Legal Ontology for Modelling GDPR Concepts and Norms

Monica PALMIRANI*, Michele MARTONI*, Arianna ROSSI*, Cesare BARTOLINI**, Livio ROBALDO**

*CIRSFID, University of Bologna.
{monica.palmirani, michele.martoni, arianna.rossi}@unibo.it

**SnT - Interdisciplinary Centre for Security, Reliability and Trust,
Université du Luxembourg
JFK Building, 29, Avenue J.F. Kennedy, L-1855 Luxembourg
{cesare.bartolini, livio.robaldo}@uni.lu

Abstract. This paper introduces PrOnto, the privacy ontology that models the GDPR main conceptual cores: data types and documents, agents and roles, processing purposes, legal bases, processing operations, and deontic operations for modelling rights and duties. The explicit goal of PrOnto is to support legal reasoning and compliance checking by employing defeasible logic theory (i.e., the LegalRuleML standard and the SPINDle engine).

Keywords. Semantic web, legal reasoning, legal ontology, GDPR.

1. Introduction

The GDPR (General Data Protection Regulation) is the new common framework for data protection that applies to the whole European Union and harmonizes the legal principles of its Member States that can thus be more effectively applied in the Digital Single Market. The Regulation places upon entities involved in the processing of personal data a number of obligations, among which is the obligation to assess the risks they could encounter and adapt their duties on the basis of the impact assessment (Article 35, GDPR), whereas specific measures for the safeguard of data subjects’ human dignity and fundamental rights are introduced. Instruments such as audits and compliance checking are intended to ensure the application of the principles of data protection by design (Article 25, GDPR) during software development (ex-ante phase), but also a punctual detection of violations (ex-post phase) when they occur. Since public administrations, enterprises and non-profit organizations alike will need to observe these newly-introduced, demanding duties, semantic web and legal reasoning techniques can offer a valuable support and ease compliance.

A legal ontology that formalizes data protection norms is therefore needed and timely. This paper introduces PrOnto [21], the privacy ontology that models the GDPR main conceptual cores: data types and documents, agents and roles, processing purposes, legal bases ex Article 6 GDPR, processing operations, and deontic operations for modelling rights (Chapter 3- articles 12-23) and duties (Chapter 4- articles 24-43). This ontology considers the GDPR as a starting point, however it is meant to be extended to the concepts and relative relations of other legal frameworks (such as
Member State laws). The explicit goal of PrOnto is to support legal reasoning and compliance checking by employing defeasible logic theory (i.e., the LegalRuleML standard [5] and the SPINDle engine [12]), as opposed to exclusively execute information retrieval. This article focuses on the analysis of deontic operators in order to manage the checking of compliance with the GDPR obligations. We use the Right to Data Portability (Art. 20) for illustrating the PrOnto benefits.

2. MeLOn Methodology

PrOnto was developed through an interdisciplinary approach called MeLOn (Methodology for building Legal Ontology), which has been successfully used to develop several legal ontologies by legal experts. MeLOn is explicitly designed for legal ontologies and the related difficulties encountered by the legal operators during the definition of a model of reality through ontological techniques, such as Protégé, or patterns design method or the foundational approach.

The MeLOn methodology iterates over ten steps: 1) Describe the goal of the ontology; 2) Evaluation indicators. PrOnto’s criteria, based on the existing state of the art, are [6]: (i) coherence, (ii) completeness, (iii) efficiency, (iv) effectiveness, (v) usability, (vi) agreement; 3) State of the art survey: PrOnto reuses existing ontologies, ontology patterns [13][14], and other existing domain vocabularies; 4) List the whole relevant terminology, extracted from legal sources, in particular legal definitions; 5) Use usable tools (such as tables, UML diagrams and the Graffoo tool); 6) Refine and optimize: an ontology expert manually adds the axioms; 7) Test the output in terms of completeness, effectiveness and usability; 8) Evaluate the ontology: OntoClean method and SPARQL queries; 9) Publish the document with the LODE tool [20]; 10) Collect feedbacks from the community in order to reach the agreement criteria. The MeLOn methodology allows to successfully work within interdisciplinary group that include engineers, lawyers, linguists, logicians and ontologists, and to model the legal knowledge rapidly and accurately while integrating the contributions of different disciplines.

3. The Right to Data Portability

Chapter 3 of GDPR lists all the rights of the data subject (right to access, right to be forgotten, right to portability, right to erasure, etc.) and Chapter 4 the duties for the controller and processor (such as the obligation to notify the data breach to the data subject). PrOnto aims at modelling the deontic operators (right, obligation, prohibition, permission) but also customize the obligations and rights for the GDPR. In particular we consider the Right to Data Portability (Art. 20). The Right to Data Portability is a complex right composed basically of two obligations:

“1. The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:

(a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and
(b) the processing is carried out by automated means.”

The Right to Data Portability involves the Controller that has the obligation to take some Steps (see Fig.1). A Step is executed by Actions and each Step commits LegalRules, in our case ObligationOfPortability. The actions involved are “Provide” personal data to the data subject, and “Transmit” the same data also to other controllers. The data type is determined by the action performed by the Controller. This means that all the personal data provided by the data subject or observed are involved in this obligation. The personal data inferred, derived or stored by the Controller are not a matter of this Right².

Figure 1 - Fragment of PrOnto concerning the ObligationOfPortability

4. PrOnto Modules

PrOnto is composed by modules, following the main structure of the GDPR legal principles: i) data and documents, ii) agents and roles, iii) processing purposes and legal bases; iv) data processing and workflow, risk management, and v) legal rules and deontic operators. Some documents and data refer to the data subject, which is a role of an agent (natural person). Data is processed following a given workflow, i.e., a plan of actions. When it is executed, each action assumes specific temporal parameters (e.g., interval of time of the processing), context (e.g., jurisdiction where the data processing

---

² Article 29 WP 242 rev.01 “In contrast, inferred data and derived data are created by the data controller on the basis of the data “provided by the data subject”. For example, the outcome of an assessment regarding the health of a user or the profile created in the context of risk management and financial regulations (e.g. to assign a credit score or comply with anti-money laundering rules) cannot in themselves be considered as “provided by” the data subject.”
is carried out, and value (e.g., place where the data processing is performed). The
processing is lawful only if a legal basis is provided. Each processing activity involves
a number of stakeholders: controller, processor, and other actors, and each has
obligations or rights (for instance, data subjects have data protection rights). Such
rights and obligations are linked to documents where the provisions appear, such as
terms of use, information, privacy policies, consent forms.

4.1. Data and Document

In the context of data protection, personal data (Article 4.1(1)) is the object of the
Regulation and the target of its protection, but data are also the information source that
regulates the relations among different agents (e.g., controller, processor, etc.) using
privacy policies, informed consent, contracts, codes of conduct, law, case-law, and any
other legal document. Since data and documents are documentary sources, the FRBR\textsuperscript{3}
onology is employed: their representation over time can thus be modelled by following
a robust design pattern that has been adopted for the publication process. Data is
organized in the categories defined in the GDPR: personal data (Article 4.1(1)), non-
personal data, anonymized data, pseudonymised data (Article 4.1(5)). The duties and
rights depend on the type of data. For instance the DPO (Data Protection Officer) is
mandatory for processing “a large scale of special categories of data” (Art. 9). This is
why a specific version of data can be detected by using the time when the event
occurred (e.g., a data breach event) and the dynamic versioning of the FRBR model is
applied also to the class data.

4.2. Agent and Role

Agents and roles are frequently mistaken in legal ontologies. PrOnto, on the contrary,
distinguishes the two classes. An agent might play multiple roles in different
processing operations or contexts (e.g., a controller could act as processor or third party
in relation to different data processing activities). Not only physical persons and
organizations are included in the agents’ class, but also IT organizations, artificial
intelligence and software, or robots. Each role is fixed in a given time period, which is
linked to the time version of the dataset and the duration of the data processing. This
implies that there is an event that assigns the role to an agent (e.g., designation of the
processor by the controller ex Article 28, GDPR). Concerning the different roles, we
have the Controller that isRepresentedBy a Representative in the European Union (Art.
27), designates a DPO and nominates a Processor.

4.3. Purposes and Legal Basis

Under the GDPR, personal data processing (Article 4.1(2)) is lawful only if
motivated by a purpose that must be supported by a legal basis (see Article 6, GDPR,
on the lawfulness of processing). This is why a lawfulness status was needed and was
thus added as a Boolean data property of the PersonalDataProcessing class, whilst each
personal data processing is based on a Purpose.

\textsuperscript{3}FRBR— Functional Requirements for Bibliographic Records
https://www.ifla.org/publications/functional-requirements-for-bibliographic-records
By modelling the knowledge in this manner, a rule engine that, for instance, is based on a rule-based language such as LegalRuleML is able to return this value after the rule reasoning process. The legal basis is also involved in the Right to Data Portability, considering that the Right is applicable only in the presence of a consent or a contract.

### 4.4. Data Processing

Human activities can be modelled through a workflow, i.e., a sequence of steps that takes some resources in input and produces certain outcomes. However, a workflow is composed of two parts: first a plan to do something is laid out (e.g., workflow), then the concrete sequence of actions is actually performed (e.g., execution of the workflow). This distinction is of utmost relevance in the GDPR framework: the plan (e.g., Impact Assessment Plan made by steps) is different from the real execution (e.g., the countermeasures acted in the event of a data breach), which is made up of a set of actions. Compliance checking presumes both a plan in line with the law, and countermeasures in the event of violations during the actual execution (e.g., remedies). For this goal, the Publishing Workflow Ontology (PWO) proved perfectly suitable as a basis to model the data processing ontology module because it includes both a workflow and an executed workflow. The workflow execution is composed by actions. An action [1] is a kind of event that is described by temporal parameters (e.g., interval) and contextual values (Time-indexed Value in Context - TVC). One of the values it can take is the place where the event occurs (e.g., within the EU borders) and the relevant jurisdiction (e.g., Regional competence). Other values and statuses can also be included to enrich the context description.

### 4.5. Deontic Operators

In order to model legal norms, deontic operators such as right, obligation, permission and prohibition are fundamental. Under the GDPR, it is also important to include violation/compliance as the status in which an obligation or a prohibition is violated or maintained. The deontic operators have temporal parameters and refer to a jurisdiction to consider those rights that are only effective in a specific domestic regulation. For all these reasons, this section of PrOnto allows to model those predicates that are necessary to implement legal rules and is an extension of the LegalRuleML meta-
model, which allows the synchronization of the legal rule language modelling with the ontology.

This module also defines the relationships among deontic rules, actors’ rights and obligations, obligations and permissions, and violation/compliance. This modelling allows the population of the ontology, or the creation of RDF triples, in order to perform queries such as “give me all the processing activities that have been violated by some actors in a given time”. This knowledge is processed by the rule engine, but it is also transformed into individuals (e.g., materialization) of the ontology (or RDF triples) without the need to perform a query on the rule engine each time.

It is also worth noticing that, within the project DAPRECO [8], PrOnto is used to formalize GDPR in reified I/O logic [22], and in the Cloud4Eu is used with defeasible logic. In both cases formulae are connected to the concepts in PrOnto via the LegalRuleML constructs. This means that PrOnto is neutral respect the type of logic adopted.

### 4.6. Rights and Obligations relationships

For each obligation there is also a right connected with the data subject (Bearer). Fig. 3 shows how the ObligationOfPortability is connected to the RightToPortability using DeonticSpecification super class.

![Diagram](image)

**Figure 3: Right to Portability and Obligation of Portability**

For implementing this fragment of ontology we have extended the LegalRuleML meta-model with several axioms. We report here some of axioms that intend to connect the right with the obligation and prohibition, and the right with the permission.

<table>
<thead>
<tr>
<th>LegalRule</th>
<th>&lt;owl:ObjectProperty rdf:about=&quot;<a href="https://w3id.org/ontology/pronto#componentOf%22%3E">https://w3id.org/ontology/pronto#componentOf&quot;&gt;</a></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;rdfs:domain rdf:resource=&quot;<a href="http://docs.oasis-open.org/legalruleml/ns/v1.0/metamodel/#DeonticSpecification%22/%3E">http://docs.oasis-open.org/legalruleml/ns/v1.0/metamodel/#DeonticSpecification&quot;/&gt;</a></td>
</tr>
<tr>
<td></td>
<td>&lt;rdfs:range rdf:resource=&quot;<a href="https://w3id.org/ontology/pronto#LegalRule%22/%3E">https://w3id.org/ontology/pronto#LegalRule&quot;/&gt;</a></td>
</tr>
<tr>
<td>Property</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>isViolatedBy</td>
<td>It defines the relationship between obligation/prohibition and violation</td>
</tr>
<tr>
<td>isFulfilledBy</td>
<td></td>
</tr>
<tr>
<td>Relationship between obligation and permission</td>
<td></td>
</tr>
<tr>
<td>Relationship between prohibition and obligation</td>
<td></td>
</tr>
<tr>
<td>Right as a subclass of Permission</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 – Some axioms of the extension of the LegalRuleML ontology

4.7. Duties and Violation

Finally, the Steps are connected with LegalRuleML that is the deontic part of the ontology, capable to model and perform reasoning with right, obligation, permission, prohibition. The violation is connected with the obligation/prohibition that is violated and the compliance states when the obligation is complied with. In this way, we are able to detect the steps that create violations of some obligations and the connected risks, along with the related measures (see Fig. 4).
5. Evaluation

The evaluation is carried out inside the Cloud4EU European project PCP, that intends to provide legal compliance checking systems for eGovernment services that are delivered across the cloud. We are currently in the phase of testing PrOnto on three different scenarios related to school services. PrOnto is also used inside the MIREL European project and the DAPRECO Luxembourgish project.

Figure 4: Violation, Compliance and Obligation, Prohibition

6. Related Work

A few privacy ontologies with specific goals [4]; [11]; have been designed, for instance the HL7 privacy ontology [13] for electronic health records. Other ontologies were created to ensure secure messaging among Internet of Things devices, whilst others are meant to manage data flows in the linked open data environment or on the blockchain. UsablePrivacy and PrivOnto [16] are more oriented to provide linguistic instruments in order to define glossary and taxonomy for the privacy domain, basically starting from the bottom-up annotation of the privacy policies (crowdsourcing annotation). GDPRtEXT [18] provides a list of concepts present in the GDPR text without really entering the modelling of the norms and the legal axioms (e.g., the actions performed by the processor, the obligations of the controller and the rights of the data subject). Moreover GDPRtEXT does not foster FRBR information for managing versioning of the legal text over the time and consequently the changes of the legal concepts due to modifications in the legal system. GDPRov aims to describe the provenance of the consent and data lifecycle in the light of the Linked Open Data principles such as Fairness and Trust [17]. The SPECIAL Project\footnote{4 https://www.specialprivacy.eu/} aims to provide tools for checking compliance in privacy domain. However, no ontology with foundational concepts, patterns, deontic operators and privacy principles has been designed to support legal
reasoning and check compliance yet. ODRL provides predicates and classes for managing obligations, permission, prohibitions, but several parts of the deontic logic are missing (e.g., right and penalty classes). ODRL is good for modelling simple policies capable to be searchable in SPARQL, but it is quite limited to manage the complex organization of the legal rules (e.g., exception in the constitutive rules or in the prescriptive rule). PrOnto is more exhaustive in this field. In order to do so, rights and obligations must be modelled through deontic operators. Moreover, actors and processing operations described in the normative prescriptions must also be included. This is why PrOnto considers and reuses existing ontologies and follows ontology design patterns [21]: ALLOT [7], FRBR [15], LKIF we use in particular lkif:Agent to model lkif:Organization, lkif:Person and lkif:Role [9], the Publishing Workflow Ontology (PWO) [10]; Time-indexed Value in Context (TVC) and Time Interval [19].

7. Conclusions and Future Work

The existing privacy ontologies presented in the state of the art (e.g., HL7 for eHealth, PPO for Linked Open Data, ODRL for modelling rights, etc.) do not integrate deontic logic models that can be used for legal reasoning. PrOnto aims at the integration of different levels of semantic representation for multiple goals: 1) document and data modelling can support information retrieval in the Semantic Web, in particular with Linked Open Data (e.g., SPARQL queries); 2) workflow and processing modelling can represent helpful tools to plan privacy policies, but also BPMN modelling can be useful in system design (e.g., privacy-by-design); rights and obligations are necessary modules to enable automated legal reasoning that employ rule languages (e.g., LegalRuleML and compliance checking); 3) and finally, human-centered approaches can allow the visualization and the presentation of data protection principles and concepts in different contexts and directed to different audiences.

The research described in these pages has a long-term goal. Our intention is that of continuing the modelling and optimization of the formal model of the ontology, but also to evaluate it with a number of use-cases. In the meantime, we deem fundamental a discussion about the ontology within a large community, in order to establish consensus and to place such results in a standardization body for future governance (e.g., OASIS, W3C). In the future, it will also become necessary to develop specific profiles, one for each specific national law or even by thematic domain (e.g., Privacy in IoT, Privacy in AI, etc.).

Acknowledgements

This work was partially supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690974 “MIREL: Miining and REasoning with Legal texts”, and by Luxembourg National Research Fund (FNR) CORE project C16/IS/1133956 “DAPRECO: DAta Protection REgulation COmpliance”.

References

