The Neurona Ontology: A Data Protection Compliance Ontology

Sergi Torralba[1], Núria Casellas[1], Juan-Emilio Nieto[1], Albert Meroño[1] Antoni Roig[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
{sergi.torralba; nuria.casellas; juanemilio.nieto; suport.projectes.idt ;antoni.roig; 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 present the process leading to the design of a legal ontology modelling data protection knowledge in the framework of the Neurona project. The ontology tries to combine simplicity and concreteness to solve the issue of how to properly classify files and notify the user whether the files are compliant or not with the Spanish data protection law. And, due to the constraints brought by the need to embed ontologies in a software application, we face the ontology design process by trying to focus on the reutilization and the introduction of new knowledge within the ontology.

Keywords: Corporate security, business intelligence, data protection, legal ontology, privacy

1 Introduction

The volume of documents and data that public and private organizations manage nowadays increases in high rates nearly day by day. This exponential growth entails the need to keep ongoing control over sensitive inputs and guard against possible misuses. To protect the rights of citizens, EU governments compel both companies and public administrations to comply with legislation on data protection.[1] Failure to comply with these regulatory provisions may cause companies to incur in significant

[1] As regards the legal framework of the Neurona project, there are two basic pieces of legislation: the EU Directive 95/46/EC on the protection of individuals with regard to the processing of personal data, and the Spanish Act 15/1999 of 13 December on the protection of personal data. The implementing Regulation of the Act was later approved by the Royal Decree 1720/2007, of 21 December.

monetary losses, either through sanctions[2] or customers' lose of trust. In 2008, the Spanish Agency of Data Protection imposed a total of 22.6 M€ in sanctions [1]. Public administrations and companies, therefore, are increasingly aware of the risks of malpractice when storing and managing files containing personal data. Nevertheless, firms and administrations may also be liable for holding huge amounts of poorly structured personal data whose implicit merging or joining also infringes the law. Even if the government provides some tools to facilitate compliant management of data protection files (i.e. the NOTA form)[3] they only help to check basic requirements and fall short of providing intelligent interpretation. In sum, there is a growing need for public and private entities to find efficient strategies to ensure the compliance of their procedures as regards privacy and data protection regulations.

The use of semantically-enabled technologies to manage sensitive personal data could provide organizations and citizens with better guarantees of proper access, storage, management and sharing of files, and then improve citizens' rights. The main goal of the Project Neurona is to develop techniques and systems to incorporate intelligence in three core areas of corporate security: legal, organizational, and technological. Such integration represent the next step in the Neurona project, and then it will focus on usiness Intelligence solutions, monitoring and reporting organizational state, corporate security and IT and asset management. Ontologies may be the key to implement these new technological advances and to facilitate knowledge search, reasoning and intercommunication.

This paper describes the design of a legal ontology for the representation of data protection in the broader framework of the Neurona project. The ultimate goal of the ontology is to automatize the classification of files containing personal data and provide early warning if their management is not compliant with the regulatory provisions.

2 Background

Corporate security is a vast research area that covers a number of IT related domains: threats and vulnerabilities, privacy policies, unauthorized access and use of information, etc. Recently, Skovira has referred to an "ecology of security" whose main components are "the organization, the networks, the operating systems, the software applications, the information, and the people involved at all these levels (ecologies)" [2]. In our view, corporate compliance with data protection is part of this ecological landscape.

There is a significant amount of research in the domain of privacy and data protection within the framework of corporate security. We can distinguish here two different approaches: privacy in the IT field and legal ontologies.

First, the research carried out in the IT field [3] [4] models the privacy aspects involved in the design anddevelopment of IT technology, linking the different legal

[2] Spanish legislation establishes sanctions ranging from 600 to 600,000 €, depending on the severity of the infringement.

[3] NOTA is the software that the Spanish Agency provides to notify which files contain sensitive data and, therefore, have to comply with the legislation concerning the Data Protection Act.

models to the real usage of the technology. From another standpoint, Mitre et al. [5] model the Spanish legal framework on data protection with the aim to preserve the privacy of users when the location information is involved: this is also a way to translate what the law says into the IT world. In a more functional point of view, Dos Santos proposes the creation of an automatic auditor [6], but it cannot be implemented globally since each country has different legislations. Another interesting approach applies ontologies to business government and to the integration of security management [7].

Secondly, we have the outcomes of research on legal semantic modelization. Some of the top legal ontologies developed so far include the Functional Ontology for Law (FOLaw) [8], the Frame-Based Ontology [9], the LRI-Core ontology [10], DOLCE+CLO (Core Legal Ontology) [11], or the Ontology of Fundamental Concepts [12] the basis for the LKIF-Core Ontology [13]. Nevertheless, most of the 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 [14], Copyright Ontology [15], CCO or Customer Complaints Ontology [16], Consumer Protection Ontology [17], Ontology of Professional Judicial Knowledge or OPJK [18], etc.). To date, the closest ontology related to our work is the one modelling Italian legislation on privacy [19].

Our approach benefits from these previous works and addresses a specific problem: how to embed our ontology in a software application. After evaluating the different approaches used for current ontologies, we decided not to use any of them and, instead, to create a domain ontology to solve our issue at hand. In other words, what we needed was a functionally oriented ontology whose objective was not to have a clear representation of the law or part of it, but to model the solution as easily as possible. Hence, the ontology is the tool, not a goal in itself.

In a nutshell, the core goal in Neurona is making an application transparent to the end user by hosting the legal knowledge on data protection inside the system. The legal knowledge is already embedded in the application and users are not required to be legal experts on privacy and data protection.

3 The Neurona Project

The main objectives of Neurona are to develop techniques and systems to incorporate intelligence in the three main areas of corporative security: legal, organisational and technological. Therefore, the project focuses on the development of a data protection compliance application that provides reports on the classification of correctness of files containing personal data. The ontological knowledge-base reasons about the correctness of the information on personal data files provided (or their lack of) as required by Spanish Agency of Data Protection, 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.

3.1 Proposed solution

Drawing in the experience obtained with previous research projects such as Iuriservice [20] or Ontomedia [21], we decided that the bottom up approach—starting from smaller parts and sub solutions to end up with global solution—was the most adequate strategy. As a first step in this project, we decided to tackle the issue by following different phases: the first one was to develop an ontology based on the most characterizable part of the Spanish Data Protection Act, the classification of the security measures. The proposed strategy consists of designing an ontology from two basic sub-ontologies: (i) the Data Protection Conceptual Ontology (DPCO) containing all the relevant concepts of the problem domain in a taxonomy-like structure, and (ii) the Data Protection Reasoning Ontology (DPRO) that includes both rules and constraints of the problem domain. This second ontology is therefore responsible for all the reasoning tasks and depends on the concepts modelled in DPCO. Thus, while DPCO is designed in a way that can be used either on its own or in conjunction with other ontologies, DPRO has least chances to be reused since it depends on the first. This neat separation allows keeping the information in a transparent way to the user, while at the same time facilitates an easy method to correct and update all the information related to concepts. DPRO is specifically tailored to the needs of the project, it is also the most functional and the one that needs a greater amount of work and validation from the experts.

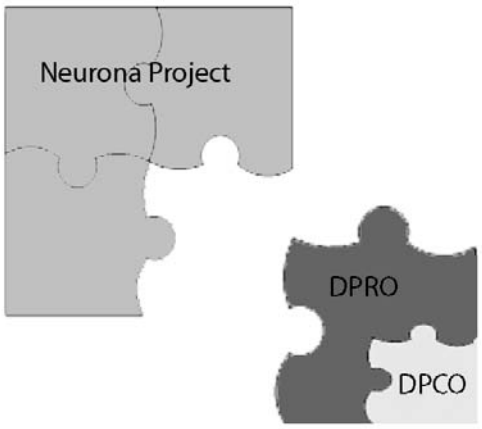

Fig. 1 Neurona ontology and sub-ontologies

4 Data Privacy and System Requirements

The 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 organisational 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: the European Council, the European Parliament, the 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.), standardised by organisations 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, as well as a team of ontology engineers and computer scientists has also taken part.

5 Knowledge Acquisition

Few ontology building methodologies give precise guidelines or recommendations regarding the knowledge acquisition stage, in particular, regarding legal or social knowledge acquisition. We then 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).

IDT-UAB and S21Sec, the two project partners, split the work into several tasks and assigned each to a working team. The first team was composed of legal experts, and their main task was to select, study, analyze and organize relevant regulations for the project's scope, related to corporative security issues. The development of the Neurona ontology is based on the knowledge acquired and organised by legal experts, and focuses on constructing several formal specifications of the required data protection concepts for the reasoning system.

Once this work was done, that team provided the computer engineers and the knowledge engineers with a pack of tables and diagrams to facilitate the comprehension of the field. In this case, therefore, the knowledge acquisition was a bottom up process: from the concrete problem (in this case data privacy compliance in a corporate environment), we try to extract the concepts, rules and properties and everything else needed to create an ontology. Then all the information had to be divided in two different blocks: on one side, all concepts in a taxonomy-like structure and, on the other side, both the relations and classes that are more specific to our precise problem and the required solution. The construction of the ontology was then focused on the acquisition of conceptual domain knowledge extracted from the legal and non-legal related documents and the interaction with the legal experts. The interaction between legal domain experts, knowledge engineers and computer engineers was kept throughout the process. This had bidirectional benefits: on one side, legal experts were aware of the needs of the engineers and helped not to do unnecessary work; on the other side, engineers were constantly advised by the legal experts, making it more difficult to neglect important legal aspects in the modelization. The combination between legal and knowledge engineers helped and reduced the correction of errors detected after the validation process.

6 Ontology Design and Formalization

While the objective of Neurona is to address a specific problem, we expanded our scope in the design of the Neurona ontology by trying to create a basis for having a reusable and scalable ontology which could be enriched or reused in conjunction to other ontologies. The modular strategy followed by separating the concepts from the reasoning part provides us with a concept basis which can be easily reused or enriched.

The Neurona ontology follows a principle of simplicity for two main reasons: first for easier use, and, secondly, if it is going to be embedded in a software product, the less complex the ontology, the faster the reasoner obtains an output (this is also why we discarded the use of top ontologies). The end user of the software solution is a company officer who may have little or no legal knowledge whatsoever. It is also better to not press the user with an excess of questions to fulfil the ontology requirements. For these reasons, the input is going to be something limited, and the possible use of a top ontology would have increased the complexity of the software solution.

The working ontology in Neurona then integrates DPCO, which contains the core concepts of the system, and DPRO, which structures the required classification reasoning towards assessing data protection compliance. As said earlier, the knowledge ontology has no specific constraints in relation with the reasoning ontology, while the reasoning ontology is fully dependent on the first. This means that any changes made in DPRO, the most suitable to be corrected or modified, do not affect DPCO. Both ontologies are being modelled with the Protégé ontology editor and using the OWL-DL ontology language. Some of the core concepts of the ontology are shown in Figure 2 below.

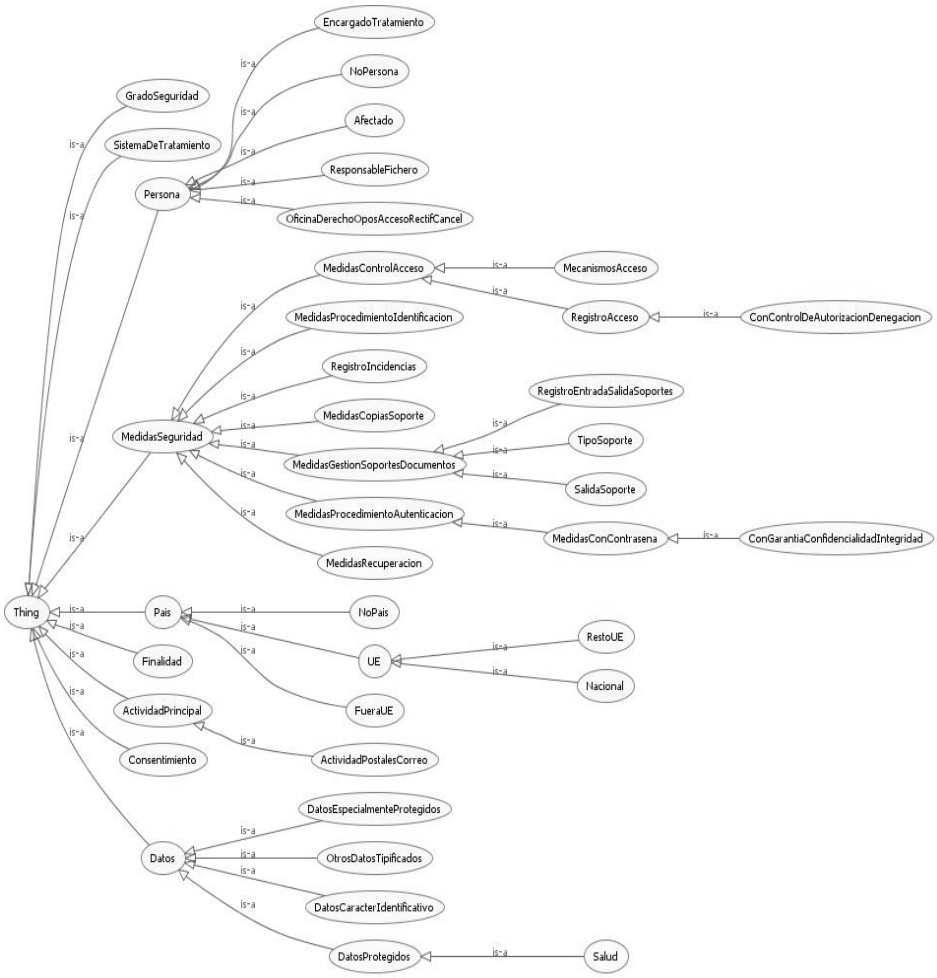

Fig. 2 Core concepts in DPCO

The concept ontology contains all the relevant terms for the classification of files, Those concepts were extracted at a primary stage from the NOTA form and the necessities of recognizing the sensitive information that requires a concrete processing. Not all the classes included in this ontology will be necessarily used by the reasoning ontology but, conversely, the references that create the majority of the classes of the reasoning ontologies are related to the concept ontology. What the reasoning ontology contains are classes that maybe have no meaning outside the project, but they are needed for a good classification. An example of the relations that appear in the reasoning ontology is shown in Figure 3.

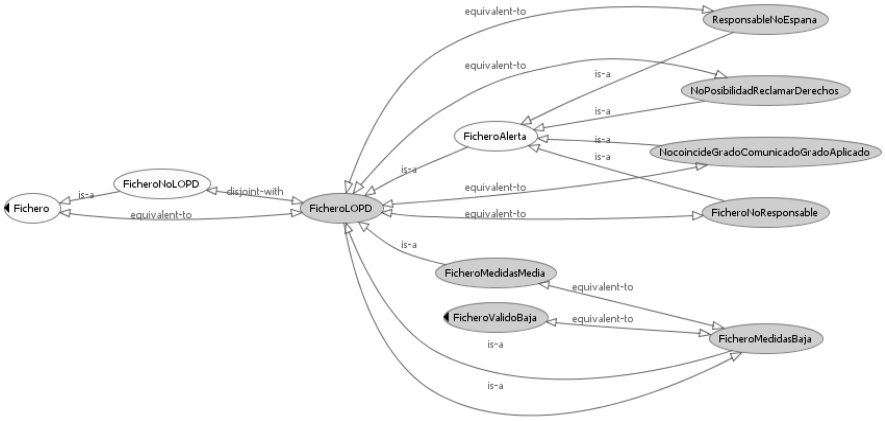

Fig. 3. Relations of one class of the reasoning ontology

6 Conclusion 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 or the lack of a responsible person defined and many other classes helped to define with the experts. But even if we are happy with the current results of the ontology and the software solution, there is a full task of development and validation pending.

In the Neurona project we have been able to create a concept ontology that can be used independently of the reasoning ontology and is a starting point to expand this project in other fields such as enterprise risk management, assets management, corporate knowledge, and many other enterprise related domains. Nevertheless, as mentioned, this project does not aim at solving the full set of issues related to personal data protection: this is just a first step that has to be expanded and refined. The creation of these supplementary ontologies with data from other fields, together with its related reasoning ontologies can provide a huge spectrum of possible plug-ins and improvements for the software application.

The aggregation of rules to the ontologies is also a possible and interesting way to improve results and make the entire project more focused. The use of rules may also provide a powerful way to simplify the reasoning ontology and improve the performance of the entire project.

Furthermore, this project goes towards determining a semi-automated way even if some aspects of the current state of a company's personal data files might not comply with the established set of regulations. But in addition to the fact that the software solution is able to tell the user more than which specific files are compliant or not with the provisions on data protection, it will be useful to know the state of the different files and see during different time periods the number of files that were problematic. Possible further work could include the implementation of a control

panel to give a vast overview of the state of the company to the executive responsible for the use of the application. It would even be more complete if we were able to enhance the concepts ontology with more knowledge fields (such as the ones expressed before).

7 Acknowledgements

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

1 Agencia española de Protección de Datos, Memoria 2008, https://www.agpd.es/portalweb/canaldocumentacion/memorias/memoria_2008/common/memoria_new_2008.pdf

2 R. J. Skovira 2007. “Framing the corporate security problem: The ecology of security”. Informing science [1547-9684] vol. 4 pages 45-52

3 Abou-Tair D, Berlik S 2006.An Ontology-Based Approach for Managing and Maintaining Privacy in Information Systems; Lecture Notes in Computer Science, Berlin: Springer Verlag,Vol. 4275 pages: 983-994

4 Lioukadis G, Lioudakisa G., Koutsoloukasa E., Tselikasa N., Kapellakia S., Prezerakosa G., Kaklamania D., Venierisa I.; 2007 A middleware architecture for privacy protection; Computer Networks: The International Journal of Computer and Telecommunications Networking, Volume 51 , Issue 16 Pages 4679-4696

5 Mitre H., González-Tablas A., Ramos B., Ribagorda A., 2006. A Legal Ontology to Support Privacy Preservation in Location-Based Services; Lecture Notes in Computer Science, Berlin: Springer Verlag, Vol 4278 pages: 1755-1764.

6 Frederick Yip, Alfred Ka Yiu Wong, Nandan Parameswaran, Pradeep Ray, "Towards Robust and Adaptive Semantic-Based Compliance Auditing," edocw, pp.181-188, 2007 Eleventh International IEEE EDOC Conference Workshop, 2007

7 Dos Santos Moreira 2008, Ontologies for information security management and governance,Information Management & Computer Security [0968-5227] vol.16 pag. 150

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

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

10. Breuker, J., Hoekstra, R. 2004. “Core concepts of law: Taking common-sense seriously”. In FOIS’04, IOS-Press, pages 210–221.

11 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.

12 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.

13 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.
14 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
15 García, R. 2006. A SemanticWeb Approach to Digital Rights Management. PhD thesis, Universitat Pompeu Fabra, Barcelona, November 2006.
16 Jarrar, M. 2005. Towards Methodological Principles for Ontology Engineering. Phd thesis, Vrije Universiteit Brussel, May 2005.
17 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.
18 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).
19 Cappeli A., Bartalesi V., Sprugnoli R. 2007 Modelization of Domain Concepts Extracted from the Italina Privacy Legislation
20 Poblet, M., Vallbé, J.J., Casellas, N., Casanovas, P.: Judges as IT Users: The Iuriservice Example. In A. Cerrillo and P. Fabra (eds.) E-justice: Using Information Communication Technologies in the Court System, pp. 38 – 56. IGI-Global, USA (2008)
21 Poblet M., Casellas N., Torralba S., Casanovas P; Modeling Expert Knowledge in the Mediation Domain: A Middle-out Approach to Design ODR Ontologies; LOAIT 2009 Workshop on Legal Ontologies and Artificial Intelligence Techniques/Workshop on Semantic Processing of Legal Texts; IDT series vol 2