Special Issue on the
Semantic Web for the Legal Domain
Guest Editors’ Editorial: The Next Step

Pompeu Casanovas\textsuperscript{a,b}, Monica Palmirani\textsuperscript{c}, Silvio Peroni\textsuperscript{d,e}, Tom van Engers\textsuperscript{e}, Fabio Vitali\textsuperscript{f}

\textsuperscript{a} pompeu.casanovas@uab.cat
Institute of Law and Technology, Universitat Autònoma de Barcelona
Campus de la UAB, 08193 Bellaterra (Barcelona), Spain

\textsuperscript{b} p.casanovasromeu@deakin.edu.au
Law and Policy Program: Data to Decisions Cooperative Research Centre, Law School, Deakin University
1 Gheringhap Street, Geelong VIC 3220, Australia

\textsuperscript{c} monica.palmirani@unibo.it
CIRSFID, University of Bologna,
via Galliera, 3, 40121 Bologna (BO), Italy

\textsuperscript{d} silvio.peroni@unibo.it
Department of Computer Science and Engineering, University of Bologna,
Mura Anteo Zamboni 7, 40126 Bologna (BO), Italy

\textsuperscript{e} vanengers@uva.nl
Leibniz Center for Law, University of Amsterdam,
Vendelstraat 8, 1012 XX Amsterdam, The Netherlands

\textsuperscript{f} fabio.vitali@unibo.it
Department of Computer Science and Engineering, University of Bologna,
Mura Anteo Zamboni 7, 40126 Bologna (BO), Italy

Abstract. Ontology-driven systems with reasoning capabilities in the legal field are now better understood. Legal concepts are not discrete, but make up a dynamic \textit{continuum} between common sense terms, specific technical use, and professional knowledge, in an evolving institutional reality. Thus, the tension between a plural understanding of regulations and a more general understanding of law is bringing into view a new landscape in which general legal frameworks—grounded in well-known legal theories stemming from 20th-century c. legal positivism or sociological jurisprudence—are made compatible with specific forms of rights management on the Web. In this sense, Semantic Web tools are not only being designed for information retrieval, classification, clustering, and knowledge management. They can also be understood as regulatory tools, i.e. as \textit{components of the contemporary legal architecture}, to be used by multiple stakeholders —front-line practitioners, policymakers, legal drafters, companies, market agents, and citizens. That is the issue broadly addressed in this Special Issue on the Semantic Web for the Legal Domain, overviewing the work carried out over the last fifteen years, and seeking to foster new research in this field, beyond the state of the art.

Keywords: Semantic Web for the Legal Domain, Ontologies, Linked Data, Semantic Web and its applications, Regulations, Rights
1. Introduction

The Semantic Web has been devoted to addressing social issues from the outset. For instance, in the field’s best-known seminal article [1], Tim Berners-Lee, Handler, and Lassila describe intelligent agents dialoguing and interacting between each other in order to solve a particular medical scenario. However, in the early years of the Semantic Web, the main effort went into background studies, where Artificial Intelligence (AI) and Knowledge Engineering (KE) were the main actors on the scene. Activities mostly involved the development of new formal languages (e.g., RDF [33], RDFS [19] and OWL [75]²), APIs (e.g., Jena [20] and OWL-API [61]), reasoners (e.g., Pellet [114] and Hermit [76]), ontology development editors (e.g., Protégé [70] and the NeOn Toolkit [116]), and visualization tools (e.g., VOWL2 [72] and LODE [91][92]).

Only later, and standing on the shoulders of the aforementioned works, did people start to approach the Semantic Web from different perspectives, such as engineering (i.e., Linked Data [12]), the social sciences (i.e., Social Semantic Web [57]), and the hard sciences (i.e., Web Science [60]). The Semantic Web then started to broaden its reach, moving from academia to other (more “applicative”) domains, such as industry, administration, and, last but not least, law—the very topic of this special issue.

In the following sections we want to spend a few words on examples of applications of Semantic Web technologies to these domains before presenting, in Section 7, five high-quality articles that have been selected for this special issue of the Semantic Web for the Legal Domain.


According to the vision provided in foundational works [109][111][95][34], Semantic Publishing is the use of Semantic Web technologies to enhance published documents (e.g., journal articles) in order to enable the definition of formal representations of their meaning, facilitate their automatic discovery, enable their linking to semantically related articles, provide access to data within articles in actionable form, and allow integration of data between papers. In particular, the ultimate aim of Semantic Publishing is to semantically represent the intellectual discourse of a document in machine-readable form, including how arguments are modelled within the text.

Since at least 2010, a number of initiatives have been proposed to promote Semantic Publishing to a broader audience, particularly triplestores (e.g., Open Citation Corpus [110][90] and the Open University Open Linked Data [126]), workshops (e.g., SePublica [45][46][47][48] and Linked Science [67][68][56]), special issues of academic journals (e.g., the Semantic Web Journal special issues on New Models of Semantic Publishing in Science [15] and on Linked Data for Science and Education [69]), academic challenges (e.g., the Semantic Publishing Challenge [101]), and research communities (e.g., Force11³ [38]). These initiatives seem to confirm that Semantic Publishing using Web standards currently constitutes one of the most interesting topics in the scientific publishing domain.

Probably due to the success the Semantic Publishing movement is having in academia, several publishing companies (e.g., the Nature Publishing Group³ and Elsevier⁴) and journals (e.g., the Semantic Web Journal⁵ [65] and Journal of Universal Computer Science⁶) are approaching Semantic Publishing by releasing concrete datasets and applications. These initiatives represent the first formal endorsement of publishing-related industries in moving towards Semantic Web technologies in real-case industrial scenarios.

On the other hand, the work carried out by Legal Information Institutes (LII)⁷ and the Free Access to Law Movement (FALM⁸) deserves a mention. Based on the right of republication, they have devoted themselves, over the last twenty years, to providing free online access to case law, legislation, treaties, law reforms, and legal scholarship.⁹ This is a significant contribution to the development of the rule of law from a global point of view.

Their success is incontrovertible, because of its global scope. For example, the Cornell LII was visited last year by about 30 million different individuals from 246 countries and territories.¹⁰ AustLII¹¹ shows similar results.¹² Total hits for 2014 exceeded 223 million and averaged 611,545 hits per day. By the end of 2015 AustLII had 708 databases; 712 in February 2016.¹³ On average, it has been adding more than one new database every two weeks since it was started in 1995.¹⁴

LII and FALM have been paying a close attention to Semantic Web developments [93] to enable easier access to legal texts and improve their readability [32]. Let Thomas R. Bruce, director of the Cornell LII, freely provide some specific examples:¹⁶

“On the technical side, we employ Semantic Web technologies in a number of our features and collections. For example, search of some portions of the US Code is enhanced with Linked Data from the DrugBank database, which directs users who search for a brand-name drug to the regulations that deal with the components of that drug (so, for example, a search for ‘panadol’ would give results similar to a search for ‘acetaminophen’). We have, in prototype, a number of features based on entity-linking techniques (so, for instance, mentions of medical conditions in the sections dealing with benefits for military veterans are linked to their corresponding MESH entries, from which the user can navigate to further medical information; in the past, we

---

8 http://www.worldlii.org/.
9 http://www.falrn.info/.
11 Cornell Legal Information Institute: https://www.law.cornell.edu/.
15 Graham Greenleaf, personal communication.
16 Tom Bruce, personal communication.

---

1 Even if the first drafts of such specifications were published starting from the end of the 20th century, here we decided to cite only their earliest versions.
2 http://www.semantic-web-journal.net/content/vowl-2-user-oriented-visualisation-ontologies.
7 Thomas R. Bruce, director of the Cornell LII, freely provide some specific examples:¹⁶

---

1 Even if the first drafts of such specifications were published starting from the end of the 20th century, here we decided to cite only their earliest versions.
2 http://www.semantic-web-journal.net/content/vowl-2-user-oriented-visualisation-ontologies.
classified agricultural regulations using AGROVOC, and linked regulations to related scientific papers classified with that ontology). These prototype features come and go as we test them for viability, but will very soon be aggregated into a system of infocards that will show the user a great deal of related, linked data having to do with the things that a particular law regulates or mentions, or that have to do with its creator or enforcer. We’ve done a lot of related work on metadata models for legislation for the Library of Congress, and some of our work on that has been adapted for use in the latest incarnation of the Government Publishing Office’s FD/SYS online publications. We’ve also done a lot with the automated extraction of statutory and regulatory definitions, with particular attention to scoping language, in order to link all defined terms in the statutes and regulations back to the relevant definitions.”

Thus, LILs set out the principles of legal information and FALM principles—the so-called Hague principles (2008)17—with Linked Open Data (LOD) [28],[51]18 to enhance the quality of search and enrich the level of the information provided. LILs have recently launched a scholarly journal on governance and models of legal publishing.19

This constitutes an example of the mixed public-private business models that will proliferate in the new Web of Data scenarios [50], changing top-down and exclusively market-based approaches into more relational and flexible ways of handling regulations, services, and rights [21].


Setting aside certain differences between civil- and common-law countries, we take public administration to be a general concept mainly referring to the public organisation and dynamic management and implementation of statutes, policies, and court decisions within a legal system within a local, regional, national, or international jurisdictional space. It thus encompasses the work carried out by professional staff in parliaments, governments, and courts. In the past few years, several countries have started to publish administrative data online as open data using open formats such as RDF/XML [40], Turtle [102], and RDFa [1].

For instance, since 2009, the US Government has started20 a process for creating and releasing governance data so as to enable public access to datasets generated by the executive branch of the federal government21 (195,033 datasets as of February 26, 2016) [59].

On the basis of that experience, the UK government’s project data.gov.uk22 [108] was launched in January 2010 with the aim of making datasets containing data from several UK government departments (27,909 as of February 26, 2016) freely available. All data are non-personal (for privacy reasons) and, in principle, available in several data formats. Each page describes a dataset and presents information about the formats used, its openness, its themes, and the timespan covered by the dataset itself. The final goal of this project is twofold. On the one hand, it aims to make governmental data freely accessible online to the public. On the other hand, it seeks to integrate such data within Open Linked Data [108].

Similarly, the London Gazette,23 which is the journal of official public record of the British government and has been keeping a record of British public life for almost 350 years, publishes all its materials as PDF files and HTML+RDFa [117] pages.24 The main part of semantic assertions described through RDFa conforms to the Gazette Ontology,25 which defines all classes and properties used for all Gazette Notices.

A sister project of the previous one, i.e., legislation.gov.uk,26 managed by the National Archives27 on behalf of the UK government,28 has recently released all UK legislation from 1267 to the present in several formats, among which RDF/XML (including Akoma Ntoso [7] [8], as we will explain shortly). The National Archives is the official archive and publisher for the UK government, and this portal is the most important official digital repository of legislation in UK.

These kinds of initiatives by the aforementioned and other local, regional, and national governments (Catalonia,29 France,30 Italy,31 the Netherlands,32 Singapore,33 etc.) show how the adoption of Semantic Web technologies in the administrative domain is widespread even for concrete uses and applications. The European Data portal34 recently launched the DCAT-AP version 1.1 (Application Profile for Data Portals in Europe)35 metadata vocabulary, a common schema for harmonising descriptions about public sector datasets in Europe, reusing, among others, Eurovoc36 thesauri now available as OWL ontologies.

23 London Gazette homepage: http://www.london-gazette.co.uk. The London Gazette is published by TSO (The Stationery Office) under the superintendence of Her Majesty’s Stationery Office (HMSO), part of the National Archives.
24 An example of an HTML+RDFa page in the London Gazette is available at https://www.thegazette.co.uk/notice/L-58664-497223.
25 The Gazette Ontology: https://www.thegazette.co.uk/def/publication#.
27 The National Archives homepage: http://www.nationalarchives.gov.uk.
29 Open public data of the government of Catalonia: http://dadesobertes.gencat.cat.
31 Italian public administration open data: http://www.dati.gov.it.
32 Dutch national open data platform: https://data.overheid.nl.
34 http://www.europandataportal.eu/.
At the European level, EUR-Lex and the Publication Office has also made great strides during the last three years.37 There is a freely accessible EU Metadata Registry.38 CELLAR, the new EU portal, serves on average some 5 million requests per day. It contains 200 million identifiers and 1100 million triples in its Oracle RDF store.39

The Re-use of Public Sector Information Directive,40 updated in 2013 (also known as the PSI Directive) sets out the conditions governing the right to reuse information resources held by public sector bodies, including provisions on non-discrimination, transparent licensing, and public accountability. This is fostering the creation of value-added information products and services (tools, apps, content) that take public sector information or data as a source (or even as their single source).41 Very recently, after a decade of promoting Open Data and PSI reuse in Europe, the ePSI Platform (created in 2006)42 shut down at the end of February 2016 and migrated to the new European Data Portal.43 In 2013, the G8 started the Open Data Charter,44 which also includes legal information as a pillar for enhancing law.

For this reason the Open Government Data movement is no longer limited to government organizations but extends its reach to other public bodies, especially to parliamentary bodies (e.g., the Italian Parliament45 and the US House of Representatives,46 the Open Government Partnership,47 and the National Library of Congress of Chile48), and the judiciary (the Open Justice Project in the UK,49 US Open Data Justice,50 the California Initiative,51 the European Court of Justice52). Moreover, thanks to the PSI Directive, the emphasis is also placed on public documents, and not only on public data, enriched with RDF metadata.

4. Semantic Web and the Law

The rise of the Web of Data, the increase of legal data, and second-generation Semantic Web applications open new scenarios for legal modelling. First-generation Semantic Web applications focused on corporate ontologies, using a single—lightweight or thick—ontology to support the integration of resources selected in a controlled environment.

The second generation instead faces a large-scale collection of distributed semantic metadata and exploits the Semantic Web as a large-scale, heterogeneous, distributed semantic resource [35]. This is changing the shape and structure of regulations and the implementation of the law—the way legal content can be accessed, circulated, used, changed, implemented, and enforced.

Early accounts of Law and the Semantic Web stressed the work done on ontologies as early as the 1990s and in the first stages of the Web [10]. Several national and EU Esprit, FP5, FP6, and FP7 projects funded the construction or refinement of upper (foundational), core, and domain and linguistic ontologies [26]. Among them are FOLaw (Functional Ontology of Law), based on normative knowledge, world knowledge, responsibility knowledge, reactive knowledge and creative knowledge [118][119]; FBO (Frame-Based Ontology of Law), based on norms, acts, and descriptions of concepts [122][123]; LRI-Core Legal Ontology, based on objects, processes, physical entities, mental entities, agents, and communicative acts [16][17]; CLO-Core Legal Ontology [41][43], based on the foundational ontology DOLCE+[2], and extended to a lexical ontology in the LOIS Project (JurWordNet) [94][42]; OPJK (Ontology of Professional Judicial Knowledge), designed to model the practical knowledge contained, in not legal documents, but in judges’ experience [29][25]; DALOS-Consumer Protection Ontology, grounded in the results of the LOIS Project and importing top ontology concepts from CLO [2][3]; Legivoc,54 a Ministry of France project conducted with the support of the European Commission.

The methodological stages of knowledge representation, the iterative lifecycle of legal ontology building, and lessons learned have been carefully explained and discussed in several monographs [62][14][27][89]. There are five dimensions of law that have been addressed through computational modelling: the structure of legal documents, norms and normative systems, concepts and legal conceptions, cases and precedents, argumentation and legal reasoning [107]. They were initially related to the development of domain-independent ontologies for knowledge-sharing and reuse, mainly as knowledge-interchange formats for knowledge-based systems. Early legal ontologies were formalised using ONTOLINGUA, LOOM, and DAML+OIL, with an increasing preference for W3C standards, Semantic Web languages (OWL DL), editors such as Protégé, and reasoners such as Pellet [18].

Scalability, reusability, and end user-centred approaches were taken into account to model specific legal domains, mainly e-commerce, e-administration, e-governance, criminal law, consumer law, mediation, drafting, and contracting. The increasing weight of Semantic Web languages has helped to

39 Fulgencio Sanmartín, personal communication.
42 http://www.epsiplatform.eu/.
43 http://www.europeanadataportal.eu/.
45 Italian Senate (http://dati.senato.it/) and Chamber of Deputies of Italy, and the project Normattiva: http://www.normattiva.it/.
47 http://www.opengovpartnership.org/groups/legislative.
49 http://open.justice.gov.uk/.
51 https://open.justice.doj.ca.gov/.
54 http://legivoc.org/.
shape a more pragmatic approach, envisaging Web services, community participation, and a growing involvement of legal experts, especially in knowledge-acquisition, validation, and implementation processes.

Thus, the way of modelling legal knowledge can be quite flexible, depending on the epistemic assumptions in the knowledge acquisition process, the selected requirements, and the final purposes of the tool, application, or system at stake. The next step is to lean on lessons learned over the past fifteen years.

In early 2005, a project funded by the United Nations, Akoma Ntoso, was launched to enable citizens to exercise their right to access African parliamentary proceedings and deliberations, while supporting Parliaments in managing legislative documentation [125][8][87][82]. This work was also immediately followed by an on-going effort to standardise its result within three OASIS technical committees [6].

Akoma Ntoso did not start in a vacuum. Almost in parallel, MetaLex emerged in 2002 from several past EU projects and was proposed as a legal XML standard. In 2006, it evolved into CEN-MetaLex as a general format for the exchange and interoperability of legal documents [14]. It adopts and adapts the concepts introduced in the FRBR specification [Functional Requirements for Bibliographic Records of the International Federation of Library Associations]. This work had a follow-up in the EU Project ESTRELLA, in which the ontology for Legal Knowledge Interchange Format-LKIF was finally created [62][18][63]. LKIF was used in several projects to model deeper legal concepts [86] and represent legal documents. Several ontologies were developed on top of the work done in CEN-MetaLex and on Akoma Ntoso: an ontology for managing a legislative text’s evolution over time and its linguistic variants [78][1]; an ontology for modifying modifications of norms [82][84]; an ontology for modelling relations between authorities, agents, and roles in the process of producing a legal document [7].

5. The Web of Data and the Law: A Step Forward

A quick look at the 2015 Legal Technology Survey Report of the ABA Legal Technology Resource Center shows the growing implications of legal professionals as a data-driven community. Trending topics are practice-management software and the nascent virtual law practice (although only 5% of lawyers in the overall sample locate themselves under this label). Here is a vignette of the “virtual lawyer” as portrayed by Joshua Poje, director of the American Bar Association Law Practice Division [99]:

“Imagine a lawyer stranded at the airport. Her flight home to visit family has been delayed, as so often happens. She’s debating how to fill her time when her Apple Watch issues a gentle tap on her wrist. It’s an alert about an email from an important client—her contract and add explanatory comments. Legal concepts are not understood. Legal concepts are not the suspicious public

With time to spare, she pulls out her iPad and jumps online using her own data plan—not the suspicious public WiFi. She quickly reviews the email, which details the client’s issues with a recently hired employee. Seconds later she has an employment contract she drafted open on the iPad, accessed from her secure cloud-based document management system.

“She uses a simple annotation app on the iPad to highlight key language in the contract and add explanatory comments. She then exports the annotated contract to her email app, types a brief memo explaining her thoughts on the matter, and sends it off to the client (with a CC: to her practice management system so the email is filed and flagged for later time entry).”

“Throughout this brief interlude, the lawyer has never left her seat in the airport terminal, and she’s used nothing more than her iPad and her Apple Watch. Her client need never know. From his perspective, she may as well have been behind a vast mahogany desk aided by a small army of legal assistants. The quality of service is the same.”

This is not an uncommon situation. The picture could be completed by adding some utilities: semantic contracts, automated structuring of the content of contracts, personal access to large databases, smart tools for reorganising the relevant legal knowledge. This is not so far off now. Shared information from heterogeneous systems, personalisation, context awareness, and interaction are features of the Semantic Web. Consumer-generated data is growing at an unprecedented pace—a 2.000% increase in global data is expected by 2020 [28] and companies, lawyers, and law firms are quite aware of this.

New emerging issues are also important in the development of markets and the evolving entwinement of law and the Web of Data. Trust, transparency, metadata markets, licensing, statistical and big-data governance, privacy, data protection, security, and intangible legal goods such as intellectual property and patents are fostering new modes of cooperation between legal experts and Semantic Web developers.

Ontology-driven systems with reasoning capabilities in the legal field are now better understood. Legal concepts are not discrete but make up a dynamic continuum between common sense terms, specific technical uses, and professional knowledge in an evolving institutional reality [120]. Thus, the tension between a plural understanding of regulations and a more general understanding of law is bringing into view a new landscape in which general legal frameworks—grounded in well-known legal theories stemming from 20th-century legal positivism or sociological jurisprudence—are made compatible with specific forms of rights management on the Web. In this sense, Semantic Web tools are not only being designed for information retrieval, classification, clustering, and knowledge management. They can also be understood as regulatory tools, i.e., as components of the contemporary legal architecture, to be used by multiple stakeholders—front-line practitioners, policymakers, legal drafters, companies, market agents, and citizens.

This has given a new boost to Digital Rights Management (DRM) [44], Rights Expression Languages (REL) [105], and Open Digital Rights Languages (ODR) [66]. They are quite diverse and may not have the same legal status.

58 https://www.w3.org/community/pua/wiki/Digital_Rights_Management.
59 https://www.w3.org/community/odrl/.
Depending on national and international jurisdictions. But all three initiatives are market, governance, and policy regulators, and on top of them SW languages can be used to generate the automated selection and aggregation of relevant information to perform legal acts. Information from disparate and heterogeneous sources often produces incompatible or contradictory results, and these kinds of problems need to be tackled in advance [53]. For instance, Natural Language Processing and the representation of norms in RuleML can be used to perform business contracts; Ontology Design Patterns (ODP) make it possible to reuse core-conceptual structures in linked data licensing [106]. At stake are rule [52][55][28] and legal compliance (by design) [30], trust [5], and the specification of standards (such as fairness or transparency). This is leading to new approaches to ethics [37] and law [81], and new research on the stages and forms of the rule of law for the Web of Data [22][23].

Consider, too, that consensus and disagreement are equally present both on the Web and in the actual law-making processes. Recent path analyses of the evolution of intellectual property rights show that international patent systems were still under construction moving into 2000 [73]. This fact changed dramatically in the 21st century, owing in part to a new extremely competitive framework, nationally and internationally alike, and to the economic success of the Web. The launching of the Creative Commons in 2001, Online Dispute Resolution (ODR) systems [71], crisis-and-disaster management platforms (such as USHAIDI), and crowd-sourced systems for constitutional and legal drafting purposes incorporate new kinds of social regulations based on participation, Open Source Intelligence (OSINT), and Social Intelligence (SI) [97]. Different private and public, individual and collective, interests have to be balanced and harmonised in a transparent and accountable manner, especially when security and data-protection principles need to be implemented to protect citizens and citizens’ rights at the same time. This objective can be advanced by forming synergies between Agreement Technologies, Normative Multi-Agent Systems (NorMAS) [4], and the Semantic Web.

6. Standardisation Trends

There is an important standardisation effort in the interplay between Web languages and the Semantic Web and legal knowledge. It is true that most behaviours on the Web seem to be implicit. Best practices and norms are not often made explicit [100]. However, even recognizing that standards and technologies make sense only to the extent that they are deployed, and that many de facto standards emerged not from standardisation bodies but bottom-up, there is huge room for improvement and rationalisation in the interface of law and the Semantic Web. Supporting such a rationale for the law actually triggers and fosters the development of the Web. Recent surveys on the quality of XML documents on the Web show that 85.4% (n=180 K) are well-formed and 99.5% of all specified encodings is correct [54]. Even so, much of the Linked Open Data available on the Web is far from the five-star level [9], and about 40% of linked datasets have been published without a license [124], i.e., without appropriate legal coverage.

Alternatives—common practices in the Web as against formal standards—are not necessarily antagonistic. Industry demands that common languages, formats, and specifications that are already in use receive the imprimatur of standardisation bodies to encourage fluent exchanges of data and digital goods. This demand is slowly shifting from XML formats to RDF models and Semantic Web technologies. One example of this is what has taken place within the MPEG-21 ecosystem, which initially had space for specifying XML schemata but more recently turned to OWL ontologies, as illustrated in this Special Issue [73].

Much of this shift can be attributed to trends in the technology world, but some can be grounded in the fact that the overall focus when dealing with legal data is shifting from a need to publish documents (on paper and online) to an effort to produce complex applications providing some kind of legal reasoning. This is certainly risky, because imposing a textual or positivistic approach on all countries, cultures, and political bodies of different natures should be avoided. The challenge is to balance the use of technical standards with the emerging patterns of social and political behaviour.

Standards for legal documents are reacting accordingly. While the first and second generations of standards for legal and legislative documents were mostly national, and mostly aimed at generating printed and online representation of legislative texts, current standards are looking to reach agreements on a more general framework.

Given that text documents are still very much the core material produced by legal professionals, and that references to text documents will remain the basis for grounded and verifiable legal reasoning regardless of the actors and technologies employed, current generation standards in the legal domain are providing a layered organization of their offerings: presentation-oriented XML is being replaced with structured XML, with ample room for metadata and annotations; naming mechanisms based on URIs and IRIs provide linkable anchors both to entire documents and to smaller fragments; and document-oriented ontologies provide the necessary glue between abstract legal reasoning and the textual pieces of supporting evidence.

Within OASIS, for instance, three Technical Committees are actively working to produce a flexible, comprehensive, and wide-reaching platform for the digital representation of legal documents and their content: the LegalDocML TC has inherited the Akoma Ntoso XML vocabulary from the early UN-based African initiative and is extending its expressive power to support the structuring of legal and legislative documents, making sure that its constructs, including metadata, are immediately expressible as RDF statements and amenable for evaluation in Semantic Web applications.

The LegalCiteM TC is providing digital representation of legal citations fostering a convergence of many existing syntaxes for legal and legislative references, including ELI, ECLI, URN-LEX, and the Akoma Ntoso Naming

---

64 https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legalxml-legislative
65 https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legalitem
68 https://tools.ietf.org/html/draft-spinosa-urn-lex-09
69 https://creativecommons.org/
60 https://creativecommons.org/
61 https://www.usahidi.com/
62 http://constitutionlab.org/
63 http://www.sintelnet.eu/
Convention,\(^6^9\) making sure that Linked Data collections of legal facts and data can refer back to the specific parts of the legal and legislative documents where they come from.

Finally, LegalRuleML TC\(^7^0\) \(^5^9^1\) is providing a framework for expressing in a formal language the constitutive and prescriptive rules and constraints of which norms and regulative texts are made. LegalRuleML also provides an RDFS meta-model for the deontic and defeasible logic operators applied in the legal domain in order to export the metadata in RDF. But this is all still under development, and must be carefully validated and tested through an on-going implementation process in different scenarios. Policies, social regulations, ethics, and law are deeply intertwined. Thus, Semantic Web technologies can also be applied on specific service-oriented approaches, focusing on social problems such as the administrative implementation of immigration laws and policies in a specific jurisdiction. In this way, methodology, ontology-building, and epistemology can be kept in separate clusters, and different dimensions can be combined ad hoc to tailor specific solutions. Legal isomorphism and the so-called scooping problem \(^1^2^1\)—the extraction of implicit meaning from general regulations with concrete aims and targets in a given context—can be tackled in an ordered and relational manner, making it possible to create scenarios with lay and expert participation alike \(^1^2^2\)[113].

So, too, new general frameworks can be added to an emerging contemporary legal landscape for the Web of Data. Customary international private law cannot be easily modelled without taking all stakeholders into account. The general balance between privacy, data protection, and security \(^5^8\) seems to broaden the legal normative scope for regulating, among other elements, linked data markets \(^1^1^5\), co-regulatory instruments \(^9^8\), self-regulated collective awareness and informed consent \(^8^8\), the behaviour of LEAs behaviour (law enforcement agents) \(^2\) \(^4\), and the use of multi-lingual and multi-jurisdictional term banks \(^1^0^3\). Distributed geospatial data, textual data, and controlled vocabularies can be combined to create interactive tools to enhance the rule of law and the specific legal information that citizens would need to perform legal acts \(^6^4\). Visualisation tools can enhance citizens’ confidence and trust, saving time and effort \(^9^6\). Legal Open Data can be elevated to the level of Linked Open Data by transforming texts into vectors, selecting suitable terms, and using a weighting function as part of the frequency calculation, increasing the degree of public accessibility \(^1^3^\).[2]

These new trends are still to be explored further, but they certainly suggest the promise of hybrid models of regulation and synergy among (and in between) legislation, Court decisions, the economy, policy, and Semantic Web technologies.

7. **Overview of This Special Issue**

This special issue on the use of Semantic Web technologies to address Legal Domain issues and scenarios brings together five high-quality contributions, out of eight submissions we originally received.

In “Semantic Model for Legal Resources: Annotation and Reasoning over Normative Provisions” \(^3^9^1\), Enrico Francesconi presents an OWL 2 DL ontology for describing normative provisions (in terms of Hohfeldian legal fundamental relations) and related axioms in order to enable advanced access to legislative documents. The discussion is supported by examples of semantic annotations of legal textual resources using RDF/OWL standards and of SPARQL queries for accessing and reasoning over provisions. This is framed on CELLAR.

In “An OWL Ontology Library Representing Judicial Interpretations” \(^3^1\), Marcello Ceci and Aldo Gangemi introduce an OWL 2 DL ontology library making it possible to describe the interpretations a judge makes of the law while engaging in the legal reasoning on which basis a case is adjudicated. This ontology library is based on a theoretical model and on some specific patterns that exploit some new features introduced by OWL 2, and it provides meaningful legal semantics, while retaining a strong connection to source documents (i.e., fragments of legal texts).

In “LOTED2: An Ontology of European Public Procurement Notices” \(^3^6^\) Isabella Distinto, Mathieu d’Aquin, and Enrico Motta describe the construction of the LOTED2 ontology for the representation of European public procurement notices. LOTED2 is a legal ontology that supports the identification of legal concepts and, more generally, of legal reasoning. In particular, it seeks to strike a compromise between the accurate representation of legal concepts and the usability of the ontology as a knowledge model for Semantic Web applications, while creating connections to other relevant ontologies in the domain.

In “PPROC, an Ontology for Transparency in Public Procurement” \(^7^7\), Jose Félix Muñoz-Soro, Guillermo Esteban, Oscar Corcho, and Francisco Serón introduce the PPROC Ontology—an ontology that enables the description of procurement processes and contracts. The authors focus in particular on making the ontology appropriate for describing the standard data relating to the tender (i.e., objectives, deadlines, awardees) and the details of the whole process of publishing and performing contracts.

In “Overview of the MPEG-21 Media Contract Ontology” \(^1^0^4^\) Víctor Rodríguez-Doncel, Jaime Delgado, Silvia Llorente, Eva Rodríguez, and Laurent Boch present the MPEG-21 Media Contract Ontology (MCO), an ontology enabling the description of contracts dealing with rights to multimedia assets and, more generally, with any content protected by intellectual property. The ontology is composed of a core model (describing permissions, obligations, and prohibitions in contracts) and a specific vocabulary representing the common rights and constraints in the audio-visual context. The paper also includes a description of the design principles and the methodology followed in developing the ontology, as well as several examples of it in RDF and a description of related tools.

8. **Acknowledgements**

We would like to thank all the authors of accepted and rejected articles, as well as the editors-in-chief of Semantic Web Journal, Pascal Hitzler and Krzysztof Janowicz, for all their support and help. A heartfelt thanks goes to all the people who have made this high-quality special issue possible. This is devoted in particular to our reviewers (in alphabetical order): Ginele Barabucci, Nick Bassiliades, Eva Blomqvist, Guido Boella, Thomas Bruce, Nuria Casellas, Paolo Ciccarese, Jaime Delgado, Stefan Dietze, Angelo Di Iorio, Miriam Fernández (two articles), Meritxell Fernández Barrera, Enrico Francesconi, Aldo Gangemi, Jorge Gracia,
Guido Governatori, Laura Hollink, Renato Iannella, Andrew Koster, José Manuel López-Cobo, Adrian Paschke, Ginevra Peruginelli, Axel Polleres, Marta Sabou, Marco Schorlemmer, Daniela Tiscornia, Serena Villata, John Zeleznikow, and Jun Zhao.

We also warmly thank Graham Greenleaf, Tom Bruce, Fulgencio Sanmartín, and Enrico Francesconi for providing us useful updated information about AustLII, Cornell’s LII, and CELLAR.

A special mention needs to go to Rinke Hoekstra: not only has he given us continuous support and encouragement, but he has also managed to review three of the articles.

The edition of this special issue on the Semantic Web for the Legal Domain has been partly supported by the national project DER2012-39492-C02-01-CRDWSOURCING, the EU projects CAPER-CA-261712 and 520250-1-2011-1-IT-ERA MUNDUS-EMJD, and the Australian Project Data to Decisions (D2D) Cooperative Research Centre (CRC).

References


