Telematics and Informatics 35 (2018) 737–752

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Semantic privacy-preserving framework for electronic health record linkage

Yang Lu [*], Richard O. Sinnott

Computing and Information System, University of Melbourne, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 3 September 2016
Received in revised form 27 April 2017
Accepted 11 June 2017
Available online 20 June 2017

ABSTRACT

The combination of digitized health information and web-based technologies offers many possibilities for data analysis and business intelligence. In the healthcare and biomedical research domain, applications depending on electronic health records (EHRs) identify privacy preservation as a major concern. Existing solutions cannot always satisfy the evolving research demands such as linking patient records across organizational boundaries due to the potential for patient re-identification. In this work, we show how semantic methods can be applied to support the formulation and enforcement of access control policy whilst ensuring that privacy leakage can be detected and prevented. The work is illustrated through a case study associated with the Australasian Diabetes Data Network (ADDN – www.addn.org.au), the national paediatric type-1 diabetes data registry, and the Australian Urban Research Infrastructure Network (AURIN – www.aurin.org.au) platform that supports Australia-wide access to urban and built environment data sets. We demonstrate that through extending the eXtensible Access Control Markup Language (XACML) with semantic capabilities, finer-grained access control encompassing data risk disclosure mechanisms can be supported. We discuss the contributions that can be made using this approach to socio-economic development and political management within business systems, and especially those situations where secure data access and data linkage is required.

1. Introduction

Data analysis and especially real-time data analysis is key to designing successful data management systems (Abello et al., 2015). Since the concept of *business intelligence* (BI) was proposed in 1990s, it has been widely-applied to describe data-centric applications with analytical functionalities. BI requires data inputs and produces data outputs (Wixom et al., 2014). Business users and services often perform analytical techniques based on data warehouse approaches, where data is collected and aggregated from different sources into a single large-scale repository. By implementing BI solutions in different domains, IT infrastructure costs can be reduced. Such an approach can also provide more accurate analytical results and save time for stakeholders and data users (Saltz, 2015). However it is often the case that this aggregation is infeasible or impossible due to security and privacy concerns. This is especially the case in the medical domain and use of data with other forms of social or environmental data for example.

Due to the seismic transformation from paper-based patient data to electronic heath records (EHRs) that is occurring, the developing health IT infrastructure allows stakeholders across health industries to appreciate significant benefits such as remote care delivery and cohort recruitment. To further improve the effectiveness, efficiency, and quality of health services,

[*] Corresponding author.
E-mail addresses: luy4@student.unimelb.edu.au (Y. Lu), rsinnott@unimelb.edu.au (R.O. Sinnott).

http://dx.doi.org/10.1016/j.tele.2017.06.007
0736-5853/© 2017 Elsevier Ltd. All rights reserved.

the adoption of BI in the health sector is regarded as an opportunity to improve health care more generally (Mettler and Vimarlund, 2009). For instance, analysts are allowed to predicate epidemics of certain communities by exploring datasets with knowledge discovery techniques (Lopez et al., 2014). Healthcare BI could save costs for both providers and patients by removing redundancy and making more accurate decisions. An example of this is long-term care in the Netherlands (Spruit et al., 2014), where it was shown that healthcare business intelligence could promote knowledge discovery based on large datasets focused on chronic care.

According to recent studies, it has been recognized that patient-centred care is now the key to high-quality healthcare delivery (Dowsett et al., 2000; Kwan and Sandercock, 2004; Levesque et al., 2013; Pulvirenti et al., 2014). By conducting patient-focused approaches it is hoped to achieve maximum biomedical value by sharing information between professionals, patients and carers (Coulter and Collins, 2011). A key point in this approach is to maintain the interoperability of the systems where doctors and health researchers may exchange clinical decisions and population-based analysis results. To make this happen, it is necessary to build systems from a holistic perspective in making decisions and treatment planning. For instance, record linkage techniques in the biomedical domain allow a more complete picture of the health of the population than previously possible. Currently linkage techniques are widely employed in Australia such as Centre for Healthcare Record (CHeReL)[1] in New South Wales, SA-NT DataLink[2] in South Australia and Victorian Data Linkage (VDL)[3] in Victoria.

Challenges in the linkage-oriented systems include access control and data anonymity. Biomedical data containing patient demographics can be too sensitive to release. To eliminate malicious or non-malicious threats upon publishing EHRs to users, data custodians have adopted a range of solutions such as requiring informed consents, de-identification and reviewing the purpose of access data as applied. Technically, the authorization languages for restricting malicious operations on data should be interoperable among autonomous organizations. Due to the modularity and extensibility, XACML has been widely implemented for distributed security. Through specifying generic vocabularies as well as domain-specific facts, access requests for linkage data can be verified against XACML policies. However, implicit associations among concepts are typically neglected in manual data linkage operations, which may lead to "authorization missing" or data leakage dangers. Whilst data may not immediately be compromised, there is a danger of gradual erosion of privacy and the risk of re-identification of individuals through linkage with other data sources. To address this issue, this paper explores how semantic web technologies can be used to discover latent disclosure risks and mitigate the chance of privacy leakage.

2. Background

The ability to share clinical data in secure ways outside of the immediate health context is essential. In the field of healthcare and clinical research, online data for secondary use allows new discoveries on drugs, treatments, and clinical benchmarking more generally (Synnot et al., 2016). For instance, through comparing clinical records with existing metrics, hidden factors of complex diseases can be ascertained (Kontos et al., 2014). Meanwhile, applying new theories in disease prediction and effective control, it is expected to improve people's health and life quality (Abdelhak et al., 2014). Since it may involve personal privacy, researchers are required to use health data in an ethical and confidential way. Typically three procedures are required before accessing and using health data (O'Keefe and Connolly, 2010).

- *Informed consent*. Biomedical data usually contains both identifying and non-identifying information. Data custodians (hospitals/clinical institutions) are often required to collect consent letters from data subjects (patients). Upon receiving the confirmation, they can allow the research use of the sensitive information.
- *Anonymity*. To protect patient privacy, some health data is required to be anonymized for secondary use. Various degrees of anonymity are often conducted in accordance with the research nature, purposes, as well as the regions and even countries. For instance, comparing influenza data and HIV/AIDS data may have different levels of restrictions and sensitivities on anonymisation.
- *Access control*. Data access should be authorized through evaluating security policies, where domain-specific knowledge is often required. Thus what roles are required to access what data and in which context. Rules are typically formalized to minimize potential security risks. These policies tend to be statically defined and inflexible to more dynamic linkage scenarios.

2.1. Data anonymity

Data anonymity is legislated by many countries to guide and mentor data processing in research activities. For instance, the National Statement on Ethical Conduct in Human Research (2007) has categorized data into individually identifiable, re-identifiable and non-identifiable. In addition, they specify in what contexts, which type of data is allowed to be collected, stored and published. In the U.S., privacy issues of health data are regulated by Health Insurance Portability and Accountability Act 1996 (HIPPA) in which de-identified data levels and guidelines on different levels of anonymity have been stressed.

[1] Centre for Health Record Linkage (CHeReL). http://www.cherel.org.au/.

[2] SA-NT DataLink. https://www.santdatalink.org.au/.

[3] Victorian Data Linkage (VDL). https://www2.health.vic.gov.au/about/reporting-planning-data/victorian-data-linkages.

Particularly a common method "safe harbours" indicates the de-identification can be achieved by removing 18 attributes such as name, address, date, biometric information, serial numbers of personal devices etc. With the increasing complexity in data usage, however, it is not always easy to distinguish the sensitivity of data items, and thus it is impossible to rely on the removal of "sensitive information". Meanwhile, employing approaches such as k-anonymity and its deviations to preserve data through generalising and suppressing the "quasi-identifier" attributes so as to satisfy the mathematical model (Sweeney, 2002; Machanavajjhala et al., 2007; Li et al., 2007; Panackal et al., 2014). With these methods, data can be de-identified by generalising the attribute values against numerical requirements. For instance, each combination of attribute values should exist with at least k individuals in the dataset. In this way, individuals are de-identified through assuming recipients have limited background knowledge and thereby cannot distinguish the target from k-1 other entities. This may or may not be the case depending on the application, and when multiple data sets are combined, it may not be easy to recognise the quasi-identifiers which should be transformed. For instance, processing address information based on the unified requirement may cause privacy threats upon patients who live in sparsely populated areas. According to de Montjoye et al. (2013), with four pieces of spatio-temporal information it is adequate to re-identify data subjects. As a result, such privacy issues need to be solved in an automatic manner. Existing methods may result in disclosing sensitive information. Privacy protection based on geo-awareness should take the semantic meaning into consideration.

2.2. Access control

Regardless of data anonymity, restricting the access to authorized individuals is also essential. The commonest access control model, Role-Based Access Control (RBAC) was originally proposed to secure access by grouping users/resources and attaching security levels to those groups (roles/clearances). On that basis, RBAC variants such as Temporal RBAC, Location and Time-based RBAC, and Spatial RBAC have been proposed and designed for different purposes (Uzun et al., 2014; Mitra et al., 2016; Baracaldo et al., 2014). In terms of health data protection, most approaches are through extending RBAC models with ethical policies and other legitimate requirements in clinical treatment and research. For example, Sicuranza et al. (2014) has designed four access control models used in EHR systems around obtaining patient consents. Other ethics-related methods also include the content such as clinician-patient associations and the request purposes (Brown et al., 2010).

For more general purposes, the attribute based access control (ABAC) is defined by incorporating attributes into access policies. Compared to RBAC, it enables a finer-grained authorization process since there is no limit on the attribute as a "role". With the attributes specified to constraint the subject, resource, action and environment, XACML framework is widely adopted to implement ABAC systems. Fig. 1 illustrates the workflow in most XACML systems. Through the Policy Administration Points (PAPs) security experts can design policies towards specific applications and write them into repositories to evaluate real-time requests (Step 0). In the most common scenario, Policy Enforcement Points (PEPs) intercept user access requests (Step 1). Then by transferring requests to the Policy Decision Point (PDP), they are evaluated against local policies (Step 2 & 3). To support policy evaluation, here the Policy Information Points (PIPs) retrieve the requested attributes and submit them to the PDPs (Step 4 & 5). Decisions are made based on the satisfaction of attributes to conditions. Through replying to the PEP (Step 6), the next-step behaviour can be either an access failure or permission to access resources/services is granted and can occur (Step 7).

Policy languages describe restrictions of access based on pre-agreed policy vocabulary and attributes. Typically, policies are expressed like "*subject x can do action y on resource z under the condition w*". Like other structured methods, XACML is defined using XML, so that it can be exchanged across distributed systems. Working with designated algorithms, access decisions can be generated by composing access control rules and policies. However, due to the syntactic representation of concepts and relationships, it can be difficult to provide dynamic access control on heterogeneous resources. In order to support more complex and flexible applications, policies should incorporate semantic meaning rather than rely on the static description and comparison. To improve such expressiveness, this paper extends XACML with semantic languages in order to eliminate barriers of understanding. For instance, through formalizing synonyms or role hierarchies by *equivalence* and *subsumption* relations, it is possible to construct a foundation for ontological reasoning. Compared with explicit and static policies in current paradigms, semantic reasoning can allow new authorization rules to be generated from existing information (Cocos and MacCaull, 2010;Finin et al., 2008; Sharma and Joshi, 2016). For more advanced authorization, XACML policies can be specified as SWRL rules (O'connor et al., 2005; Rahmouni et al., 2010; Hsu, 2013; Yu et al., 2013; Addas and Zhang, 2014; Zhang et al., 2014; Orlando et al., 2015). However, these semantic models are hard to generalise due to lack of formulation and inference on XACML vocabulary and structures. Towards this issue, this paper focuses on a semantic-based policy framework through analysing core mechanisms and semantically formalizing components, relationships and algorithms. Finally, a general approach by which linked datasets can be released subject to adaptive protection leveraging enhanced anonymisation of data linkages (Kalloniatis et al., 2014).

2.3. Semantic web technology

Due to the autonomous management of distributed systems, it is essential to improve current mechanisms so that they can handle more dynamic access requests. Previous work shows how semantic technologies could augment current security solutions to improve the security and trust underpinning clinical collaborations (Lu and Sinnott, 2015). It is often the case

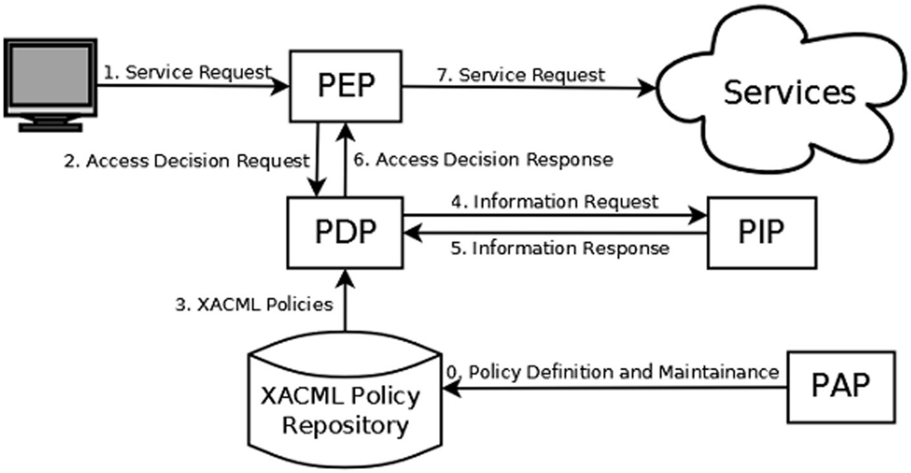


Fig. 1. Workflows in XACML-based systems.

that an ontology is used for policy development, compliance checking and knowledge discovery. Through the Ontology Web Language (OWL), XACML policies can be formalized as classes, instances and properties (Carroll et al., 2015). With the amount of heterogeneous resources increasing over time, it is necessary to deal with complex scenarios based on diverse regulations. To this end, the Semantic Web Rule Language (SWRL) has been established based upon extending semantic languages with *if-then* logic (Orlando et al., 2015). Access rules are specified as SWRL rules based on ontological concepts, which allow them to be recognized by semantic reasoners and thus enforce access control.

We focus explicitly on three procedures: policy formalization, compliance checks and knowledge discovery. Section 4 presents a scenario focused on type-1 diabetes patient records and how they can be securely linked with external geospatially coded social science data. Through standardizing the policy framework, we show how semantic technologies can be applied to infer implicit and risky relations that could jeopardise the privacy of released data.

3. Semantic based access control framework

The XACML standard covers the basic policy components such as **PolicySet**, **Policy** and **Rule**. Specifically, **Rule** is the smallest unit that describes access requirements in **Targets** and **Conditions**, and produces **Effects**, which are predefined as **Deny** and **Permit** to each access request. A **Target** is the bridge between requests and rules (policies). Through describing attributes for Subject, **Action**, **Resource** and **Environment**, requests can match with target elements and then locate applicable rules (policies). As long as applicable rules are located, certain decisions can be made based on the satisfaction of requirements in **Condition**. A **Policy** is composed of one or more rules, and optionally, some special operations described in the **Obligation** to be conducted on certain effects. Initiated by access requests with attributes, policy evaluation is resulted as one final decision, calculated from rule effects and algorithms. For instance, an example XACML policy about patient record access is presented in Listing 1, which speculates that "*patient records are allowed to be read by clinicians for research, if and only if they already obtained the consent letter from data subjects*".

As mentioned, the lack of semantic expressiveness impedes the finer-grained authorization. For instance, if a legal user presents the role "Doctor" for resource access, only "Deny" can be replied in this case – even though these two groups have the same level of trust and function in a given context. With heterogeneous security information and data abounding, the improved access control model is demanded in dynamic scenarios. To this end, Rahmouni et al. (2014) suggested using sematic web technologies (OWL + SWRL) to handle heterogeneous resources that are not merely syntactic. To support semantic reasoning, it is essential to map policy vocabulary, domain knowledge and internal logic into ontological concepts. Firstly, they simplify the mechanism of XACML systems as Semantic Rule (1). Through dividing the authorization rules into antecedents and consequences, policies can be described in the *If-then* form, the basic logic of SWRL. For each policy file, the **Condition** and **Obligation** are not the must.

$$\mathrm{Match(Target)(\wedge Match(Condition))\longrightarrow Effect(\wedge Obligation)} \tag{1}$$

According to the structure and key concepts Policy_1, its ontological elements can be correspondingly defined as Table 1. Specifically, **Generic vocabulary** reflects the policy structure while **Domain specific knowledge** is specific to domain contexts. Both parts are mapped as Classes, Instances and Properties, to which Domains and Ranges are specified. Generally, policies are structured based on generic contents, such as hasRule(Policy_1, Rule_1), hasTarget(Rule_1, Target_1), hasEffect(Rule_1, Permit) etc. In order to respond to access requests, it is necessary to incorporate domain

Table 1
Semantic Interpretation of XACML policy.

	Instances (Classes)	Properties	Domain	Range
Generic concepts	Policy_1 (Policy)	hasRule	Policy	Rule
		hasObligation		Obligation
	Rule_1(Rule)	fromPolicy	Rule	Policy
		hasEffect		Effect
		hasTarget		Target
	Target_1 (Target)	hasSubject	Target	Element
		hasAction		
		hasResource		
		hasEnvironment		
	Subjust_1 (Element)	hasAttribute	Element /Attribute	Attribute
	Resource_1 (Element)			
	Action_1 (Element)			
	Environment_1(Element)			
Domain specific knowledge	Clinician (Role->Attribute)	isRoleOf	Element	Attribute
	Read (Action->Attribute)	isOperationOf	Element	
	PatientRecord (Data->Attribute)	isSourceFrom	Element	
	ForResearch (Purpose->Attribute)	isPurposeFor	Element	
	Project-01 (Project->Attribute)	isActivityOf	Element	
		isProjectOf	Attribute	

*Classes are printed in bold; and class hierarchies are expressed by "Class->Subclass"

knowledge into the framework. For instance, specific attributes such as Clinician, PatientRecord are introduced. Through integrating these concepts within the framework, policies are able to evaluate relevant requests for accessing services/resources.

In order to generate new facts from existing knowledge, semantic reasoning is applied to improve the scalability and flexibility of data linkage solution. Policies specified as semantic notions provide the foundation for conducting semantic inference for policy formalization, compliance checks and domain knowledge management.

3.1. Policy formulation

Originally, policies were developed by statically assigning the generic variables with domain knowledge. However the ever-growing application contexts may result in a large range of reconstruction of policies, demands for dynamic protection have appeared in distributed environments. As mentioned, through formalizing policy concepts in OWL, policy can be dynamically generated based on inference results. Compared with the tightly-coupled pattern, the semantic approach allows dynamic policy development by associating and dissociating policy elements. Semantic Rule (2) is defined to describe XACML policy development. Specifically, it implies that assigning attribute y to the subject variable x, the attribute can be transmitted to the target via hasSubjectAttribute. Based on the facts in Table 1, results can be produced based on such rules on Subject, Resource, Action and Environment elements, such as hasSubjectAttribute(Target_1,Clinician), hasResourceAttribute(Target_1,PatientRecord) etc.

$$\mathrm{Target}(?\mathrm{a}) \wedge \mathrm{hasSubject}(?\mathrm{a},?\mathrm{x}) \wedge \mathrm{hasAttribute}(?\mathrm{x},?\mathrm{y}) \longrightarrow \mathrm{hasSubjectAttribute}(?\mathrm{a},?\mathrm{y}) \tag{2}$$

The reasoning process depicted in Fig. 2 reflects attribute transmission from targets to rules, policies and policy sets. Targets belong to rules (policies) used for applicability checking. Therefore Semantic Rule (3) embodies how target attributes dynamically transfer to Rules (Policies), expressed as hasSujectAttribute(Rule_1, Clinician). With a layer of generic variable names (e.g. Subject_1), a loosely-coupled policy formulation can be realised between the policy framework and domain specific knowledge. Compared with traditional methods where specified contents are designed in a top-down pattern, the semantic-based policy is formulated with basic elements. As a result, policy re-engineering is allowed in supporting diverse security requirements. By leveraging semantic reasoning in policy development, generic variables can be dynamically related to domain knowledge.

$$\mathrm{Rule}(?\mathrm{r}) \wedge \mathrm{hasTarget}(?\mathrm{r},?\mathrm{a}) \wedge \mathrm{hasSubjectAttribute}(?\mathrm{a},?\mathrm{y}) \longrightarrow \mathrm{hasSubjectAttribute}(?\mathrm{r},?\mathrm{y}) \tag{3}$$

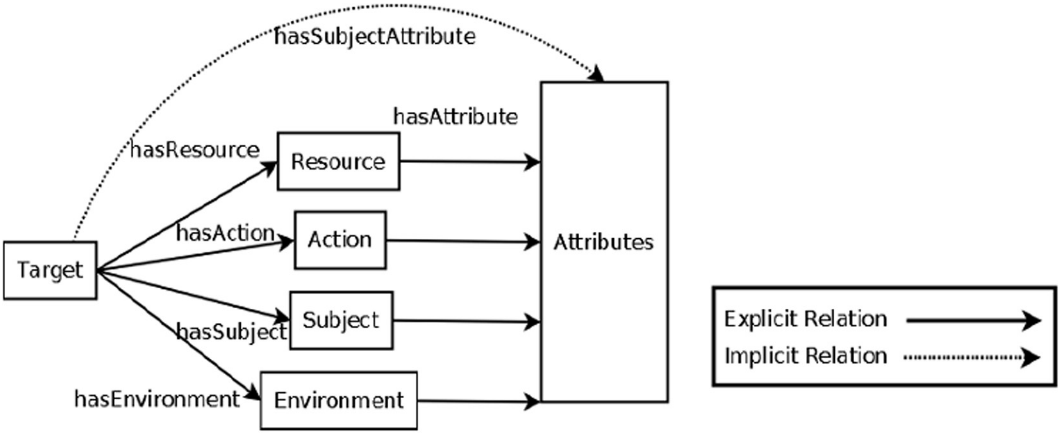


Fig. 2. Reasoning Categorized Attributes.

3.2. Compliance checking

In addition to formulation, semantic technologies can be used to achieve compliance checking between access requests and policies. Listing 2 shows an access request in the XACML framework. With the attributes in subject, resource, action and environment, the request is contextualised like *can clinicians read the patient records by participating in the project-01?*

Context handlers are established in requests for policies evaluation (Rissanen, 2013). At the policy side, targets package a set of attributes as constraints. Such attributes are pre-defined through mutual agreement but many bilateral agreements can be reached using different concepts and relationships. This cannot be detected by traditional methods due to the syntactic limitation. Additionally, attributes in different contexts may have different interpretations. As a result, policies with associated semantic meanings are proposed for applicability checking in distributed systems. To illustrate this, requests can be modelled by semantic concepts and associations. As shown in Fig. 3, the request context `RC_1` can be initialised by using pre-defined attributes in such predicates like `hasSubjectCategory (RC_1,Clinician)`, `hasResourceCategory (RC_1,PatientRecord)`, `hasEnvironmentCategory (RC_1,Project-01)` and `hasActionCategory (RC_1, Read)`.

The XACML policy evaluation is shown in Fig. 4. Upon receiving requests from PEPs, PDPs will initiate a two-stage checking: the request context is required to match to at least one target, in which all the attributes must be satisfied. To ensure it is successfully completed, the functions *AnyOf*(disjunction) and *AllOf*(conjunction) are applied at different levels. It is noted that attribute conformance is checked from the types and values, e.g. subject attributes are only compared with subject constraints. The aim here is to find at least one fully matched target to the submitted request.

Based on the semantic request and policy, Semantic Rules (4)–(7) are defined for checking attribute compliance on both sides. To guarantee the type consistency, rules are described with pairwise properties. For example, `Project-01` and `ForResearch` associated by `hasEnvironmentCategory` or `hasEnvironmentAttribute` can be compared. Once matched the intermediate results such as `candidateRuleE(RC_1, Rule_1)` are generated as conditions of Semantic Rule (8). For instance, it is required to compare `RC_1` and `Rule_1` regarding subject, resource, action and environment attributes to determine the applicable rule.

$$\begin{aligned}&\mathrm{Rule}(?a) \wedge \mathrm{RequestContext}(?b) \wedge \mathrm{hasSubjectCategory}(?b,?su) \wedge \mathrm{hasSubjectAttribute}(?a,?sub)\\&\quad \wedge \mathrm{match}(?su,?sub) \longrightarrow \mathrm{candidateRuleS}(?b,?a)\end{aligned} \tag{4}$$

$$\begin{aligned}&\mathrm{Rule}(?a) \wedge \mathrm{RequestContext}(?b) \wedge \mathrm{hasResourceCategory}(?b,?re) \wedge \mathrm{hasResourceAttribute}(?a,?res)\\&\quad \wedge \mathrm{match}(?re,?res) \longrightarrow \mathrm{candidateRuleR}(?b,?a)\end{aligned} \tag{5}$$

$$\begin{aligned}&\mathrm{Rule}(?a) \wedge \mathrm{RequestContext}(?b) \wedge \mathrm{hasActionCategory}(?b,?ac) \wedge \mathrm{hasActionAttribute}(?a,?act)\\&\quad \wedge \mathrm{match}(?ac,?act) \longrightarrow \mathrm{candidateRuleA}(?b,?a)\end{aligned} \tag{6}$$

$$\begin{aligned}&\mathrm{Rule}(?a) \wedge \mathrm{RequestContext}(?b) \wedge \mathrm{hasEnvironmentCategory}(?b,?en) \wedge \mathrm{hasEnvironmentAttribute}(?a,?env)\\&\quad \wedge \mathrm{match}(?en,?env) \longrightarrow \mathrm{candidateRuleE}(?b,?a)\end{aligned} \tag{7}$$

$$\begin{aligned}&\mathrm{Rule}(?a) \wedge \mathrm{RequestContext}(?b) \wedge \mathrm{candidateRuleS}(?b,?a) \wedge \mathrm{candidateRuleR}(?b,?a) \wedge \mathrm{candidateRuleE}(?b,?a)\\&\quad \wedge \mathrm{candidateRuleA}(?b,?a) \longrightarrow \mathrm{applicableRule}(?b,?a)\end{aligned} \tag{8}$$

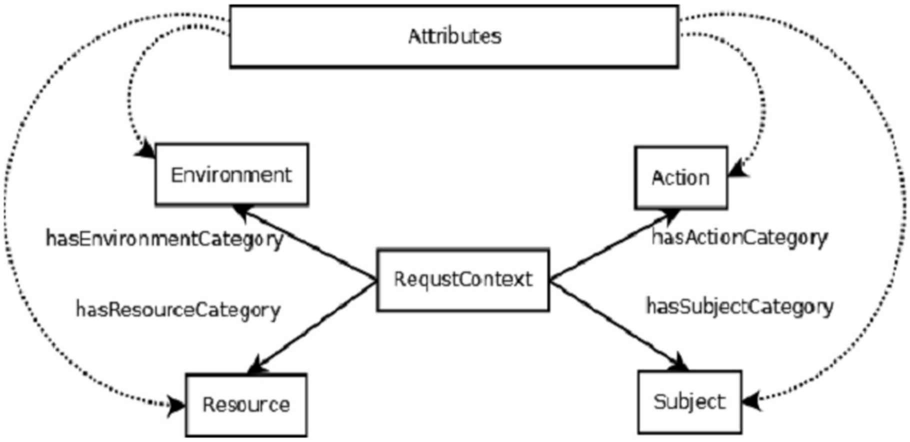


Fig. 3. Initialising semantic requests to Rule_1.

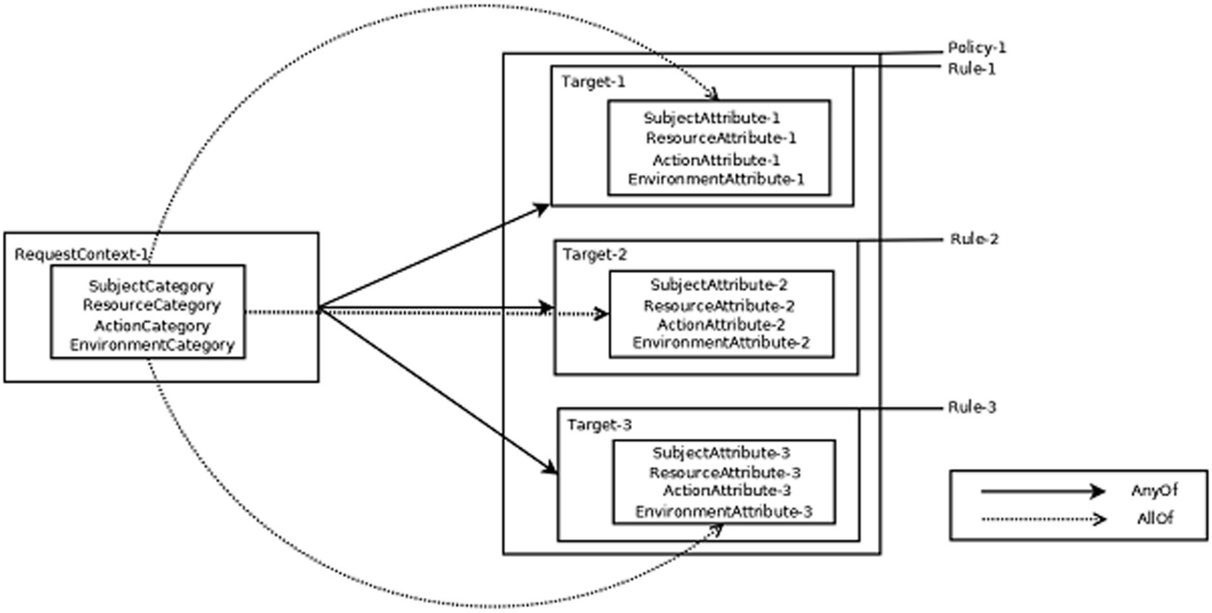


Fig. 4. Matching between XACML requests and policies.

Since the semantic web language depends on the Open World Assumption (OWA) (Walter et al., 2014) `Permit` can be used as the default access control effect. As a result, once achieving such results as `applicableRule(RC_1, Rule_1)`, the access request is permitted while potentially has obligations to conduct. Assume several rules are applicable to `RC_1`. It is necessary to compute the final effect by using combination algorithms (Xu et al., 2015). For example, algorithm "den y-unless-permit" is leveraged by this model since rules define `Permit` as the policy effect. As a result, final effect can be inferred positive as long as applicable rules are located via Semantic Rule (9). Through using this approach in policy evaluation, the access request can only have one final effect. In particular, `fromPolicy` is the inverse function of `hasRule`, and thus `hasRule(Policy_1, Rule_1)` is semantically same as `fromPolicy(Rule_1, Policy_1)`. It is noted that `hasEffect (Rule_1, Permit)` is defined in policy development while `finalEffect(RC_1, Permit)` results from the effect combination.

$$\text{RequestContext}(?x) \wedge \text{applicableRule}(?x,?a) \wedge \text{hasEffect}(?a,?b) \wedge \text{fromPolicy}(?a,?c) \longrightarrow \text{finalEffect}(?x,?b) \wedge \text{finaEffectFrom}(?x,?c) \tag{9}$$

As mentioned, associated obligations are triggered by certain effects. In this regard, the result `finaEffectFrom(RC_1, Policy_1)` support locating obliged operations. Through reasoning via Semantic Rule (10), the obligation is confirmed from the consequence `hasObligation(RC_1, Obligation_1)`.

$$\text{RequestContext}(?x) \wedge \text{finaEffectFrom}(?x,?y) \wedge \text{hasObligation}(?y,?z) \longrightarrow \text{hasObligation}(?x,?z) \tag{10}$$

As shown in Fig. 5, semantic-based compliance checking is depicted as a two-stage process: attributes at the bottom layer are compared with intermediate results; then applicable rules will be determined based on the collection of results. Different from the syntactic comparison, semantic methods facilitate policy evaluation where evolving security demands on ever-growing information can be satisfied.

3.3. Inferential disclosure control

Multiple information sources are used to satisfy different applications while resulting in heterogeneous policies and data. Therefore, a finer-grained reasoning becomes more demanding than simple categorisation. In order to reuse basic policy concepts and domain knowledge, Finin et al. (2008) proposed using ontology-based medical vocabulary where semantic reasoning is allowed. For example, "Clinician" in an international clinical system may have synonyms such as "Doctor", "Arzt" (German) and "Docteur" (French) due to languages used. This can be handled through *subsumption* reasoning based on class hierarchies or identical individuals, depending on which model (class-as-role or individual-as-role) is applied (Finin et al., 2008). However, context information may become more complex than before. For instance in using `Policy_1` and `RC_1`, the environment attribute `ForResearch` is completely unmatched with `Project-O1` unless other implicit relationships can be found. Suppose `Project-O1` is a medical research project on brain tumour treatment. Since it is known as a research-oriented project from extra facts `isProjectOf(Project_O1, ForResearch)` in Listing 3, results `match (Project_O1, ForResearch)` can be generated from Semantic Rule (11) to determine the authorisation.

$$\text{Projects}(?a) \wedge \text{Purposes}(?b) \wedge \text{isProjectOf}(?a,?b) \longrightarrow \text{match}(?a,?b) \tag{11}$$

4. Case study – Semantic record linkage between ADDN and AURIN

4.1. ADDN and data schema

In order to make advance in diabetes treatment and prevention, the Australasian Paediatric Endocrine Group (APEG) and the Juvenile Diabetes Research Foundation (JDRF) jointly established a centralized repository: the Australasian Diabetes Data Network (ADDN – www.addn.org.au) By sourcing de-identified patient records from hospitals and research institutes, it has been established to support clinical trials and research on type-1 diabetes management (T1DM) within Australia and New Zealand. At present the system includes information on over 8000 patients across Australia. ADDN Policies in ADDN embody the regulations on health information, such as collecting patient information with informed consent. Typically, policies are formalized based on a set of roles (Clinicians, Coordinators and Researchers) and geographical clearances (Centre, Multiple Centre and All) about data and users, such as "*clinicians can read the records that are collected from their own centres*" or "*coordinators can read the records that are collected from the collaborating institutes*".

As introduced, these can be managed through semantic approaches. Over 200 data points such as demographic details, treatment information, genetics details and availability of bio-samples are collected, whereas only disease types, ethnicities and zip codes are selected to represent for simplicity. Since patient records are assigned a unique identifier, a patient-centred model is adopted to represent data views where security policies are defined by imposing conditions, i.e. what roles can access that view of data. Examples in Listing 4 are the semantic models corresponding to different sets of patient records.

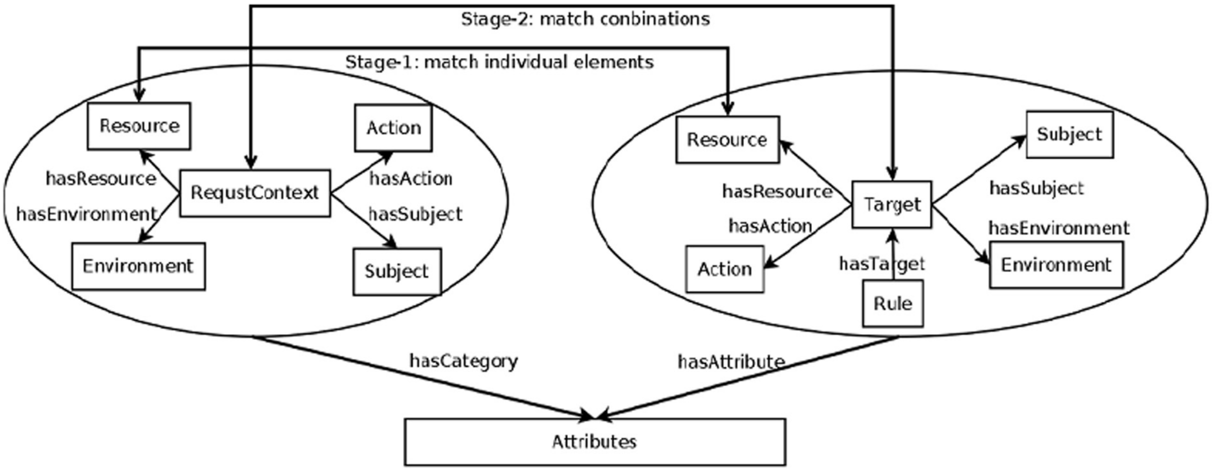


Fig. 5. Policy evaluating by individual and collective match.

Respectively, `Patient_x` is described at the class level to express the entire ADDN T1DM patient database. To enforce special requirements, data models can be enriched with further details. For instance, `Patient_y` is specified with properties such as `German` and `3128` as the concrete ethnicity and postcode. Semantic data models can facilitate policies description on data subsets, where the security mechanisms are embodied. It is noted that both `Patient_x` and `Patient_y` are applied in the following case study. However in the practical application more views can be created.

4.2. Collaboration with AURIN

In order to show how a semantic approach offers dynamic protection of integrated datasets, we consider a public health scenario in which the incidence of T1DM and the distance between residences to bottle shops are correlated. Therefore, it is necessary to visualize the locations of bottle shops and T1MD patients based on the same interface. The Australian Urban Research Infrastructure Network (AURIN – www.aurin.org.au) is established as a comprehensive research platform to support seamless and secure access to a wide array of data. One such data set is official details about premises licensed to sell alcohol (bottle shops) in Victoria, which is available through the Victorian Commission for Gambling and Liquor Regulation (VCGLR - www.vcglr.vic.gov.au). Fig. 6 illustrates the location of bottle shops in the area Box Hill.

Additionally, AURIN allows importing external data resources according to the demand. Through uploading ADDN data (postcodes and patient population) into AURIN, the population distribution can be illustrated on the Australian map. Since AURIN system supports data aggregation at different geographical levels, e.g. Statistical Area Levels (SA4-SA1) (Australian Bureau of Statistics, 2011a), patient numbers reduce proportionally with the spatial scaling, which can lead to privacy issues arising and the potential to identify individuals. As shown in Fig. 7, there are 4 patients labelled with Box Hill, denoted by the postcode 3128. Suppose one patient was born in Germany. Regardless of the de-identification, it is necessary to impose access policies to restrict risky disclosure. Assume policies are defined based on some general security guidelines such as lowering down the risk of identity disclosure. The obligation rules can be included in the policy set such as "*given a geospatial range, the patients identified from a minor group (population size less than 100) can be disclosed only if they are at least 5% of the population*".

There are various data sources available in AURIN, which can be leveraged to find more valuable information/trends. For instance, Fig. 8 illustrates the distribution of German speakers in Victoria, Australia (Australian Bureau of Statistics, 2011b). Through selecting the region Box Hill, we can see the minor group - German speaker includes 84 individuals. Since languages can sometimes reflect more background information, such as where they are from originally and what ethical groups they belong to, we can estimate the proportion of German patients (approximately 1.2%) in this area. According to the obligation, the access request to patient data will incur a privacy protection when enforcing security operations (e.g. generalisation). Since it is difficult to predicate all possible attributes in the linkage application, semantic reasoning is used to detect implicit associations from existing knowledge.

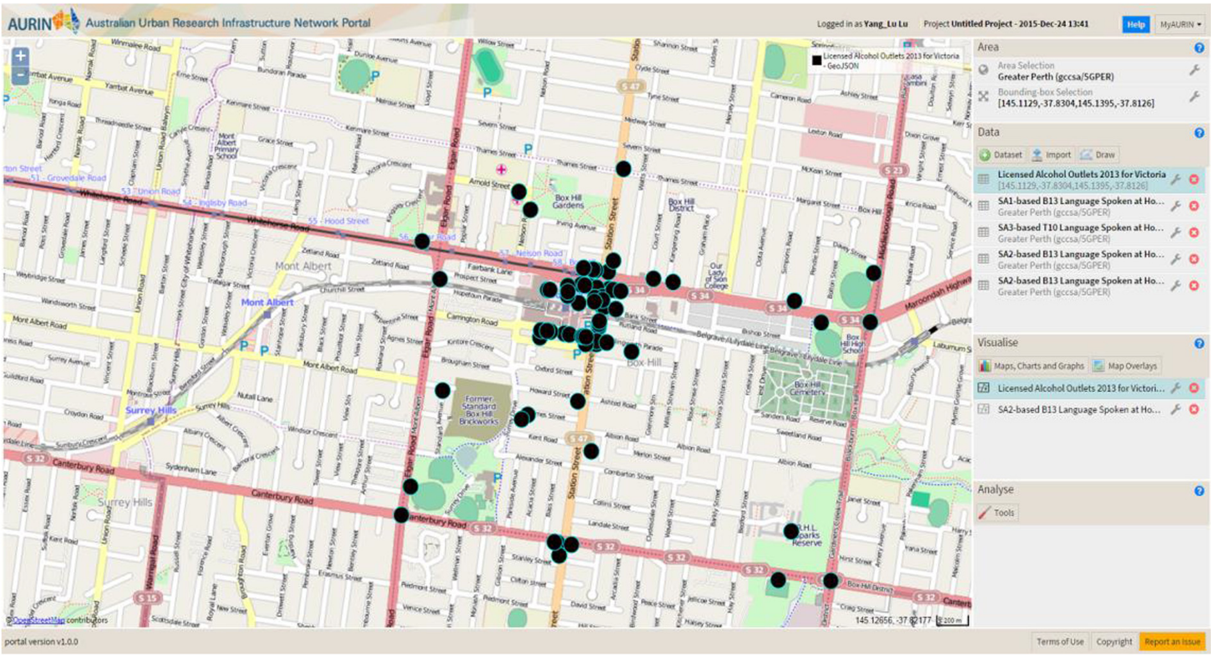
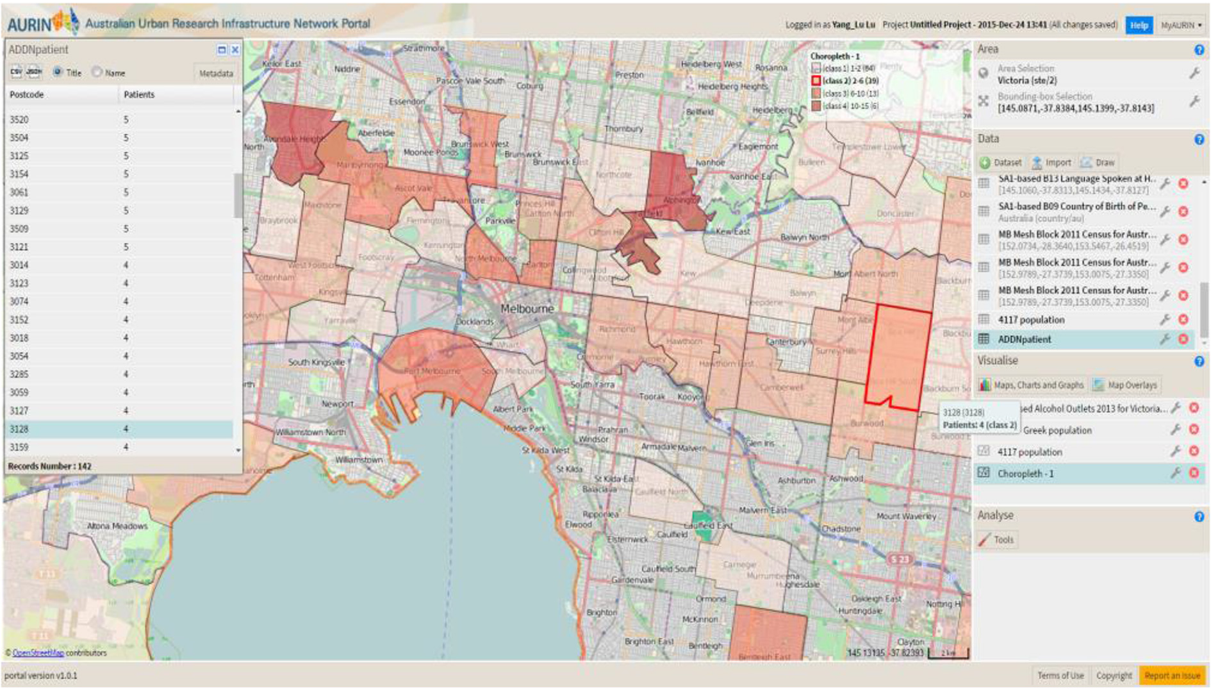

AURIN Australian Urban Research Infrastructure Network Portal
Logged in as Yang_Lu Lu
Project Untitled Project - 2015-Dec-24 13:41
Help
MyAURIN
Licensed Alcohol Outlets 2013 for Victoria - GeoJSON
Area
Area Selection
Greater Perth (gccsa/5GPER)
Bounding-box Selection
[145.1129,-37.8304,145.1395,-37.8126]
Data
Dataset
Import
Draw
Licensed Alcohol Outlets 2013 for Victoria
[145.1129,-37.8304,145.1395,-37.8126]
SA1-based B13 Language Spoken at Ho...
Greater Perth (gccsa/5GPER)
SA3-based T10 Language Spoken at Ho...
Greater Perth (gccsa/5GPER)
SA2-based B13 Language Spoken at Ho...
Greater Perth (gccsa/5GPER)
SA2-based B13 Language Spoken at Ho...
Greater Perth (gccsa/5GPER)
Visualise
Maps, Charts and Graphs
Map Overlays
Licensed Alcohol Outlets 2013 for Victori...
SA2-based B13 Language Spoken at Ho...
Analyse
Tools
Box Hill
Mont Albert
Surrey Hills
© OpenStreetMap contributors
145.12656, -37.82177
200 m
portal version v1.0.0
Terms of Use
Copyright
Report an Issue

Fig. 6. Visualizing bottle shops in the Box Hill, Victoria, Australia.



Logged in as **Yang_Lu Lu** Project **Untitled Project - 2015-Dec-24 13:41** (All changes saved)

ADDNpatient

● Title ○ Name

Postcode	Patients
3520	5
3504	5
3125	5
3154	5
3061	5
3129	5
3509	5
3121	5
3014	4
3123	4
3074	4
3152	4
3018	4
3054	4
3285	4
3059	4
3127	4
3128	4
3159	4

Records Number : 142

Choropleth - 1
- (class 1) 1-2 (84)
- (class 2) 2-6 (39)
- (class 3) 6-10 (13)
- (class 4) 10-15 (6)

3128 (3128)
Patients: 4 (class 2)

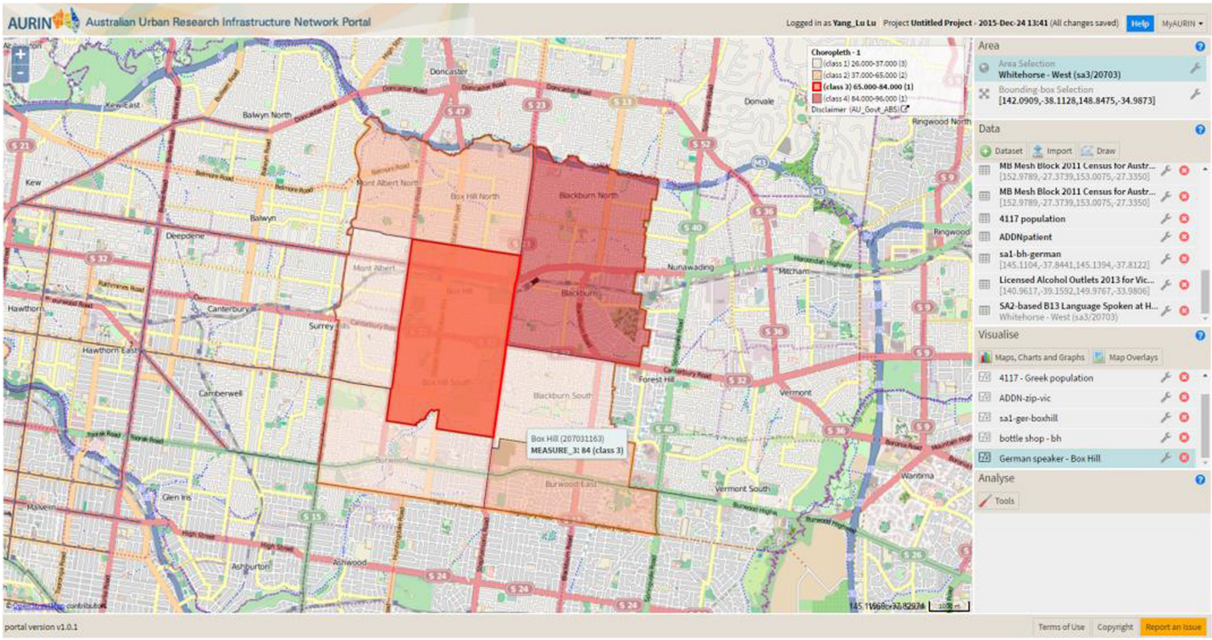
Area Selection
Victoria (ste/2)

Bounding-box Selection
[145.0871,-37.8384,145.1399,-37.8143]

- SA1-based B13 Language Spoken at H... [145.1060,-37.8313,145.1434,-37.8127]
- SA1-based B09 Country of Birth of Pe... Australia (country/au)
- MB Mesh Block 2011 Census for Austr... [152.0734,-28.3640,153.5467,-26.4519]
- MB Mesh Block 2011 Census for Austr... [152.9789,-27.3739,153.0075,-27.3350]
- MB Mesh Block 2011 Census for Austr... [152.9789,-27.3739,153.0075,-27.3350]
- 4117 population
- ADDNpatient

- ...sed Alcohol Outlets 2013 for Victoria...
- ... Greek population
- 4117 population
- Choropleth - 1

Fig. 7. Distribution of T1DM patients in Victoria, Australia.



Logged in as **Yang_Lu Lu** Project **Untitled Project - 2015-Dec-24 13:41** (All changes saved)

Choropleth - 1
- (class 1) 26.000-37.000 (3)
- (class 2) 37.000-65.000 (2)
- **(class 3) 65.000-84.000 (1)**
- (class 4) 84.000-96.000 (1)

Disclaimer (AU_Govt_ABS)

Box Hill (207031163)
MEASURE_3: 84 (class 3)

Area Selection
Whitehorse - West (sa3/20703)

Bounding-box Selection
[142.0909,-38.1128,148.8475,-34.9873]

- MB Mesh Block 2011 Census for Austr... [152.9789,-27.3739,153.0075,-27.3350]
- MB Mesh Block 2011 Census for Austr... [152.9789,-27.3739,153.0075,-27.3350]
- 4117 population
- ADDNpatient
- sa1-bh-german [145.1104,-37.8441,145.1394,-37.8122]
- Licensed Alcohol Outlets 2013 for Vic... [140.9617,-39.1592,149.9767,-33.9806]
- SA2-based B13 Language Spoken at H... Whitehorse - West (sa3/20703)

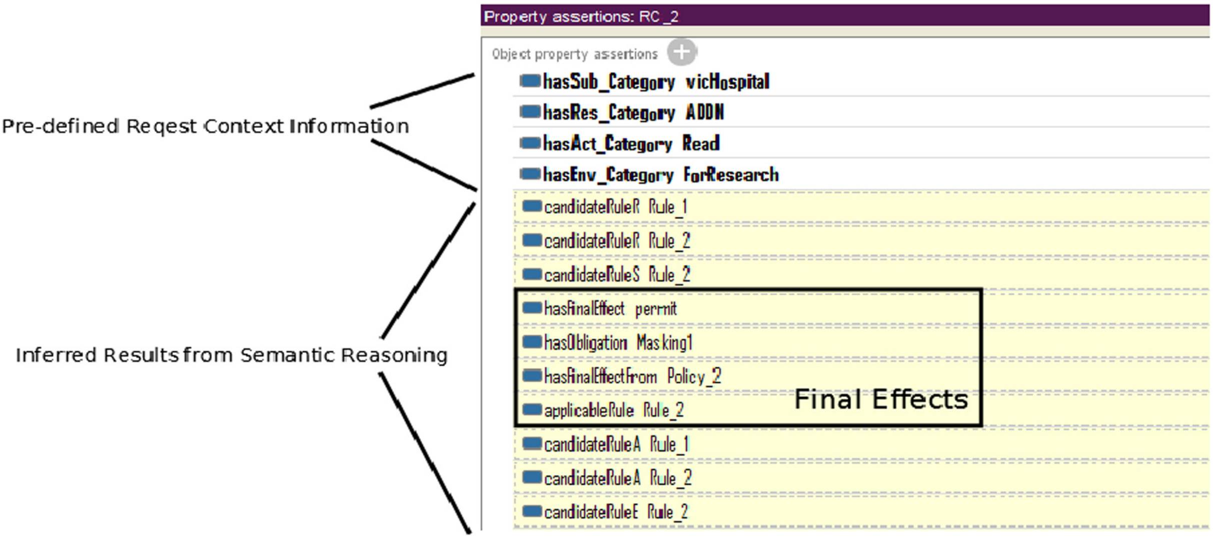
- 4117 - Greek population
- ADDN-zip-vic
- sa1-ger-boxhill
- bottle shop - bh
- German speaker - Box Hill

Fig. 8. Distribution of German speakers in Victoria, Australia.

To show this approach is feasible in avoiding privacy policy violations, we consider a scenario in which Victorian Hospital users are permitted to access the platform resources for research purposes. Particularly, patient ethnicities are specified with the reference to Australian Standard Classification of Cultural and Ethnic Groups, a three-level reference system about the ethnicity (Australian Bureau of Statistics, 2011c). For instance, "2306 German" is the subset of "23 Western Europe", which is further included by the "2 North-west European". With the condition at a higher level, the population magnitude grows and thus privacy threats can be reduced. As shown in Listing 5, the policy in this scenario is designed based on previous discussions. It identifies that Victoria Hospital members are allowed to use ADDN contents for research, while conducting the obligation, `Masking1`.

Suppose a request `RC_2` is sent to ADDN to access patient records. Based on the context information, the access decision can be made by reasoning about the attributes submitted. As shown in Fig. 9, both the intermediate and final effects are produced through compliance checking. Based on the results such as `candidateRuleA(RC_2, Rule_1)`, only `Rule_2` is found applicable to the `RC_2`.

It has been estimated that data with certain properties (e.g. `German` and `3128`) is a risk to ADDN, and thus the privacy policy is specified in the obligations. Therefore, Semantic Rule (12) is defined to further retrieve the concrete operation

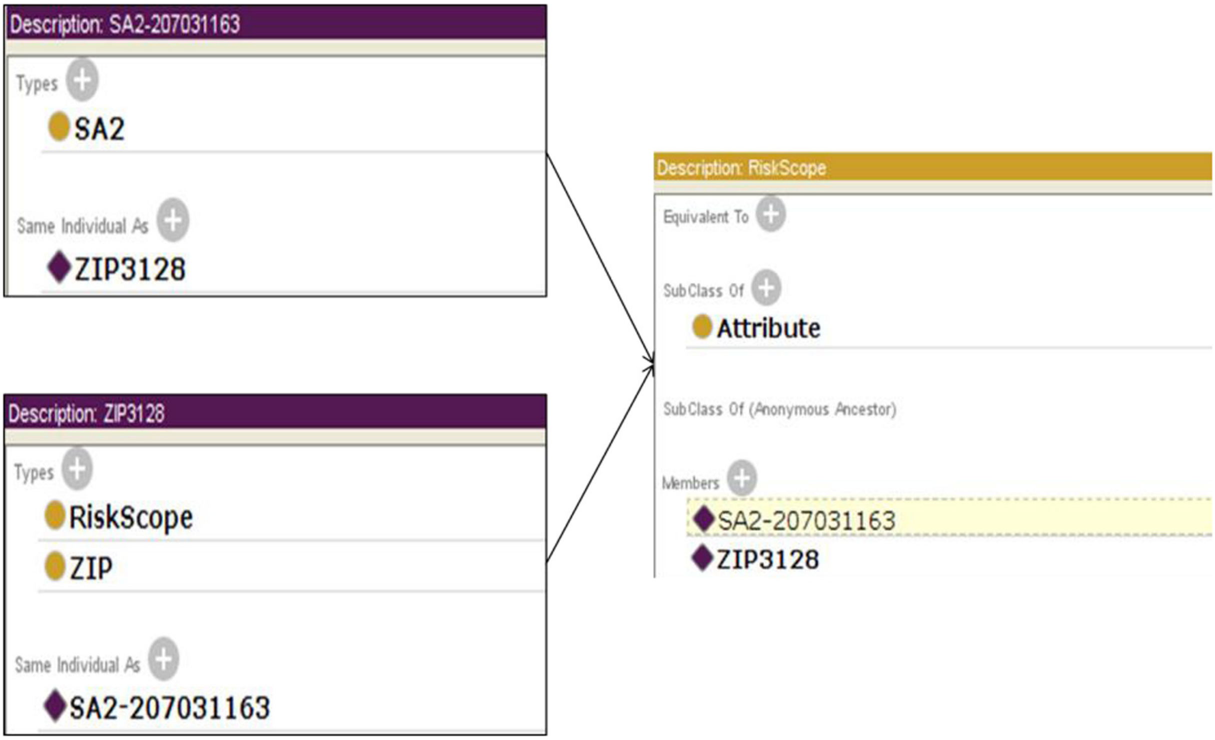

Pre-defined Reqest Context Information
Inferred Results from Semantic Reasoning
Final Effects

Property assertions: RC_2

Object property assertions

- hasSub_Category vicHospital
- hasRes_Category ADDN
- hasAct_Category Read
- hasEnv_Category ForResearch
- candidateRuleR Rule_1
- candidateRuleR Rule_2
- candidateRuleS Rule_2
- hasFinalEffect permit
- hasObligation Masking1
- hasFinalEffectFrom Policy_2
- applicableRule Rule_2
- candidateRuleA Rule_1
- candidateRuleA Rule_2
- candidateRuleE Rule_2

Fig. 9. Results of semantic policy evaluation.



Description: SA2-207031163

Types
- SA2

Same Individual As
- ZIP3128

Description: ZIP3128

Types
- RiskScope
- ZIP

Same Individual As
- SA2-207031163

Description: RiskScope

Equivalent To

SubClass Of
- Attribute

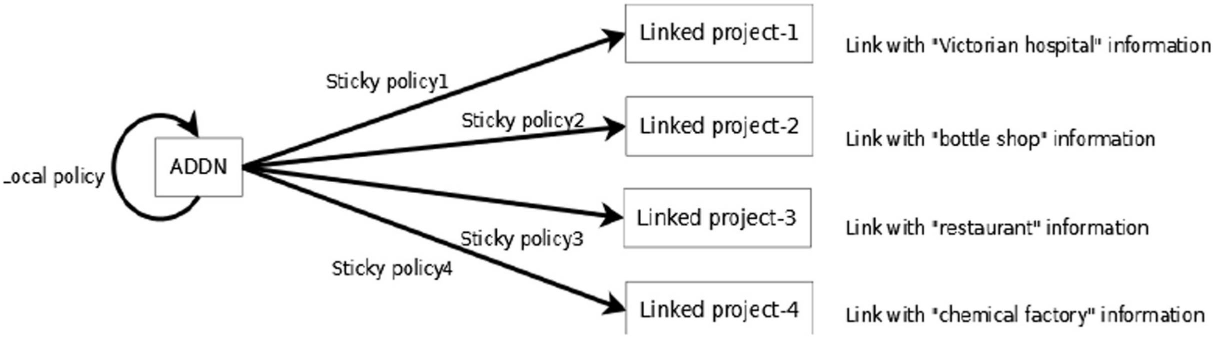
SubClass Of (Anonymous Ancestor)

Members
- SA2-207031163
- ZIP3128

Fig. 10. Management of domain-specific knowledge.

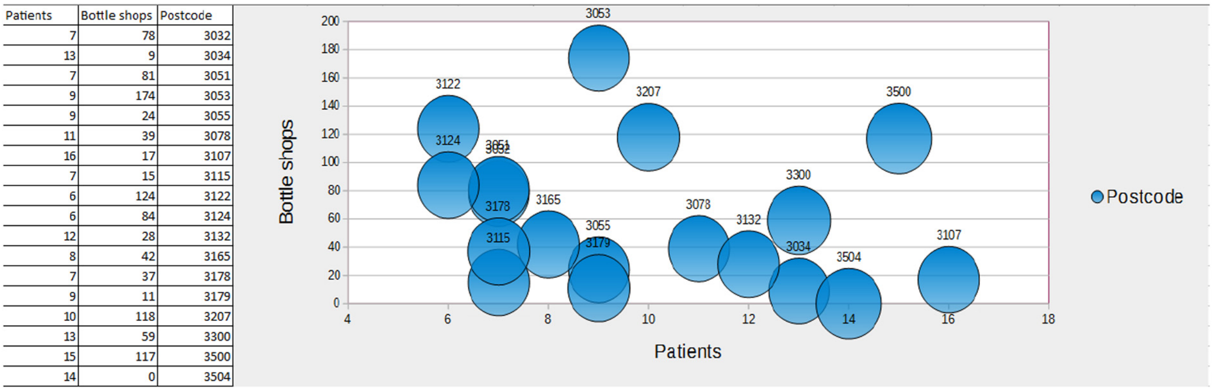
Table 2
Query results of ADDN patients.

Patient-ID	Ethnicity	ZIP
10001	Australian	3128
10002	Australian	3128
10003	Australian	3128
10004	Western Europe (German)	3128



Local policy
ADDN
Sticky policy1
Sticky policy2
Sticky policy3
Sticky policy4
Linked project-1
Linked project-2
Linked project-3
Linked project-4
Link with "Victorian hospital" information
Link with "bottle shop" information
Link with "restaurant" information
Link with "chemical factory" information

Fig. 11. Centralized management of ADDN policy on data-linkage projects.



Patients	Bottle shops	Postcode
7	78	3032
13	9	3034
7	81	3051
9	174	3053
9	24	3055
11	39	3078
16	17	3107
7	15	3115
6	124	3122
6	84	3124
12	28	3132
8	42	3165
7	37	3178
9	11	3179
10	118	3207
13	59	3300
15	117	3500
14	0	3504

Fig. 12. Correlation between bottle shops and patient distribution.

```
<Policy PolicyId= Policy_1>
 <Rule RuleId= Rule_1 Effect= Permit>
  <Target TargetId = Target_1>
   <Attribute>
    <Id = Role Category = Subject AttributeValue = Clinician/>
    <Id = Data Category = Resource AttributeValue = PatientRecord/>
    <Id = Action Category = Action AttributeValue = Read/>
     <Id= Purpose Category = Environment AttributeValue = ForResearch/>
    </Attribute>
   </Target>
   </Condition>
 </Rule>
 <Obligation>
  <Id = State-of-consent FulfillOnEffect = Permit Algorithm = string-equal>
   <AttributeAssignment Id = State Value = Agreed/>
 </Obligation>
</Policy>
```

Listing. 13. XACML policy example (key concepts are printed in bold).

```
<Request RequestId= RC_1>
 <Attribute Category = “Subject”>
  <AttributeId= “role”/>
  <AttributeValue= Clinician/></Attribute>
 <Attribute Category = “Resource”>
  <AttributeId= “data”/>
  <AttributeValue= PatientRecord/></Attribute>
 <Attribute Category = “Action”>
  <AttributeId= “operation”/>
  <AttributeValue = Read/></Attribute >
 <Attribute Category = “Environment”>
  <AttributeId= “project”/>
 <AttributeValue = Project-01/></Attribute>
</Request>
```

Listing. 14. XACML request example.

```
<owl:NamedIndividual rdf:about=”&untitled-ontology-11; Project_01 “>
 <rdf:type rdf:resource=”&untitled-ontology-11;Project”/>
 <projectOf rdf:resource=”&untitled-ontology-11;ForResearch”/>
</owl:NamedIndividual>
```

Listing. 15. Semantic-based knowledge management.

and object. Based on the applicable rule (`Rule_2`) and the relevant data (`ADDN`), the need for operations like `Mask(ADDN)` can be inferred.

$$\mathrm{RequestContext}(?x) \wedge \mathrm{hasObligation}(?x, \mathrm{Masking1}) \wedge \mathrm{applicableRule}(?x,?y) \wedge \mathrm{hasResource}(?y,?z) \longrightarrow \mathrm{Mask}(?z) \quad (12)$$

In dealing with multiple resources, it is essential to estimate implicit meanings and then manage risks effectively. For instance, the concept of **`RiskScope`** is defined to group sensitive geographical information. This is to say, patient records of such features should be processed and checked before release. For instance postcode 3128 is the instance of **`RiskScope`** based on previous analysis. Since German patients are at risk of being identified within this scope, the ethnicity `German` can be related to `3128` by `hasSensitiveEthnicity(3128, German)`. Once selecting certain regions, concrete operations and objects can be inferred based on Semantic Rule (13). In this case, the result `generaliseItem(Patient_y, German)` can be generated based on existing views (e.g. `Patient_x` and `Patient_y`) and domain knowledge. In this scenario, the rule consequence implies that the ethnical value **`German`** should be generalised before disclosing `Patient_y`.

$$\begin{aligned}&\mathrm{Mask}(?x) \wedge \mathrm{hasPatient}(?x,?p) \wedge \mathrm{hasZip}(?p,?l) \wedge \mathrm{RiskScope}(?l)\\ &\quad \wedge \mathrm{hasSensitiveEthinicity}(?l,?e) \longrightarrow \mathrm{generaliseItem}(?p,?e) \qquad (13)\end{aligned}$$

Data resources can be linked through different attributes, such as postcodes and SAs. Instead of making policies on specific spatial levels, semantic inferences can facilitate dynamic privacy protection in distributed environments. For instance, ADDN records can be linked with the bottle shop database by **`ZIP`** and **`SA2`** since they are equivalent geographically. As shown in Fig. 10, through associating `SA2-207031163` and `3128` with equivalence, the function range of `Policy_2` is extended due to inferred results like `RiskScope(SA2-207031163)`. Through this approach, requests specified at the SA2 level could be recognized and evaluated without re-defining any new policy.

Based on the ASCCEG, the ethnical information is specified with a 3-level classification in the ADDN database. The resulting data is partially shown in Table 2. To show the function of semantic technologies, risky records like Patient-10004 are generalised (from “German” to “Western Europe”). Through detecting the privacy risk by semantic reasoning, the framework is able to protect and obfuscate the data to more general and hence less risky levels.

5. Discussion

In this paper, we have shown how the adoption of semantic privacy-preserving framework linkage techniques has the potential to promote patient-centred healthcare and population health research. This approach can be used in many scenarios, e.g. where patients are treated in different locations and by different GPs and hospitals. In such cases, it is often essential for health staff to make clinical decisions by accessing the complete historical information of the patient. To treat chronic conditions, it may well be the case that clinicians request record linkage from linkage centres where discrete EHRs can be linked across hospitals. In addition, applying linkage in public health research activities aims to improve health services based on large-scale analyses. In general, the idea of extending privacy preservation in XACML policy is similar to sticky pol-

```
Patient_x:
<owl:NamedIndividual rdf:about="&untitled-ontology1;Patient_x">
   <rdf:type rdf:resource="&untitled-ontology1;Patients"/>
   <hasEthnicity rdf:resource="&untitled-ontology1;Ethnicities"/>
   <hasType rdf:resource="&untitled-ontology1;Types"/>
   <hasZIP rdf:resource="&untitled-ontology1;ZIPs"/>
</owl:NamedIndividual>

Patient_y:
<owl:NamedIndividual rdf:about="&untitled-ontology1;Patient_y">
   <rdf:type rdf:resource="&untitled-ontology1;Patients"/>
   <hasEthnicity rdf:resource="&untitled-ontology1;German"/>
   <hasType rdf:resource="&untitled-ontology1;Types"/>
   <hasZIP rdf:resource="&untitled-ontology1;3128"/>
</owl:NamedIndividual>
```

Listing. 16. OWL formalization of ADDN Records.

```
   Policy: <owl:NamedIndividual rdf:about="&untitled-ontology-11;Policy_2">
            <rdf:type rdf:resource="&untitled-ontology-11;Policy"/>
            <hasRule rdf:resource="&untitled-ontology-11;Rule_2"/>
            <hasObligation rdf:resource="&untitled-ontology-11;Masking1"/>
           </owl:NamedIndividual>

   Rule:   <owl:NamedIndividual rdf:about="&untitled-ontology-11;Rule_2">
            <rdf:type rdf:resource="&untitled-ontology-11;Rule"/>
            <hasTarget rdf:resource="&untitled-ontology-11;Target_2"/>
            <hasEffect rdf:resource="&untitled-ontology-11;Permit"/>
           </owl:NamedIndividual>

   Target:  <owl:NamedIndividual rdf:about="&untitled-ontology-11;Target_2">
             <rdf:type rdf:resource="&untitled-ontology-11;Target"/>
             <hasAction rdf:resource="&untitled-ontology-11;Action_2"/>
             <hasEnvironment rdf:resource="&untitled-ontology-11;Environment_2"/>
             <hasResource rdf:resource="&untitled-ontology-11;Resource_2"/>
             <hasSubject rdf:resource="&untitled-ontology-11;Subject_2"/>
            </owl:NamedIndividual>
Target-Subject:
            <owl:NamedIndividual rdf:about="&untitled-ontology-11;Subject_2">
             <hasAttribute rdf:resource="&untitled-ontology-11;vicHospital"/>
            </owl:NamedIndividual>
Target-Resource:
            <owl:NamedIndividual rdf:about="&untitled-ontology-11;Resource_2">
             <hasAttribute rdf:resource="&untitled-ontology-11;ADDN"/>
            </owl:NamedIndividual>

            <owl:NamedIndividual rdf:about="&untitled-ontology-11;ADDN">
            <rdf:type rdf:resource="&untitled-ontology-11;Data"/>
            <hasPatient rdf:resource="&untitled-ontology-11;Patient_x"/>
            <hasPatient rdf:resource="&untitled-ontology-11;Patient_y"/>
           </owl:NamedIndividual>
Target-Action:
           <owl:NamedIndividual rdf:about="&untitled-ontology-11;Action_2">
            <hasAttribute rdf:resource="&untitled-ontology-11;Read"/>
           </owl:NamedIndividual>
Target-Environment:
           <owl:NamedIndividual rdf:about="&untitled-ontology-11;Environment_2">
            <hasAttribute rdf:resource="&untitled-ontology-11;ForResearch"/>
           </owl:NamedIndividual>
```

Listing. 17. OWL-based policy for the linked project

icy (Pearson and Mont, 2011; Spyra et al., 2015). Beyond the existing boundary between systems, sticky policies are expected to give extra protection on data sharing. As shown in Fig. 11, through knowing the scope of references in every linked project, this centralized framework helps define sticky policies with semantic technologies, so as to facilitate dynamic security protection in the scaling environment.

In addition to merging EHRs based on ID mappings, linkage is allowed to be constructed based on common attributes. As one example displayed in Fig. 12, is there a correlation between the distribution of places to buy alcohol (bottle shops) and T1DM incidence within specific postal areas? As shown in Fig. 12, through merging the aggregated information (at the same postcode level), we are able to investigate such possible correlations. Through visualising such aggregated data we can directly observe such correlations. Such localised analysis can be augmented by considering other factors, such as comparing the average distances to bottle shops or patient income levels.

The linkage component here is also more broadly suited to other linkage scenarios encompassing socio-economic, multicultural and potentially political domains, e.g. who will vote for a particular party and what factors might impact their decision. Other examples are also feasible, e.g. answering the question ‘‘whether the alcohol consumption relates to violence” or ‘‘whether the frequency of exercise of individuals is impacted by air quality”, policy makers and stakeholders are able to conduct effective investment and localised governance in situational contexts. However, inferential disclosure of background knowledge is the major reason causing privacy leakage while dealing with unit level data – even when combined with aggregate level data. Since an adversary may have access to other data and background knowledge, there is a demand for security solutions that can detect leakage with arbitrary linkages. Nevertheless, the ability to reason about risks beyond the context of a single authorisation decision for a given organisation is essential. This work represents a significant step forward in the mechanisms of achieving this and experiences gained in applying the work in major national projects.

6. Conclusions

In this work, we identify how record linkage can help form the complete profile of patients and thus add value to existing healthcare systems. Through the introduction of data anonymity, access control techniques as well as semantic technologies, we show how privacy preservation can be satisfied through specifying background knowledge and further restricting the access to certain data. To prevent privacy risks with arbitrary linkages, we propose a semantic framework including policy formalization, compliance checking and knowledge discovery. Based on a real-world scenario where Type-1 Diabetes patient records are linked with geospatially coded social science data we show how the semantic extension helps to identify implicit risks relations that could jeopardise data privacy.

The next stage of this work is to explore associated attributes in different types of datasets. For this reason, more scenarios where patient-centric business intelligence is involved should be explored to facilitate distributed authorisation and trust negotiation. In this work we mainly focus on related geospatial concepts however real-world linkage applications are constructed with arbitrary contents from different data sources. Given a semantic framework with reasoning capacities, it is necessary to measure the ‘‘relatedness” among a variety of attributes and benefit from other existing approaches like k-anonymity.

Acknowledgement

The ADDN project is funded by the Juvenile Diabetes Research Foundation. The AURIN project is funded by the Australian Government Department of Innovation. We gratefully acknowledge their support.

References

Abdelhak, M., Grostick, S., Hanken, M.A., 2014. Health Information: Management of a Strategic Resource. Elsevier Health Sciences.

Abello, A., Romero, O., Pedersen, T.B., Berlanga, R., Nebot, V., Aramburu, M.J., Simitsis, A., 2015. Using semantic web technologies for exploratory OLAP: a survey. IEEE Trans. Knowl. Data Eng. 27 (2), 571–588.

Addas, R., Zhang, N., 2014. An enhanced linkable anonymous access protocol of the distributed electronic patient records. 2014 Ninth International Conference on Availability, Reliability and Security (ARES). IEEE, pp. 146–151.

Australian Bureau of Statistics, 2011a. Australian Statistical Geography Standard (ASGS). Retrieved October 15, 2015, from http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2901.0Chapter23102011.

Australian Bureau of Statistics, 2011b. SA2-based B13 Language Spoken at Home by Sex. Retrieved October 20, 2015, from http://stat.abs.gov.au/Index.aspx?DataSetCode=ABS_CENSUS2011_B13_LGA.

Australian Bureau Statistics., 2011c. Australian Standard Classification of Cultural and Ethnic Group (ASCCEG). Retrieved October 15, 2015, from http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/1249.0main+features22011.

Baracaldo, N., Palanisamy, B., Joshi, J., 2014. Geo-Social-RBAC: a location-based socially aware access control framework. International Conference on Network and System Security. Springer International Publishing, pp. 501–509.

Brown, I., Brown, L., Korff, D., 2010. Using NHS patient data for research without consent. Law Innovation Technol. 2 (2), 219–258.

Carroll, J., Herman, I., Patel-Schneider, P.F., 2015. OWL 2 web ontology language RDF-based semantics. W3C Recommendation.

Cocos, C., MacCaull, W., 2010. An ontological implementation of a role-based access control policy for health care information. In: Proceedings of the Workshop of Ontologies in Biomedicine and Life Sciences.

Coulter, A., Collins, A., 2011. Making shared decision-making a reality. No decision about me, without me, London: King’s Fund.

De Montjoye, Y.A., Hidalgo, C.A., Verleysen, M., Blondel, V.D., 2013. Unique in the Crowd: the privacy bounds of human mobility. Sci. Rep. 3.

Dowsett, S.M., Saul, J.L., Butow, P.N., Dunn, S.M., Boyer, M.J., Findlow, R., Dunsmore, J., 2000. Communication styles in the cancer consultation: preferences for a patient-centred approach. Psycho-oncology 9 (2), 147–156.

Finin, T., Joshi, A., Kagal, L., Niu, J., Sandhu, R., Winsborough, W., Thuraisingham, B., 2008. ROWLBAC: representing role based access control in OWL. Proceedings of the 13th ACM symposium on Access control models and technologies. ACM, pp. 73–82.

Hsu, I.C., 2013. Extensible access control markup language integrated with Semantic Web technologies. Inf. Sci. 238, 33–51.

Kalloniatis, C., Mouratidis, H., Vassilis, M., Islam, S., Gritzalis, S., Kavakli, E., 2014. Towards the design of secure and privacy-oriented information systems in the cloud: identifying the major concepts. Comput. Standards Interfaces 36 (4), 759–775.

Kontos, E., Blake, K.D., Chou, W.Y.S., Prestin, A., 2014. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J. Med. Internet Res. 16 (7), e172.

Kwan, J., Sandercock, P.A., 2004. In hospital care pathways for stroke. The Cochrane Library.

Levesque, J.F., Harris, M.F., Russell, G., 2013. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. Int. J. Equity Health 12 (1), 18.

Li, N., Li, T., Venkatasubramanian, S., 2007. t-Closeness: privacy beyond k-anonymity and l-diversity. IEEE 23rd International Conference on Data Engineering, 2007. ICDE 2007. IEEE, pp. 106–115.

Lopez, D., Gunasekaran, M., Murugan, B.S., Kaur, H., Abbas, K.M., 2014. Spatial big data analytics of influenza epidemic in Vellore, India. 2014 IEEE International Conference on Big Data (Big Data). IEEE, pp. 19–24.

Lu, Y., Sinnott, R.O., 2015. Semantic security for e-health: a case study in enhanced access control. 12th Intl Conf on Autonomic and Trusted Computing. IEEE, pp. 407–414.

Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., 2007. L-diversity: privacy beyond k-anonymity. ACM Trans. Knowl. Discovery Data (TKDD) 1 (1), 3.

Mettler, T., Vimarlund, V., 2009. Understanding business intelligence in the context of healthcare. Health Inf. J. 15 (3), 254–264.

Mitra, B., Sural, S., Vaidya, J., Atluri, V., 2016. Mining temporal roles using many-valued concepts. Comput. Secur. 60, 79–94.

O'connor, M., Knublauch, H., Tu, S., Grosof, B., Dean, M., Grosso, W., Musen, M., 2005. Supporting rule system interoperability on the semantic web with SWRL. International Semantic Web Conference. Springer, Berlin Heidelberg, pp. 974–986.

O'Keefe, C.M., Connolly, C.J., 2010. Privacy and the use of health data for research. Med. J. Aust. 193 (9), 537–541.

Orlando, J.P., Musen, M.A., Moreira, D.A., 2015. User extensible system to identify problems in OWL ontologies and SWRL rules. International Symposium on Rules and Rule Markup Languages for the Semantic Web. Springer International Publishing, pp. 112–126.

Panackal, J.J., Pillai, A.S., Krishnachandran, V.N., 2014. Disclosure risk of individuals: a k-anonymity study on health care data related to Indian population. 2014 International Conference on Data Science & Engineering (ICDSE). IEEE, pp. 200–205.

Pearson, S., Mont, M.C., 2011. Sticky policies: an approach for managing privacy across multiple parties. Computer 9, 60–68.

Pulvirenti, M., McMillan, J., Lawn, S., 2014. Empowerment, patient centred care and self-management. Health Expect. 17 (3), 303–310.

Rahmouni, H.B., Solomonides, T., Mont, M.C., Shiu, S., 2010. Privacy compliance and enforcement on European healthgrids: an approach through ontology. Philos. Trans. R. Soc. London A: Math. Phys. Eng. Sci. 368 (1926), 4057–4072.

Rahmouni, H. B., Mont, M. C., Munir, K., Solomonides, T., 2014. A SWRL Bridge to XACML for Clouds Privacy Compliant Policies. In: Proceedings of 4th International Conference on Cloud Computing and Service Science.

Rissanen, E. (Ed.), 2013. eXtensible Access Control Markup Language (XACML) Version 3.0. Edited by Retrieved October 13, 2015, from http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html.

Saltz, J.S., 2015. The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. 2015 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2066–2071.

Sharma, N.K., Joshi, A., 2016. Representing attribute based access control policies in owl. 2016 IEEE Tenth International Conference on Semantic Computing (ICSC). IEEE, pp. 333–336.

Sicuranza, M., Esposito, A., Ciampi, M., 2014. A patient privacy centric access control model for EHR systems. Int. J. Internet Technol. Secured Trans. 5 (2), 163–189.

Spruit, M., Vroon, R., Batenburg, R., 2014. Towards healthcare business intelligence in long-term care: an explorative case study in the Netherlands. Comput. Hum. Behav. 30, 698–707.

Spyra, G., Buchanan, P.W.J., Ekonomou, D.E., 2015. Sticky policy enabled authenticated OOXML for Health Care. BCS Health Informatics Scotland Research Conference 2015, pp. 1–6.

Sweeney, L., 2002. K-anonymity: a model for protecting privacy. Int. J. Uncertainty Fuzziness Knowledge-Based Syst. 10 (05), 557–570.

Synnot, A.J., Hill, S.J., Garner, K.A., Summers, M.P., Filippini, G., Osborne, R.H., Shapland, S.D., Colombo, C., Mosconi, P., 2016. Online health information seeking: how people with multiple sclerosis find, assess and integrate treatment information to manage their health. Health Expectations 19 (3), 727–737.

Uzun, E., Atluri, V., Vaidya, J., Sural, S., Ferrara, A.L., Parlato, G., Madhusudan, P., 2014. Security analysis for temporal role based access control. J. Comput. Secur. 22 (6), 961–996.

Walter, T., Parreiras, F.S., Staab, S., 2014. An ontology-based framework for domain-specific modeling. Softw. Syst. Model., 1–26

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., Kulkarni, U., Mooney, John G., Phillips-Wren, G., Turetken, O., 2014. The current state of business intelligence in academia: the arrival of big data. Commun. Assoc. Inf. Syst. 34 (1), 1.

Xu, D., Zhang, Y., Shen, N., 2015. Formalizing semantic differences between combining algorithms in XACML 3.0 policies. Proceedings of IEEE International Conference on Software Quality, Reliability and Security. IEEE, pp. 163–172.

Yu, B., Yang, L., Wang, Y., Zhang, B., Cao, Y., Ma, L., Luo, X., 2013. Rule-Based Security Capabilities Matching for Web Services. Wireless Pers. Commun. 73 (4), 1349–1367.

Zhang, R., Liu, L., Xue, R., 2014. Role based and time-bound access and management of EHR data. Secur. Commun. Networks 7 (6), 994–1015.