Ontological Semantics for Data Privacy Compliance: The NEURONA Project

Núria Casellas(1), Juan-Emilio Nieto(1), Albert Meroño(1), Antoni Roig(1), Sergi Torralba(1), Mario Reyes(2), Pompeu Casanovas(1)

(1) Institute of Law and Technology (IDT-UAB), Universitat Autònoma de Barcelona
Faculty of Law, Building B, Campus UAB, Bellaterra (08193), Spain
nuria.casellas; juanemilio.nieto;suport.projectes.idt;antoni.roig;sergi.torralba;pompeu.casanovas@uab.es
(2) S21sec
C/ Alcalde Barnils 64-6, Bg. Testa, D, 1ª floor (08174), Sant Cugat del Vallès, Spain
mreyes@s21sec.com

Abstract

In this paper, we describe the knowledge acquisition process devoted to the analysis of Data Protection requirements in the Spanish legal system towards the development of a legal ontology for the representation of data protection knowledge in the framework of the NEURONA project. The use of legal ontologies could provide legal professionals and citizens with better access to legal information.

Introduction

The increasing need for legal information and content management caused by the growing amount of unstructured (or poorly structured) legal data managed by legal publishing companies, law firms and public administrations, or the increasing amount of legal information directly available on the World Wide Web, have created an urgent need to construct conceptual structures for knowledge representation to share and manage intelligently all this information, whilst making human-machine communication and understanding possible.

The use of semantically-enabled technologies for legal knowledge management could provide legal professionals and citizens with better access to legal information (acts and regulations, judgments, information from other legal bodies, etc.), and improve the conditions in which citizens may participate in public affairs.

Legal ontologies may be the key to implement these new technological advances and to facilitate legal knowledge search, reasoning, and intercommunication, as formal legal ontologies make explicit the underlying assumptions and the formal definitions of the components of legal knowledge. Some of the top legal ontologies developed so far include the Functional Ontology for Law (FOLaw) [Valente, 1995], the Frame-Based Ontology [van Kralingen, 1995], the LRI-Core ontology [Breuker, 2004], DOLCE+CLO (Core Legal Ontology) [Gangemi et al., 2003], or the Ontology of Fundamental Concepts [Rubino et al., 2006] the basis for the LKIF-Core Ontology [Breuker et al., 2007]. Nevertheless, most legal ontologies are domain specific ontologies, which represent particular legal domains towards search, indexing and reasoning in a specific domain of national or European law (IPRONTO [Delgado et al., 2003], Copyright Ontology [García, 2006], CCO or Customer Complaints Ontology [Jarrar, 2005], Consumer Protection Ontology [Tiscornia et al., 2008], Ontology of Professional Judicial Knowledge or OPJK [Casellas et al, 2007, Casellas, 2008], etc.)

In this paper, we describe the knowledge acquisition process devoted to the analysis of Data Protection requirements in the Spanish legal system towards the development of a legal ontology for the representation of data protection knowledge in the framework of the NEURONA project. First, we will briefly describe the aims of the project and the data protection requirements for the NEURONA system. Second, the legal knowledge acquisition process will be outlined together with a description of the design and development process of the NEURONA Data Protection Ontology. Finally, some issues for discussion and further work will be presented.

The NEURONA Project

The NEURONA project has been financed by the Spanish Ministry of Industry, Tourism and Commerce, and is developed by the security company S21sec[1] and the legal researchers of the Institute of Law and Technology[2] (IDT-UAB) of the Universitat Autònoma de Barcelona. The project's general goal is the development of techniques and systems to incorporate intelligence in the three main areas of corporative security: legal, organizational and technological. Such integration may represent the next step in corporative security and IT asset management.

Therefore, the project focuses on the development of a data protection compliance application that offers reports regarding the classification correctness of files containing personal data. The ontological knowledge-base reasons about the correctness of the information regarding personal data files provided (or their lack of) according to the information required by the *Agencia Española de Protección de Datos* [Spanish Data Protection Agency], and the correctness of the measures of protection applied to these data files. This is a first step towards determining whether some aspects of the current state of a company's personal data files might not comply with the established set of regulations.

Data Privacy and System Requirements

This semantic knowledge encoded in the application ought to represent, not only the most relevant legal data protection concepts in the Spanish legal system (and their relationships), but also their correspondent corporate or organizational concepts together with their technological counterparts. As legal knowledge is not only contained in acts and regulations, but also case law interpretations, guidelines from independent authorities for advice and regulatory compliance, and international or professional standardization bodies, several domains of knowledge intertwine.

In particular, data protection regulations include both top traditional normative concepts (validity, obligation, prohibition, responsibility, etc.), and context-dependent normative concepts (personal data, notification, security measure, etc.). And these regulations come from several bodies with distinct normative functions: European Council, European Parliament, National Parliament (and Regional competent bodies), Data Protection Agencies, courts of justice, etc. Moreover, data protection and privacy issues are not only regulated directly, but also included in other more general domain normative statutes: Business Law, Medical Law, etc. Furthermore, conceptual knowledge regarding data protection is further enriched or modified by the intertwining of normative concepts and professional technology auditing concepts (data, file, etc.), standardized by organizations such as ISACA (e.g. COBIT model of corporate governance), and others.

Towards the development of the ontological knowledge-base, a team of legal experts has selected and analyzed relevant documents in order to extract the knowledge to be encoded in the NEURONA ontology a team of ontology engineers and computer scientists.

Legal Knowledge Analysis and Acquisition

IDT-UAB and S21sec have split the work into several tasks and assigned each to a working team. A first team is set up by legal experts, and their main task is to select, study, analyze and organize relevant regulations for the project's scope, related to corporative security issues. Nevertheless, the main interest is on the study and analysis of Spanish personal data protection regulations.[3]

Data Protection Law

The Organic Act 15/1999 of 13 December on the protection of personal data (PPDA) constitutes the main piece of legislation regarding this matter in Spain. The PPAA adapted the Spanish legal system to the provisions of Directive 95/46/EC, of the European Parliament and of the Council, of 24 October 1995, on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The implementing Regulation of the PPDA was later approved by the Royal Decree 1720/2007, of 21 December (PPDR). This legislation applies to any kind of personal data recorded on a physical support which makes them capable of being processed, and to any type of subsequent use of such data by the public and private sectors. Notwithstanding, certain types of files are excluded from the scope of the PPD legislation[4]. In addition, the legal regime applicable to the processing of certain types of personal is laid down by specific provisions[5].

As for the implementing institutions, it's worth noting that the *Agencia Española de Protección de Datos*

[1] http://www.s21sec.com

[2] http://idt.uab.es

[3] Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD), you may find an English version of the act at: https://www.agpd.es/portalweb/canaldocumentacion/legislacion/estatal/common/pdfs_ingles/Ley_Orgnica_15-99_ingles.pdf (July 12th, 2009).

[4] Those created or maintained by a natural person in the exercise of activities which are exclusively personal or domestic; those subject to the legislation on the protection of classified materials; those established for the investigation of terrorism and serious forms of organized crime.

[5] (a) Files regulated by the legislation on the electoral system; b) Those used solely for statistical purposes and protected by central or regional government legislation on public statistical activities; c) Those intended for the storage of the data contained in the personal assessment reports covered by the legislation on the personnel regulations of the armed forces; d) Those contained in the Civil Register and the Central Criminal Register; e) those deriving from images and sound recorded by videocameras for the security forces in accordance with the relevant legislation

[Spanish Data Protection Agency (DPA)] is the public law authority overseeing compliance with the legal provisions on the protection of personal data.

The provisions of both PPDA and PPDR together with related case-law and DPA's interpretation, set the conceptual framework when designing and formalizing the ontology. The paragraphs below briefly describe the key elements deriving from the Spanish legal regime regarding data protection that are relevant for the NEURONA project.

a) Personal data: PPDA defines personal data as any information concerning identified or identifiable natural persons. The PPDR specifies that personal data may consist on any alphanumeric, graphic, photographic, acoustic or any other type of information pertaining to identified or identifiable natural persons. The PPD legislation establishes a sort of categories of personal data with special protection: i.e. data regarding ideology, religion, trade union membership, racial origin, health or sex life. These categories have a direct impact in the legal regime applying to their collection, consent, and processing.

b) Consent. In accordance with article 6 of the PPDA, processing of personal data shall require the unambiguous consent of the data subject, except in certain cases regarding the exercise of the functions proper to public administrations, among others. The type of consent (e.g. explicit, written, etc.) depends on the category of data.

c) Purpose. In general terms, personal data may be collected for processing, and undergo such processing, only if they are adequate, relevant and not excessive in relation to the scope and the specified, explicit and legitimate purposes for which they were obtained. In any case, personal data subjected to processing may not be used for purposes incompatible with those for which they were collected. Further processing of the data for historical, statistical or scientific purposes shall not be considered incompatible. Finally, it must be noted that files created for the sole purpose of storing personal data which reveal the ideology, trade union membership, religion, beliefs, racial or ethnic origin or sex life remain prohibited.

d) Data security. The PPDA establishes requires the controller or, where applicable, the processor to adopt the technical and organizational measures necessary to ensure the security of the personal data and prevent their alteration, loss, unauthorized processing or access. These must have regard to the state of the art, the nature of the data stored and the risks to which they are exposed by virtue of human action or the physical or natural environment. In this sense the PPDR, establishes three levels of security (basic, medium and high), depending on the specific categories of personal data, and sets out different security requirements applying to the processing of such data. In general terms, these requirements have to do with the functions and responsibilities of the controllers, access control, record of incidents, identification and authentication of the users, etc.

Ontology Design and Formalization

The development of the NEURONA ontology is based on the knowledge acquired and organized by legal experts, and focuses on constructing several formal specifications of the required data protection concepts for the reasoning system.

The construction of the ontology was, thus, focused on the acquisition of conceptual domain knowledge extracted from the legal and related documents and the interaction with the legal experts. Few ontology building methodologies give precise guidelines or recommendations regarding the knowledge acquisition stage, in particular, regarding legal or social knowledge acquisition. We based our knowledge acquisition step in the selection of the relevant knowledge sources and the use of adequate knowledge acquisition techniques from experts. This development followed, as established by most ontology methodologies nowadays, 1) a preparatory phase (specification of ontology requirements), 2) a development phase (knowledge acquisition –experts, documents, reuse–, Conceptualization –classes, relations, properties, instances–, expert validation and formalization), and 3) an evaluation phase (internal consistency, requirements, competency questions, and expert evaluation).

The design of this modular ontological system is based on a central Data Protection Knowledge Ontology, which contains the core concepts of the system, and a Data Protection Reasoning Ontology, which structures the required classification reasoning towards assessing Data Protection compliance. Both ontologies are being modeled with the Protégé ontology editor and using the OWL-DL ontology language. Some of the core concepts of the ontology, represented in Figure 1 below, are: `Data`, `Consent`, `Purpose`, `Security_Measures`, `Person`, `Treatment_Process`, and `Security_Degree`.

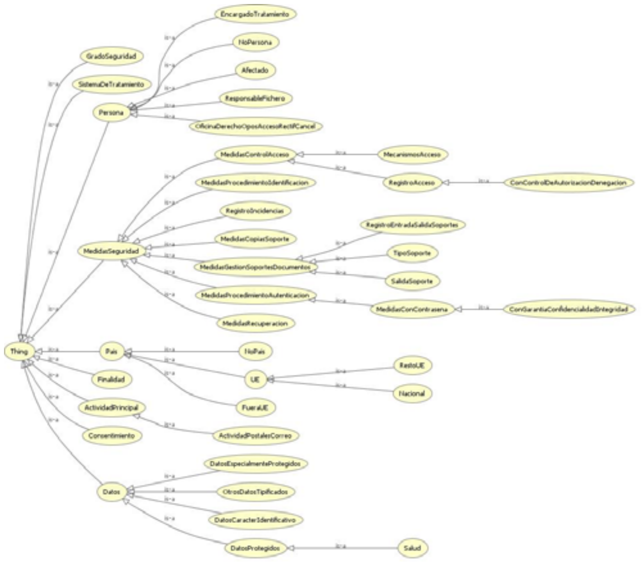

Figure 1: Screenshot of the Data Protection Knowledge Ontology

Issues for Discussion and Further Work

The NEURONA project develops a data protection application which classifies files containing personal data into different categories regarding their compliance with, within others, the required measures of protection. The knowledge is encoded in two OWL-DL ontologies (the Data Protection Knowledge Ontology and the Data Protection Reasoning Ontology). Nevertheless, as mentioned, this is just a first step towards determining in a semi-automated way whether some aspects of the current state of a company's personal data files might not comply with the established set of regulations. Further knowledge will need to be formalized (e.g. legal domain knowledge, corporate knowledge, etc.).

At the moment, the ontological knowledge contained in both modules is under revision by the team of legal experts and its content will be evaluated using *usability* measures towards ontology refinement [Casellas, 2009].

Acnowledgements

NEURONA. Development of an intelligent system to control data protection compliance Funder: Ministerio de Industria, Turismo y Comercio, Spain (TSI-200100-2008-134), 2008-2010 (AVANZA).

References

Breuker, J., Hoekstra, R. 2004. "Core concepts of law: Taking common-sense seriously". In *FOIS'04*, IOS-Press, pages 210–221.

Breuker, J. Hoekstra, R., Boer, A., van den Berg, K., Rubino, R., Sartor, G., Palmirani, M., Wyner, A., Bench-Capon, T. *D.1.4 OWL ontology of basic legal concepts (LKIF-core)*. Deliverable d.1.4 of the ESTRELLA European Project.

Casellas, N. 2008. *Modelling Legal Knowledge through Ontologies. OPJK: the Ontology of Professional Judicial Knowledge*, PhD thesis, Universitat Autònoma de Barcelona, Spain (available at: http://idt.uab.es/~ncasellas/nuria_casellas_thesis.pdf).

Casellas, N. 2009. "Ontology Evaluation Through Usability Measures. An Experiment with the SUS Scale in the Legal Domain". In Ceravolo, P. *4rth International Workshop on Ontology Content at OnTheMove Conferences & Workshops (OTM'09)*, 4-5 of November, 2009, Algarve, Portugal [to be published as Springer Verlag LNCS].

Casellas, N., Casanovas, P. Vallbé, J.J., Poblet, M., Blázquez, M., Contreras, J., López-Cobo, J.M., Benjamins, V.R. 2007. "Semantic enhancement for legal information retrieval: Iuriservice performance". In *ICAIL'07*, ACM, pages 49–57.

Delgado, J., Gallego, I., Llorente, S., García, R. 2003. "Ipronto: An ontology for digital rights management". In Bourcier, D. (ed.), *JURIX'03*, Amsterdam: IOS Press, pages: 111-120

Gangemi, A., Sagri, M.T., Tiscornia, D. 2005. "A constructive framework for legal ontologies". In Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (eds.) 2005. *Law and the Semantic Web*, Lecture Notres in Computer Science, 3369, pages 97–124.

García, R. 2006. *A SemanticWeb Approach to Digital Rights Management*. PhD thesis, Universitat Pompeu Fabra, Barcelona, November 2006.

Jarrar, M. 2005. *Towards Methodological Principles for Ontology Engineering*. Phd thesis, Vrije Universiteit Brussel, May 2005.

Rubino, R., Rotolo, A., Sartor. G. 2006. "An OWL ontology of fundamental legal concepts". In van Engers, T.M. (ed.) *JURIX'06*, volume 152 of Frontiers of Artificial Intelligence and Applications: IOS Press.

Tiscornia, D., Francesconi, E. 2008. "Building Semantic Resources for Legislative Drafting: The DALOS Project". In Casanovas, P., Sartor, G., Casellas, N., Rubino, R. Computable Models of the Law, 4884, Lecture Notes in Artificial Intelligence, Berlin: Springer Verlag, pages: 56-70.

Valente, A. 1995. *Legal Knowledge Engineering; A Modelling Approach*. IOS Press: Amsterdam.

van Kralingen, R.W. 1995. *Frame-Based Conceptual Models of Statute Law*. Kluwer Law Intl.