A semantic information model based on the privacy legislation

Kheira Bekara, Maryline Laurent

A Semantic Information Model based on the Privacy Legislation

Kheira. Bekara[#1], Maryline. Laurent[#2]

[#] *Institut Telecom,Telecom SudParis, CNRS Samovar UMR 5157*

9 rue Charles Fourier, 91011 Evry, France

{[1]Kheira.Bekara, [2]Maryline.Laurent}@it-sudparis.eu

***Abstract*— Users' concerns regarding their privacy have a negative impact on their confidence into e-services, and tend to slow down the widespread adoption of online services. Until today, the protection of personal data is mainly left to the legislation by means of guidelines. This paper aims to increase the perceived control by users over their data and to bring down into the technological reality the legislative data protection principles. To do so, it discusses the main concepts involved in the legislative privacy principles, and deduces a privacy semantic information model, i.e. a privacy ontology. This model serves to build users' privacy preferences and SP's privacy policies.**

***Keywords*— Privacy, legislative requirements, Access control, Ontology.**

I. INTRODUCTION

More than a century ago, after the first essay identifying privacy as a fundamental human right [1], never before in history the citizens have been more concerned about the privacy of their personal data, and the privacy threats caused by emerging technologies [2]. These privacy-related issues are intensified by the ubiquity, the invisibility, and the processing power of computation and communication afforded by today's Information and Communication Technologies.

Privacy protection, as a social issue [3], is still left regarding its protection to the legal framework and to Service Providers (SP) self regulation. Privacy policies [4] are defined by SPs and so far they are displayed to users under a literal form with abstract terms that are difficult to understand by most of the users. As such, to simplify personal data privacy protection enforcement, there is a strong need to bring down legislative requirements into the technological reality and to design new technical solutions. Note that these solutions must be adapted to the transaction context, i.e. it must take into account all the following elements:

- The type of the service requiring the users' data,
- The type of the required data element,
- The applicable policy rule to that context.

The Platform for Privacy Preferences (P3P) W3C specification [4] has been the first initiative towards this direction, providing a way for a web site to encode its relevant practices and to communicate them to the visiting users. P3P formalizes SPs' privacy commitment but is limited to the following aspects: Purpose (for which purposes the SP is requesting data?), Recipient (with whom the SP is authorized to share the collected data?), and Retention (how long data will remain stored at the SP?).. Also P3P does not permit to specify the type of the service requiring users' personal data, nor the type of the requested personal data item.

The challenge for enforcing privacy requirements has been widely examined. Research and development efforts resulted in several frameworks proposed by HP [5], IBM [6], and OASIS [7]. These frameworks mainly focus on enterprise environments and provide means for automating the enforcement of the privacy policies. However, the privacy policies specified in the context of these frameworks cannot be efficiently audited to verify their consistency and legislative compliance. By definition, the expression of privacy policies in these frameworks is not based on the whole privacy legislative requirements. In fact, automating the enforcement of privacy policies is done by applying privacy-aware access control mechanisms, i.e. the traditional Role-Based Access Control (RBAC) models are enriched with additional privacy related aspects [8].

Gandon and Sadeh [9], Rao et al [10], and Jutla et al [11] investigate using the semantic web technologies to support privacy in e-commerce. They choose ontology languages, e.g. OWL and ROWL, to represent users' privacy preferences and contextual information. Garcia in [12] proposes a privacy ontology to support translation of privacy policies expressed using a P3P vocabulary into assertions that are used to control access to personal data. Although these solutions manage and address the access control to personal data by means of privacy policies enforcement, they focus on the user's privacy preferences specifications and not legislation based privacy policies. Also, the specifications of the privacy policies don't take into account neither the type of the service, nor the type of the requested personal data item.

In the light of above limitations of current approaches, this article presents a new privacy semantic information model to better enforce the privacy access control framework satisfying the privacy legislation requirements. The main idea behind the semantic information model is the formal and detailed specification of the main concepts of the legislative frameworks.

This paper is organized as follows. Section II introduces legislative principles for personal data protection. Section III

briefly presents the privacy legislative requirements. Section IV describes our proposed privacy semantic information model and the usage of it by the user and the SP to define their preferences and policies. Section V presents our improved XPACML approach [13] for translating that model into XACML, and supporting the storage and exchange of privacy policies/preferences. Section VI gives conclusions.

II. Legislative Personal Data Protection

Since its acknowledgement as a fundamental human right by the Universal Declaration of Human Right of the United Nation in 1974 [14]. The personal data are protected by the relevant legislation in many countries around the world.

The first influential legal framework was the US Privacy Act [15] adopted by the Congress in 1974. Nowadays, the European directive 95/46/EC related to the protection of physical persons and the processing of personal data [16], is the main legislative piece in the European Union in terms of privacy protection. This Directive found its source in the OECD (Organisation for Economic Co-operation and Development) privacy protection guidelines [17] and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [18]. The former, was a significant milestone for unifying the protection of privacy. The latter was at that time a legislative framework for European Union members regarding the personal data protection. Therefore, the collection, processing and dissemination of personal data from that time are regulated by the relevant legislation in all the democratic countries.

Later the European Directive 95/46/EC was completed with the directives 02/58/EC [19], 2006/24/EC [20], which are examples of some sectoral applications of the privacy principles of the 1995 directive (related to the privacy preserving processing of personal data in the electronic communication sector).

Hereafter, there is a summary of the fundamental privacy principles with respect to the lawfulness and fairness of personal data collection and processing. These principles constitute the basis of the functional requirements defined in section III.

1. Fairness and lawfulness of the processed data

The principle of fairness imposes that the data processing cannot be performed with a malicious intent or with the objective to cause harm to the data subject.

The data processing should comply with the applicable privacy law, and also with all applicable laws and legislations.

2. Explicitness and specification of purposes

This principle is commonly known as the purpose principle. Since the beginning of the data collection, the data processing should be marked with the intended purpose. The purpose principle is aimed at guaranteeing to the data subject an effective control over the processing of his data.

The data collected and processed for specified, explicit and legitimate purposes may not further be processed for purposes that are incompatible with these for which they have been collected.

3. Necessity of data collection and processing

Data processing is widely defined in the Directive 95/46/EC and covers any operation or set of operations performed, automatically or not, over personal data, including collection, recording, storage, adaptation, deletion, etc.

Article 7 of the Directive 95/46/EC provides the criteria for the legitimate data processing. It states that all the personal data must be fairly and lawfully processed. This requirement is amplified by a number of rules prescribing criteria as pre-conditions to legitimate processing. Therefore, the processing will be legitimate only if one or more of the following conditions is satisfied:

- The data subject has consented to the processing,
- The processing is necessary to perform the contract requested by the data subject, or to comply with the later request by the subject,
- The processing complies with a legal obligation,
- The processing enables to protect the vital interests of the data subject,
- The processing is necessary for the administration of justice,
- The processing is necessary for the legitimate interests of the SP to which the data are disclosed, except when it is unwarranted because it is harmful to the interests of the data subject.

4. Information, notification and access right of the users

The information requirement is commonly held as the basic requirement for any data processing activity. Article 10 of the directive 95/46/EC specifies a list of information that should be given to the data subject prior to starting the processing activity. Article 11 tackles the case where information is not obtained directly from the data subject but from third parties. The information statement that the system has to give to the data subject must contain at least the following information, except when already known by the data subject:

- The identity of the SP,
- The purposes of the processing for which data are requested,
- The recipients or categories of recipients to which data are likely to be delivered,

- Information whether provision of personal data is mandatory or optional, with their consequences when claimed attributes are not delivered.

The SP has also to provide the data subject, with the aforementioned set of information, when the SP intends to communicate data to third parties.

5. Security and accuracy

The security issue is considered as a prerequisite for a lawful data processing. The SP has to take the necessary steps and actions in order to protect data both in the static and dynamic phases of the data processing.

Indeed, Article 17 of the Directive 95/46/EC provides that the SP must implement appropriate organizational, physical and technical measures to protect personal data against accidental or malicious destruction, accidental loss, alteration, unauthorized disclosure or access, and any other unlawful forms of processing, in particular when the processing requires transmitting data over a network. Such measures shall ensure the appropriate security level matching the risks represented by the processing and the nature of the data to be protected.

Article 16 of the Directive 95/46/EC deals with third parties providing attributes.

Any person acting under the authority of the SP, including any entities which have access to personal data, must process them according to the legal instructions only.

6. Supervision and sanction

An independent Privacy Authority has to be designated and should be responsible for supervising privacy provisions. If violation of the provisions of privacy legislation occurs, criminal (or other) penalties should be envisaged. In that respect, the SP should provide to the Privacy Authority the means for controlling every action of personal data collection and processing.

The Directive 95/46/EC states that each Member State shall apply its own privacy legislation, resulting from the instantiation of the above listed privacy principles of the Directive 95/46/EC. Data protection legislation worldwide, where available, naturally defines some exceptions, exemptions and restrictions concerning the scope of the aforementioned principles. In general, for the purposes of national security and defence, for public security, for prevention, investigation, detection, and prosecution of crimes and for other reasons, the collection and processing of personal data might be enforced by the authorities. Lawful interception is currently a common practice for all the legislative frameworks for the protection of privacy.

III. LEGAL REQUIREMENTS

The principles listed in section II form the basis for defining the functional requirements for a system intended to manage the privacy protection of data subjects. Based on these functional requirements, we extracted the main concepts needed to build a common information model regarding the legislation privacy principles.

Hereafter, we give a summary of the privacy requirements with respect to the fundamental privacy principles defined in section II.

1. The data processing must be fair, and lawful

Regarding principle (1), as defined in the previous section, the system should be able to examine whether the data processing complies with applicable laws and legislations.

2. The data processing purposes must be specific, explicit and legitimate

The SP must detail the reasons for which data are processed, as mentioned in principle (2) of the previous section. This is solved through policies.

At the user side, the system should provide means for identifying the data processing purposes, and make explicit to the data subject what are the pursued processing purposes.

3. Data must be adequate, relevant, and not excessive with regard to the purposes for which they are collected and/or further processed

This law provision has to make adequacy between data and the processing purpose.

The system should be able to guarantee that the required personal data are necessary for the claimed purpose or processing. There should be periodic audits in order to verify adequacy, relevance, and no excessive usage of users' personal data.

4. Identifiable data

Data might be kept in an identifiable form by the system, only for the period of time necessary to perform the processing purpose. Once the processing purpose is achieved, data must be either deleted or made anonymous.

5. Notification

The notification is a formal communication from the SP to the data subject in which the SP provides specific and detailed information about the features of the running data processing. The principle (4) of section II requires the system to communicate this information to the user.

6. Information to be given to the data subject

According to principle (4), the system should be able to provide means to inform the data subject that his data are processed according to the applicable data protection legislation, prior to starting the data processing.

7. Consent of the data subject

The consent of the data subject is necessary to make the data processing legitimate. However, exceptions can apply. According to principle (4), the system should provide means that enable SP to get the data subject's consent, and his privacy preferences for his requested personal data.

8. Data subject's rights

Still according to principle (4) of the previous section, the system should enable the data subject to use his rights according to the applicable data protection legislation related to information and intervention. These rights include rights to obtain information on the data processing, rights to be active in the processing by asking data rectification, erasure, blocking, and the right to deny the data processing.

9. Security and Confidentiality of processing

The system should be secure in order to guarantee the confidentiality, integrity, and availability of the processed data. Also, the system should provide that any kind of interception or communication, and the related traffic data may be performed only with the data subject's consent or when permitted by applicable legislation for public interest purposes.

10. Access limitation

The system should provide an authorization procedure that entails different levels of access to data.

11. Special categories of data

The Directive 95/46/EC provides for a set of limitations related to the processing of sensitive and judicial data. These data are related to the very intimate and personal sphere of individuals, hence they need a higher degree of protection and confidentiality.

Sensitive data are those revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and data concerning health or sex life.

The SP should provide that sensitive data may be processed only with the explicit consent of the data subject. Moreover, the SP's system should be able to guarantee that the processing of special categories of data is performed with compliance to the specific requirements formalized by the applicable data protection legislation.

IV. LEGISLATION MODELLING

This section presents the legislation privacy ontology deduced from the legal requirements listed in section III, and describes which usage of it can be made by the SP and users.

A. Designing the Privacy Ontology

We define a semantic information model with elementary components on which a legislative access decision to users' personal data is based. The context of a transaction, i.e. the type of the service requiring the data item, and the data item type, is essential.

The main idea behind our approach is to model these main concepts using a semantic information model that associates personal data and services with explicit legislative rules. To that respect, we use W3C Web Ontology Language (OWL) [21] to implement the sub-graphs related to data types, service types and related legislation-based policies. The resulted privacy information model is shared between the user and the SP as a common information model that contains, as much detailed as possible, the vocabulary related to data types and service types. The objective is for the user and the SP to know how to handle the requested data type. As such, for identifying the rules that should apply during a transaction, the type of the SP service requiring the data item needs first be identified so it is then possible to delimit which data categories are permitted to be delivered under which privacy conditions.

Several OWL classes are defined in Fig. 1 for implementing this semantic information model. The DataType, and ServiceType classes are intended to structure data types and service types as categories in a hierarchical way with well defined inheritance rules that enable our defined framework [13] to associate privacy related decisions to semantically specified notions.

Regarding the personal data subgraph, all the types are defined as instances of the DataType OWL class. Relationships between personal data are defined using OWL properties.

Our privacy model defines the following properties:

1. inheritsFromData

This property expresses the inheritance relationship between general data types and specific ones. It is implemented by means of inheritsFromData object OWL property.

2. hasMoreDetailed/hasLessDetailed

This property supports different levels of revelation. As such, in case there is a privacy policy conflict between the SP and the user about a data item, it is possible to substitute the data item by another one with a higher level of abstraction.

3. containsType/isContainedToType

This property expresses the complexity of a data item (e.g.: FullName contains the FirstName, LastName and MiddleName). This relationship is implemented by means of containsType/isContainedToType OWL property, which in essence, defines a tree hierarchy.

A similar pattern is adopted for the service's subgraph. The various types of services are defined as subclasses of the ServiceType class. In accordance with DataType subgraph, properties (1) and (3) are implemented for the classes of service.

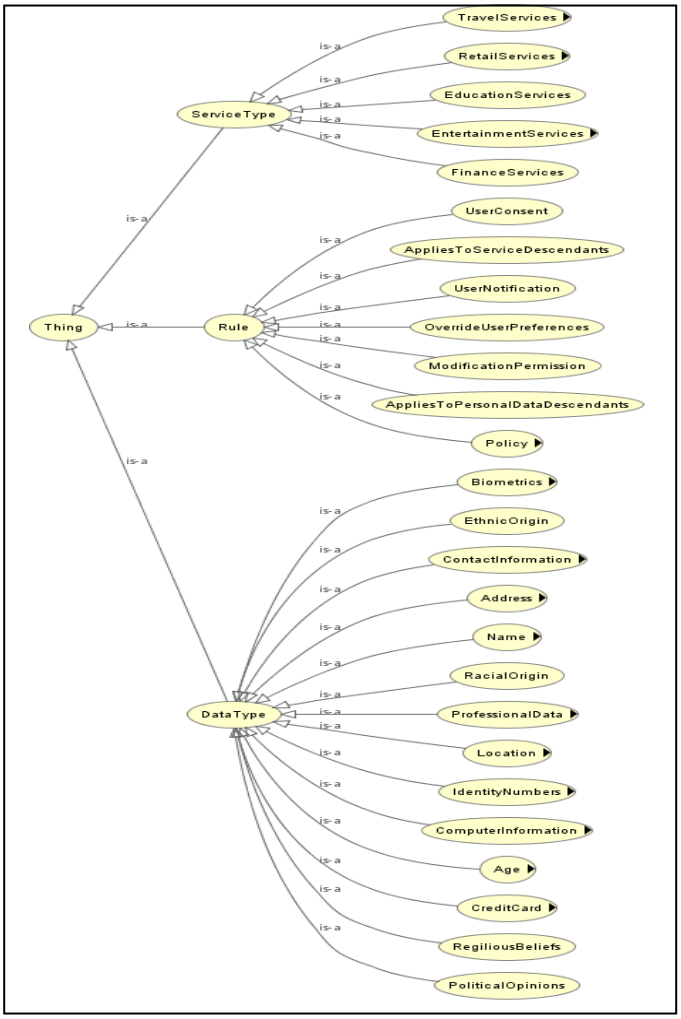

Fig1. Privacy Ontology

B. Usage of the Privacy ontology by SPs and users

Based on the above privacy ontology, the SP is directly enabled to build its own full privacy policy by defining instances of the Rule class to express the data retention, recipient and purposes, and assigning these rules to the requested personal data attributes.

Also the user can directly benefit from the privacy ontology to define his preferences. That is, the user has to create instances of the Rule class, DataType class and ServiceType class so he can express the permitted and denied access control rules over certain data type attributes and for a given service type.

Note that the ontology serves also for the user to interpret unambiguously the privacy policy received from the SP during a transaction. Note that this policy is received under the XPACML format described in section V.

V. XPACML LANGUAGE

Our XML-based language, namely eXtensible Privacy Access Control Markup Language (XPACML) was originally described in [13]. In this section, it is extended to better fit our previously defined ontology. Note that the objectives of XPACML are twofold. It serves for the SP to send its privacy policy to the user, and for both the SP and the user to locally store their own privacy policy and preferences.

The XPACML syntax is XML-based [22] and contains the appropriate elements for the specification of the personal data attributes and the corresponding rules/preferences. With the use of OWL annotation properties, every rule contains the following elements:

- `DATA_TYPE`: Expresses the type of the data in question; its values come from the domain defined by the DataType within the ontology. It can also take the value `ALL`, covering all the types of data.
- `SERVICE_TYPE`: Expresses the type of the service for which some data are about to be disclosed or processed; its values come from the domain defined by the ServiceType of the ontology. It can also take the value `ALL`, covering all the services.
- `Effect`: Determines whether the data of `DATA_TYPE` for the service of `SERVICE_TYPE` should be disclosed to the provider or not.
- `PURPOSE`: Specifies the purposes for which for the data of type `DATA_TYPE` is requested by the SP.
- `RECIPIENTS`: Expresses the entities intended for collecting the data of type `DATA_TYPE`.
- `RETENTION_PERIOD`: Specifies the retention period for the data of type `DATA_TYPE`.
- `Abstraction_LEVEL`: Determines the level of precision for the data of `DATA_TYPE` for the service of `SERVICE_TYPE`.

While the above sub-elements define the core of the rule, additional properties specify the complementary actions that might be executed:

- `MODIFICATION_PERMISSION`: Determines whether the service provider has modification privileges of `DATA_TYPE` data during the provision of a `SERVICE_TYPE` service.
- `NOTIFICATION`: Determines whether the user should be notified for some action on his data of `DATA_TYPE` for the service of `SERVICE_TYPE`.
- `CONSENT`: Determines whether the user should be asked for his consent for some action on his data of `DATA_TYPE` for the service of `SERVICE_TYPE`.
- `DATA_TYPE_DESCENDANTS`: Denotes whether the defined rule is applicable by inheritance to the descendants of the specified `DATA_TYPE` in the

class hierarchy of the DataType of the ontology. It can take the value `YES` or `NO`.

- `SERVICE_TYPE_DESCENDANTS`: Denotes whether the defined rule is applicable by inheritance to the descendants of the specified `SERVICE_TYPE` in the class hierarchy of the ServiceType of the ontology. It can take the value `YES` or `NO`.

VI. CONCLUSIONS

In this paper, a semantic information model for supporting personal data protection is presented under the form of a privacy ontology. This work is based on the European legislation framework, and has the objective to bring as many legislative requirements (regarding data privacy handling) as possible into a privacy ontology. Doing so makes us believe that the setting done by users and SPs about the privacy handling is compliant to what is required by legislative authorities.

REFERENCES

[1] Warren, S. D., and Brandeis, L. D., Dec 1890. *The Right to Privacy*. Harvard Law Review, Vol. IV, No. 5, pp. 193–220, Dec. 1890.

[2] The European Opinion Research Group, Dec. 2003. *European Union citizens' views about privacy*. Special Eurobarometer 196,

[3] Laufer R. et Wolfe M. (1977), Privacy as a concept and a social issue: a multidimensional developmental theory, Journal of Social Issues, 33, 22-42

[4] Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., and Reagle, J. The platform for privacy preferences 1.0 (P3P1.0) specification, Apr. 2002. W3C Recommendation, http://www.w3.org/TR/2002/REC-P3P-20020416/.

[5] Casassa Mont, M., and Thyne, R. 2006. *A Systemic Approach to Automate Privacy Policy Enforcement in Enterprises.* 6th Workshop on Privacy Enhancing Technologies, Lecture Notes in Computer Science, Vol. 4258, Springer-Verlag.

[6] Ashley, P., Hada, S., Karjoth, G., Powers, C. and Schunter, M., 2003. *TheEnterprise Privacy Authorization Language (EPAL), EPAL 1.2 Specification*. IBM Research Report,

[7] Ding, W. and Ma Organization for the Advancement of Structured Information Standards (OASIS), 2004. *eXtensible Access Control Markup Language TC*.

[8] Ni, Q., Bertino, E., Lobo, J., Brodie, C., July 2010. *Privacy-Aware Role-Based Access Control.* ACM Transactions on Information and System Security Volume 13 Issue 3.

[9] Gandon, F. and Sadeh, N, 2004. *Semantic Web Technologies to Reconcile Privacy and Context Awareness"*. In Web Semantics Journal. Vol. 1, No. 3, pp. 241-260.

[10] Rao J., Dimitrov D., Hofmann P. and Sadeh N, 2006. *A Mixed Initiative Framework for Semantic Web Service Discovery and Composition.* In Proceedings of the IEEE International Conference on Web Services (ICWS 2006), 7 pages.

[11] Jutla D.N., Bodorik P, Zhang Y. 2006. *PeCAN: An Architecture for Privacy-aware Electronic Commerce User Contexts.* In Elsevier's Information Systems Journal, Vol. 31, Issue 4-5, pp. 295-320.

[12] Garcia, D.Z.G.; Toledo, M.A. 2008. *Web Service Privacy Framework Based on a Policy Approach Enhanced with Ontologies*. 11th IEEE International Conference on Computational Science and Engineering Workshops, pp. 209 – 214.

[13] Bekara, K., Ben Mustapha, Y., and Laurent, M. *XPACML eXtensible Privacy Access Control Markup Language*., Second International Conference on Communications and Networking (ComNet'2010), Tozeur, Tunisia, Nov. 2010.

[14] United Nations, "Universal Declaration of Human Rights", http://www.un.org/Overview/rights.html.

[15] U.S. Public Law No. 93-579, Dec. 31, 1974, 5 U.S.C. 552a.

[16] European Parliament and Council, "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 of the European Communities*, No. L 281, pp. 31–50, Nov. 1995.

[17] Organization for Economic Co-operation and Development, "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data", Sep. 1980.

[18] http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?NT=108&CL=ENG

[19] European Parliament and Council, "Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)", *Official Journal of the European Communities*, No. L 201, pp. 37–47, Jul. 2002.

[20] European Parliament and Council, "Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC", *Official Journal of the European Communities*, No. L 105, pp. 54–63, Apr. 2006.

[21] The World Wide Web Consortium (W3C), "Web Ontology Language (OWL)", online: http://www.w3.org/2004/OWL/.

[22] http://www.webreference.com/xml/reference/standards.html