PrOnto: Privacy Ontology for Legal Compliance

Monica Palmirani¹, Michele Martoni¹, Arianna Rossi¹, Cesare Bartolini² and Livio Robaldo²

¹University of Bologna, CIRSFID, Bologna, Italy
²SnT - Interdisciplinary Centre for Security, Reliability and Trust, Université du Luxembourg, Luxembourg

Abstract: This paper introduces PrOnto, a legal ontology of the GDPR with the goal of providing a legal knowledge modelling of its conceptual cores: privacy agents, data types, types of processing operations, rights and obligations. This recently introduced Regulation places upon entities that process personal data the obligation of assessing the risks they could encounter and of adapting their duties on the basis of the impact assessment, including specific measures that intend to safeguard the data subject’s human dignity and fundamental rights. In this paper, we argue and show how legal compliance and privacy-by-design can be supported and eased by Semantic web technologies and legal reasoning tools. A specific focus is placed on the Risk Analysis ontological module: we intend to demonstrate that PrOnto is capable of supporting compliance checking between risks and measures. The methodology used here is based on legal theory analysis joined with ontological patterns.

Keywords: semantic web, legal reasoning, legal ontology, compliance checking

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 that process personal data the obligation of assessing the risks they could encounter and of adapting 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 like audits and compliance checking are intended to ensure the application of the principles of privacy-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 administrati ons, companies 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 privacy and data protection norms is therefore needed and timely. 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 ex Article 6 GDPR, processing operations, and deontic operations. Such ontology considers the GDPR as starting point but is meant to be extended to the concepts and relative relations of other legal frameworks. The explicit goal of PrOnto is to support legal reasoning and compliance checking by employing defeasible logic theory (i.e., the LegalRuleML standard (Athan et al., 2015) and SPINDle engine (Governatori et al., 2016)), as opposed to exclusively execute information retrieval.

This article focuses on risk analysis in order to demonstrate the ability of PrOnto to manage the compliance checking with the GDPR obligations.

2. Risk management process scenario

One of the most important innovations introduced by the GDPR is the concept of accountability (Chapter IV, GDPR) and self-assessment of the risks in general, with particular attention to those that affect the rights and freedoms of individuals (Article 32, GDPR). The instrument selected by the GDPR is the Risk Analysis that implies

---

¹ GDPR, REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
an as-is evaluation of the personal data processing and the classifications of the risks using levels (e.g., high, medium, low). This analysis has also to clearly and deeply identify the specifics impacts of each risk and measures have to be identified for mitigating risks. All this complex assessment is described in a document called DPIA (Data Protection Impact Assessment) (Article 35, GDPR). Article 29 Data Protection Working Party has delivered some guidelines (Article 29 Working Party, 2017) for achieving these goals. This approach imposes a continuous and dynamic update of DPIAs according to the evolution of the personal data processing, and periodic audits have to be performed by the controller, by the DPO (Data Protection Officer, Article 37, GDPR) and also by supervisory authorities. This scenario stresses the importance to have automatic tools that can support the compliance checking of the norms related to risks, measures, and violations. PrOnto is designed for supporting this goal. Frequent queries by a possible auditor could be: give me all the risks connected with rights and freedoms of individuals, give me the measures implemented for a category of risks (unlawful access, modification, destruction (Article 29 Working Party, 2017)) connected with a particular personal data processing (Article 35.7(d) and Recital 90), give me the possible impacts described in the DPIA related to data breach risk (Article 29 Working Party, 2018) (see paragraph 5) happened at a given time t. PrOnto intends to help the end-users to answer these queries.

3. 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. MeLOn is explicitly designed for legal ontologies and the related difficulties encountered by the legal community during the definition of a model of reality through ontological techniques, such as Protégé, the glossary method and the foundational approach.

The MeLOn methodology is composed of ten recursive steps:

- **1. Describe the goal of the ontology.** PrOnto’s goals are:
  - to model data protection legal norms starting form legal texts but including also social norms, practitioners’ opinions or social behaviours;
  - to build a legal ontology that is usable for legal reasoning;
  - to build a legal ontology that is usable for data and information retrieval.

- **2. Evaluation indicators.** PrOnto’s criteria, based on the existing state of the art, are (Bandeira et al., 2016): (i) coherence, (ii) completeness, (iii) efficiency, (iv) effectiveness, (v) usability, (vi) agreement.

- **3. State of the art survey:** PrOnto reuses existing ontologies, ontology patterns (Hitzler et al., 2016), and other existing domain vocabularies.

- **4. List the whole relevant terminology:** extracted from legal sources, in particular legal definitions.

- **5. Use usable tools:** e.g. tables, UML diagrams and the Graffoo tool2.

- **6. Refine and optimize:** an ontology experts 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 tool3 (Peroni et al., 2012).

- **10. Collect feedbacks** from the community in order to reach the agreement criteria.

The MeLOn methodology permits to work with success inside interdisciplinary group that include engineers, lawyers, linguists, logicians and ontologists, and to model the legal knowledge rapidly, accurately and integrating the contributions of different disciplines.

---

2. [http://www.essepuntato.it/lode](http://www.essepuntato.it/lode).
4. PrOnto modules

PrOnto is composed of the following modules: i) data and documents, ii) agents and roles, iii) processing purposes and legal bases; iv) data processing and workflow, and v) legal rules and deontic operators.

Some documents and data refer to the data subject, which is a role of an agent (physical 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 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 controller, a processor, and other actors and each of them 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 (ex Article 4.1(1)) is the object of the regulation and the target of its protection, but 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 documental sources, the FRBR ontology 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, pseudonymized data (Article 4.1(5)). The DPIA is a specific document defined in Article 35 of the GDPR, which evolves over time (Article 29 Working Party, 2017). This is why the appropriate version of the DPIA 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.

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. Furthermore, 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).

4.3 Purposes and legal basis

Under the GDPR, personal data processing (ex 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. 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.

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 a workflow, and also an executed workflow. The workflow execution is composed by actions. An action (Abrams, 2014) 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 data processing that has been violated by some actors in a given time”. This knowledge is processed by the rule engine, but it is also transformed into individuals of the ontology (or RDF triples) without the need to perform a query on the rule engine each time.

5. Risk analysis, risks and measures

The risk analysis is one of the most relevant parts of the innovations brought forth by the GDPR. It is a step that precedes the DPIA. It aims to evaluate if the personal data processing has high risk to affect rights and freedom of individuals and consequently to develop a plan of measures for mitigating the risk (e.g., reduce, tolerate, annul). The risk analysis is a dynamic activity that depends on organizational, management, technological and legal factors but in particular on the frequency of the risk and the magnitude (risk=frequency*magnitude). For this reason, the risk analysis should be monitored and evaluated periodically. PrOnto includes a specific module for managing risks, measures and management of the risk checking using OWL.

5.1 Risk management process as workflow

The DPIA is an internal document that evolves over time using the FRBR model that tracks the versioning. The versioning is fundamental for applying the correct DPIA in case of audit related to the infringement that happened at a given moment in the past. The DPIA describes risk analysis and is a kind of workflow (description of a plan composed by steps). A step of the risk analysis entails a risk that has a level of risk (high, medium, low) and a Boolean attribute for recording the riskiness for rights and freedoms. A step introduces also measure.
The risk is classified in different subclasses: legal risk, ethics risk, digital risk, and organizational risk. The risk is mitigated by one or more measures.

Figure 2: Risk classes

5.2 Workflow execution

It is very important to distinguish the plan to do something from the real execution of the proposal. The execution could be also discrepant or inconsistent with the original design. In this manner, we can check the compliance of what happened with what was programmed. The workflow is executed in real time through actions. Action has a breachness attribute that states if the action is prone to the risk of data breach (see Fig. 3).

The actions are divided in subclasses for grouping the main types of activities managed in the GDPR. The actions on the data are: provide, store, derive, infer, observe, delete, transmit. Those categories are important also for correctly applying some rights such as the right to data portability and the right of access. The right to data portability (Article 20, GDPR) is applicable to the data provided by individuals, in active or passive way, but not to the data inferred or derived during the activity of the controller. In the right of access (Article 15, GDPR) the “controller shall provide a copy of the personal data undergoing processing”, including also derived and inferred data. The category delete is subdivided in different other subclasses for distinguishing the level of deletion: permanent, destruction, anonymized. Other actions concern the communication, consent, etc. (see Fig. 4).

Figure 3: Workflow execution
5.3 Detection of violations

The steps are connected with LegalRule that is the deontic part of the ontology capable to model and make reasoning with right, obligation, permission, prohibition. The violation is connected with the obligation 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.
6. Evaluation

The evaluation is carried out inside the Cloud4EU European project PCP\(^4\) 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\(^5\) and the DAPRECO Luxembourgish project\(^6\). Some examples of the use of PrOnto related to Risk Analysis are hereafter presented.

| a. Give me the DPO of public-body(X) when X is a controller. | SELECT ?dpo WHERE { ?dpo a :DPO . [ a :Public_Body ; :plays [ a :Controller ; :designates ?dpo ] ; rdfs:label "X" ] . } |

Those queries test the first three criteria defined by the MeLOn methodology:

- coherence: the axioms of the ontology can’t create inconsistency or contradictions;
- completeness: the domain is adequately covered by the ontology and the main concepts are included;
- efficiency: the ontology is technically sound, concise and the reasoning is computable in reasonable time, and it is based on patterns.

The remaining criteria should be tested with end-users and large number of data:

- effectiveness: the ontology covers the most important queries about the domain and the end users find it helpful to resolve applicative situations;
- usability: the end users find the ontology clear, understandable, easy to use, close to the main terminology used inside of the community, self-explained.
- agreement: the degree of agreement and acceptance of the ontology in the legal expert community.

We intend to proceed in the next term with this second phase using concrete use-cases.

7. Related work

A few privacy ontologies with specific goals (Ashley, 2017; Gharib et al., 2017; Sacco and Passant, 2011; Samavi and Consens, 2018) have been designed, for instance the HL7 privacy ontology\(^7\) (Health Level Seven

---

\(^4\) [http://www.agid.gov.it/cloudforeurope](http://www.agid.gov.it/cloudforeurope)
\(^5\) [http://www.mirelproject.eu/](http://www.mirelproject.eu/)
International, 2015) 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\(^8\), PrivOnto (Oltremari, 2016) 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 (Pandit, 2018) provides a list of concepts present in the GDPR text without really enter in 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). Morover 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. GDPR\(^9\) 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 (Pandit, 2017). SPECIAL Project\(^5\) 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\(^10\) 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 policy 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:

- **ALLOT**: this ontology implements the Akoma Ntoso Top Level Classes (TLCs) as a formal OWL 2 DL and allows to connect the data and document classes with the FRBR ontology (Barabucci et al., 2010).
- **FRBR**: FRBR is an ontology that implements the FRBR model (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1996).
- **LKF Core**: Action.owl is an ontology that represents actions in general, i.e., processes that are performed by an agent. We use in particular lkif:Agent to model lkif:Organization and lkif:Person (Breuker et al., 2007).
- **LKF Core**: Role.owl is an ontology to describe typologies of roles (epistemic roles, functions, person roles, organisation roles). We use in particular lkif:Role (Breuker et al., 2007).
- **The Publishing Workflow Ontology (PWO)** is a simple ontology written in OWL 2 DL for the characterization of the main stages in the workflow associated with the publication of a document (e.g., being written, under review, XML capture, page design, publication on the Web). We reuse the workflow pattern to model the different types of processing of personal data (Gangemi et al., 2017).
- **Time-indexed Value in Context (TVC)** is an ontology pattern that allows to describe scenarios in which someone (e.g., a person) has a value (e.g., a particular role) during a particular time and for a particular context. We use this portion of ontology to connect the event with value, context and time parameters (Peroni et al., 2017).
- **Time Interval (TI)** is an ontology design pattern that enables the description of periods of time that are characterised by a starting date and an ending date. We use this ontology to manage the time interval (Peroni et al., 2017).

8. Conclusions and future work

None of 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.) is integrated with deontic logic models that can be used for legal reasoning. On the contrary, the ontology presented in this article, PrOnto, aims at the integration of different levels of semantic representation for multiple goals: 1) documents’ 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 modeling can represent helpful tools to plan privacy policies, but also BPMN modelling can be useful.

---

8 [https://usableprivacy.org/](https://usableprivacy.org/)
9 [https://www.specialprivacy.eu/](https://www.specialprivacy.eu/)
10 [https://www.w_org/ns/odrl/](https://www.w_org/ns/odrl/)
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-centred 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.).

9. Acknowledgements

This work was partially supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 690974 “MiREL: Miining and REasoning with Legal texts”, and by Luxembourg National Research Fund (FNR) CORE project C16/IS/11333956 “DAPRECO: DATA Protection Regulatory Compliance”.

10. References


Article 29 Working Party, 2017. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679 (No. wp248rev.01).


