An Ontology-based Compliance Audit Framework for Medical Data Sharing across Europe

Hanene Rahmouni[1,3], Kamran Munir[1], Intidhar Essefi[3], Marco Mont[2], and Tony Solomonides[4]
[1]Department of Computer Science and Creative Technologies, University of the West of England, UK
[2]Hewlett-Packard Labs, Cloud & Security Lab, UK
[3]University of Tunis el Manar, the Higher Institute of Medical Technologies of Tunis Research Laboratory of Biophysics and Medical Technologies Tunis, Tunisia
[4]Outcomes Research Network, Research Institute, NorthShore University Health System, USA

Abstract: *Complying with privacy in multi-jurisdictional health domains is important as well as challenging. The compliance management process will not be efficient unless it manages to show evidences of explicit verification of legal requirements. In order to achieve this goal, privacy compliance should be addressed through "a privacy by design" approach. This paper presents an approach to privacy protection verification by means of a novel audit framework. It aims to allow privacy auditors to look at past events of data processing effectuated by healthcare organisation and verify compliance to legal privacy requirements. The adapted approach used semantic modelling and a semantic reasoning layer that could be placed on top of hospital databases. These models allow the integration of fine-grained context information about the sharing of patient data and provide an explicit capturing of applicable privacy obligation. This is particularly helpful for insuring a seamless data access logging and an effective compliance checking during audit trials.*

Keywords: *Privacy, regulation, verification, audit, compliance, ontology, SWRL, health data, public clouds, GDPR.*

Received June 24, 2019; accepted April 15, 2020
https://doi.org/10.34028/iajit/18/2/4

1. Introduction

Privacy protection in geographically distributed healthcare and medical research systems is subject to policies dictated by conflicting jurisdictions and legislations [10, 31, 43]. Enforcing these policies at system level is necessary but does not provide sufficient guarantees to privacy protection [7]. Indeed, there is a tremendous need to audit privacy compliance in order to enhance social acceptance of these systems. To achieve this goal, we need to incorporate better privacy assurance mechanisms that would let patients across the European Union (EU) benefit from both enhanced healthcare strategies and better medical research outcome while having less concerns about their privacy [40]. It is therefore essential to adopt mechanisms ensuring the robustness of privacy enforcement processes and proofs that the systems managing patient data can provide further assurance to the enforcement. We believe this could be achieved through the monitoring and auditing of past data sharing events within the healthcare context. In this paper, we contribute to the field of privacy compliance technologies by presenting a novel mechanism that allows an auditor to audit the compliance with privacy regulation within healthcare IT Systems. This is by allowing them to check the validity of a set of privacy guidelines given to the user against a set of formal data sharing policies previously enforced at data access run time. First, our proposed framework can capture the characteristics of a previous event of data manipulation recorded in an audit log. It can then generate a set of privacy rules related to the event in question and finally measures the users' compliance to the identified set of rules. In fact, despite all automation, the risk of breaching patient privacy continues to exist, particularly, in the context of health data management. This could be linked to the usage of exception-based access to data, which allow users to bypass system controls due to emergency cases and other unforeseen reasons [15, 29, 44]. In addition, system controls might be hacked or twisted by intruders when the systems are exposed to Internet and cyber security risks [48]. The risk could also be related to the simple reason of system failures. Hence, introducing audit and governance mechanisms is highly required when deploying healthcare computer applications. In this context, the guidelines provided to the user should match the constraints that have been initially tested for conformance at access control run time. These efforts involve reviewing all users' past data sharing events and the permissions they were granted in each data usage attempt. However, these reviews are extremely costly and difficult in the absence of an automated solution.

The ideas in this paper are structured as follows: section 2 sets the background through a study of some related work. Section 3 presents our compliance

checking approach; in particular, we explain here the different functionalities offered by the compliance framework. Afterwards, the system architecture is presented in section 4 and a grounding scenario is presented in section 5. In section 6, we discuss the paper results and contribution with relation to state-of-the-art work in the field of systemic healthcare privacy compliance auditing. Finally, the paper conclusions are presented in section 7.

2. Background and Related Work

Privacy is considered of high priority for compliance, mainly for disciplines that rely on the processing of personal data to achieve success. In Europe most member states have specific legislations to rely on for issues related to citizen's privacy. But these legislations were divergent and presented potential conflicts regarding a free flow of information across Europe. The European commission realized the need for the harmonization of data protection legislations within Europe and published the EU 95/46/EC Directive on the protection of personal data in 1995 (EC, 1995). By May 25, 2018, the General Data Protection Regulation (GDPR) came to supersede the Data Protection Directive and become national law for all EU Member States. The GDPR has rather more global scope and has brought to obligation more specific data protection requirements, stiffer enforcement as well as non-compliance penalties [19].

As an impact of major digital transformation, many healthcare IT systems now highly involve the use of patient data including medical heath records, medical images or even information about patient lifestyle, which could be shared through cloud services. This data falls within a special category of personal data [33] which is sensitive data as specified and defined by privacy legislation. Policy makers in the field of healthcare, have made a great effort to balance between the limited use of such category of data for the general public good and the protection of patient privacy and autonomy. From a technical and an operational perspective, privacy is considered as the main challenge for security developers working on large scale integrated healthcare systems. The analysis made in [33] showed that no future for such technology if the legal requirements are not carefully implemented in large and heterogenous health IT infrastructures. Moreover, the increasing complexity of electronic patient records raises a requirement for forms of mandatory access control. Different parts of each patient's medical records should be accessible by varying sets of people: administrative staff, nurses, general practitioners, specialists. For purposes of epidemiology and other research, anonymised or pseudonymised records should be available. In addition, the management of genetic data introduces more problems and difficulties due to the potential re-identification of data. Although attempts are being made to keep the access rules simple, issues of consent implied that some detailed fine-grained rules are required. If an element of a patient record is stored in an encrypted form, the decrypted form should only be available to individuals in selected roles, and where suitable facilities for recordkeeping and audit exist. Intelligent audit systems need to be put in place to prevent a clinician receiving sensitive information from passing it to a third party. At the moment, no strong mechanisms exist to enforce these requirements. As a result, the UK National Health System (NHS), for example, uses large separated networks, which drives up costs, and can nevertheless not give strong guarantees of separation and complete audit [47].

We believe that an effective and useful way of contributing to the problem of data governance and privacy compliance in our context is by working on the integration of both business strategies and automated processes responsible for specifying and managing organizational policies. It is important that harmony and unification of these both components are optimized. It is highly recommended for organizations to start thinking about minimizing dependency on manpower for compliance management tasks and saving money and time by adopting automated solutions. However, reaching a parallelism between the organizational high-level policies and its operational processes is still a challenge to be dealt with. For all healthcare IT environments; where personal data is subject to processing or sharing, it is important and necessary for audit trials to continue to operate; both externally and internally in order to detect occurrences where a mismatch between policies and processes is obvious. In addition, this is also important in order to continuously trace for breaches of regulations not only when claims are made and compensations are required. The ability to proactively trace the access and disclosure of personal data is also essential in maintaining an effective compliance program [2].

The literature includes some works focusing on approaches to automated compliance auditing. As an early stage work we cite for example Hippocratic Databases (HDB) compliance auditing [2] which was extended thereafter in [18, 23]. This approach allows enterprises to log information about past queries in the form of metadata constituting a database log. Auditors could therefore track past data usage attempts by relying on the query log where information such as: the identity of users who have accessed a particular item, the date and time of queries, the purpose of access and the final recipients of the information could be determined. Moreover, hippocratic databases compliance auditing adopts a way of auditing disclosures by relevance ranking which is based on tracking the origin of sensitive data that was subject to

misuse by comparing it to the outputs of past queries saved in the backlog [21]. The query with high similarity to the disclosed data will have the high relevance rank. Similarity is calculated though specific proximity measures [21]. HDB along with other work such as that in [24] have recently tackled the issue of databases compliance audit, have undoubtedly greatly enhance the accountability of database systems by allowing enterprise to save past events of data access in order to verify its compliance with high level policies. However, by adopting these approaches the task of compliance verification will continue to involve heavy human-based verification. This is because the proposed solution does not incorporate mechanisms to automatically evaluate the handling of specific, fine-grained legal requirements. Regulatory compliance is very much related to vocabularies, meanings and interpretations [6, 28, 37]. It is therefore meaningful and necessary to adopt an approach that looks at the integration of semantics in the way privacy policies are specified and enforced. This will facilitate the validation of formal data usage policies against regulations that are written in natural languages.

In relation with semantic policy-based compliance checking, the work in [4, 22, 46] provides a solution to some aspect of compliance auditing. The work focuses on the adaptation of semantic rules dictating requirements of compliance applicable to a given security domain with another set of rules referring to another security domain. This work is more useful when tackling the problem of policy interoperability in heterogeneous domains. For example, in order to check whether the policies adopted respectively by an organization A and an organization B are compliant, the structure of the rules in both sets (the reference set and the target set) are supposed to be similar. In our case the policy compliance checking approach is designed to make a map check compliance between two sets of rules that have different structures and used for different purposes: one to indicate requirements of compliance and the other is used to enforce compliance while inferring access control decisions and obligations. In other terms, we are testing whether the requirements generated by one set of rules are satisfied by the other set of enforcement rules. The works in [4, 22], build on other previous works in [37, 38, 46]. It suggests the use of fuzzy sets theory in order to deal with limitation of The Semantic Web Rule Language (SWRL) for modelling uncertainty and dealing with missing compliance information in rule-based policy specification. Moreover, in relation to auditing compliance in business processes, the work in [39, 42], encodes business process models into semantically-annotated digraphs then defines ways of measuring proximity relation of the process to another process predefined as compliant. It also provides mechanisms that use compliant process models and compliance patterns based on heuristic guidelines for detecting and resolving non-compliant processes.

As previous studies in the same context, we have described in [37, 38] how semantic web technologies were used to classify the resources that we would like to protect. At that stage the resources were specified using the metadata captured within a privacy and data protection ontology we have designed and implemented. We also showed how different scenarios of data/resource sharing are modelled within the same ontology. In [36], we described extensions to the previous model (with necessary metadata added) and extend the data sharing scenarios to include privacy policy contexts applied to particular medical or healthcare scenarios. We then showed how this allows a better specification and enforcement of security and authorisation policies in the context of cloud computing as described in [5, 6]. This work did not stop at the control of access to data, but has rather focussed on ways of controlling any type of data usage even after the data were shared with a party belonging to an external domain. In addition, the work allowed the disclosure of data handling policies to external parties receiving personal data. In [13] we have developed models for sensitive data discovery in hospital business processes. The management of patient data is therefore easy to govern by applying GDPR and the Health Insurance Portability and Accountability Act (HIPAA) rules as shown in [35]. Moreover, details on the ontology-driven relational query formulation using the semantic and assertional capabilities of OWL-DL are presented in [30]. In this paper we look at how to make healthcare IT systems more self-disciplined and trustworthy by integrating a privacy audit dimension to patient data management services.

3. The Compliance Checking Approach

Our previous work as presented in [36, 37] has resulted in the production of models of data sharing along with abilities to generate and record legal requirements for privacy protection. Moreover, it included the ability to assess the context presented by each case of sharing and generates applicable access control permissions and obligations at the time the sharing request is submitted for validation to an access control decision-making system. In this paper, we are taking this work further by including an automated framework allowing auditors to test for compliance between heterogeneous domain (eHealth cloud domain)'s high-level policies and their enforced privacy preserving controls. This solution helps to verify that the data usage decisions taken by the adopted security middleware are in-line with the requirements and guidelines dictated by the organizational policies. We would like to highlight again the fact that we only deal with requirements and

obligations that have been considered for enforcement through a policy-based approach. Our approach suggests benefiting from the contextual information and past data sharing events recorded in our ontology in order to allow an auditor to track many useful information including:

1. An instance of data sharing that has previously been executed by the system described based on specific models capturing all required context information.
2. The set of privacy requirements applicable for the instance of data sharing in question.
3. The disclosure authorisation decision the system made at run time.

Afterwards, the system could then automatically check if a compliant case was in place the time the actual sharing was authorised in the past. This is effectuated through an automated mapping, using a rule-based approach between the data sharing context information and the privacy requirements interpreted from the legislation. We end up with two possibilities either

1. The requirements were satisfied; therefore, the system reports a compliant data sharing to the auditor.
2. The system indicates that alternative proof of compliance is required.

Hence the auditor pursuits investigating compliance of the data-sharing event as described in Figure 1:

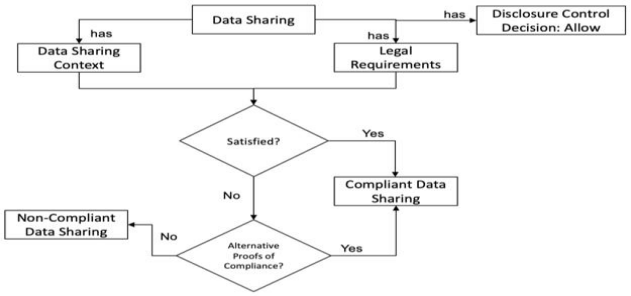


Figure 1. Compliance auditing model.

For the logging of past data sharing events, we aim to use a medical data sharing ontology as an audit trail log, which is explained in forthcoming sections. For future, we are already extending this to systematically link our ontology to a relational database. This will allow real time usage of our framework.

3.1. Illustrative Scenario

We describe the features of the compliance checking framework with an example from healthcare. This example presents a "data sharing" scenario where the data is shared between two nodes of a European Health Grid infrastructure. It involves the sharing of a patient mammogram collected in the *Breast Cancer Screening* programme at a UK hospital from a patient "Mrs. White" for the purpose of diagnosis and treatment. Mrs. White has provided her consent at the time of collection. "Dr House" the radiologist who is taking responsibility for Mrs. White's case needs a second opinion from "Dr Casa" a colleague at an Italian hospital. The mammogram cannot be fully anonymised before sending, as some personal data about Mrs. White are necessary in order to make an accurate judgment about 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.
- If the secondary purpose is compatible with the original purpose of "Breast cancer diagnosis and treatment".

3.2. Logging of Data Sharing Events to be used for Compliance Auditing

In most of the existing audit approaches, data logging usually occurs after performing specific queries to the data stored in the database. Unlike these approaches, our logging tasks are not executed as database triggers. Instead, they are meant to be executed as access control systems obligations and would only be triggered if a specified access control was permitted. Here, the data that is to be logged is same as used/generated by the privacy aware access control process. In particular, the logged data consists of the context of the access control request and the access control response for all the past data access and disclosure that have been allowed by the system as well as the time those actions were performed. For example, a data sharing can have the following context as shown in Figures 2 and 3:

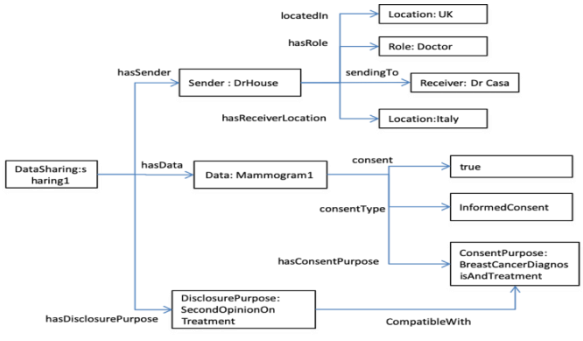


Figure 2. Example of a data sharing context model (instantiated).

In addition, the following access control decision:

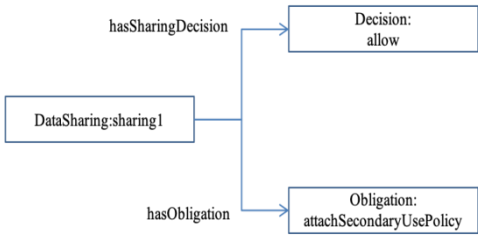


Figure 3. An Example of data sharing decision.

3.3. Generating Legal Requirements for Past Data Sharing Events

Our semantic web rules as presented in [36, 37] are compliant with the privacy and data protection European law, as they have been designed by explicitly translating high-level natural language (English) policies interpreted from text law. This has been achieved through the use of SWRL [32, 41] rules by annotating each rule with its reference in a text law. The rules allow the reasoning about data sharing context information in order to infer the applicable compliance information as dictated in high-level policies. Through this inference the information or requirement generated is being recorded in the ontology as properties of the specific case of sharing where its information is given as facts to the rule engine. The generated requirements could be classified into two sets; enforceable requirements and non-enforceable requirements. To distinguish between enforceable and non-enforceable requirements, we create two disjoint classes in our compliance management ontology named *EnforceableRequirement* and *NonEnforceableRequirement* respectively. Both of these *classes* are defined as *sub-classes* of the *LegalRequirement* class. An example of non-enforceable requirement (that could be found in the requirement of fair processing represented in article 8 of the European directive on data protection [11] indicates that consent should be provided in a fair way where the patient was not under some kind of threat or pressure when they provided their consent. We believe it is hard to automatically test for the satisfaction of this requirement. Therefore, we consider it as a non-enforceable requirement. Moreover, other non-enforceable requirements are the ones that are difficult to be enforced through a policy-based approach but could be enforced through other mechanisms. For example, the requirement of complying with high-level policies that require the anonymisation of some category of data before the data could be shared for research purposes. The organization would be required to incorporate data anonymisation tools conforming to specific obligations in order to be considered compliant. Through the use of policy-based approaches, we can only perform a check that the data has been anonymised before it could be sent. For example, by checking if the data protection status of the item to be shared is set to *anonymous data*. But we cannot guarantee that the data has been anonymised. We can then consider the specific requirement of anonymising the data as difficult to enforce or non-enforceable. In the light of the above arguments we end up with two sets of requirements Enforceable Requirements (EReq) and Non-Enforceable Requirements (NEReq) as described in Figure 4.

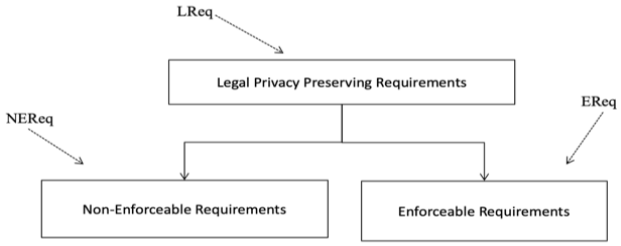


Figure 4. Model of Legal Requirement Hierarchy

We only focus on EReq when we test for matching between guidelines and privacy aware access control rules. In the following section, we present the rules we use for verifying legal requirements satisfaction.

3.4. Testing for Requirements Satisfaction

In the previous section, we have explained how to generate legal requirements for a given data sharing event selected from the events log. We have also showed how our ontology distinguishes between enforceable requirements and non-enforceable ones. In this section we describe our approach to test for requirement satisfaction at the time of requesting a data access. In other words, we want to check whether the user has complied with legal requirements while accessing or disclosing personal data in the past. For this purpose, we only need to consider the enforceable requirements. In fact, this process is not applicable for the non-enforceable requirements.

We adopt a rule-based approach in order to accomplish the task of compliance checking. We implement our approach through the use of SWRL rules that has as antecedent a conjunction of two Web Ontology Language (OWL) axioms, for more details on the use of ontologies for effective knowledge modelling and information retrieval see [16]. As described in Figure 5 the first axiom is representing a specific legal requirement (we call it a requirement axiom), which would be generated by executing the group of legal requirements rules from our knowledge base and added by the rule engine as property of the data sharing instance that is subject to audit. The other axiom is an OWL axiom testing for the satisfaction of the legal requirement (we call it a tester axiom). This axiom does already exist as part of the semantics describing the data-sharing event being audited.

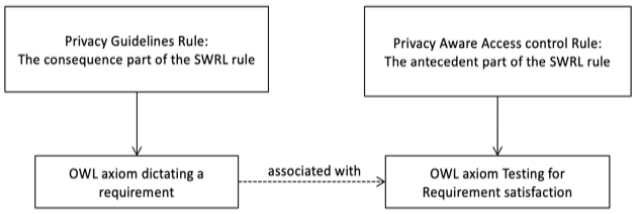


Figure 5. Extracting compliance knowledge from business rules.

For example, the OWL axiom HasConsentNecessity(DataSharing, Necessity) is presented as follows:

OWL Axiom: *hasConsentNecessity (DataSharing, Necessity)*
OWL Property: *HasConsentNecessity*
Property Type: *Object Property*
Property Domain: *DataSharing*
Property Range: *Necessity*

We associate the following OWL axiom: *consent(DataSharing, Boolean literal)*needed for testing the satisfaction of the requirements dictated by the earlier axiom which models the provision of consent for the same case of sharing.

OWL Axiom: *consent(DataSharing, Boolean literal)*
OWL Property: *Consent*
Property Type: *DataType Property*
Property Domain: *DataSharing*
Property Range: *A value of Boolean type{true, false}*

The consequence part of the rule contains an OWL axiom indicating whether the requirement was satisfied. We call this the *satisfaction axiom*.

OWL Axiom:
ConsentNecessitySatisfaction(DataSharing, Boolean literal)
OWL Property: *ConsentNecessitySatisfaction*
Property Type: *DataType Property*
Property Domain: *DataSharing*
Property Range: *A value of Boolean type{true, false}*

Our SWRL rules will then follow the following template named as 'compliance rule template':
T: Requirement Axiom ^Tester Axiom → Satisfaction Axiom

Based on the axioms presented above we can build the SWRL rule model which follows the *compliance rule template* described above. This rule checks whether the requirement of consent necessity for a specific data sharing was satisfied before allowing it to happen in the past. The requirement of consent necessity is part of our privacy requirements conceptual model (ontology) presented in previous work [37]. It is used to model the legal obligation with relation to obtaining the patient consent before using their medical data.

R: hasConsentNecessity(?DataSharing, Necessity) ^ Consent (?DataSharing, Boolean literal) →ConsentNecessitySatisfaction(?DataSharing, Boolean literal)
Some instantiations of the rule would be as follows:

- First, for the case where obtaining consent is a necessary obligation and where the consent for sharing was obtained from the patient before using the healthcare information system to disclose the data:

  - *Rule* 1: hasConsentNecessity (dataSharing1, Necessary)
    ^ Consent (dataSharing1, true)
    →ConsentNecessitySatisfaction(dataSharing1 ,true)

- The second instantiation of the rule is for the case where the consent is a necessary obligation but it was not obtained before the sharing was allowed and therefore this requirement was not complied with or in other terms has not been satisfied:

  - *Rule* 2: hasConsentNecessity (dataSharing2, Necessary)
    ^ Consent (dataSharing2, false)
    →ConsentNecessitySatisfaction(dataSharing2 , false)

- The third instantiation of the rule is for the case where the consent necessity obligation was not implemented in the access control policy model used by the healthcare data sharing application. By consequent no tester axioms were used in the access control rules applicable to the sharing context (this is applicable to the case of dataSharing1) but it was not obtained before the sharing was allowed and therefore this requirement was not complied with or in other terms has not been satisfied i.e.

  - *Rule* 3: hasConsentNecessity (DataSharing1, Necessary)
    →ConsentNecessitySatisfaction(DataSharing1, false)

We break an instant here to remember the fact that OWL is governed by an Open World Assumption. This means that the absence of a tester axiom among the OWL semantics of the data sharing which is subject to reasoning, does not certainly mean that the requirement was not satisfied unless we explicitly state this in the ontology. Hence, in order to be able to infer that: "if the tester axiom is not found therefore the requirement would be considered unsatisfied", we are out to close the Open World Assumption through other means. We suggest the following assumption and algorithm in order to solve the problem: For the example stated above Set-1 will contain:

- *Rule* 1: hasConsentNecessity (dataSharing1, Necessary)

^ Consent (dataSharing1, true)
→ ConsentNecessitySatisfaction(dataSharing1, true)

- *Rule* 2: hasConsentNecessity (dataSharing2, Necessary)

^ Consent (dataSharing2, false)
→ ConsentNecessitySatisfaction(dataSharing2, false)

And Set-2 will contain:

- *Rule* 3: hasConsentNecessity (dataSharing1, Necessary)

→ ConsentNecessitySatisfaction(dataSharing1, false)

- **Assumption**: We distinguish between 2 sets of rules (requirements-satisfaction checking rules):
  - *Set* 1: Consists of rules where the antecedent contains both clauses (a requirement clause and a tester clause).
  - *Set* 2: Consists of rules where the antecedent contains only a requirement clause.

Algorithm 1: Requirements Satisfaction Testing Algorithm

For each instance of requirement, the rule engine will first run Set-1.
If any of the rules fire up therefore the engine will consider the knowledge inferred by the consequence of the rule.
Else the engine will run the second set of rules (Set-2).

Moving to Set 2 we assume that the tester clause was not found among the semantics describing the data Sharing context therefore the operational controls used to allow this sharing to happen, did not enforce the given requirement. As stated above Set 2 gathers the set of rules that allows to infer; for a given requirement; a non-satisfaction status. As a result, the engine will set the value of requirement satisfaction property to non-satisfied.

4. The Privacy Compliance Audit Architecture

The compliance management framework has an audit mechanism module. In this regard, Figure 6 provides its high-level view, which contains the following components:

- **Privacy Compliance Audit (PCA)**: It allows a privacy compliance auditor to generate audit reports about selected data-disclosures recorded in the ontology log. The audit report reveals information about data disclosure occurrences to the auditing officer allowing him to decide on the degree of compliance. The data in the report specify the legal requirements for complying with privacy legislation and whether these requirements were met at the time of allowing disclosure of the data (or also called policy evaluation time).
- **Legal Requirements Generator (LRG)**: Once invoked this component generates a set of legal requirements that are relevant to a specified event of data sharing. The LRG answers requests from the audit component PCA by generating legal requirements associated with the facts provided as input. The output is a set of OWL axioms of the instance of sharing in question describing legal requirements applicable. The output is sent back to the PCA requestor instead of medical users.
- **Requirements Satisfiability Checker (RSC)**: This component will take charge of checking whether the set of requirements generated by the legal requirements generator LRG were satisfied, before allowing it to happen. This is achieved through effective assessment of the access control request and response produced through the access control policy enforcement process. The assessment involves first, parsing the properties of the data sharing instance described as attributes of the access control policy context and then, evaluating them against the set of legal requirements generated for the same context. Information describing both the access control request and answer is saved in the ontology as a systematic action of running the jess rule engine on the SWRL access control rules.

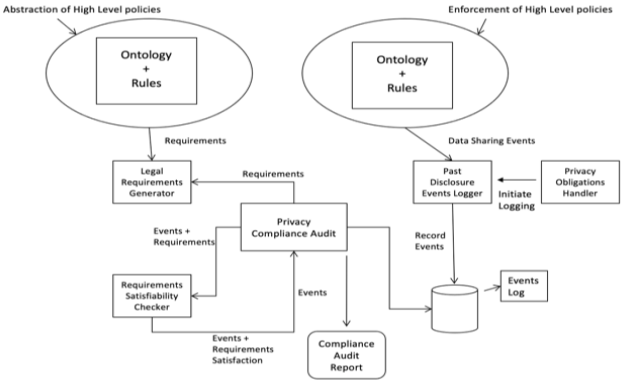


Figure 6. High level architecture of compliance auditing framework.

5. Compliance Checking Scenario

In this section we present a use-case scenario built on the illustration scenario presented earlier. We are using this use-case to show the outcome of the proposed system when the use-case is executed for a specified set of accessed data and for another context information. We consider a set of compliant accesses and also few failures cases. In order to perform a routine audit, we suppose that a procedure will be in place that would allow an auditor to submit an audit query. The query aims to generate a list of all the disclosures of a mammogram performed the previous month between the UK and Italy. As a result, the sharing of the mammogram between Dr House and Dr Casa (presented in the illustration scenario) was one of

the returned past disclosures. Recalling more details about this data sharing case described before, Dr House has obtained an informed consent from Mrs White allowing him to share her data with Dr Casa for the purpose of obtaining a second opinion for diagnosis and treatment procedures. A record of Mrs White's consent form and consent letter was kept in the database along with her signature as a proof for compliance with the legal and ethical requirement of explicit informed consent. As a result, a flag was added in the database indicating that explicit informed consent was obtained. At the time of disclosure, the following information about the sharing operation was generated from the ontology and logged in a backlog.

The SWRL rule presented below is the rule responsible for inferring the access control decision for the scenario described above. The rule states that "When sharing a UK patient mammogram between the UK and Italy for the purpose of getting a second opinion on treatment; if the patient consent has been originally given for breast cancer diagnosis and treatment purposes with which the purpose of processing is compatible then the sharing is allowed and the secondary usage policy of the data should be disclosed along with the data"

dataSharing(sharing1)
- ∋ hasSender(sharing1, Dr House)
- ∋ hasReceiver(sharing1, DrCasa)
- ∋ hasPurpose(sharing1, SecondOpinionOnTreatment)
- ∋ locatedIn(DrHouse, UK)
- ∋ locatedIn(DrCasa, Italy)
- ∋ concerning(sharing1,M1)
- ∋ belongsTo(M1, UK)
- ∋ patientConsent(M1, true)
- ∋ hasConsentPurpose(M1, BreastCancerDiagnosisAndTreatment)
- ∋ compatibleWith(p1, p)

→ hasSharingDecision(sharing1,allow)

∋hasObligation(sharing1, attachSecondaryUsePolicy)

Following the approach in [37] which focuses on designing the rules for inferring legal requirements; we present the rule presented below as an applicable rule for our scenario. The rule states: *"a specific patient consent is necessary for the sharing or the sending of UK patient data from the UK"*.

dataSharing(?x)
∋hasSender(?x, ?s)
∋hasReceiver(?x, any)
∋locatedIn(?s, UK)
∋sharedData(?x, ?d)
∋belongsTo(?d, UK)
→ consentNecessity (?x, Necessary)
∋consentSpecificity(?x, SpecificConsent)
∋con_Explicitness(?x, any)

The auditor can then choose to generate an audit report about the selected sharing from the audit backlog. The report describes the context of sharing as along with the decisions made by the system as a disclosure control decision, any obligations that was required from the sender or the system. The report includes as well all the legal requirements for the audited context of sharing and an indication of whether each requirement was satisfied or not. Detailed report is provided in Table 1.

Table 1. Data sharing compliance report.

Data-Disclosure-ID	Legal Requirements	Requirement Enforceability	Satisfied?	Disclosure-Control-Decision
sharing1	Consent Informed-Consent Explicit-Consent Fair-Consent	EReq EReq EReq NEReq	Yes Yes Yes No	Allow
sharing2	Consent for primary Purpose Secondary-Purpose-Compatibility	EReq EReq	Yes Yes	Allow

After examining the data sharing compliance report shown in Table 1 the auditor notices that some of the legal requirements have not been satisfied however, the decision of the access control system was "***allow***". For this scenario, the sharing would be considered as non-compliant if and only if the non-satisfied requirements are enforceable. If the non-satisfied requirements were of type non-enforceable, then the organization or the sender of the data will be asked to show other (non-automated) proofs of compliance. Visiting again our example of reports described above in Table 1, the requirement that has not been satisfied is the requirement of fair consent implementing article 8 principles [11] of data protection of the EU Directive [11] and the distinguishable purpose for consent required by the EU General data protection regulation [12] and minimal data processing required by the article 24 of the EU general data protection regulation [25, 27, 31]. This requirement was set as non-enforceable which means that there are no means of enforcing or implementing this requirement in an automated way or the means do exist but were considered by the hospital as not practical. For this case, Dr House or another appointed body from the hospital management team will be asked to show other non-automated proofs for complying with this requirement. The proof for complying with such requirements could be the general hospital procedure of obtaining consent from patients for example for special cases of disclosure i.e., for disclosing patient data outside the UK. This procedure would involve the existence of two witnesses one from the hospital and one appointed by the patient. Both witnesses need to sign a form declaring that consent was taken in a fair and lawful way where the meaning of fair and lawful consent is explained in the consent form.

The audit report for sharing2 highlights two legal requirements necessary for the lawfulness of the data sharing. These are consent-for-primary-purpose and secondary-purpose-compatibility. The Consent-for-Primary-Purpose represents the legal requirement for collecting personal data from patient. As required by

most national and European regulations patient consent should be taken for the collection of private and medical data from a patient. The consent should be linked to a specific purpose of collection and processing. The requirement of Secondary-Purpose-Compatibility is for the case where the data is being used for a second purpose of processing. As specified in most national and European regulation the secondary purpose should be compatible with the purpose the patient has consented for otherwise a new request of consent is necessary. Both generated requirements for sharing2 where enforceable. In addition, the report indicates that both requirements were satisfied by the data sharing requestor. Based on these facts the auditor will decide that the data sharing was compliant.

7. Discussion

Our system could distinguish between enforceable and non-enforceable requirements. If the requirement is enforceable, the system will check whether it was complied with at access control run time. According to the received output, an auditor could reflect on whether a past data-sharing event would be considered compliant with regulations. If the requirement is not enforceable, we suggest that the organization should provide other proofs for compliance. Our approach is built on the use of SWRL rules for specifying:

1. Privacy aware access control.
2. Privacy requirements generation.
3. Privacy audit rules.

These rules were tested for correctness and were executed using the JESS rule engine [32] plugged in Protégé [34] 3.4 platform. The concepts used within the SWRL rules were specified in OWL ontologies edited in Protégé [34] 3.4. As a result of the execution of the audit rules and more precisely the privacy requirement satisfaction rules; the satisfaction of the requirements was checked and the result was recorded back in the ontology as OWL properties of the data sharing event in question.

Our work does not provide a complete automation of the process of privacy compliance audit. This is because we believe a human intervention is always necessary to investigate cases such as checking for compliance with non-enforceable requirements. Besides, the issue of exposing private data about users of the system to be exploited by our audit mechanism might raise some privacy preserving concerns. For the time being, we have concentrated more on protecting the privacy of individuals (patients) their data is stored in healthcare systems storage. More work needs to be done to insure the auditors are allowed to check only necessary information about users who were involved in a non-compliant data sharing. It is necessary to minimize the amount of data logged in audit logs about users of the system and organizations' employees. Also, it is necessary to minimize the amount of exposed data about users and employees who were only involved in compliant data sharing. For this category of users, no further procedures should be taken by the auditor or the organisation in order to get them identified.

Many existing works were found in the literature, which seem a big interest in privacy management in healthcare [14, 17, 20, 26]. The topics of access control for cloud systems in healthcare were broadly tackled for example in [1, 3, 5, 9, 45]. The work in [44] suggested an enhancement to privacy management throw the use of Block chain. Some works have also looked at privacy compliance management and audit [4, 8, 24, 25, 29]. To the best of our knowledge we couldn't find in the literature an approach that similarly test for compliance with explicitly specified requirements dictated by regulations in healthcare. However, some interesting issues are still open. First, the model of logged data has been designed in a way that allow an informative analysis of the data log by auditors. With the fact that data logging is a demanding process, the model should therefore be optimised in future work in order to allow only the necessary data to be recorded. Second, high level policies are by nature mutable and dynamic. Some of the designed policies in this work could either change or be overridden by other newly issued policies (dictated by data protection law or by other law interfering with the medical and healthcare domain as medical law [6]. The impact of these possible changes on the outcome of the audit mechanism we are proposing in this paper needs to be studied and tackled in future work.

8. Conclusions

In this paper, we have addressed the problem of audit for the purpose of privacy compliance assurance in healthcare heterogeneous domains such as eHealth Cloud systems. Particularly, we have introduced a framework for privacy compliance checking. It could be used for checking if the access control decision making has incorporated a proactive verification adherence to the requirements dictated by high level legal and ethical policies. In a future work, we think this framework should be extended to allow only the necessary data to be recorded about the users while recording previous data sharing events in the audit log. We will be also looking on how to allow policy overriding and refreshing in order to comply with newly issued data protection regulation in the future.

References

[1] Al-Muhtadi J., Shahzad B., Saleem K., Jameel W., and Orgun M., "Cybersecurity and Privacy Issues for Socially Integrated Mobile Healthcare

Applications Operating in A Multi-Cloud Environment," *Health Informatics Journal*, vol. 25, no. 2, pp. 315-329, 2017.

[2] Agrawal R., "Privacy Enhancing Techniques for Database Systems," *in Proceedings of the 9th International Conference for Extending Database Technology*, Greece, 2004.

[3] Asharov G., Halevi S., Lindell Y., and Rabin T., "Privacy-Preserving Search of Similar Patients in Genomic Data," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 4, pp. 104-124, 2018.

[4] Atymtayeva L. and Kozhakhmet K., "Development of Expert System for Information Security Audit," *International Journal of Computer Research*, Huttington, vol. 22, no. 4, pp. 399-433, 2015.

[5] Belaazi M., Rahmouni H., and Bouhoula A., "Towards a Legislation Driven Framework for Access Control and Privacy Protection in Public Cloud," *in Proceedings of 11th International Conference on Security and Cryptography (SECRYPT)*, Vienna, pp. 1-6, 2014.

[6] Belaazi M., Rahmouni H., and Bouhoula A., "An Ontology Regulating Privacy Oriented Access Controls," *in Proceedings International Conference on Risks and Security of Internet and Systems*, Mytilene, pp. 17-35, 2016.

[7] Bender E., "4 BIG QUESTIONS," *Nature*, vol. 527, no. 7576, p. S19, 2015.

[8] Brodin M., "A Framework for GDPR Compliance for Small- and Medium-Sized Enterprises," *European Journal for Security Research*, vol. 4, no. 2, pp. 243-264, 2019.

[9] Clarke N., Vale G., Reeves E., Kirwan M., Smith D., Farrell M., Hurl G., and McElvaney N., "GDPR: an Impediment to Research?" *Irish Journal of Medical Science*, vol. 188, no. 4, pp. 1129-1135, 2019.

[10] EU Directive 2011/24/EU of the European parliament and of the council on the application of patients' rights in cross-border healthcare. Official journal of the European Union. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:088:0045:0065:EN:PDF, Last Visited, 2020.

[11] EU Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML, Last Visited, 2020.

[12] EU Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of European Union, L119/1. http://eur-lex.europa.eu/eli/reg/2016/679/oj, Last Visited, 2020.

[13] Essefi I., Rahmouni H., and Ladeb M., "Sensitive Data Discovery in Care Pathways Using Business Process Modelling and HL7-CDA," *International Journal on Advances in Life Sciences*, vol. 11, no. 1&2, 2019.

[14] Feltus C., Grandry E., Kupper T., and Colin J., "Model-Driven Approach for Privacy Management in Business Ecosystem," *in Proceedings of the 5th International Conference on Model-Driven Engineering and Software Development*, pp. 392-400, 2017.

[15] Fernández-Alemán J., Señor I., Lozoya P., and Toval A., "Security and Privacy in Electronic Health Records: A Systematic Literature Review," *Journal of Biomedical Informatics*, vol. 46, no. 3, pp. 541-562, 2013.

[16] Munir K. and Anjum M., "The Use of Ontologies for Effective Knowledge Modelling and Information Retrieval," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 116-126, 2018.

[17] Gope P. and Amin R., "A Novel Reference Security Model with the Situation Based Access Policy for Accessing Ephr Data," *Journal of Medical Systems*, vol. 40, no. 11, pp. 242, 2016.

[18] Grandison T., Johnson C., and Kiernan J., *in Handbook of Database Security*, Springer, 2008.

[19] Information Commissioner's Office, 2017, Consultation: GDPR consent guidance. Available at: https://ico.org.uk/media/abouttheico/consultations/2013551/draft-gdpr-consent-guidancefor-consultation-201703,pdf, Last Visited, 2020.

[20] Iwaya L., Giunchiglia F., Martucci L., Hume A., Fischer-Hübner S., and Chenu-Abente R., *IFIP International Summer School on Privacy and Identity Management*, Springer, 2015.

[21] Johnson C. and Grandison T., "Compliance with Data Protection Laws Using Hippocratic Database Active Enforcement and Auditing," *IBM Systems Journal*, vol. 46, no. 2, pp. 255-264, 2007.

[22] Kalaiprasath R., Elankavi R., and Udayakumar R., "Cloud Security and Compliance-A Semantic Approach in End to End Security," *International Journal of Mechanical Engineering and Technology (Ijmet)*, vol. 8, no. 5, pp. 987-994, 2017.

[23] Kirchberg M. and Link S., "Hippocratic Databases: Extending Current Transaction Processing Approaches to Satisfy the Limited Retention Principle," *in Proceedings of 43rd Hawaii International Conference on System*

Sciences, Honolulu, pp. 1-10, 2010.

[24] Kwon J. and Johnson M., "Security Practices and Regulatory Compliance in the Healthcare Industry," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 44-51, 2013.

[25] Kwon J. and Johnson M., "Proactive Versus Reactive Security Investments in the Healthcare Sector," *MIS Quarterly*, vol. 38, no. 2, pp. 451-472, 2014.

[26] Liu W. and Park E., "E-Healthcare Cloud-Enabling Characteristics, Challenges and Adaptation Solutions," *Journal of Communications*, vol. 8, no. 10, pp. 612-619, 2013.

[27] Mahmood S. and Power L., "Getting to Know the General Data Protection Regulation, Part 6-Designing for Compliance. Privacy Law Blog. Available at: http://privacylawblog.fieldfisher.com/2016/getting-to-know-the-general-data-protection-regulation-part-6-designingfor-compliance/, Last Visited, 2020.

[28] Maxwell J., Antón A., Swire P., Riaz M., and McCraw C., "A Legal Cross-References Taxonomy for Reasoning About Compliance Requirements," *Requirements Engineering*, vol. 17, no. 2, pp. 99-115, 2012.

[29] Mohammed D., "U.S. Healthcare Industry: Cybersecurity Regulatory and Compliance Issues," *Journal of Research in Business, Economics and Management*, vol. 9, no. 5, pp. 1771-1776, 2017.

[30] Munir K., Odeh M., and McClatchey R., "Ontology-Driven Relational Query Formulation Using the Semantic and Assertional Capabilities of OWL-DL," *Knowledge-Based Systems*, vol. 35, pp.144-159, 2012.

[31] Negrouk A., Horgan D., Gorini A., Cutica I., Leyens L., Halfmann S., and Pravettoni G., "Clinical Trials, Data Protection and Patient Empowerment in The Era of The New EU Regulations," *Public Health Genomics*, vol. 18, no. 6, pp. 386-395, 2015.

[32] O'Connor M., Knublauch H., Tu S., and Mark M., "Writing Rules for the Semantic Web Using SWRL and Jess," *in Proceedings of the 8th International Protégé with Rules workshop collocated with Protégé*, Madrid, 2005.

[33] Olive M., Rahmouni H., Solomonides T., Breton V., Legré Y., Blanquer I., Hernandez V., Andoulsi I., Herveg J., and Wilson P., "SHARE Roadmap 1: Towards A Debate," *Studies in Health Technology and Informatics*, vol. 126, pp. 164-73, 2007.

[34] Protégé, The Protégé Ontology Editor and Knowledge Acquisition System, http://protege.stanford.edu/ Last Visited, 2020.

[35] Rahmouni H., Essefi I., and Ladeb M., "Enhanced Privacy Governance in Health Information Systems through Business Process Modelling and HL7," *Procedia Computer Science*, vol. 164, pp. 706-713, 2019.

[36] Rahmouni H., Munir K., Mont M., and Solomonides T., "Semantic Generation of Clouds Privacy Policies," *in Proceedings of International Conference on Cloud Computing and Services Science*, Barcelona, pp. 15-30, 2015.

[37] Rahmouni H., Solomonides T., Mont C., Shiu S., and Rahmouni M., "A Model-Driven Privacy Compliance Decision Support for Medical Data Sharing in Europe," *Methods of Information in Medicine*, vol. 50, no. 04, pp. 326-336, 2011.

[38] Rahmouni H., Solomonides T., Mont M., and Shiu S., "Privacy Compliance And Enforcement on European Healthgrids: An Approach Through Ontology," *Philosophical Transactions of the Royal Society A*, vol. 368, pp. 4057-4072, 2010.

[39] Sapkota K., Aldea A., Younas M., Duce D., and Banares-Alcantara R., "Automating The Semantic Mapping Between Regulatory Guidelines and Organizational Processes," *Service Oriented Computing and Applications*, vol. 10, no. 4, pp. 365-389, 2016.

[40] Shi X. and Wu X., "An Overview of Human Genetic Privacy," *Annals of the New York Academy of Sciences*, vol. 1387, no. 1, pp. 61-72, 2017.

[41] Straccia U., *Foundations of Fuzzy Logic and Semantic Web Languages*, Chapman and Hall/CRC, 2016.

[42] Ternai K., "Semi-Automatic Methodology for Compliance Checking on Business Processes," *in Proceedings of International Conference on Electronic Government and the Information Systems Perspective*, Valencia, pp. 243-256, 2015.

[43] Townend D., *Implementation of the Data Protection Directive in Relation to Medical Research in Europe*, Routledge, 2017.

[44] Wang H. Song Y., "Secure, Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain," *Journal of Medical Systems*, vol. 42, no. 8, pp. 152, 2018.

[45] Yimam D. and Fernandez E., "A Survey of Compliance Issues in Cloud Computing," *Journal of Internet Services and Applications*, vol. 7, no. 1, pp. 5, 2016.

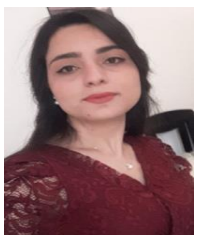
[46] Yip F., Wong A., Parameswaran N., and Ray P., "Semantic-Based Fuzzy Reasoning for Compliance Auditing," *in Proceedings of IEEE International Conference on Semantic Computing*, Santa Clara, pp. 299-306, 2008.

[47] Zerlang J., "GDPR: a Milestone in Convergence for Cyber-Security and Compliance," *Network Security*, vol. 2017, no. 6, pp. 8-11, 2017.

[48] Zhang J., Guo Y., and Chen Y., "Collaborative Detection of Cyber Security Threats in Big Data," *The International Arab Journal of Information Technology*, vol. 16, no. 2, pp. 186-193, 2019.

Hanene Rahmouni, holds a BSc (Hons) and a PhD in computer science from the University of the West of England (UWE – Bristol, United Kingdom). She is actually a Senior Lecturer in Information Science at the Faculty of Environment and Technology, UWE. Her research experience includes many aspects of design, development and implementation of distributed information systems using (databases, data science, knowledge representation, ontologies and semantic web technologies, artificial intelligence, security and data protection) technologies. Hanene was involved in various research projects funded by the European Commission at (UWE – Bristol, United Kingdom) and the HealthGrid organisation, France. Shehas also given numerous presentations to scientific audiences, technologists and decision makers.

Kamran Munir is Associate Professor in Data Science in the Department of Computer Science and Creative Technologies. Dr. Munir's research projects are in the areas of Data Science Big Data and Analytics Artificial Intelligence and Virtual Reality mainly funded by the European Commission (EC) Innovate UK and British Council. In the past he has contributed to the various CERN (the European Organization for Nuclear Research) and EC funded projects e.g. CERN WISDOM CERN CMS Production EC Asia Link STAFF EC Health-e-Child EC neuGRID and EC neuGRID4You (N4U) in which he led the Joint Research Area and the development of Data Atlas/Analysis Base Big Data Integration and Information Services.



Intidhar Essefi (BSc,MSc) is a consultant in healthcare systems interoperability. She was involved in many projects in the area of medical informatics, precisely in the area of data security and clinical process modelling. Intidharis a member of the Research Laboratory of Biophysics and Medical Technologies at the Higher Institute of Medical Technologies of Tunis, Tunisia.

Marco Mont is a Principal Cyber Security Consultant at BMT, Defence and Security, UK. He was a Principal Cyber Security Architect and R&D Scientist at HP Labs, where he worked for more than 20 years (where the work in this paper was initiated). He holds a MSc and BSc in Computer Science. He is a Senior IEEE member and a ISC)2 CISSP certified professional. His areas of expertise includes cyber security, big data analytics, AI/ML applied to cyber security, cyber threat analysis & detection, risk management, enterprise security solutions, security architectures and cloud. He is the author/co-author of more than 50 peer reviewed papers published at international conferences and journals.

Tony Solomonides, PhD MSc (Math) MSc (AI) FAMIA, has been active in biomedical informatics since the mid-1980s. In the early 2000's he played a leading role in the MammoGrid and other healthgrid projects before the policy-oriented project "SHARE," which aimed to establish healthgrids as the preferred infrastructure for biomedical research. In the context of SHARE and after the end of that project, he worked with Hanene Rahmouni to address patient data sharing under divergent jurisdictions, using ontology to translate declarative legal knowledge to a logic of permissions and obligations for data transfers. This work was presented in 2010 to the United States Institute of Medicine (now the National Academy of Medicine), initiating his transition from the UK to the US. In the past eight years he has been active in Clinical Research Informatics as well as in Ethical, Legal and Social Issues in a succession of roles at NorthShore University HealthSystem's Research Institute.