint work with partners from the SHARE project in [7].

3.1. ***Patient Consent***

To qualify as legitimate, the processing of medical data has to be covered by one of seven hypotheses listed in Article 7 of the Directive 95/46/EC (the first hypothesis being patient consent) [7]. Article 8-2(a) of the Directive thus provides that the data subject's explicit and valid consent constitutes the very first source of the legitimacy for the processing of his medical data. The standard of consent is defined in Article 2 of the directive as: "*the data subject's consent* shall mean any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed." Consent for the processing of personal data must be given unambiguously. Consent for the processing of sensitive data must be explicit. The directive does not explain or define what a specific consent means, which creates opportunities for different interpretations from different member states. Article 6 of the directive permits the collection of personal data only for "specified purposes". This might be an indication that it is meant by "specific consent". In this context, specific consent is given only when the purpose of processing has been specified to and acknowledged by the data subject so as to allow him to accept or reject it. However, the required degree of specificity is still left unqualified and open to two different interpretive approaches. The first assumes that the data processor knows the different processing tasks in fine detail. The second interprets specific purposes for relatively broad sectors such as "commercial purposes" or "scientific purposes". The directive also adduces a principle on compatibility of purpose: once the processing of the data has been established as legitimate for a specific purpose, it may be further processed for a compatible purpose, as well as for any historical, statistical or scientific purpose.

At the national level, regulatory frameworks have addressed consent obligations either within data protection law or in other legislation, or both, including e.g. Common Law and Case Law. Member States' frameworks have highlighted various requirements for consent. These include the necessity, expressiveness, specificity and form of consent. Some Member States have modified this set of requirements by devoting separate sections within their data protection acts to particular issues; for example, the Italian legislation has simplified the ways consent is collected and should be recorded, as well as the determination of practicability of consent. [4].

Based on detailed analysis in [5], some Member States do not distinguish between consent and explicit consent (e.g. Poland), while in others consent must always be explicit informed consent, although this does not mean it has to be written (e.g. Estonia). The Czech Republic distinguishes between consent to the processing of personal data and consent to the processing of sensitive data which must be explicit and written. UK Law requires explicit consent when sensitive data is to be processed; this requires active communication between the relevant parties, but this may be other than written. The period of validity of consent also differs from one state to another.

The SHARE Project [7] investigated different European and national legal frameworks on consent for the processing of patient data and found that some general themes are repeated in most of these:

- necessity of consent to the processing of the data;
- explicit (or express) patient consent;
- specificity of consent (specific or general);
- way in which consent must be collected (verbal, written);
- who may contact the data subject to get his consent;
- how consent should be documented (electronic, printed)
- legal competence of the data subject;
- who may give consent instead of the data subject (next of kin, proxy or legal representative)
- lifetime of consent validity;
- practicability of consent (practicable, impracticable); and
- Miscellaneous others of narrower scope.

The vocabularies of most national frameworks, whether legal or ethical, include most of these topics. However, harmonization of these requirements is not complete, not only because some Member States have omitted certain requirements, but also because of the diversity of definitions and interpretations. Hence, we consider that consent requirements in Europe should be classified under a standard taxonomy where the local description or definition of each entity in the taxonomy is allowed to differ from one Member State to another.

3.2. ***Personal Data Anonymization***

Data protection legislation in Europe is mainly designed to govern and control the processing of personal data. While they mostly ban the processing of personal data, they do allow conditional lawful processing of such data in certain circumstances. If the conditions or circumstances do not allow, the only way to process personal data is by de-personalizing them first. We are interested in patients' medical data which is normally classified as "sensitive" in data protection law. Research involving anonymous data does not require patient consent, provided data controller and processor commit to special safeguards to ensure complete anonymity. However, if anonymised data can still be considered indirectly nominative (e.g. through correlation with other data), consent is generally required and further safeguards must be adopted to protect the privacy and confidentiality of individuals [8]. De-personalization of data can take one of two forms, *anonymisation*, where all data that can potentially identify the data subject (the patient) are masked or eliminated, and *pseudonymisation*, where the identifying data are reversibly mapped onto and replaced by non-identifying codes appropriate to the circumstances. The degree of required anonymity varies from one Member State to another, not least because of differences in the definition of "personal data" within the different legal frameworks. Further possible conflicts of interpretation also arise between Member States; these are discussed in [5].

3.3. ***Specific Purposes of Processing***

According to Article 6-1.b of the Directive 95/46/EC, data may 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".

Therefore, there is an assumption of compatibility between the original (collection) purpose 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, of the identity of the controller and the purpose of the processing. This duty to inform may be lifted only if the 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 Directive considers the disclosure of personal data to third parties as a processing operation, and thus subject to usual legal provisos. Transfer of data, as a particular form of disclosure, 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 certain special purposes, such as protection of the vital interests of the subject or of the security of the state, or where the subject has manifestly already made that data public.

In the light of the Data Protection Directive, if healthgrids are to be used for risk detection, disease monitoring and preventive care, legal guidelines should be established that clarify the circumstances in which professionals can make further use of personal data related to health in the interests of public health [7]. Such guidelines should allow for secondary uses even where such uses could not have been foreseen at the time of data collection.

4. Modelling Privacy Requirements: OWL plus Rules

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 shall use the term "privacy requirements" to mean all those obligations that must be fulfilled by all parties involved in the process of sharing and processing sensitive patient data for medical purposes (by which we embrace both healthcare and medical research) to preserve the patient's privacy. This term therefore encompasses patient consent, anonymisation or pseudonymisation, the rights of the data subject including his right to dissent and to be notified. Our approach deals only with the requirements that could be enforced using a policy-based approach and does not include the cases where the intervention of ethical committees is essential. Our model should rather reflect similarity and possible conflicts between the EU Member States in the specification and the provision of these requirements. In the following paragraphs we present our attempt to model and to automate privacy requirements in the context of medical data disclosure in Europe.

Our approach uses the Web Ontology language (OWL) [9] to represent privacy obligations in the context of medical data disclosure. OWL allows us to model the conceptual domain of "data sharing" or "data disclosure" and its components as hierarchies of classes/subclasses and of properties to represent the relationships between them. As shown in Figure1, privacy requirements (e.g. *Consent*) may be modelled as OWL classes and assigned to the "dataSharing" resource as object properties.

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* may be termed *Express Consent* in another model but both concepts have the same meaning.[2]

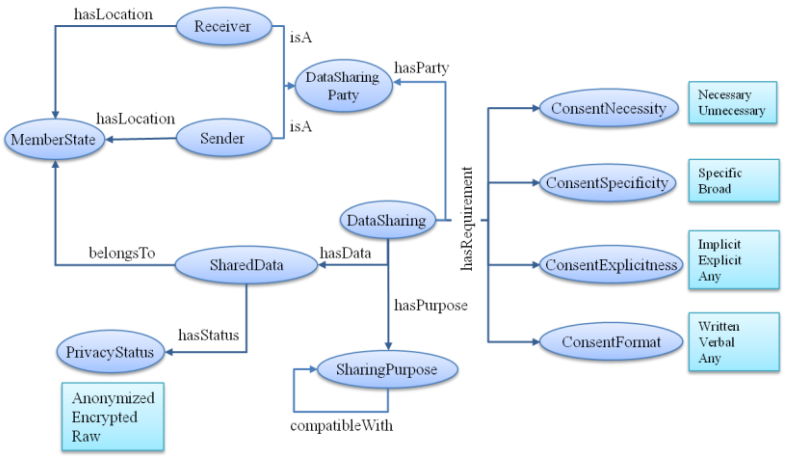

Figure 1: Ontology of privacy requirements for the sharing of patient data in Europe.

[2] We evade here the linguistic issue: e.g. is "explicit" in English *the same* as "express" in French (i.e. an exact linguistic translation) or are they equivalent concepts?

In complex legal domains, we need to model relationships that cannot be expressed in OWL, whose logic for describing properties is not rich enough. Legal rules are usually expressed as *if–then*-like rules. For example, we want to model a rule stating that if the data belongs to the UK then patient consent is necessary for any processing. Expressing this kind of rule requires the use of a semantic web rule language to allow sets of rules to be built up in terms of the different concepts of the sharing process (as described in the ontology) and properties of those concepts. This allows us to reason with the relevant set of rules and ontology classes in order to infer privacy requirements for different possible instances of sharing from the real world. The Semantic Web Rule Language (SWRL) [10] satisfies our criteria for this task. The following example is a SWRL representation of the rule stating that *patient consent is necessary for the sharing of a UK medical data item that is anonymized*. Thus,

$$\begin{aligned} &\text{dataSharing}(?x) \wedge \text{hasSender}(?x, ?s) \wedge \text{hasReceiver}(?x, \textit{any}) \\ &\qquad \wedge \text{locatedIn}(?s, \text{UK}) \wedge \text{hasSatus}(?d, \text{Anonymized}) \\ &\qquad\qquad \rightarrow \text{hasConsentNecessity}(?x, \text{Necessary}) \end{aligned}$$

In the next section we describe how the OWL ontology we have created and the semantic rules we have defined can be used to provide decision support for medical users to help them share patient data on a healthgrid in a privacy-aware manner.

5. Decision Support for Clinicians to Enhance Privacy Compliance

Our system has to reason on the privacy requirements model and knowledge base, described in the previous section, to generate protocols for medical users to guide them through the different processing tasks. For this purpose, we have developed a semantic web application that allows users to specify details of the different entities that constitute a sharing process and receive, as output, appropriate privacy management guidelines. This includes, for example, requirements regarding patient consent, such as establishing the necessity of consent and the required type of consent. The work we did in the previous section using the *Protégé* toolkit would be useful for ontology developers, e.g. to test and verify the usability of their models in decision making tasks, but it cannot be used by non-technical users such as clinicians. Our application allows clinicians and other medical users to enter a description of the data processing they would like to undertake in a standard fashion. A graphical user interface has been provided to allow non-technical users to enter descriptions in a standard way. Our application then processes the data under consideration in the order portrayed graphically in Figure 4.

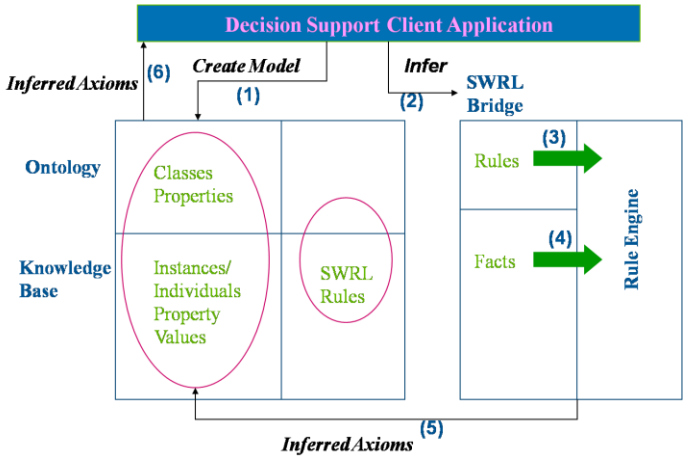

Figure 2: Architecture of the privacy decision support application.

With the OWL and SWRL model ready to receive concrete data (cf (1) in Figure 2), the user enters data and these are matched (2) to individuals stored in the knowledge base. The application then creates an instance of a SWRL rule engine (3, 4). As with the *Protégé* toolkit, we have chosen *Jess* as a rule engine. *Jess* is usually accessed through the SWRL rule engine bridge. When working programmatically, we have first to create a SWRL rule engine bridge. In our case we need to explicitly set the rule engine name to *Jess*. This is because the bridge is specialized for each rule engine implementation. However, interaction with the bridge should be the same irrespective of the underlying rule engine implementation. An implementation for the *Jess* rule engine is supplied with the standard Protégé-OWL distribution in a Java archive (JAR) called `swrl-jess-bridge.jar`. A class in this repository called SWRLJessBridge contains the *Jess* implementation. The constructor for this class takes an instance of the OWLModel class, representing the OWL knowledge base with its associated SWRL rules, and an instance of a *Jess* `Rete` object, which represents an instantiation of the *Jess* rule engine. The following code snippet shows the creation of a *Jess* bridge. It assumes that the user knows how to create an instance of an OWL model using the *Protégé*-OWL API.

```
OWLModel owlModel = ... // Create using normal Protege-OWL mechanisms.

SWRLRuleEngineBridge bridge = BridgeFactory.createBridge("SWRLJessBridge", owlModel);
```

A SWRLRuleEngineBridgeException will be thrown if any errors occur during the bridge creation. Once the *Jess* bridge is created, the public methods it inherits from the SWRLRuleEngineBridge class can be used to interact with it.

Once an instance of a jess SWRL bridge is created we invoke the infer() method (a method of the class SWRLRuleEngine) in order to load the facts and rules into the rule engine, do all the necessary transformations before and after running the rule engine and record the newly inferred axioms into the ontology (5).

6. Case-Study and Results

In this section, we explain how privacy requirements are integrated within some real world workflows of medical data sharing. We focus on a real world grid scenario, the MammoGrid project [12], whose aim was 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: (a) uploading the data from local resources to the grid, and (b) sharing the data on the grid. Data sharing for this project involve organizations from two EU Member states, UK and Italy. With a substantial grid node at CERN to support communication, we suppose that France is a grid party as well.

6.1.1. Uploading Data on the Grid

When a user requests to upload data from the hospital database to the federated grid database, the system must first generate the set of privacy obligations that the user needs to comply with before the data is uploaded to the grid. These requirements are generic and do not depend on the geographic location of the entities that would have access to it or share it in the future. In other terms, the national legal and ethical framework would be the primary reference for identifying privacy requirements for this task. Requirements could include anonymisation, pseudonymisation, data de-identification including image scrambling, consent for storing the data in the grid and obligations related to the quality of the data including data provenance, accuracy and relevance. To achieve this goal, a local version of the framework must be deployed as part of the local resources at each hospital or medical research centre participating in the grid.

The data that is subject to processing for this project are patient breast mammograms along with other data revealing the age and somebody 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 shared grid database, our system will automatically generate a set of privacy guidelines to assist her through her data uploading task. For example, the following rule will be inferred in order to indicate to a radiologist at a French hospital that more than an express and specific patient consent is required in order to share data on the grid:

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

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:

dataSharing(?x) ∧ hasSender(?x, ?s) ∧ locatedIn(?s, Italy)
∧ concerns(?x, ?d) ∧ belongsTo(?d, Italy)
→ consentNecessity (?x, Necessary)
∧ con_Explicitness(?x, Any)
∧ consentSpecificity(?x, SpecificConsent)

The following schema demonstrates the kind of output the user gets when they interact with the decision support prototype application, when they choose the type of the data to be shared as mammogram and select the consent requirements report button.

Sharing Subject	**Compliance**
Sharing Data Type: *mmx*	Member State: all
	UK Necessity: Necessary Specificity: Specific Explicitness: Any Who can contact: GP only
	Italy Necessity: Unnecessary Specificity: Broad Explicitness: Any Who can contact: Research team
	France Necessity: Necessary Specificity: Specific Explicitness: Express Who can contact: Research team

Figure 3 Schematic report of privacy requirements per Member State.

Downloading Data from the Grid

In our application, the grid system is not fully open and data may be shared only on request. When a user within Member State A requests to access data belonging to another Member State B, the system should generate the relevant set of requirements which are just the additional safeguards that Member State B would usually ask users in Member State A to guarantee before sharing medical data with them. Allowing access to the data would be subject to some security policies that are not part of our focus and also to the privacy assurance the user provides when requesting the access. In order to control data disclosure when downloading data from the grid, a distributed version of the framework is required. As shown in Figure 6.7, this application will be deployed as a component of a general privacy compliance framework we are working on. This allows the management of sharing requests coming from all nodes participating on the grid in an appropriate manner.

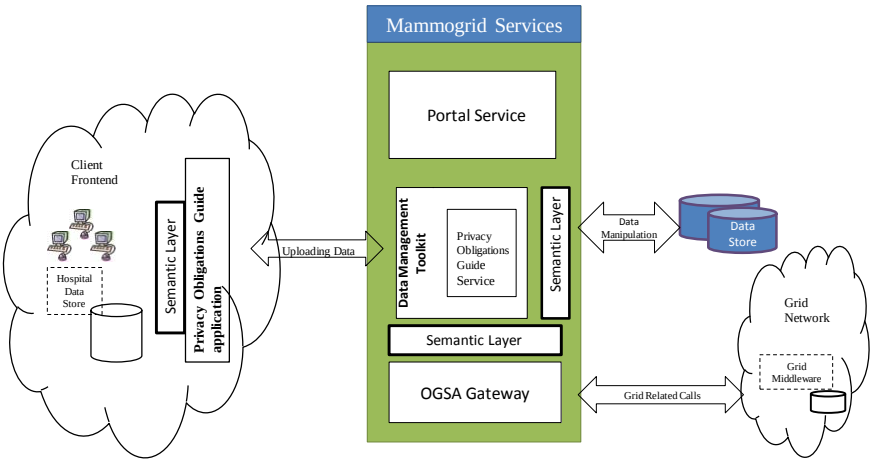

Figure 4. Architecture integrating the privacy decision support application in MammoGrid services layer.

The data now is uploaded to the grid and ready to process for the specific medical purposes of the MammoGrid project. 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 use 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:

$$\begin{aligned} &\text{dataSharing}(?x) \wedge \text{concerns}(?x, ?data) \wedge \text{belongsto}(?data, UK) \\ &\quad \wedge \text{about}(?data, ?patient) \wedge \text{hasPurpose}(?x, ?p) \\ &\quad \wedge \text{isa}(?p, SecondaryPurpose) \wedge \text{generalPractitioner}(?gp, ?patient) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \rightarrow \text{consentPointofContact}(?gp) \end{aligned}$$

and for Italy:

$$\begin{aligned} &\text{dataSharing}(?x) \wedge \text{concerning}(?x, ?data) \\ &\quad \wedge \text{belongsto}(?data, Italy) \wedge \text{about}(?data, ?patient) \\ &\quad \wedge \text{hasPurpose}(?x, ?p) \wedge \text{isa}(?p, SecondaryPurpose) \\ &\quad \wedge \text{hasRequestor}(?x, ?r) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \rightarrow \text{ConsentPointofContact}(?r) \end{aligned}$$

and for France:

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

7. Related Work

There has been some other work involving a legal decision support mechanism in sharing biomedical data. Notable in the literature is the work of the *caBIG* project [20]. *caBIG*, funded and led by the National Cancer Institute's Center for Bioinformatics, has as its goal the delivery of innovative approaches for the prevention and treatment of cancer. Its vision is the implementation of infrastructure and tools with broad utility and reusability within and outside the cancer community. These tools are designed to support the sharing and reuse of large volumes of research data created by high throughput genomics and proteomics technologies. The legal, regulatory and security requirements for data sharing were studied [21] and specialized tools are being developed in order to address these challenges. Among the tools being developed and adopted by *caBIG* infrastructure is the Data Sharing and Security Framework (DSSF) [22]. The *caBIG* DSSF can be used as a decision support tool to facilitate data sharing by determining which data can be shared and under which type of access, data security and regulatory controls. This requires the user to assess the sensitivity of the data by using the Framework's *Privacy, Confidentiality and Security Considerations* element [23]. For example, the framework asks the user to select the category of sensitivity that best describes the data he wants to share. The user can choose from three categories: (a) Low Sensitivity (i.e. de-identified or anonymised data set), (b) Medium (coded or limited data set), or (c) High Sensitivity (identifiable data). By doing so the framework can answer legal questions related to Privacy and Security, such as the sample question, ***Do federal or state laws or your institution's policies prohibit or restrict disclosure?***

This framework, if automated and adopted, would certainly have a key impact on enhancing the task of data sharing while complying with diverse legislation. The ambiguity around legal issues of privacy would be better clarified by providing specialized answers to users' concerns. However, we have noted certain concerns: first, leaving the responsibility to individual medical users to assess the sensitivity of data presents a risk of inconsistent assessments and diverse judgements for the same data, possibly because of lack of experience or expertise in the privacy domain. Second, the DSIC Knowledge Center [24] is working on automating the DSSF decision support tools. The work is in progress and information about methodology, architecture and techniques adopted has not yet been published.

8. Conclusion and Future Work

Privacy requirements for the sharing of medical data between European Member States can be described within a semantic model. Once it is rich enough, the model can form a

knowledge base for an inference engine to reason about the duties of medical users as imposed by different European and national legislation in order to preserve patient privacy. The new inferred knowledge generated by the inference engine can provide guidelines and protocols to help clinicians and other medical users across Europe to share medical data while complying with relevant regulatory frameworks. Our work has mainly focused on the requirement of patient consent, but we believe other requirements could be modelled in the same way, including anonymization, role-roaming, etc.

In the literature, several research projects have addressed the problem of privacy management for sharing identifiable data across European borders including [13] and [14]. However, they have tackled this problem as only a system process through designing a system and access controls that are privacy aware. We have similarly addressed these issues in [15]. In contrast, the work in this paper stresses the importance of considering privacy management and compliance as a human process through more effective teaching of privacy policies and by providing users with automated support to help minimizing unintentional breaches of privacy principles.

In future work, we will extend our semantic model of privacy requirements by classifying privacy requirements rules under two main categories allowing the users to differentiate between legal and ethical guidelines. It would also be valuable to adduce a measure of confidence in any given decision, using, for example, different authoritative rankings of statutes and rules to weight alternative decisions. In addition, we are looking at integrating non-European policies such us the US Health Insurance Portability and Accountability Act (HIPAA).

9. References

[1] EU Directive 95/46/EC: The Data Protection Directive. Available from: http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf.

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

[3] Iversen A. et al. Consent, confidentiality, and the Data Protection Act, British Medical Journal 332(7534):165–9, 2006.

[4] Decreto legislativo 30 giugno 2003, n. 196 *Codice in materia di protezione dei dati personali* Italian Personal Data Protection Code, Legislative Decree No. 196 (June 30 2003).

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

[6] Breton, V. et al. The HealthGrid White Paper. Proceedings of the Third HealthGrid Conference; 2005. Available from: http://initiative.healthgrid.org/fileadmin/whitepaper/HealthGrid_whitepaper_full.pdf

[7] SHARE Deliverable D4.3 [Internet]. Legal, social & economic challenges component roadmap II. Available from: http://www.eu-share.org/about-share/deliverables-and-documents.html

[8] Fond de la Recherche en Santé du Quebec [Internet]. A Governance framework For Data Banks and Biobanks Used for Health Research.2006 Dec. Available from: http://www.frsq.gouv.qc.ca/en/ethique/pdfs_ethique/Sommaire_groupe_conseil_anglais.pdf

[9] W3C Recommendation [Internet]. McGuinness, D.L., van Harmelen, F. OWL Web Ontology Language Overview2,. February 2004. Available from: www.w3.org/TR/owl-features/.LA.

[10] Joint US/EU ad hoc Agent Markup Language Committee [Internet]. SWRL: A Semantic Web rule language combining OWL and RuleML. 2004. Available from: www.w3.org/Submission/2004/SUBM-SWRL-20040521/.

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

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

[13] PRIME [Internet]. Privacy and Identity Management for Europe. PRIME Architecture . 2007; Version 2. Available from: https://www.prime-project.eu/prime_products/reports/
[14] PRIMELife.eu [Internet]. Privacy and Identity Management for Europe for Life. Available from: http: //www.primelife.eu/
[15] Rahmouni, H.B.; Solomonides, T.; Mont, M.C.; Shiu, S. Privacy Compliance in European Healthgrid Domains: An ontology-based approach. Proceedings of the 22nd IEEE International Symposium of Computer-Based Medical Systems (CBMS). Summer 2009.
[16] Hanene Boussi Rahmouni, Tony Solomonides, Marco Casassa Mont, and Simon Shiu. Privacy compliance and enforcement on European healthgrids: an approach through ontology. Phil. Trans. R. Soc. A *September 13, 2010 368:4057-4072; doi:10.1098/rsta.2010.0169*
[17] The Guide to Data Protection. UK. Available From http://www.ico.gov.uk/upload/documents/library/data_protection/practical_application/the_guide_to_data_protection.pdf
[18] Protégé official website. Available at: http://protege.stanford.edu/
[19] E. Friedman-Hill. Jess in Action: Java Rule-based Systems, Manning Publications Company, June 2003, ISBN 1930110898, http://herzberg.ca.sandia.gov/jess/
[20] The caBIG official Website. Available from: https://cabig.nci.nih.gov/
[21] The caBIG Data Sharing Policy. Available from https://cabig.nci.nih.gov/working_groups/DSIC_SLWG/data_sharing_policy
[22] The caBIG Documentation and Training Workspace in cooperation with the Data Sharing and Intellectual Capital Workspace and the caGrid Knowledge Center . An Introduction to caGrid Technologies and Data Sharing Prepared for the caBIG® Community. May 2010. Available at: https://wiki.nci.nih.gov/download/attachments/24271074/Intro_GridTech_DataSharing.pdf
[23] caBIG Privacy Decission Support Tool. Available at https://cabig-kc.nci.nih.gov/DSIC/KC/index.php/Privacy_Decision_Support
[24] caBIG Data Sharing and Intellectual Capital (DSIC) Knowledge Center. Available from: https://cabig-kc.nci.nih.gov/DSIC/KC/index.php/Main_Page