Towards a Legal Ontology for the Digital Preservation Domain

Proceedings of the International Conference on ICT LAW 2013

Marzieh Bakhshandeh 1
Hossein Miri 2
Barbara Kolany-Raiser 3
Silviya-Aleksandrova Yankova 3
Mykola Galushka 4
Artur Caetano1
Jose Borbinha1

ABSTRACT

Digital preservation is all about long term preservation of data substance and operability. That enables future users not only to benefit from today’s knowledge, but also to actually use stored data and rerun the preserved processes. Furthermore, Law is becoming an essential application domain for technology developments. For example, in the case of digital preservation, all kinds of copyright protected data is an exclusive right of the copyright holder which, every process of a digital preservation system may violate this right, if it stores copyright protected material. This paper presents a Legal Ontology that provides a hierarchical overview of how legal constraints and obligations (e.g. IP rights and licensing issues) could be implemented in an automated process of a DP system. In simple words, the problems with legal taxonomies arise when the creators and the users don’t share the same perspective and it usually happens when the creators of the taxonomy are lawyers and the users are not lawyers. Legal taxonomies for digital preservation can be represented with ontologies which are an explicit account of a shared understanding in any domain and can improve communication which, in turn, can give rise to greater reuse, sharing, transparency, and inter-operability. An inherent element of every DP activity is ensuring the authenticity and legitimacy of the performed actions and processes. The correctness of our legal ontology is validated with a set of competency questions defined in a specific case study. The aim is to obtain a clearer taxonomical view of the necessary legal knowledge that will address the concerns of industrial use-case DP stakeholders. We propose using the Legal Ontology for the DP domain, in order to integrate different legal perspectives and perform reasoning and inference over legal knowledge and information.

Keywords: Digital Preservation, Ontology, Legal Ontology, Legal taxonomies

1 Instituto de Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento (INESC-ID), Rua Alves Redol 9, 1000-029 Lisbon, Portugal
2 Karlsruher Institute of Technology (KIT), TecO, Vincenz-Prießnitz-Straße 1, 76131 Karlsruhe, Germany
3 Institut für Informations-, Telekommunikations- und Medienrecht (ITM), Leonardo-Campus 9, 48149 Münster, Germany
4 SAP, The Concourse, Queen's Road, BT3 9DT Belfast, United Kingdom
INTRODUCTION

Digital Preservation (DP) is nothing new, though an imperative challenge for the information society. It has existed for a long time, but in a domain where the focus has been on the preservation of static digital objects and artifacts. A novel approach has been introduced [17] which seeks to leverage this expertise pool to implement solutions capable of preserving dynamic digital objects, interactive media, and entire business processes and services. Efforts are being put into describing whole processes and capturing their complete inter-dependencies and constituent components, along with their configurations, in order to re-deploy them in the future. The intent is to do this in a way which allows future interaction with them. An inseparable element of every DP activity is ensuring the authenticity and legitimacy of the performed actions and processes. The ontological approach to organize legal information and requirements could help with the legal perspectives and concerns – making it a pivotal element of any DP system. The need for addressing legal issues and obligations in the DP domain is manifest: almost every process of a DP system may infringe a right, among other legal requirements and constraints, e.g. contracting issues and licensing. Ontologies can help us with this problem and, therefore, could be a sensible solution to achieve our goal of creating a common understanding of the meaning of legal concepts and terms, thus mitigating the risk of misinterpretation, particularly in legal applications. They could, effectively, fulfill this objective by providing contextual explanation and precise legal information.

The rest of this paper is organized as follows: First, we briefly introduce the concept of Legal Ontology Engineering in the domain of DP, followed by presenting some of the related works done in this area. Then, we point out the drawbacks of these works, and highlight our methodology to address these shortcomings. We, subsequently, concentrate our efforts on showing the novelty and merits of our developed Legal Ontology by employing a recent case-study in e-Health. Finally, we illustrate concrete validation steps taken to evaluate our work, summarize our contributions and conclude the paper.

LEGAL ONTOLOGY FOR DIGITAL PRESERVATION

Gruber defined ontology as “an explicit specification of a conceptualization” [8]. Put differently, ontology is an explicit formal specification of the terms in a domain and relations among them – essentially akin to a taxonomical representation of a class hierarchy in a given domain. Ontologies describe structure and hierarchy. These characteristics could benefit the legal domain, as they could help legal research and information organization tremendously. For example, taxonomy in the form of ontology can be instantiated to form a knowledge-base, which would then allow a DP expert to acquire and express inferred legal knowledge through the contents of the ontology. Such an advantage makes a Legal Ontology a beneficial source of hierarchical knowledge to the experts and stakeholders, particularly in the domain of DP that has an inherent and inseparable legal-compliance element to it. It is, also, noteworthy that taxonomies facilitate knowledge engineering, knowledge extraction, and consistency/conformity checking. Put simply, they could help us to properly retrieve what we have stored, e.g. specific rules and laws. Whether one is a common lawyer or an expert in a non-legal field, they could greatly benefit from browsing through a legal taxonomy and its hierarchical classification. The ontology-based approach to capture and formalize legal information is, therefore, about shifting power towards business experts, domain experts, and industrial use-case owners, thus representing a business-centric approach.
One of the key concepts of DP is to preserve data substance and information operability, to enable future users not only to benefit from today’s knowledge and technology, but also use the stored data and information to re-run whole preserved processes. Hence, there are two main parties involved in DP processes: The DP User willing to preserve the information, and the DP provider who either carries out DP himself or provides means for it. For this reason, the reproduction of the stored data and information is inevitable. This includes re-production of copyright-protected data and software, as well as intellectual-property materials, which is always an exclusive right of the copyright-holder; e.g. software and service contracts that protect the links between pieces of code with a formally verifiable interface [1]. It is, therefore, crucial that every organization that deploys a DP system is aware of possible legal infringements. This need has been accentuated in recent years, as legal information needs to be presented consistently to avoid confusion and potential infringements. Legal constraints and requirements (e.g. IP rights and licensing issues) could help stakeholders, business analysts, and IT experts speak the same language and use the same vocabulary. In other words, they would gain a common understanding of the meaning of concepts and terms, thus mitigating the risk of misinterpretation, which plagues many legal applications. This advantage is a common goal in developing legal ontologies for the domain of DP. Such drawbacks as lack of a common understanding of the meaning of terms and entities could be addressed effectively by providing additional contextual explanation and precise legal information.

In order to avoid legal conflicts and infringements, we initially designed, implemented, and published [9] exhaustive clauses regarding DP’s legal issues, including Digital Escrow Services that ensure that software systems can survive their providers [16]. In this paper, however, we present the Legal Ontology that we have designed and developed alongside this implementation and publication. Moreover, it formalizes the key parameters of legal contexts to be captured and reasoned about. This legal taxonomy, essentially, serves as a unifying conceptual framework in the domain of DP, enabling the identification of common legal ideas, terms, issues, entities, concepts, and their relationships to each other. This unifying Legal Ontology is intended to function as a lingua-franca to facilitate the translation and mapping between different perspectives, as well as reasoning and inference over legal information.

In any Legal Ontology, there must exist a hierarchical and structured overview of legal concepts and entities, as well as the relations between them. Moreover, it must accommodate legal information based on the degree of expertise that is required in a specific domain. The ontology could, then, provide an integrated view on the legal domain, providing specifics that are vital to be recognized and conformed to. We propose using the constructed Legal Ontology for the DP domain, in order to integrate different legal perspectives and perform reasoning and inference over legal knowledge and information. This goal is realized through developing a legal taxonomy that can ensure consistency/conformity checking, traceability and inter-operability through ontology reasoning on domain competency queries. The novelty of our proposed Legal Ontology lies in its focus on the DP of whole business processes and services.
RELATED WORK

Outstanding work has been conducted by many researchers, such as the one presented by Mommers [15]. He explained that the question of how knowledge can be accommodated in ontology of Law basically induces two perspectives. First, the knowledge perspective that refers to the way in which knowledge is acquired and justified (epistemology). Second, the existence perspective that refers to what entities exist, and the way in which they exist (ontology). Therefore, in building taxonomy of the Law, we need to distinguish between these two perspectives, in order to create an integrative view of the legal domain for our specific purposes, as well as a general legal taxonomy. Mommers further suggested that a Legal Ontology should consist of six basic types: entities, ontological status layers, epistemic roles, relations, acts, and facts – an advice that we have adhered to. He argued that these basic types may be used to reflect certain views on the existence of entities in the legal domain ontology as well as the ways in which we can qualify some of these entities as knowledge.

Work has also been done on formal specification of legal ontologies, e.g. Visser and Bench-Capon [21]. They discussed four issues encountered while formalizing an informally-described ontology, and presented an ONTOLINGUA specification of the ontology (used also by Valente [18]). They pointed out that researchers in the legal domain (e.g. Moles and Dayal [14]) argue that the Law community should study the implicit assumptions being made about the nature of Law when making legal knowledge systems. The Law community has not, however, shown a considerable interest in explicitly documenting such assumptions until recently. Two well-known ontologies have been proposed for the legal domain: A Functional Ontology of Law by Valente [18] and A Conceptual Frame-Based Ontology for the Law by Van Kralingen [19] and Visser [20].

Particularly in the last few years, many works have been carried out to develop and implement legal ontologies. There is a vast literature in this domain, which was reviewed by Breuker et al. in 2009 [2]. There is even knowledge modeling approaches for the legal domain, e.g. DILK-DK approach [4, 5, 6, and 7] that aim at keeping domain knowledge distinct from its legal perspective. This particular succession of works also includes an automatic approach based on Machine Learning and NLP Techniques to support bottom-up knowledge acquisition from legislative texts within the DILK-DK framework.

The closest work to our Legal Ontology development is the EU DALOS project (Drafting Legislation with Ontology-Based Support) [3, 25]. DALOS aimed at providing legislators with control over legal concepts and the corresponding vocabulary across several European languages. Its domain ontology represented the consumer Law. Their proposed learning approach for legal knowledge acquisition provides several benefits, including suggesting concepts for hand-crafted ontologies.

Most legal ontologies do not yet take context into account, even though in the domain of DP capturing contextual legal constraints is absolutely crucial. In this paper, however, we focus on competency questions (reasoning queries) obtained directly from the case-study owners that take context into consideration. We identified all the necessary legal concepts and constructed a conceptual map based on the identified entities and relationships among them. We, then, added necessary constraints and rules to our ontology, and finally validated it using an e-Health use-case.
METHODOLOGY

The ontology building process is a craft, rather than an engineering activity [22]. Every development team usually follows its own set of principles, design criteria, and phases in the ontology development process. However, there are a series of well-known methodologies that have been proposed for building standard ontologies. The methodology employed in this work is an adaptation of the one defined by Horridge [23]. The steps include: (1) identification of the concepts and concept hierarchy (2) identification of the disjoint concepts (3) modeling composition (4) addition of all the relationships between concepts (5) identification of definitions (6) addition of annotations (7) and refinement of the ontology through various iterations of the above steps. Most ontology building methodologies propose iterative approaches in order to allow formalization to be accomplished progressively. In this work, we follow an iterative approach by using conceptual maps as a “bridge” between the legal taxonomy and the formal specification. For the first phase, the concepts and their relationship were drawn in a Conceptual Map model Figure 1 depicts a representation of the conceptual map used to develop our Legal Ontology. We have used the XMind tool [26] to progressively detail the model.

In figure 1 we can see a conceptual map of the legal perspective. In this description the concepts are written in bold and the relationships are in italic. As we live in a society where there are legal rules for the conduct of Legal Persons, their Actions NeedToComplyWith the Legal Requirements imposed by the law. A Legal Requirement means generally everything that is demanded or imposed as an obligation by law. As a matter of course, Legal Requirements DifferAccordingTo the Location where Legal Persons carryOut their Actions because the legal rules in each country are different and depend on the national legislation.

DigitalPreservation as such an Action NeedToComplyWith Legal Requirements as well. Regarding DigitalPreservation, the most relevant Legal Requirements are Data Protection, IP-Rights, ObligationsToPreserve and Contracting. In order to lawfully preserve BusinessProcesses each Legal Person has to be aware of legal restrictions, conduct law-abiding and fulfill its legal obligation.

For example, legal ObligationsToPreserve which require DigitalPreservation already exist. Such ObligationsToPreserve can be generally found especially in the areas of tax law (annual balances, invoices, etc) or medical law (the health records of patients) where it appears essential that specific Data files need to be archived for a long period of time.

Artifacts like Software, Databases or other types of Data CanBeProtectedBy Copyright. In order to be able to digitally preserve them without any infringement of IP-Rights a Legal Person has to be aware how far the protection of these Artifacts reaches and whether preservation Actions/Methods like Migration or Porting are allowed. While Software is usually a subject of Copyright protection, Data and Databases need to fulfill more specific criteria to be protected by IP-Rights. Databases for example CanBeProtectedBy either Copyright if they constitute the author's own intellectual creation; or if that is not the case, they have simply ProtectionSuiGeneris if their maker has made a substantial investment. According to the differing scope of protection different methods and techniques for DigitalPreservation are permissible.

The scope of IP-Right protection CanBeDefinedBy not only national law or European directives but by Contracts as well. Due to the fact that Legal Persons AreRightholderOf Software, they CanGrant RightOfUse to other Legal Persons by signing (CanSign) a Contract. These Contracts are usually Licenses or Sale Contracts and Software canBeDeliveredOnTheBasisOf of these Contract types. Thus, not only the author and original rightholder of the Software but other Legal Persons as well can be authorised to use the Software and obtain RightofUse. In this sense, some aspects of Copyright like the RightofUse
canBeDeterminedIn Contracts. For example, the licensor CanGrant the licensee the right to freely modify or migrate the Software in a License Contract and thus make Actions necessary for the execution of DigitalPreservation legally feasible. In case that one JuridicalPerson like a company offers DigitalPreservation as a service for other JuridicalPersons, they CanSign a ServiceContract and define the particular parameters of appointed service level in an annex to the Contract called ServiceLevelAgreements.

Data CanBe related to an identified or identifiable NaturalPerson and therefore CanBe PersonalData or even SensitiveData. Such Data needs legal protection from any acts of DataProcessing which are unwelcomed by the NaturalPerson to whom the PersonalData belongs. This due to the principle that every NaturalPerson has the right of informational self-determination and the right of privacy. Therefore, privacy security and Data Protection are essential Legal Requirements and the compliance with them is monitored by public authorities.

Thus, if Data is digitally preserved it has to be guaranteed that the Actions necessary for DigitalPreservation are compliant with (NeedToComplyWith) the rules of Data Protection. One basic concept of DataProtection is that DataProcessing requires the ConsentOfDataSubject. The NaturalPerson to whom the PersonalData belongs is called in this sense Datasubject. The ConsentOfDataSubject has to be given in advance regarding the specific DataProcessing process and cannot be generic. A way to be compliant with the rules of DataProtection can be to “hide” the personal component of Data as well as the connection between the certain Datasubject and its PersonalData by transforming the Data to AnonymousData or EncodedData. Table 1 shows summary of Classes and relationships in the Legal conceptual map.

![Figure 1: Legal Conceptual Map](image-url)
The legal conceptual map was used as the input to create the ontology, using the OWL representation. The concepts contained in the concept map were mapped into OWL classes. Relations were mapped into OWL ObjectProperties, and restrictions were added into those properties: InverseObjectProperties and SuperObjectProperties axioms were added to the OWL ontology. Cardinalities were also added to some of the concepts and relations. Furthermore, some DataProperties were defined. Additionally, the concept descriptions in Error! Reference source not found. were added as annotation (rdfs:comment) for each concept in the ontology. This ontology was built in Protégé tool [27].

Error! Reference source not found. shows the OWL representation of the legal ontology with the Software class highlighted on the left pane and the annotation description and restriction in the middle pane and respective object properties and data properties in the right pane.

Figure 2: Legal Ontology

CASE STUDY

Our case-study for the validation phase is an e-Health scenario. Each prescription drug package sold in Europe must contain information about how it works and what the intended effect is. It also has to contain a description of side effects, instructions, and cautions for its use; including warnings about possible allergies. During a patient’s visit, the medical practitioner tries to identify the best treatment strategy, which may include a prescription of one or more drugs. A prescription of a combination of drugs may cause Adverse Drug Reaction (ADR) [9, 11, 12]. ADR describes harm(s) caused by taken medications at a normal dosage during normal use. A more generic term, Adverse Drug Event (ADE), refers to any injury caused by drugs, whether they were used at normal dosage and/or due to overdose, and any associated harm(s) [13, 10].

Our case-study is concerned with addressing the ADR problem by providing a web-based solution for discovery and search of ADE rules used by doctors and pharmacists for prescribing drugs. Drug prescription mistakes based on incorrect ADE search results can cause serious complications to a patient’s health, which may lead to complex and expensive lawsuits. The follow-up investigation process includes a complete re-construction and re-run of the discovery business process carried out in the past. It must be re-run on the same hardware/software stack.
and use the same input data as the original business process. This requires constant monitoring and long-term preservation of the discovery business process.

A high-level overview of the e-Health case-study, which provides doctors and pharmacists with information about potential ADEs, is shown in Figure 3. The e-Health business processes include three companies: DrugFusion, DataMole and SemanTech. They also utilize two external Web-services: Central Medical Repository for Drug Prescriptions (CMRDP) and Pharmaceutical Company (PhC). Long-term digital preservation helps companies to fulfill legal obligations and ensure the overall reliability of the drug prescription processes within the European market. All companies involved in this use-case are real; however their names were changed due to privacy protection. The involved processes can be split into discovery of business processes and search for business processes. The discovery process analyses the source data of drug usage collected on quarterly bases and generates ADE rules. The search process performs indexing and retrieval of relevant ADE rules for the requested patient’s conditions. The legal risks in this case-study include Non-Compliance with Data Protection Obligations, Non-Compliance with Licenses and Contracts (which exist between the companies involved in the business processes), Liability for Incorrect Information and Liability for Damages. The latter, specifically, includes Contractual Liability in Civil Law, Non-Contractual Liability in Civil Law and Penal Law Liability. This scenario was insatiable in the legal ontology as individuals (instance) for the concepts.
For example, from *Error! Reference source not found.*. DrugFusion, DataMole, SemanTech, Pharmaceutical companies are all instances of the Juridical Person concept. Drug Pre-Processing, Drug Adverse Effect Sequence Discovery, Search Engine, Result Validation are all instances of the Software concept. The Drug Adverse Effect Discovery and Drug Adverse Effect search are all instances of the Business Process concept. Also AER Rules Database and Drug Usage Databases are all instances of the Database concept. Drug Usage Data and Drug Adverse Effect rules are all instances of the Data concept. Germany, UK, Spain are instances of the Location concept. Finally, the contracts between the DrugFusion and DataMole, DrugFusion and SemanTech, and Doctors and DrugFusion are all instances of the Service Contract concept.

**VALIDATION OUTCOMES**

After implementing the scenario in the legal ontology we had to validate the legal ontology by performing reasoning and inference over legal knowledge and information. According to [24] one of the ways to determine the scope of the ontology is to sketch a list of questions that a knowledge base based on the ontology should be able to answer. We have applied reasoning queries (competency questions) to our Legal Ontology. The goal here is to ensure consistency/conformity and attain specialized legal information for the DP of whole business processes and services. A set of predefined competency questions were used in order to validate ontology. Some of the competency questions defined to validate the legal ontology (due to the lack of space we could not represent more) is composed by the following questions:

1. Which database is protected by ProtectionSuigeneirs?
2. Which are the basic legal requirements regarding digital preservation?
3. Which software has 70 years’ time protection by copyright?
4. Who has the exclusive right of the copyrightholder for the Drug Instruction database?
5. What is the business process that exists between the drugfusion & datamole company?

![Figure 4: Query result](image)
Figure 4 to Figure 8 depict the results of these reasoning processes. As you can see, our formal legal ontology approach was able to successfully answer all the competency questions.
CONCLUSION AND FUTURE WORK

This paper proposes using ontologies to integrate law perspective with digital preservation domain. Ontologies describe a domain model by associating meaning to its terms and relations. The importance of this technology is evidenced by the growing use of ontologies in a diversity of application areas. A legal ontology was made for the digital preservation domain. This unifying Legal Ontology is intended to function as a lingua-franca to facilitate the translation and mapping between different perspectives, as well as reasoning and inference over legal information in the domain of digital preservation. Next, the legal ontology was validated by a set of competency questions through a specific case study. This validation was processed with reasoning methods. Future work will focus on the application of this approach to new scenarios in order to discover the analysis possibilities, considering the usage of different reasoning and querying techniques.

ACKNOWLEDGEMENTS

This project is partially supported by the European Commission under the 7th Framework Programme (FP7/2007-2013) under grant agreement 269940 TIMBUS project (http://timbusproject.net) and also supported by national funds through FCT – Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2013.

REFERENCES


3. DALOS EU project http://www.dalosproject.eu/


17. TIMBUS EU project [http://timbusproject.net](http://timbusproject.net)


26. XMind tool: [http://www.xmind.net](http://www.xmind.net)