

# Modelization of Domain Concepts Extracted from the Italian Privacy Legislation

A. Cappelli\* V. Bartalesi Lenzi R. Sprugnoli  
CELCT, Povo (TN), Italy  
{cappelli,bartalesi,sprugnoli}@celct.it  
\*ISTI-CNR, Pisa, Italy  
{amedeo.cappelli}@isti.cnr.it

C. Biagioli  
ITTIG, Firenze, Italy  
{biagioli}@ittig.cnr.it

## Abstract

In this document we present an ongoing experiment for the creation of a specialized ontology, ontoPrivacy, for the Italian Privacy legislation domain. Taking into consideration some previous work in the legal domain, the aim of our project is to give, by using a bottom-up approach, a conceptual organization of a selected glossary extracted from the Italian Personal Data Protection Code. The connection between relevant terms and ontoPrivacy will facilitate the retrieval process and the consultation and understanding of the law code.

## 1 Introduction

In recent years there have been several initiatives for the development of legal ontologies in order to offer a solid support for the acquisition, sharing and re-use of legal knowledge. In particular, we have taken into consideration two core ontologies:

1) LRI-Core Ontology [5]: it covers the main concepts that are common to all legal domains. LRI-Core has five main categories: Physical-Entity (e.g. substances and physical objects), Mental-Entity (e.g. emotions), Abstract-Entity (e.g. numbers), Role (e.g. social and communication roles), and Occurrence (e.g. events).

2) Core Legal Ontology [7]: it organizes juridical concepts in classes of entities that are specializations of the DOLCE+ foundational ontology [6].

The three basic categories are Region (space and time), Endurant (object or substance, either physical or not), and Perdurant (events, states, or processes).

At the same time, we have taken the Legal-RDF Vocabularies [9] as a reference. These Vocabularies have been developed by John McClure (Legal-RDF director<sup>1</sup>) for the annotation of narrative legal documents in the Semantic Web perspective. They are organized according to the classes and properties defined by the W3C’s Resource Description Framework (RDF) and Web Ontology Language (OWL). The goal of this modeling is to reflect the real world to the maximum extent possible. Seven classes are defined: Actors (individuals and groups), Dramas (events, divided into discrete actions and open-ended activities), Props (products and legal properties), Roles (participant, occupations, and legal and family roles), Scenes (place and time), Scripts (document types), and Themes (topics of a script or drama).

## 2 An Ontology for the Italian Privacy Legislation

OntoPrivacy has been created for the modelization of a glossary of keywords extracted from the Italian Personal Data Protection Code [1]. This glossary has been manually created by an expert in the legal field and it is made up of both specific terms of the Public Administration domain (e.g. *atto amministrativo/administrative act*) and generic words (e.g. *razza/race*). Synonymy, hypernymy and hyponymy relations are identified among such terms (e.g. *administrative act* and *judicial act* are kinds of *act*).

Taking the relevant terms contained in the glossary as starting points, we have followed a bottom-up approach to create the ontology. Vocabulary and ontology are, in fact, closely tied with a two-way relation: a lexicon can be the basis for a well-built ontology and an ontology can serve as foundation to lexicon organization [8].

The classes in ontoPrivacy have been defined by analysing the lexical entries of our glossary and the relations among them. In Figure 1 we show a tree where the upper classes of ontoPrivacy are presented: each box represents a different concept, where “Thing” is the universal class that includes everything which is identified. The “is-a” arrows relate a subclass with its superclass. There are five main classes: (i) Events: something that happens at a given place and time; (ii) Scenes: the place and time where an event occurs; (iii) Non-Physical Objects: mental objects, like conceptualizations; (iv)

---

<sup>1</sup><http://www.legal-rdf.org>

Physical Objects: things that are observable and can be experienced through sensory input; (v) Roles: functions played by individuals and groups.

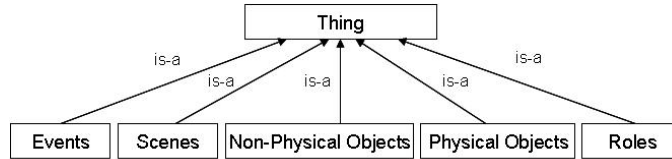


Figure 1: The upper part of ontoPrivacy

Figure 2 shows the Physical Objects class together with its dependent elements, as an example of the second level of ontoPrivacy. Physical Objects is divided in two subclasses: Artifacts and Natural Objects. Instruments is the subclass of Artifacts: e.g. *cavo/cable* and *fibra ottica/optical mean* are instances of Instruments. Natural Objects is the superclass of Natural Persons; Natural Persons is further categorized as Individual and Group. As Natural Persons performs social and functional roles, Roles class denotes occupations (e.g. *pediatra/paediatrician* is a Natural Person of type Individual performing the role Worker) and legal roles (e.g. *ministro/Ministry* is a Natural Person of type Group performing a Government role).

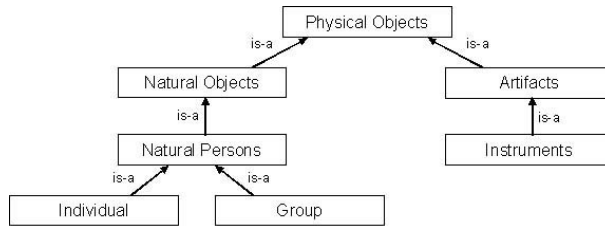


Figure 2: The Physical Objects class

### 3 Ontology Application: metaSearch

OntoPrivacy has been created to support the metaSearch (mS) software, a tool developed at ITTIG<sup>2</sup>, that allows to query the functional profile of legislative data [4]. So far, mS has been tested on the Italian Personal Data Protection Code, whose text has been annotated following the DAO (Dispositions, Arguments and Operators) model: this annotation scheme,

<sup>2</sup>Institute of Legal Information Theory and Techniques, CNR-Italy

defined at ITTIG, permits to deeply and widely describe legislative acts, making the semantic information explicit and enabling the user to directly retrieve the relevant rules [3].

Taking advantage of the knowledge deriving from ontoPrivacy, mS allows the user to formulate a better detailed query selecting a keyword from the structured glossary. Such ontology will not only be used in the query composition phase, but also in the presentation of the results: the final user will be able to widen his/her search by including terms which are in ontological relation with the chosen one.

The creation of mS is part of the Law Making Environment (LME) project, that is focused on the development of a software for the recovery, the review and the formulation of laws and for the guided planning of new acts [2]. The idea for the future is to provide the LME tool with the possibility for the user of inserting links to ontology concepts, so as to produce a semantic indexing of normative documents. On the other hand, the use of an ontology will allow to obtain additional general information on the entities that are involved in a new act. This work will have ontoPrivacy as starting point, trying to apply the same conceptual organization to new regulated areas.

## References

- [1] Decreto legislativo 30 giugno 2003, n. 196. In *GU n.174 del 29-7-2003 - Suppl. Ordinario n.123*. 2003.
- [2] C. Biagioli and E. Francesconi. A visual framework for planning a new bill. In *Proc. 3rd Workshop on Legislative XML*, 2005.
- [3] C. Biagioli and S. Pietropaoli. Towards self-explaining texts. In *Proc. 3rd Workshop on Legislative XML*, 2005.
- [4] C. Biagioli and F. Turchi. Model and ontology based conceptual searching in legislative xml collections. In *Proc. of LOAIT Workshop, ICAIL 2005*.
- [5] J. Breuker. Constructing a legal core ontology: Lri-core. In *Proc. of WONTO-2004, Workshop on ontologies and their applications*, 2004.
- [6] A. Gangemi and P. Mika. Understanding the semantic web through descriptions and situations. In *Proc. ODBASE*, 2003.
- [7] L. Gangemi, M. Sagri, and D. Tiscornia. Constructive framework for legal ontologies. In *Law and the Semantic Web*. Springer Verlag, 2005.
- [8] G. Hirst. Ontology and the lexicon. In *Handbook on Ontologies in Information Systems*. Springer Verlag, 2003.
- [9] J. McClure. Legal-rdf: Annotated laws on the semantic web, <http://www.hypergrove.com/legalrdf.org/inventory.html>.