The Role of Pragmatics in the Web of Data

in: Francesca Poggi, Alessandro Capone (Eds.) ***Pragmatics and Law. Practical and Theoretical Approaches***, Berlin, Heidelberg: Springer Verlag, 2016 (forthcoming)

Springer link: http://www.springer.com/series/11797?detailsPage=titles

Pompeu Casanovas[1], Víctor Rodríguez-Doncel[2], Jorge González-Conejero[3]

[1,3] UAB Institute of Law and Technology, Autonomous University of Barcelona (UAB) , Barcelona, Spain
[2] Department of Artificial Intelligence, Polytechnic University of Madrid (UPM), Spain

Abstract: This Chapter is an introduction to the Semantic Web, the Web of Data, regulatory models, and the law. It does not take anything for granted. The first part of the Chapter describes the languages of the Semantic Web, and shows how the perspective of the Web of Services and Linked Data is related to the conditions under which services can be offered, managed and used. The Web has been massively populated with both data and services. Semantically structured data, the Linked Data Cloud, allows and fosters human-machine interaction. Linked Data aims at creating ecosystems to facilitate the browsing, discovery, exploitation and reuse of datasets for applications. Licensed Linked Data is offered along with information about the rights involved. Rights Expression Languages are able to regulate half-automatically the use and reuse of content.

The second part of the Chapter shows that the nature of law is experiencing a deep transformation in the cloud. What links the information flow, social intelligence, rights management, and modelling in the Web of Data is the pragmatic approach —what we call the *pragmatic turn.* I.e. the representation of users' needs and contexts to facilitate the automated interactive and collective management of knowledge. Both ontology building and knowledge acquisition share this perspective. The Web of Data brings about new challenges on agency, knowledge, communication, and the coordination of actions. Institutions can regulate both human and machine behaviours within these new environments. *Licensed Linked Data, Licensed Linguistic Linked Data, Right Expression Languages, Semantic Web Regulatory Models*, *Electronic Institutions*, *Artificial Socio-cognitive Systems* are examples of regulatory and institutional design (Regulations by Design). In the cloud, regulatory systems become more complex, in order to be simpler.

Keywords: Semantic Web, Pragmatics, Web of Data, Licensed Linked Data (LLD), Right Expression Languages (REL), Semantic Web Regulatory Models, Artificial Socio-Cognitive Systems

1 Introduction

According to (Kasper and Rose 1999) pragmatics has two roles in socio-linguistic analysis, (i) it acts as a constraint on linguistic forms and their acquisition (representing a type of communicative knowledge); and (ii) on a par with morphosyntax, lexis, and phonology it focuses on learners' knowledge, use, and acquisition of pragmatics in a second language (interlanguage pragmatics).

However, if pragmatics is seen through the prism of human-machine interaction considering the cognitive and computational perspectives, we may go a bit further. Machine-languages (models) are a kind of second language, too. We will use functional and cognitive pragmatics as a part of a broader methodology. In this sense, stemming from technology and law, a third role comes into play, as pragmatics is able to bridge common and expert knowledge into iterative cycles of legal practical analysis and computational modelling.

First, we will focus on the Web of Data, setting the contextual requirements for such an aim — not in vain, the Web is the space where most human-machine communication acts take place. In particular, we will analyse the Linked Data cloud, as a semantised subset of the Web of Data. Second, we will describe the crucial role of pragmatics to help debugging the so-called "*knowledge acquisition bottleneck*" (Feigenbaum, 1980, 1992). It is our contention that such an approach can better define the situated knowledge that is needed to set platforms and build up legal ontologies, functionalities, applications and mash-ups within social ecosystems. This constitutes a particular theoretical trend. Third, we will state that this kind of bridge requires the construction of intermediate institutions —*Semantic Web Regulatory Models* (SWRM)— to anchor the reuse of ontologies, datasets, and general knowledge into specific contextual legal settings. At present, there are coordination and cohesion problems between the regulatory instruments —law, policies, standards, and ethical principles— and the scenarios set forth within the Web of Data. Fourth, we will concentrate on licenses for Linked Data on the Web. The lawful consumption of online datasets across different jurisdictions constitutes a good example of regulatory problem, and the mere idea of having Licensed Linked Data (data semantically described with a machine-readable rights expression) deserves our analysis. Fifth, we will discuss the use of autonomous agents and Electronic Institutions to behave and to perform legal acts in this new environment. In the end, we will introduce some contextual and anchorage problems of Artificial Socio-cognitive Systems (ASCS).

At the implementation level, the reuse of knowledge patterns on a Web of Linked Data needs to be much more specific about design conditions, strategies and guidelines. This is what Hoekstra (2010) has called "*the knowledge reengineering bottleneck*". We think that this chapter can contribute to clarify and partially solve the cost of control in the new scenarios raised by the Web of Data.

2 The Web of Services and Linked Data: a broad landscape

The Web is not only the place where the most exciting information lives, but also a scenario where things *happen*. Besides passive information, actions can be triggered, either having a physical effect in the real world or a purely informational effect. The regular provision of these actions is organised through *services*, which are invoked by humans and machine alike. One informational service may offer the current temperature in Barcelona, another service may place a purchase order in the stock market.

In the last few years the Web has been massively populated with both data and services. Data has been offered by public and private organizations, by sensing devices (Internet of Things) and by the own internet users in social networking sites and other user generated content. A small fraction of this Web of Data is the so called *Linked Data cloud*, with special interest for its semantic foundation. The Linked Data cloud is perhaps the most visible result of the Semantic Web (SW) vision, an endeavour aimed at providing structure to the Web so that machines can better understand it. This section analyses the Service Web, the Linked Data cloud and a further fragment of it: the *Licensed* Linked Data (LLD), namely, Linked Data that is offered along with rich semantic rights information.

2.1 The Web of Data

2.1.1 The Semantic Web

The World Wide Web is estimated to contain information in the order of zettabytes, but raw data is still a small share of it: many Web pages are built based on databases which are not directly accessible, and most information is in an unstructured form difficult to parse by computers. The inventor of the World Wide Web, Tim Berners-Lee, conceived the idea of connecting all the data on the Web, establishing relationships between data published by different entities. Information would then build the Giant Global Graph, a unique global database, a *Web of Data*.

For this information to be comparable, data on the Web had to be structured and intelligible for computers. He was thinking of the *Semantic Web*; in his own words: "*The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines.*" (Berners-Lee 1999: 191)

This vision would open up vast amounts of data to be mined by Artificial Intelligence processes, enabling a smarter use of the resources, increasing the productivity of the society and improving the understanding of our world. Public institutions, companies and individuals were invited to dump their data in open standard formats, to build the Web of Data as the WWW had been.

These open standard formats would eventually be published by the W3C, the guardian of the WWW essences. One of the grounds of the Semantic Web was RDF (*Resource Description Framework).* RDF is a standard model for data interchange on the Web that organises information as small units called *RDF triples*: one subject, one predicate and one object. Each of the elements in this triple can be related to elements in other triples, weaving a web of relationship –an RDF graph.

By the time RDF appeared, XML was at its peak. XML is a good serialisation format, namely, a good way of encoding information so that two computers can understand it. But these two computers must have agreed on the structure of the XML document. There is no XML format, but many XML formats. Different computer programs use different XML structures and XML merely defines a syntax. Grammars can be defined with other standards in the XML family, like DTD or XML Schema. Semantics are outside the realm of the XML specification. However, semantic interoperability with RDF is much easier. In RDF all objects are independent entities and mapping between two RDF descriptions are straightforward. So even if two organizations want to exchange data in different RDF models, aligning the models (either manually or automatically) is a much easier task than with XML.

The way of identifying objects in the Semantic Web is using URIs (Uniform Resource Identifier). While XML grammars never aimed at being universal, RDF willingly accepts that two different modellers referring to the same concept use the same identifier.

Computer scientists like enumerating two alternatives to declaring semantics: *procedural* and *declarative*. In the procedural approach, the meaning of an expression is given by describing the effect that some procedure (e.g. a computer program) will have on it. In the declarative approach, the meaning of an expression is given by describing its properties or the conclusions that can be derived from it. The latter approach (which is the one followed by RDF) is considered to be a more shareable and extensible form of declaring semantics.

Further, the RDF models can be supported by OWL ontologies (Ontology Web Language) [1]. OWL facilitates reasoning (a simple deductive reasoning) through sets and properties that model formally concepts, relationships and instances. The result drives to information management and processing as *knowledge —hypertext* links, connection of objects, and information retrieval from the Web using not keywords (terms), but concepts. RDF (hence OWL) can be serialized as XML, but also as Turtle or N3; they are different encodings of the same model.

We have described URIs, XML, RDF and OWL. They are the key components of the Semantic Web technologies[2]. At least, they were in the Semantic Web stack of technologies as initially proposed. Table 1 summarizes these technologies. It distinguishes between Berners-Lee's hypertext original ones, and those adding structure and meaning to the information, transforming it through *metadata* from "tagged" or "marked" data into a set of linked objects of knowledge. The use of XML and related technologies (XML Schema, XPATH) to publish linked data is declining fast in favour of the simpler JSON (JavaScript Object Notation) structures, but the overall ideas still remain valid several years after the publication of the SW technologies stack.

Table 1. Semantic Web Technologies.

	Hypertext technologies
Web 1.0/2.0	IRI [*Internationalised Resource Identifier*], generalization of URI [*Uniform Resource Identifier*], it facilitates the identification of semantic resources to manipulate and handle them.

[1] For a technical introduction, cfr. Antoniou and van Harmelen (2004, 2012), cfr. also Domingue et al. (2011).

[2] See the last version of the famous SW 'cake' or stake of languages by T. Berners-Lee at http://www.w3.org/2007/03/layerCake.png

Web 1.0/2.0	*Unicode* is a codification standard of features of multiple technical and natural languages (including the ancient ones). It drives to uniformity, unicity and universality of their representation.
Web 1.0/2.0	XML [*Extensible Mark-up Language*] is a tag or mark-up language (a meta-language in fact) defining a set of documents in a readable format language, either by humans or by machines.[3]
Web 1.0/2.0	*XML Schema* is a schema language that constraints the structure and content of XML documents, adding more abstraction into it.[4]
Web 1.0/2.0	*XML Namespaces* furnish elements and attributes with a single name to a XML sequence; this sequence may contain element names or attributes coming from a more than one vocabulary. It is used to identify and single out entities to be referred to without any ambiguity.[5]
	Standard Semantic Web Technologies (meaning for structured data)
Web 3.0	*Resource Description Framework* [RDF] is a knowledge representation language to create sentences as triples. It makes their conversion into expressions with the form subject-predicate-property (resource-property/relationship-property/property value-relationship) and facilitates its representation as a graph.[6]
Web 3.0	*Resource Description Framework Schema* [RDFS] extends RDF vocabulary allowing the description of taxonomies of classes and properties. It also sets the domain and range of properties and relates the RDF classes and properties to taxonomies (hierarchies).[7]
Web 3.0	OWL [*Web Ontology Language*] adds more structures to describe the semantics of RDF sentences (cardinality, restrictions of values, transitivity….). It is based on descriptive logics, and provides reasoning capabilities to the SW. It defines sets, properties, instances and operations through the construction of *ontologies*.[8]
Web 3.0	SPARQL [***S****PARQL* ***P****rotocol* ***a****nd* ***R****DF* ***Q****uery* ***L****anguage*] allows the search of structured data (in RDF, RDFS and OWL). It is a W3C recommended search language specifically devised to retrieve and manage data stored in RDF, to perform graph queries, and to build up SW applications.[9]

2.1.2 Linked Data

The Semantic Web ideas have been embraced relatively slowly. More than a decade had elapsed and these technologies did not abandon the academic circles, not being embraced by the crowd —like other technologies like HTML or web programming. However, in recent years the Semantic Web has been incarnated under the form of Linked Data, and RDF is at last being adopted as a data publication model.

Linked Data still represents a negligible fraction of the information in the Web, but it attracts our attention for three reasons: (i) it is highly *semantised* data, (ii) it is growing at a very fast pace, (iii) and it is supported by the W3C recommendations described before.

Linked data might be defined as data published in a structured manner, in such a way that information can be found, gathered, classified, and enriched using SW annotation and query languages. It is almost a commonplace by now to quote the Linked Data principles laid down by

[3] http://www.w3.org/XML/
[4] http://www.w3.org/XML/Schema
[5] http://www.w3schools.com/XML/xml_namespaces.asp
[6] http://www.w3.org/RDF/
[7] http://www.w3.org/TR/rdf-schema/
[8] http://www.w3.org/TR/owl-guide/
[9] http://www.w3.org/TR/rdf-sparql-query/

Tim Berners-Lee in a famous speech note at W3C in July 2006: (i) Use URIs (Uniform Resource Identifiers) as names for things, (ii) use dereferenceable[10] HTTP URIs, so that people can "look up" those, (iii) when someone looks up a URI, provide useful information, using the web standards like RDF or SPARQL (SPARQL Protocol and RDF Query Language), (iv) include links to other URIs, so that they can discover more things. We will go back in a moment to what these languages mean and how they are related (see Table 2).

It is worthwhile to notice that Linked Data principles are described therefore from a pre-normative point of view, as they consist of a set of *best practices* for publishing and connecting structured data on the Web. We should distinguish the process from the result, but both are embedded into an ongoing and interactive design planning, massively followed by a great plurality of users using a great plurality of distributed and diverse sources:

"*In summary, Linked Data is simply about using the Web to create typed links between data from different sources. These may be as diverse as databases maintained by two organisations in different geographical locations, or simply heterogeneous systems within one organisation that, historically, have not easily interoperated at the data level. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets*." (Bizer et al. 2009: 256)

Independent Linked Data producers publish their datasets as collections of RDF. These RDF datasets are mutually interlinked —or at least encouraged to be so. There is an increasing number of linked datasets that are available on the Web, growing very fast. Heterogeneous data publishers participate in this race, including individuals, private and public institutions. If we represent datasets as nodes, and arrows as links between the nodes, we may represent the connectivity between datasets in a graph-like diagram. This diagram changes every day — datasets are not necessarily static, and an analogy has been made calling this diagram the Linked Data Cloud.

Considering the best quality datasets, Jentzsch, Cyganiak, and Bizer used the descriptions of the data sets in this group to generate the Linked Open Data Cloud diagram at regular intervals. The latest of these diagrams[11] was made by Schmachtenberg (Schmachtenberg et al. 2014), and it is shown (slightly modified) in Figure 1.

The size of the circles in Figure 2 indicates their valency, the degree or number of edges incident to the vertex. Thick arrows indicate a greater degree of linkage. Bidirectional arrows point at the coexistence of elements.

[10] "The act of retrieving a representation of a resource identified by a URI is known as *dereferencing* that URI. Applications, such as browsers, render the retrieved representation so that it can be perceived by a user. Most Web users do not distinguish between a resource and the rendered representation they receive by accessing it" (Lewis 2007).

[11] http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/LODCloudDiagram.html

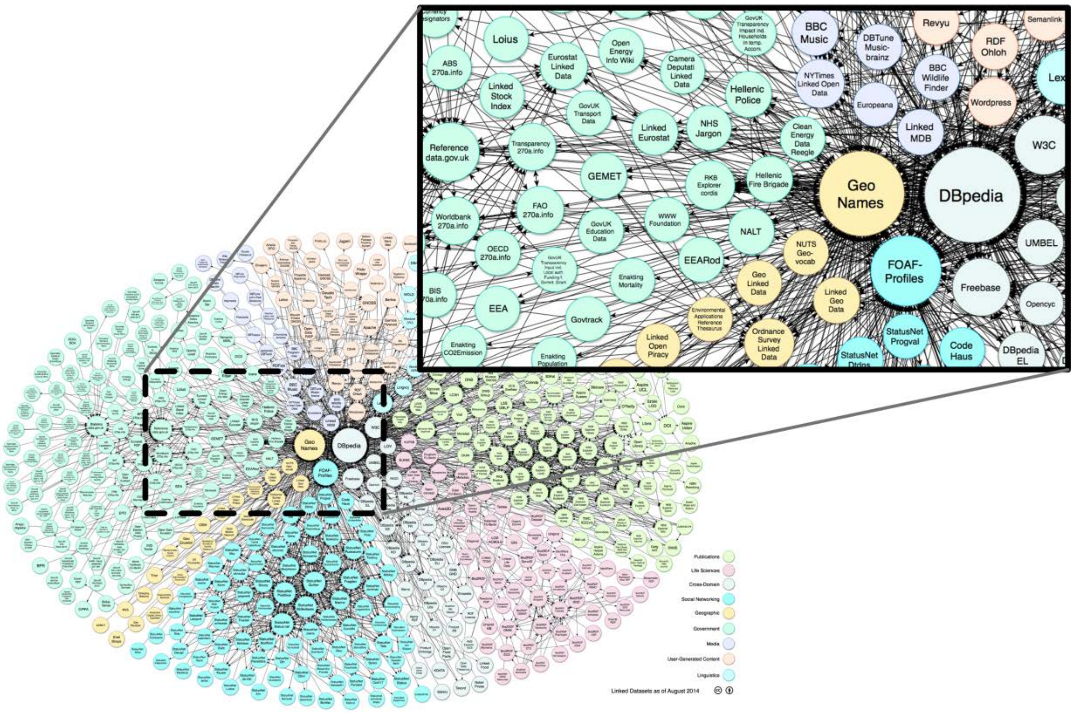


Figure 1. Linked Open Data cloud diagram (2014), by M. Schmachtenberg, C. Bizer, A. Jentzsch and R. Cyganiak. Source: http://lod-cloud.net/

Perhaps one of the most popular projects (at the centre of Figure 2) is the DBpedia project.[12] Since 2007 onwards, it offers the key information present in the Wikipedia pages as Linked Data by regularly making an automatic extraction. This extraction parses the boxes in the HTML articles and information in the natural language text, and transforms it into structured RDF. This RDF is connected to other datasets, like GeoNames about geographical names (on the left of DBPedia in the figure), which in turn is linked by many others etc.

The topic of the datasets is disparate, and colours in the Figure 2 represents a vague categorisation: geographic in orange, governmental in green, purple the multimedia etc. In the magnified box of Figure 2, the relevance of some datasets is captured: Eurostat, UK government data, BBC music, Govtrack (Federal US legislation), etc. Table 2 shows an overview of the amount of triples as well as the amount of RDF links, in dataset of the different domains coloured in Figure 2.

Table 2. Source: Jentzsch, Cyganiak, and Bizer (2011) http://lod-cloud.net/state/

Domain	Number of datasets	Triples	%	(Out-)Links	%
Media	25	1,841,852,061	5.82 %	50,440,705	10.01 %
Geographic	31	6,145,532,484	19.43 %	35,812,328	7.11 %
Government	49	13,315,009,400	42.09 %	19,343,519	3.84 %
Publications	87	2,950,720,693	9.33 %	139,925,218	27.76 %
Cross-domain	41	4,184,635,715	13.23 %	63,183,065	12.54 %
Life sciences	41	3,036,336,004	9.60 %	191,844,090	38.06 %
User-generated	20	134,127,413	0.42 %	3,449,143	0.68 %

[12] http://dbpedia.org/About

content					
	295	31,634,213,770		503,998,829	

Table 2 shows data from 2011. Five year later, numbers have increased dramatically — 85,567,007,302 triples from 3426 datasets.[13] This huge amount of data entails a cultural and organisational shift. This semantic data is a drop of water in the immense ocean of the general Internet. Every second 30,000 gigabytes of data (1 gigabyte= 10^9, 1,000,000,000 bytes) are shed on the web (Marz and Warren, 2014: 1). What is happening now? A multiple and converging technological process in which the different fields of big data —data mining, advanced statistics, predictive analytics, data engineering, and visualization— meet the second Web Semantic generation, Natural Language Processing (NLP), and a plurality of theoretical approaches in Robotics and Artificial Intelligence. We have stated elsewhere that this process is not natural or neutral, but deeply political and literally crossed by global strategies and punctual decisions at all levels (Casanovas 2015a, 2015b).

Some of the datasets in Figure 2 belong to the language domain, or can be useful for the community of linguists. The so called Linguistic Linked Data Cloud is such a fragment of the Linked Data Cloud. This subset is of special interest for the purpose of this chapter. First because these datasets are being used in applications facilitating the human-machine communication in natural language, second because these datasets are likely to be traded in commercial exchanges, requiring a closer legal attention.

The need for consuming linked datasets in a lawful manner did not trigger the first requirement of having licensing information. It had happened before, with the wide acceptance of Creative Commons among the Web surfers. Many RDF datasets were also licensed, and as a matter of 'good behaviour', Web users generally give credence to the author and respect the license even if they know that they will not suffer effective prosecution.

2.1.3 Use of licenses for Linked Data

We have defined in detail Linked Data irrespective of its legal status in the previous pages. Some of the linked datasets have been published as Linked *Open* Data, referring to the fact that they are open —being either in the public domain or licensed under quite liberal terms, e.g. under Creative Commons Share-Alike licenses. Some institutions also publish data with more restrictive licenses or even in a completely closed fashion.

But, how much data is already being licensed? The licenses used in linked datasets have been systematically studied (Rodríguez-Doncel et al. 2013b), finding a notable amount of datasets published with a license. The most recent snapshot of the licenses in the Linked Data Cloud of Figure 1, finds an even bigger use of licenses if not in absolute numbers at least in terms of important datasets covered. Figure 2 colours the nodes according to the type of license.

[13] http://stats.lod2.eu/

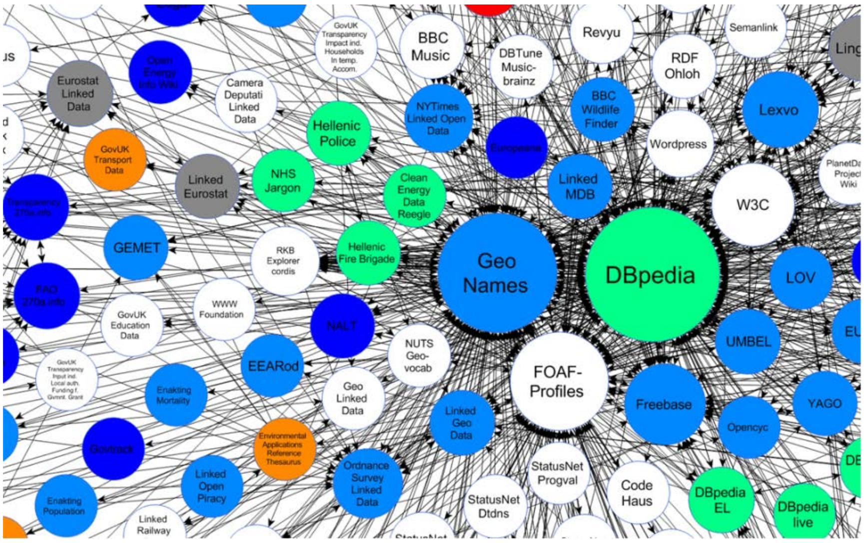


Figure 2. RDF Datasets coloured by license type. Green indicates “share-alike” style (cc-by-sa, gfdl, odc[14]-odbl licenses), blue nodes indicate in the public domain (expressed with cc-zero or odc pddl) or only attribution (like cc-by, odc-by licenses), orange nodes indicate with ‘some restrictions’, like “non-commercial” or “no derivatives”. Source: http://lod-cloud.net/

But the numerical results of the license analysis, shown in Figure 3, reveal that a significant number of datasets still had been published without a license or under restrictive terms.

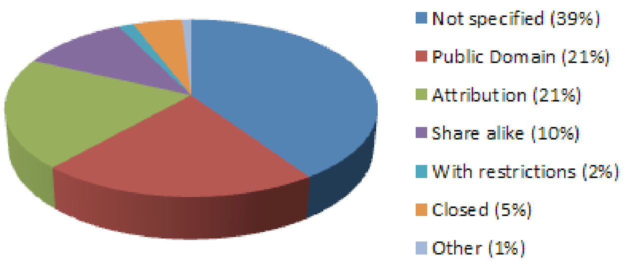


Figure 3. Kind of licenses used in datasets annotated present in the LOD cloud. Source: Vila-Suero et al. (2014)

The main idea of Linked Data is to create an ecosystem that facilitates the browsing, discovery and exploitation/reuse of datasets for applications. The issue of whether datasets can be actually used for a specific purpose is thus a crucial one; so that understanding the conditions under which a certain dataset has been licensed is also crucial. Such licensing information should be expressed ideally in a machine-readable fashion to facilitate automated reasoning by end applications on the conditions of use of a dataset.

2.1.4 Rights Expression Languages (REL)

Leaning on the perspective described above, we can identify six challenges for legal services (Casanovas et al. 2010: 5–7): (i) the relationship between the Social Web (Web 2.0) and the Web of Data (broadly, Web 3.0); (ii) the construction of evolving and contextual *legal*

[14] Odc stands for Open Data Commons

ontologies (and their relationship with *folksonomies*)[15]; (iii) the construction of Legal Semantic Web Services (LSWS); (iv) the bridge between Information Technology (IT) law (e.g. intellectual property), and IT for lawyers (computational support); (v) the convergence between Web 2.0 and Web 3.0 to grasp the changing and evolving nature of regulations; (vi) the addition of reasoning —e.g. defeasible and non-monotonic logic, among others— to facilitate users' exchanges and legal operations on the Web.

But this is only half of the story. If we open up the objective of the lens to broaden the angle of sight we will realise that Linked Data Best Practices (LDBP) propose that every dataset should provide provenance and licensing information, data-set level metadata, and information about additional access methods (Schmachtenberg et al. 2014). To sum up, each one should provide self-describing metadata. The other way around, everyone should practically use this information in a "pro-sumer" (consumer/producer) way, i.e. creatively, adding, using, and reusing content. The link between Wikipedia and DBpedia is an example of dynamic *social intelligence*[16], in which *crowdsourcing* —the aggregation of multiple and massive individual behaviour— and the structural framework where actions are carried out by users, bring about a collective output that keep evolving over time and causes the growth of the web. *Personalisation*, user-centred approach, is one of the main trends of the second-generation Semantic Web (D'Aquin et al. 2008). We will come back to this later, after exploring some structural constraints.

As already shown, licensing is crucial for the use and reuse of information in the new Web environment. However, there are some differences between the way licenses are traditionally defined in commercial law treaties and how they are conceived in computer science. Traditionally, licensing is conceived as a specific type of commercial *contract* between a licensor (owner or possessor of rights) and a licensee (who receives them) under certain *conditions* settled by the agreement under a mandatory national or international *law*. In computer science, licensing is "the act of transferring limited *rights* to another party, under certain terms and conditions, for *use and reuse* of *content*" (Iannella 2003).

Since the very beginning, then, the scope of the so called Digital Rights Management technologies had a broader focus, which could not be simply adapted to a single jurisdiction or, so to speak, a "legal" conception of the law. As Innella (ibid.) put it:

"*Digital Rights Management (DRM) technologies covers the broad area of intellectual property management by providing secure and trusted services to control the use and distribution of content. DRM technologies consists of a mix of business models, social issues, legal conformance, and technical capabilities. This implies that DRM Technologies are an integral part of the entire end -to - end content management lifecycle, not just a single service that exists in isolation. Hence the key to successful DRM technologies is that it is not be seen as a separate 'DRM system', but as part of the overall content management and consumption framework.*"[17]

Nowadays DRM is a term with bad connotations for the Internet community (Casanovas 2015b). But, for the past twenty-five years, *Rights Expression Languages* (REL) have been a key element in these systems. Thus, there is a shift between law conceived as a set of centralised

[15] Folksonomies relate to crowdsourcing. They are usually understood as a system in which users apply public tags to online items.We will define ontologies and come back in the next sections on this particular problem.

[16] http://www.sintelnet.eu/

[17] http://www.nicta.com.au/pub?doc=764

normative systems, and law conceived as set of enacting rights empowering at the same time stakeholders (such as big companies) and users (consumers, citizens).[18] REL models focus on permissions (usages allowed over the content: play, print, sell...), constraints (limits to permissions, e.g. time-based restrictions), and requirements (obligations needed to exercise the permissions). Figure 4 shows the most important elements in a ODRL policy —although similar schemas are found in other policy languages and rights expression languages.[19]

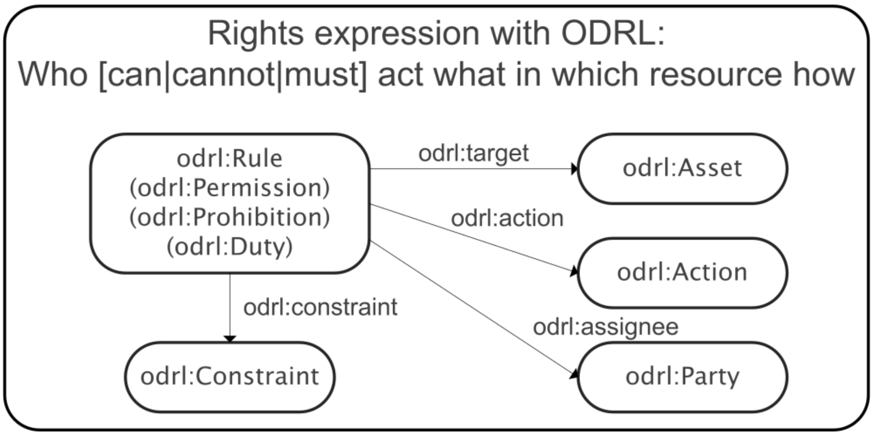


Fig. 4 Main elements in ODRL policies. An ODRL policy is made up of rules. These rules declare which actions are permissible, obligatory or forbidden effected over an asset by a certain party under certain constraints.

Open Digital Rights Language (ODRL) supports open publishing, distribution, and consumption of content, applications and services on the Web. This model respects and puts in the users' hands the management of their rights: it requires their explicit permission. Otherwise, moves are forbidden by default. This model sets a common vocabulary to allow reusing by a plurality of communities, and an ontology to allow interoperability built by McRoberts and Rodriguez-Doncel (2014). Some rules, such as "permission", are introduced and grouped with basic restrictions and conditions.[20]

As example, we should stress the success of half-automated Creative Commons licenses, massively adopted by now.[21] CC licenses makes use of RELs. They incorporate a three-layer design. The first one is a standard legal description of rights ("legal code"). The second one ("Commons Deed") summarises and explain in plain non-technical language the content of the licence. The third layer is machine-readable, *CC Rights Expression Language* (ccREL): "*a specification describing how license information may be described using RDF and how license*

[18] REL were born in the early nineties, when Mark Stefik developed at Xerox PARC a language that would became the eXtensible Right Markup Language (XrML). Permissions and restrictions can be modelled according to Creative Commons principles (ccREL), Open Digital Rights Language (ODRL), MPEG-21[18], or national copyright protections. There is no universal Right Expression Language, but many —among the more relevant: ODRL, MPEG-21 REL, XACML, ccREL, MPEG-21, MVCO and WAC (Rodríguez-Doncel et al. 2013a, 2013b).

[19] https://www.w3.org/community/odrl/model/2.1/ See a recollection of models at http://delicias.dia.fi.upm.es/~vrodriguez/pdf/Poster-LicensingPatternsForLinkedData.pdf

[20] https://www.w3.org/community/odrl/model/2.1/#section-References

[21] http://creativecommons.org/licenses/

information may be attached to works". People can easily find CC open content using search engines.

2.2 A Web of Services

2.2.1 The *Service Web*

An overview of data on the Web has been given in the above section —particularly addressing *semantic data* on the Web. The present one explores its proactive components, the Services that are offered and consumed within Web technologies.

Electronic communications between companies had happened long before, but the first generation of services in the Web was born around 2000. The term "Web Service" is defined by the W3C as "*a software system designed to support interoperable machine-to-machine interaction over a network*"[22]. The technologies on which Web Services are based have several purposes: describing the data model syntax (XML), the transference protocol (SOAP, Simple Object Access Protocol), the available services (WSDL, Web Services Description Language) and the services discoverability through a directory (UDDI, Universal Description, Discovery and Integration). SOAP Web Services gained wide acceptance, but UDDI, which was supposed to be the "yellow pages" of Web Services, never took off. The programmer of a system consuming Web Services (*client*) should know and learn about the Web Service existence before using it.

Few years ago, Semantic Web developers were envisaging that the combination of semantic technologies and Service Oriented Architectures (SOAs) would lead to the creation of a Service Web "*allowing billions of parties to expose and consume billions of services seamlessly and transparently, where all types of stakeholders —from large enterprises to SMEs and singleton end users— engage as peers, consuming and providing services within a network of equals*" (Davies et al. 2010). Semantic Web services would be eventually discoverable and understood by machines, and their consumption automatable.

However, they were also aware that SOAs could not scale without incorporating several regulatory requirements: (i) principles for a worldwide communication infrastructure (already satisfied by the standard Web Services); (ii) significant mechanisation of service lifecycle activities (location, negotiation, adaptation, composition, invocation and monitoring, service interaction requiring data, protocol and process mediation); (iii) and a balanced integration of services provided by humans and machines. Figure 5 summarises this framework, the so called *Service Web*, as drawn by Davies et al. (2010), combining social dimensions, Web 2.0 and semantic technology. The lower layer of the rectangle (the Web) reaches the upper one (the Services) by means of combining semantic technology and Web Services (Semantic Web Services); and by combining Web 2.0 and Web Services, the possibility of creating Service Communities arises.

[22] W3C Working Group Note, Web Services Glossary, http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/#webservice

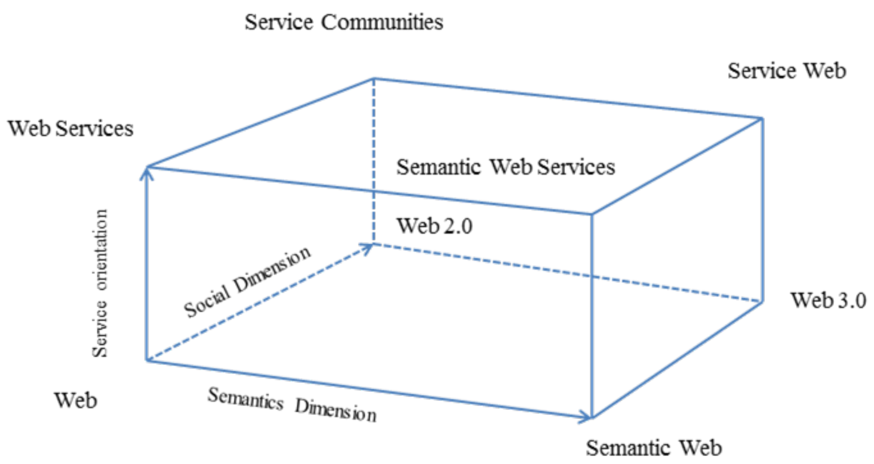


Figure 5. Service Web Pillars (2.0/3.0). Source: Davies et al. (2010)

The Service Web, as figured out by J. Davies, C. Pedrinaci, D. Fensel, and many other players in computer science and industry, is still under development. It has to be fleshed out at many levels of abstraction and operating languages. Syntactic, semantic and pragmatic dimensions are involved in this broad landscape. The problem we are dealing with is how they should be represented and combined to produce effective results in specific settings and different contextual levels.

The Web of Services has experienced an exuberant growth in the last few years, and the service orientation paradigm has been massively adopted under the form of online APIs (Application Programming Interface) using standard protocols (HTTP) in simple stateless REST (Representational State Transfer) services.

As of today, the provision of services is a very lucrative business. Consumers pay for the usage of services under different modalities and this flourishing industry is expected to grow and be supported by the cloud computing paradigm. In fact, the peak of inflated expectations for *cloud computing* has been surpassed. According to Gartner[23], cloud computing will be found on the plateau of productivity in a couple of years.

3 The pragmatic turn

Some lessons can be drawn from the last section. In the cloud, not only law but the nature of law is experiencing a deep transformation. The difference lies on that which the regulation of data and metadata would require for an effective, secure, and fair functioning. It is our contention that law alone as it is generally defined —the production of legislative and parliamentary documents, administration or court-ruling— is not well-adapted to take into account the dynamicity of social intelligence and big data environments.

The old way of drafting and enacting regulatory models assumed a simple ontology, in which human knowledge could be treated as separated knowledge *about* human behaviour (understood as either inner or outer experiences). However, the structuring of data by means of metadata

[23] Gartner's 2014 hype for emerging technologies, https://www.gartner.com/doc/3100227.

incardinate action *and* knowledge at the same time in a more complex dynamic flow (action, knowledge, shared knowledge, meta-knowledge) —i.e. it is endowed through an *intelligent* flow. How could this social intelligent flow be described, controlled, monitored, and guided?

The clear-cut line between human norm-regulated external actions and normative legal systems, at least in the usual positivistic way as they are assumed by lawyers and jurisprudence, is becoming blurred in information-processing environments. What kind of connection can be constructed in between? And what does it link agency, agents' behaviour, be they digital or human, with the normative design of regulatory planning?

We are pointing at: (i) the interface between virtual and non-virtual *contexts* (a set of scenarios in social ecosystems); (ii) intentional communicative acts understood as the exercise of rights (e.g. intentional acts of selecting a license); (iii) regulatory performative acts constrained by standards or guidelines embedded into computer design (e.g. Privacy by Design, PbD), (iv) the collaborative coordination between agents' decisions, intentional acts and planning. Summing up, the asymmetric relationship between the exercise of rights and the regulatory power of norms. Self-described data (metadata) are bounded by the general course of agents' behaviour in a specific ecosystem.

3.1 Pragmatics

3.1.1 Some preliminary thoughts

Pragmatics is often defined as the study of language as it is used in a social context, including its effect on the interlocutors (utterers and interpreters). Context is placed at the centre, as pragmatics studies the meaning in actual situations assuming an interactional and dynamic notion of context. Equivalently, legal pragmatism emphasizes the need to include a more diverse set of data than the one usually considered from a positivistic view, and again the specific context at hand plays a central role in judicial decision-making. The evaluation of the circumstances on a particular issue is a genuine human task. Machines are fast and accurate when performing specific procedures, but the evaluation of context requires the comprehensive understanding of a human being.

However, the latest technologies might improve the capabilities of machines in this regard, possibly bringing practical tools that include context for decision-making in a near future. These promises are brought by means of a combination of new developments in the fields of Big Data, the Semantic Web and Natural Language Processing (NLP).

Advances in NLP increase machine understanding of natural languages. Disambiguation algorithms are among technologies which have experienced a larger improvement. Disambiguation is the capability of correctly identifying the referred entity when a piece of text (word or words) make reference to different things without any additional context. Disambiguation is carried out considering the text as a whole, taking into account past experience and the social perception of it. Once this hurdle is falling, machines will be able to safely establish relationships between unrelated pieces of text. Undoubtedly, this task will be accomplished in a smoother manner by using a new breed of resources based on semantic technologies. In what it follows, we will link them with the knowledge acquisition process, ontology building, institutional analysis, and other developments in Normative Multi-agents Systems (norMAS).

3.1.2 The *Pragmatic Web*

Pragmatics always played a role in SW developments. Invoking the metaphor of Polya's inventors paradox —"the more ambitious plan may have more chances of success"— M.P. Singh wrote a seminal paper launching the "Pragmatic Web" in 2002.

"*The core mission of the semantic web, namely, to enable information to be shared across the web, faces significant challenges. The challenges come from the difficulty of capturing semantics in a manner that is reusable across applications, the priority of process over data, the importance of interaction, and the critical need for accommodating user context. Overcoming these challenges takes us from the realm of semantics and brings us into the realm of pragmatics, which* we view as supervenient upon the semantics [our emphasis]. *We claim that when the vision of the semantic web is realized, it will be via the pragmatic web.*" (Singh 2002: 82)

Several Conferences have been organised on the topic following this trend.[24] There is an active research community, taking inspiration from semiotics and philosophy of language —C.S. Peirce, C. Morris, P. Grice, J.R. Searle. This perspective has been already applied to several SW fields —multi-agent systems, interaction design, self-organizing communities of practice, and Web Services (Bonancin et al. 2013). Speech-acts theory, illocutionary force, intended meaning, implicatures, and communicative acts and processes have been used as bases for computer design with interesting results.

This approach usually stems from the classical semiotic distinction between syntax, semantics, and pragmatics. E.g., as applied to the Web, de Moor (2005) distinguishes between (i) the *Syntactic Web* (interrelated syntactic information resources linked by HTML references), (ii) the *Semantic Web* (a collection of semantic resources, mainly in the form of ontologies), (iii) and the *Pragmatic Web* (a set of pragmatic contexts of semantic resources). He contends that meaning negotiation can be based on a set of fundamental *pragmatic patterns*, which can made available in a *meta-ontology*. Pragmatic patterns are defined as "template definitions that can be used as the basis of conceptual definitions used in meaning negotiation and other meaning evolution processes" . Communities may use and extend them for a plurality of purposes.

Core pragmatic patterns include (de Moor, ibid.) : (i) *Pragmatic context* ("a pattern that defines the speakers, hearers, type of communication, and identifiers of the individual and common contexts of a community"); (ii) *Individual context* ("a pattern that defines an individual community member, individual context parameters and an identifier of the individual context ontology"); (iii) *Common context* ("a pattern that defines the common context parameters and an identifier of the common context ontology of a community"); (iv) *Individual pragmatic pattern* ("a meaning pattern relevant to an individual community member", "an individual context ontology consists of the total set of meaning patterns relevant to that individual"); (vi) *Common pragmatic pattern* ("meaning pattern relevant to the community as a whole").

[24] http://www.csw.inf.fu-berlin.de/pragweb/

Paschke et al. (2007) focused on the infrastructure for collaborative human-computer networks. Using RuleML[25], they aimed at extending "the Semantic Web to a rule-based S*emantic-Pragmatic Web* which puts the independent micro-ontologies and domain-specific data into a pragmatic context such as communicative situations, organizational norms, purposes or individual goals and values".

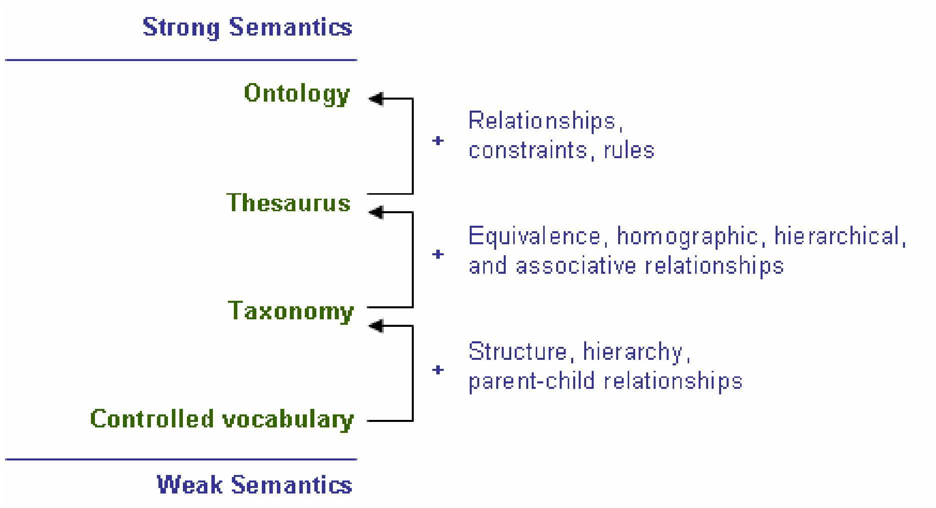
Following the three semiotic distinctions proposed by Liu (2009) —who groups illocutions into three dimensions: *time* , *invention,* and *mode*— Bonacin et al. (2013) have recently built *CactO*, a communication act ontology that links acts, agents and behaviour patterns. In the same vein, applying Sperber and Wilson's (1995) relevance theory on implicit inferences and contextual inferred meaning, Mazak and Wally (2014) make use of meta-information and weighted graphs for ontology building and aligning.

All these approaches deal with and deepen into contextual constraints, interactive communication and collaborative work. But in the field, *semantics*, *ontologies*, at a different level of depth and granularity, play a crucial role. The "Pragmatic Manifesto" addresses particularly ontology-building:

"*The vision of the Pragmatic Web is thus to augment human collaboration effectively by appropriate technologies, such as systems for ontology negotiations, for ontology-based business interactions, and for pragmatic ontology-building efforts in communities of practice. In this view, the Pragmatic Web complements the Semantic Web by improving the quality and legitimacy of collaborative, goal-oriented discourses in communities.*" (Schoop et al. 2006)

3.1.3 Ontologies and the pragmatic approach

Leaning on the same distinction between the three linguistic dimensions, Cardoso (2006) proposed a gradual understanding of semantics and he placed (formal) ontologies on top of it. He distinguishes four main representations that can be used to model and organize concepts to describe terms: controlled vocabularies, taxonomies, thesaurus, and ontologies (Figure 6).



[25] RuleML is a markup language to express rules in XML for inferential purposes and reasoning. Vid. an introduction to LegalRuleML principles in Athan et al. (2015).

Fig. 6. Levels of Semantics. Source: Cardoso (2006:8)

In the same way that the Semantic Web is not a separate web, but an extension of the current one, the Pragmatic Web does not entail a different web, but an extension of the Semantic Web.

From a computational point of view, Cardoso (2006) differentiated between three kind of different approaches to bring semantics to Web services: (i) by mapping concepts in a Web service description (Web Semantics Descriptive Language specification) into ontological concepts; (ii) by using OWL-S to describe Web services, mapping OWL-S into WSDL operations; (iii) by using Web Service Modelling Ontology (WSMO). It includes Web services, goals, ontologies and mediator concepts. He proposed to combine the functionalities of WSMO and OWL-S to create SW processes, allowing complex interactions among organisations to be represented in all stages of their lifecycle.

Each of these three alternatives placed a burden on the shoulders of developers that they were not willing to accept, and Semantic Web Services never took off. The latest efforts in this line have been directed towards describing a simpler vocabulary to document APIs: the Hydra Core Vocabulary[26]. But even if Hydra fails to succeed, massive adoption of structured API documentation is taking place. Technologies like APIBluePrint or Swagger's describe JSON structures to derive automatic human-readable documentation, but they could also be automatically parsed by semantically enabled tools.

The pragmatic basket is full of phenomena analysed from different theoretical backgrounds — rhetoric, discourse analysis, dialectics, argumentation theory, ethnomethodology, cognitive anthropology, socio-linguistics, and cognitive linguistics, among others. Actually, the pragmatic approach to semantics can be faced also gradually, according to the different operational stages in ontology-building, at the knowledge acquisition level (KA) and at the representation level.

Pragmatism, the attitude to learn and building knowledge in a machine-operable way, has been a *common motto* for SW researchers since the beginning. The common assumption is that there is no clear difference between pragmatic and semantic conceptualisations of meaning. There is a *continuum* between the actual use and the fixed content of language. The difference lies on the specific objectives and weight allocated onto the interactive view (e.g. community-building), or onto the structural one (eg. upper-top, core or domain ontology building). There is a large room for positioning and selection in between. This means that there is a *hybrid* approach. E.g. E. Motta, M. d'Aquin, and A. Gangemi , among many others, differentiate between *conceptual coverage* and *pragmatic sustainability* (d'Aquin and Gangemi 2011), and they apply this distinction to reach a *trade-off design* (Distinto et al. 2014).

This is consistent with Gruber's (1992:1) original and most quoted definition of what a computational ontology consist of: "An *ontology* is an explicit specification of a conceptualization."[27] In Studer et al. (1998) words: an ontology is "a f*ormal, explicit*

[26] https://www.w3.org/community/hydra/, spec. published by the W3C Hydra Community Group

[27] "For knowledge-based systems, what "exists" is exactly that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational *vocabulary* with which a knowledge-based program

specification of a shared conceptualization". Portability, clarity, certain degree of formality, and, most of all, the quality of being *shareable and reusable*, are attributes of computer ontologies. In this sense, they are not dependant on the philosophical usage of the term (Guarino and Giaretta 1995). The term *ontologia* does not possess Greek origins. It was coined at the beginning of 16 c. to express the quality for an object (being) to be *known* (Øhrstrøm et al. 2008).

Gruber (2000) himself assessed many times the pragmatic understanding of what ontologies are built for: e.g. "Ontologies are not about truth or beauty. They are agreements, made in a social context, to accomplish some objectives. It's important to understand those objectives, and be guided by them". Thus, *ontological commitment* expresses this kind of epistemic agreement within a community —lay people, computer scientists, and experts pertaining to a particular domain of knowledge. This goes back to the problems raised in knowledge acquisition processes, which is something particularly familiar to computer scientist that were working on expert systems in the eighties and nineties.

3.1.4 The Knowledge Acquisition perspective: some remaining and new questions

Knowledge Acquisition (KA) is a well-trodden path in computer science. Starting in 1986, several Workshops have been organised over the last thirty years with regular attendees, setting a working group to discuss epistemological, problem-solving and ontological problems. According to Musen (2013: 196), this community achieved remarkable consensus on a number of issues: (1) the rejection of the notion of knowledge as a commodity to be transferred from one locus to another, (2) an acceptance of the situated nature of human expertise, (3) emphasis on knowledge acquisition as the modelling of problem solving, and (4) the pursuit of reusable patterns in problem solving and in domain descriptions that can facilitate both modelling and system implementation.

M. Musen recalls W. Clancey going to the point in a Workshop Keynote in 1997: knowledge acquisition is not hard because it is difficult to get domain experts and knowledge engineers to speak the same language; knowledge acquisition is hard because *the domain experts and the knowledge engineers together must create a language to define a model of professional expertise* —a model that never existed previously in any formal sense (Musen, ibid. 197).

The "situated cognition" approach developed by Clancey stated that *experts do not provide knowledge that covers all possible contexts*. He pointed at the level of abstraction of modelling and simulating work processes in organisational settings. Individual work practice, collaboration, communication, 'off-task' behaviours, multi-tasking, interrupted and resumed activities, informal interactions, use of tools and movements were described and integrated in BRAHMS, a programme developed at NASA Ames Research Center. The influential Brahms design was geared towards modelling people's activity behaviour (Clancey 1998, Suit 2007).

---

represents knowledge. Thus, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names are meant to denote, and formal axioms that constrain the interpretation and well-formed use of these terms." (Gruber, 1992: 1)

After thirty years of research in knowledge acquisition this problem has been addressed in many ways. I.e. Compton (2013: 184) figured out how to bridge problem solving, contextual and data approaches, and situated meaning through "Ripple-down rules". I.e. if people distinguish between different conclusions in different contexts, they do so because they can identify features that distinguish the contexts. This type of case or context differentiation could be readily integrated with a range of knowledge acquisition frameworks. *Ripple-Down Rules* would provide a way to gather a large amount of knowledge for a system, simply identifying such contextual knowledge.

The shift from intensive problem solving approach and the early Semantic Web, to knowledge management and the Web of Data raises many questions and issues. Some of them are related to ontology building, use and reuse, and have had a strong continuity with the early KA issues. As Aussenac-Gilles and Gandon (2013) put it: Which kind of knowledge can be modelled? How important is the distortion between the computer interpretation of an ontology and the human interpretation of the associated knowledge? What kind of intelligent behaviour and tasks can formal representations be useful for? Others issues are new, as the initial problem of having too few sources has shifted towards the problem of too many, big, diverse and diverse sources.

After the experiences of WATSON[28] as iFAQ and SIRI[29] as i-Virtual Personal Assistent, T. Gruber (2013: 194) asks for an even more radical change:

"*As computers get faster and bigger they can look through more data, but they are still being told what to look for. For knowledge acquisition systems, the inductive bias is the endowment from nature that guides how the agent adapts to its nurturing environment. [...] For our field, a fundamental question for knowledge acquisition is this: to learn from collective human knowledge, what kind of structure applied to what kind of knowledge in what form can lead to learning genuinely new knowledge? Will the structure be in the patterns to look for, the relationships among things, the statistical properties of collections, or other inductive biases? Will the golden discoveries be found in the human generated content, structured data, graphical or auditory sources, query streams, or other breadcrumb trails of human experience? What will be the most useful measures of interestingness when learning from everything? The answers to this inquiry will guide us into the next level, which is*: can this process be applied recursively to itself? That is, can our machine learners apply what they learn to improve their own learning? [...] Is there an epigenetics of knowledge? [Our emphasis]"

Summing up, there is a possible loop in the self-reflective descriptions provided within and across the Web of Data. The ongoing openness, participation and creative thinking of the many are seen as essential to sustain the Web. Closeness should be avoided. This means that heuristics matters, and the plurality of human languages and human cultures should be taken into account when "shared knowledge" is at stake.

We can distinguish both the formal and the cultural side of social intelligence. On the one side, as stated by Hitzler and Harmelen (2010), "reasoning for the SW "should be understood as *shared inference*, which is not necessarily based on deductive methods." [30] On the other, the

[28] http://www.ibmwatson.com/
[29] http://siri.com
[30] "Model-theoretic semantics (and sound and complete reasoning based on it) functions as a gold standard, but applications dealing with large-scale and noisy data usually cannot afford the required runtimes. Approximate methods, including deductive ones, but also approaches based on entirely different methods like machine learning or nature inspired computing need to be investigated, while quality assurance needs to be done in terms of precision and recall values (as in information retrieval) and

work by Breuker (2013) provides some useful examples of the modelling issues stemming from social intelligence and cognitive foundations. Breuker leans on previous pragmatic studies to show some difficult decisions in ontology-building —especially for upper-top or foundational ontologies. E.g. the notions of "space" and "position" are not uniform. Levinson (2003) distinguished three frames of reference for locating objects: absolute, relative, and intrinsic. As summarised by Breuker (2013: 180):

"*The intrinsic reference frame locates an object A relative to another object B. For instance, in ‘‘the dog is in front of the car’’, the position of the dog is defined by what is considered the ‘front’ of the car. In other words, in (many) intrinsic references a property of object B, in particular its ‘form’, connects here object properties with spatial locations. Spatial reference frames can be viewed as what has recently become an important topic in discussing ontology engineering: (design) patterns or frameworks. Frameworks go beyond concepts as simple building stones, fit for any combination. It restricts far more the semantics than simple lists of properties, but emphasizes essential relationships that can be used for reasoning. For instance, the relative spatial reference frame defines a triple relationship between an observer, object A and object B for which transitivity can be claimed. These frameworks are important because they enable specialized inference.*"

Therefore, there are cultural shifts and nuances about time, space, distance, positioning... at different levels of knowledge. Cultures differ in which notions can be used, and usually they use more than one. Depending of what the ontology is aiming at, it should reflect these differences and diverse "common knowledge" cultural features (Breuker, 2013: 180).

However, the level of abstraction in which the Web of Data operates make this "grasping" tasks a bit more difficult. Again, trade-offs and intensive exchanges appear to be essential to reach acceptable results.

One of the classical players in the KA field, Gaines (2013: 133) identifies the different tiers for what he calls a *developing symbiosis science*. The first four belonging to computer science, the remaining ones to knowledge science: (i) the underlying digital electronics; its application in computer architectures; (ii) the programming of general-purpose computers through software; (iii) the development of computer–people and computer–computer interactivity; (iv) the representation of human knowledge; (v) the acquisition of additional knowledge from interaction with the world, people and stored knowledge; (vi) the development of goal-directed autonomous knowledge creating processes; (vi) the increasing coupling of knowledge processing entities in social networks; (vii) the development of techniques to facilitate the synergy between human and computer knowledge processes; (viii) the synthesis of both into a unified system.

3.2 Institutions

3.2.1 The Institutional Turn

Let's deepen into the subject of a hybrid, symbiotic approach. In this section, we will recover the legal thread we left ready in Sections 1 and 2. Let's start with Feingebaum's "knowledge

---

not necessarily in terms of soundness and completeness of the underlying algorithms." (Hitzler and van Harmelen, 2010: 39).

acquisition bottleneck". At the acceptance speech for the establishment of the *Feigenbaum Medal* presented at the World Congress on Expert Systems at Orlando in December 1991, he recalled how he invented the dynamically growing decision trees, under Herbert Simon's mentorship. This was an example of what he termed the "performance motive" of Artificial Intelligence —"less theory, more system building"— or learning by practicing. The process of hypothetic induction constituted one of the main problems to be tackled on the Newell and Simon's path. Envisaging the future, he said:

"*Knowledge sharing, to these researchers, means more than just the knowledge bases of several expert systems interoperating to solve a problem. Knowledge sharing also means the computer-facilitated cooperation of many people in the building of a large body of codified knowledge.* The vision is that hundreds or thousands of knowledge base builders would cooperate [our emphasis] ".

Well, times are most ripe for such a vision, and perhaps the lesson to be drawn nowadays is that dealing with billions of structured data and metadata would require more attention to cooperative, common processes and single, detailed, performances. In a way, this is what Hoekstra (2010) is pointing at with his reformulation of Feigenbaum's dictum. *The knowledge engineering bottleneck*: Semantic Web content can no longer be assumed to have been produced in a controlled task-independent environment. When reused in a non-controlled task environment , "Semantic Web content needs to be remoulded, refiltered and recurated for a new task". In contrast to Feigenbaum's problem, the knowledge reengineering bottleneck refers "to the general difficulty of the correct and continuous reuse of pre-existing knowledge for a new task ". We do believe that this is the new problem we are facing now. How are we going to proceed with legal and political tasks and content in the Web of Data?

Certainly one feasible way of doing it is remaining stuck to legal documents and texts. Organising at present huge legal libraries, such as the millions of documents gathered in the EURLex repository, constitutes a real challenge (Francesconi et al. 2015). Structuring legal knowledge is the aim of successful projects on LegalXML interchange formats for legal resources, such as Akoma-Ntoso[31] and CEN Meta-Lex[32] —the later one due partially to Hoekstra's own work. But this is not solving the raised problem of contextual and *personalised* reuse of such repositories. One of the main problems is that which both approaches are taking for granted —law and legal knowledge are assumed to be contained into documents, liable to be extracted and self-described through metadata and a set of reusable ontologies.

These are useful trends. However, again, this is only half of the story. Dynamic and specific legal acts, performances and knowledge are being produced daily in many global places —e.g. markets; political and public scenarios— through the conflictive and cooperative behaviour of millions of players alike in specific and changing social scenarios. People are using licenses and referring to intellectual property or fundamental rights in a wide range of different contexts in a conflicting world. In a way, *rights are becoming independent from the coded content of legal norms* in this new environment. To bring some order into the social scenarios of the Web of Data, ethical principles, governance and law should be worked out at the same time in regulatory instruments, as stated by several recent EU reports (Schlinder et al, RAND co. 2012).

---

[31] http://www.akomantoso.org/
[32] http://www.metalex.eu/

This constitutes a challenge to be faced in the immediate future, at different levels of abstraction and formalisation, through different kind of regulatory instruments. We will advance in this Section the notion of Semantic Web Regulatory Models (SWRM) to cope with these objectives, compatible with electronic institutions, Multi-agent Systems (MAS), platforms and APIS. They can bridge users' needs, decisions and protections with big lexical repositories and RELs.

3.2.2 Iterative Pragmatic Cycle (IPC)

The aim of field research is to elicit and describe all the processes, interactions, patterns of behaviour, and scenarios in which end users get involved. As a result, the expert knowledge obtained through knowledge acquisition processes is modelled as "situated knowledge" —in Clancey's sense—, allowing an active participation of end users in the ontology-building process. End users are integrated into the process from the beginning to the end. Accordingly, this Section introduces (i) the dynamic knowledge acquisition stage carried out to collecting and transforming knowledge (e.g. into ontologies) (ii) to be integrated into a regulatory model set for specific purposes to be embedded into a particular social ecosystem.

Our ontology-building methodology finds its theoretical roots and follows the *middle-out approach* —in contrast with top-down and bottom-up ones— advanced by MethOntology, the engineering perspective set by Gómez-Pérez et al. (2004), and refined and updated to adapt it to networked ontologies on the web exploiting the NeON toolkit (Suárez-Figueroa et al. 2012). Building ontologies has become a more complex task now, in which several levels of formality and different engineering groups are involved.

As described by Corcho et al. (2015): (i) Upper-level ontology engineers (formal ontologies: DOLCE, BFO, GFO, SUMO); (ii) heavyweight ontology engineers being domain experts (for example, graduates in biology, geography, or law) or computer scientists (OWL-based domain and application ontologies developed by this profile may reuse axioms, properties and concepts from upper-level ontologies); (iii) lightweight ontology engineers (they develop vocabularies to be used in the linked data context, usually written in RDF Schema or in OWL profiles with little expressivity, e.g. OWL Lite); (iv) SKOS [*Simple Knowledge Organization System*] concept scheme developers are interested in developing thesauri and other types of classifications (library sciences); (v) Web developers contributing to *Schema.org*[33] or using it to annotate websites from the cluster of Schema.org vocabulary developers (HTML, RDFa and JSON-LD).

What it is worth to notice is that ontology-building can be integrated into a broader institution-building cycle. From our perspective, the process comes to an end only when a regulatory instrument or set of instruments have been built, delivered and put into practice into a specific organisation, setting or community. We call it an *Iterative Pragmatic Cycle* (IPC). The final output can be a judicial iFAQ —such as IURISERVICE (Casanovas et al. 2006; Casellas 2011); a mobile application —such as MEDIAKIDS (Poblet et al. 2015); a platform for Online Dispute Resolution (ODR) —such as CONSUMEDIA (Suquet et al. 2014), or a set of ontologies and ethical and policy principles to monitor and control the information flow of a security platform —such as CAPER (González-Conejero et al. 2014, Casanovas et al. 2014). Sometimes the lifecycle cannot be completed. Eventually, the regulatory instrument does not reach the implementation stage.

[33] https://schema.org/

Many different methodologies have been already described to build legal ontologies at different levels (Sartor et al. 2011). Casellas (2011) shows and explain a full array of them. Our starting point is an *Iterative Knowledge Acquisition Process* (IKAP) stemming from a socio-legal approach (Casanovas et al. 2011), using all kind of linguistic and sociological methods (quantitative and qualitative). Hua (2008) discusses different families of techniques that are specifically devised to elicit and analyse knowledge acquired from experts. We propose an IKAP based on five stages: *elicitation, collection, analysis*, *modelling and validation* (see a schematic description in Fig. 7).

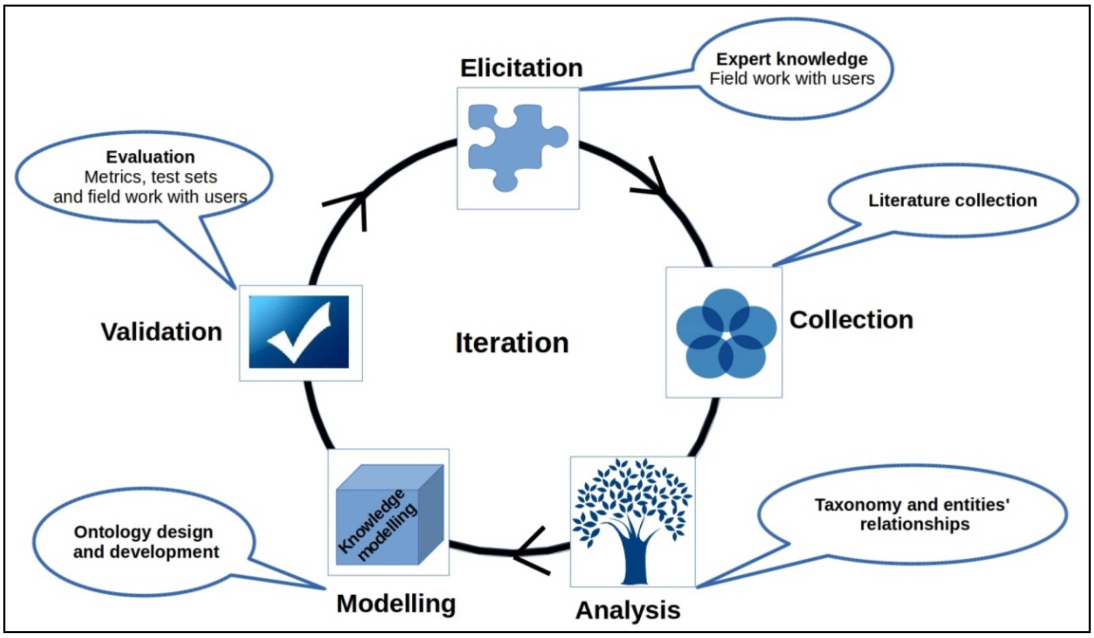


Fig. 7. Iterative Knowledge Acquisition Schema (IKAP schema)

However, how to link this bulk of acquired social knowledge with ethical, governance and legal issues constitute a separate problem. Fig. 8 draws the filtering position of regulatory models that sets forth the requirements to be taken into account in a legal modelling process.

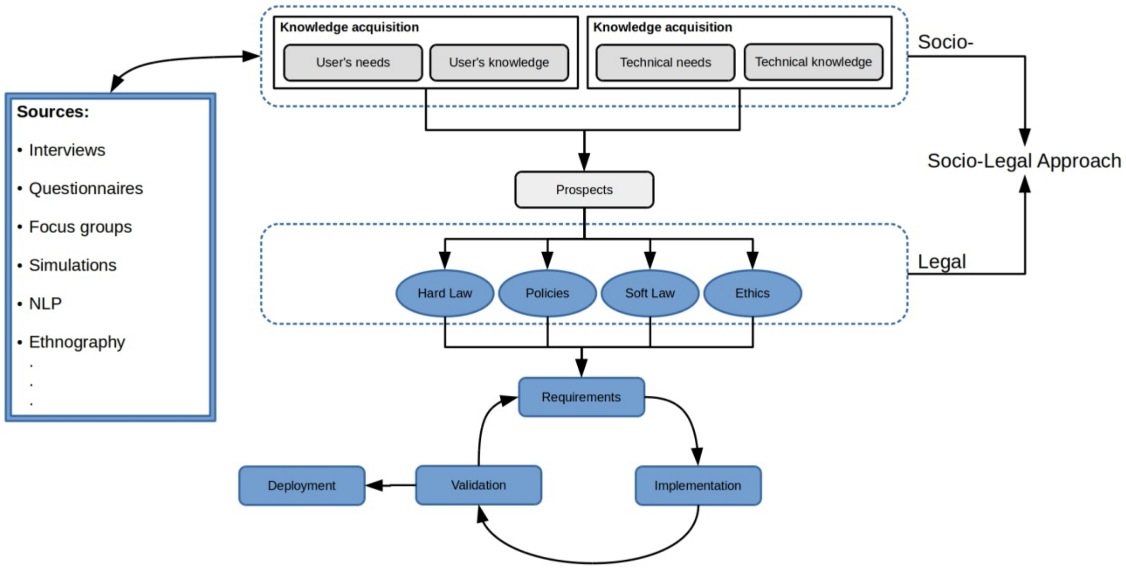


Fig. 8. Pragmatic socio-legal approach

3.2.3 Semantic Web Regulatory Models (SWRM)

Regulatory Models (RM) can be defined as a set of values, patterns, rules, norms and principles conceptually represented in a certain natural, half-artificial or artificial language. They can be represented as constraints and conditions for agency and the performance of moral, political and legal rights. They can be implemented into the social ecosystem in which Right Expression Languages and Digital Rights Management operate. They can be strategically used to regulate the informational flow on platforms, and to implement Data Protection, security, privacy and compliance by design. Figure 9 draws a simple schema, a general meta-model, in which Ethics (e.g. Fair Information Practices, FIPs) play a regulatory role, along with Hard Law (national and international acts and court decisions), Governance (national and international agency policies), and Soft Law (transnational standards, professional guidelines, corporate models) (Casanovas et al. 2014b, Casanovas and Zeleznikow 2014, Casanovas 2015a, 2015b).

These regulatory sources are located according to their distance to *binding power* and *social dialogue*, the two converse axes of the Rule of Law. Legality or validity of a particular regulatory system, as a whole, depends onto the degree of compliance of their components. It is a second-order property. According to a certain threshold for the degree of enforceability, efficiency, effectiveness, and justice, the system becomes stable, and a third axis of trust and institutional strengthening (security) emerges as a result. This is an example of strategic collective rationality, as trust cannot be directly commanded or induced.

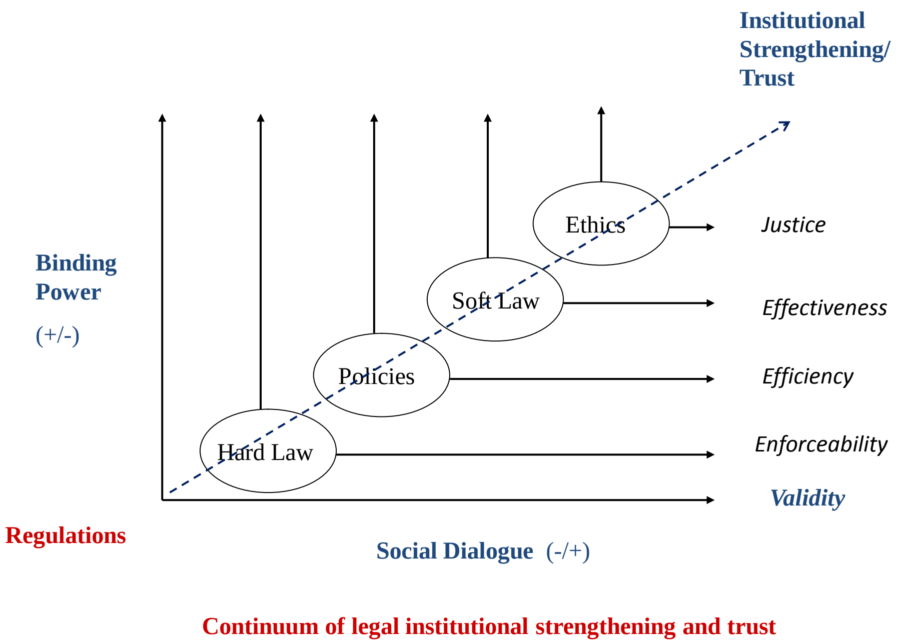


Fig. 9. Meta-model of SWRM. Source: Casanovas (2015a)

When this structure is applied to ODRL policies (Fig. 6) or to the W3C Core-model for rights and policies governance (Iannella 2014)[34], the result is a *Meta-Rule of Law*, a set of conditions to be taken into account and restrictions for the possible normative inferences drawn from data and meta-data models (Casanovas 2015b). Thus, this is a way to enhance democratic values on the new global environment of the Web of Data, linking micro and macro scenarios to a common understanding of regulations.

SWRM can be applied to a great deal of different legal problems and situations, operating on the web at a transnational level through Semantic Web languages. Different dimensions (the type of regulation) and jurisdictions (related to public powers -e.g. where courts are located) can be distinguished. SWRM are conceived as *intermediate institutions* between the enhanced positive norms of a particular legal system and the actual performances of technological devices (platforms, APIS, mashups...). This is why they are able to be modelled as electronic half- or self-regulated systems, embedded into the design of technological devices, and help fasten decision-making when it is needed (Fig. 10).

[34] https://www.w3.org/community/odrl/model/2.1/

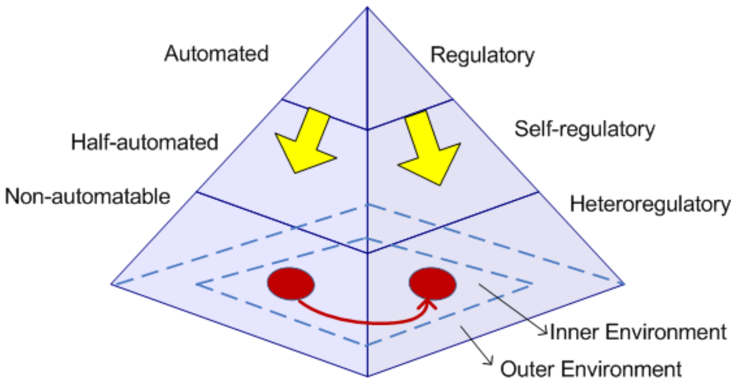


Fig. 10. Layered Regulatory Automation

SWRM seem well suited to build the institutional outer context for normative Multi-agent Systems (norMAS) operations. I.e. linking social, legal and governance scenarios with the particular ecosystem that electronic institutions are aiming to produce. Electronic Institutions (EI) are a way to implement interaction conventions for agents —human or software— who can establish commitments on an open environment (within and outside organisations, administrations or companies). [35] The dynamic entrenchment between agreement technologies (Ossowski 2013) and such environment constitute the place in which SWRMs come to play, as they furnish the structured and consistent information that is needed for EI to perform legal acts—such bidding, contracting or mediating.

3.2.4 Artificial Socio-cognitive Systems (ASCS)

Challenges for Artificial Socio-cognitive Systems (ASCS) (Noriega et al. 2014) are two-fold and can be understood in two directions. The first one goes from the system to the users, as it sets forth the conditions in which the system can be managed, imposing some conditions to users. In the other way around, from the users-to-system perspective, both human and artificial agents are embedded into a broader socio-legal context (see Fig. 11) that it must be taken into account for the effective implementation and acceptance of ASCS performances. The "valid" or "legal" acts carried out by ASCS should be effectively legal.

Pragmatics matters at all layers, as open multi-agent systems can be effectively designed and implemented as electronic institutions (d'Inverno et al. 2012). In 1997, Noriega proposed building computational environments —agent-mediated institutions— that allow heterogeneous agents to interact successfully by imposing appropriate restrictions on their behavior—<speaker,role; hearer set, role; action; timestamp>. Moves are represented as illocutionary acts in a dialogical framework, but carrying out complete performative acts (Noriega, 1999).[36] The final result of a negotiated process is intended to be a full trade or a settled agreement.[37]

E.g. we can define the framework for an EI for mediation setting a *conceptual model* to describe the institution, a *computational model* that explain how it is enacted, and a *pragmatic model* that establishes how it is implemented. Interactions between agents within EI are speech acts that *count as* actions in the world, repetitive and organised as they were *scenes in a play*. Again, (i)

---

[35] http://e-institutions.iiia.csic.es/

[36] tuple DF= <O, L, I, CL, *Time*>, where O stands for Ontology, L stands for language for domain content, I is the set of Illocutionary particles, CL is the agent (communication) language, *Time* is a discrete and partially ordered set of instants.

[37] Cfr. Ossowksi (20013) for the state of the art in agreement technologies. An example of electronic institution for mediation purposes can be found at the Annex of the *Catalan White Book on Mediation* (Spanish version), cfr. Noriega and del Toro (2010), http://www.llibreblancmediacio.com/ . For the full study, cfr. Poblet et al. (2010a), and Casanovas et al. (2009).

the dialogical frame specifies which speech acts are admissible; (ii) the performative structure is formed by a network of scenes, understood as conversation protocols or dialogue games (specified as directed graphs were arcs are labelled by speech acts *schemata*, and nodes as *institutional states*). Actions are *illocutions* made by individuals who are playing a given role, constrained by rules of behaviour that establish the normative positions of commitments that arise from agent interactions (Noriega and López del Toro, 2009).

ASCS create their own institutional and social context, their specific social ecosystem, as shown by the meta-model drawn in Noriega et al. (2013). However, some interesting aspects still remain to flesh out ASCS full capabilities, mainly related to their social use, implementation and effects. As advanced by Noriega and d'Inverno (2014), electronic institutions might be open to be *anchored* in a specific environment (Fig. 12).

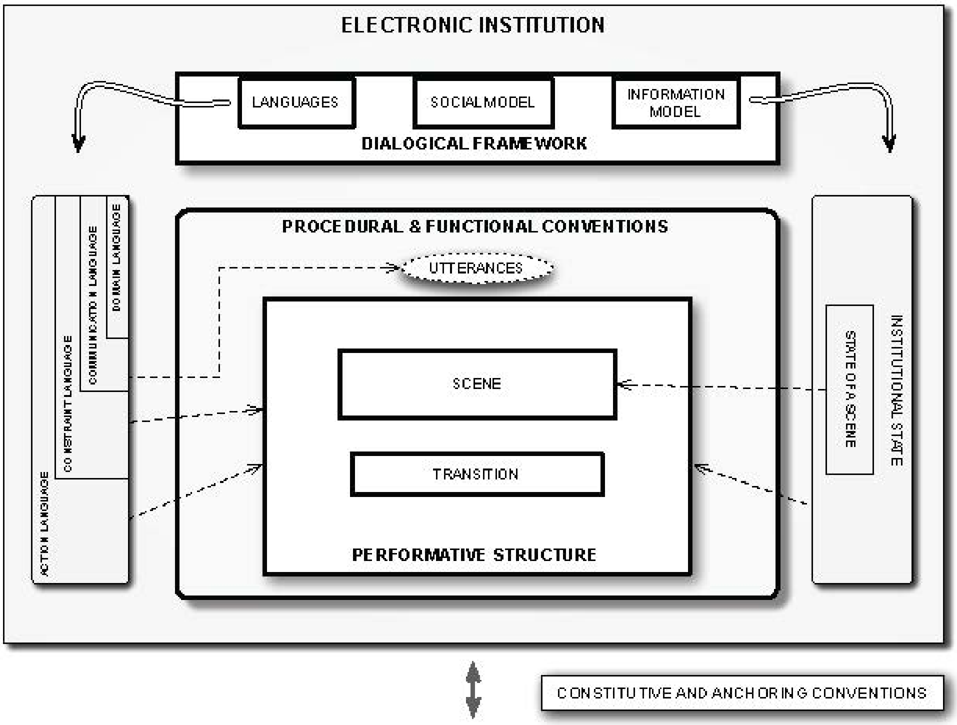


Fig. 11. Conceptual Model of Electronic Institutions. Source: Noriega and d'Inverno (2014).

Thus, they cannot be completely settled in a situated context without selecting and taking into account some salient treats of the outer ecosystem. They should be *accommodated* into a broader social environment. Still, this is a dynamic and two-fold process, as there is a bidirectional flow between the ASCS creation of their own ecosystems, and the outer ones, which put their own constraints (to be understood, negotiated and eventually balanced). In this sense, it is our contention that ASCS pragmatic cycle could be completed making use of SWRM. Some *metapragmatic awareness* (Verschueren, 1999: 196) is needed to trigger the systems. The pragmatic attitude in building ASCS, the representation of interactions as speech acts, and actions (moves) as illocutionary acts —or "utterances" (d'Inverno et al. 2012)— could be completed by filtering the outcomes as full *perlocutionary acts* through SWRM.

The statement assumed so far — "Actions in the computational systems *count as* actions in the real world where they are subject to appropriate regulations" (Noriega et al. 2013: 95)— could be fleshed out more explicitly. What do "appropriate regulations" mean? Things are usually a bit more complex in the real world. We cannot simply assume that EI are "mirroring" regulations. The issue of coupling (and decoupling) the ASCS generated ecosystems with the already existing ones should be carefully addressed, as it seems richer than merely assuming *ex hypothesi* the correspondence between facts and actions in the physical world, and institutional facts and actions in the IE space.[38] What links an EI with the outer (legal, political) contexts in which their outcomes will be re-evaluated and implemented?

Governance and legal implementation of existing statutes and regulations constitute a complex process by its own. The norMAS community does not ignore the problem. Researchers have begun collecting and describing specific experiences involving software systems, people, and organizations (Singh et al. 2013), i.e. *hybrid* systems (Noriega and Sensoy 2015).[39]

Let's put an example of the importance of the outer environment to create social ecosystems. In emergencies, crisis, terrorist attacks, and natural disasters (bushfires, floods, hurricanes, earthquakes...), things run very fast. Conveying true and quick information among the population, all volunteers, organizations and authorities is vital. Mobile technologies are massively used (Poblet 2011). Victims, relatives and friends have a common need for information, but they may be located in different countries *under a different metadata and data protection regime*. That is to say, *existing privacy laws and policies can create legal barriers* to the way missing persons organizations collect, use and share information. However, there is a common agreement that anticipation, cooperation, and adjustment can balance all interests involved to achieve a proportional result (Reidenberg et al. 2013).

Up to date, after the many experiences gathered internationally in many of these events (2002 Bali bombing, 2004 Boxing Day tsunamis, 2005 Katrina etc.), governments have started to change *manually* the legislation for these kind of situations. In the 2011 Christchurch earthquake, the New Zealand Privacy Commissioner reacted within 24 hours of the emergency declaration and issued a "temporary information sharing code" to assist in the relief effort, granting broader discretion to emergency services and government agencies. However, this kind of general personal data and metadata disclosure may create unintended effects. What about media storing all these information? What about the aggressive behaviour shown by some companies or organizations collecting and using later on this information for commercial or

---

[38] This is the approach proposed by Searle (1995) —*X count as Y in context C.* According to Searle, the underlying principles in any society are quite simple: (i) collective intentionality, (ii) the assignment of function, (iii) constitutive or institutional structures: function-status assignments that take the form X counts as Y in context C. "*What is an institution? An institution is any collectively accepted system of rules (procedures, practices) that enable us to create institutional facts*" (Searle, 2005). Cfr. Searle (2010) for further developments of the same thesis. This perspective stems from philosophy of language, not from empirical social sciences. Tuomela (2011) contends that the emergence of intentional collective action goes beyond Searle's individualistic account based on speech act theory. Cfr. Bolander et al. (2014).

[39] See especially the example furnished by M. Singh (2013: 192 and ff.) as part of the NSFfunded Ocean Observatories Initiative (OOI), a thirty-year $400 million project, with thousands of stakeholders (ocean scientists, resource providers, technicians, operators, policy makers, application developers, and the general public). "*OOI provides novel capabilities for data acquisition, distribution, modelling, planning and control of oceanographic experiments, with the main goal of supporting long-term oceanographic and climate research*".

political purposes? What about people with special protection needs (not that uncommon: e.g., the localization of threatened women protected by a residence restraining order)?

It is our contention that general provisions, exceptions and even conditions for particular cases can be better managed through SWRM and REL, which in addition: (i) have the property to coordinate the agency of different related powers (Legal Enforcement Agencies, DP agencies, supervisory commissions, local and national bodies...), (ii) can facilitate the common interoperability of norms, rights, and concepts —as for this case e.g. the legal concept of *permitted purpose* for the disclosure—, (iii) may offer a general framework to *coordinate* the legal actions to be taken before, during and after the exception. *Visualisation* of law and regulatory systems is one of the main problems in order to induce participation and an easy coordination of collective behaviour (Poblet 2013).

EI, ASCS for crisis and disaster management could help to managing the overall situation in a faster and more efficient way, but they should take into account the experiences and lessons learned.

4 Conclusions and further work

The previous example leads to these preliminary conclusions, as the use and reuse of ontologies and big lexical repositories come into play and should be encompassed with the use of ASCS. Enactment, protection and exercise of rights, i.e., the actual *empowerment* of individuals and communities on the Web of Data is the next step. The implementation of Artificial Socio-cognitive Systems, Semantic Web Regulatory Models, Rights Expression Languages, and Linked Licensed Open Data could help to reach this objective.

What is the role of pragmatics in the Web of Data? This chapter brings about the following conclusions: (i) Pragmatics has always been the main perspective in the Semantic Web area; (ii) this approach is maintained in the methodology to be applied, especially in knowledge acquisition process and in ontology building; (iii) we should look closely at the notions of situational context, environmental context and social ecosystems when creating and implementing institutions; (iv) this pragmatic approach can bridge the gap between legal systems and the kind of computational models which are needed to manage data and metadata; (v) especially hybrid systems, encompassing human and artificial agents, deserve a special attention in the immediate future.

Besides, the application of LD technology to the publication of linguistic data promises to alleviate issues related to the integration and aggregation of dataset stemming from heterogeneous sources and using different vocabularies. This has lead to coining the term Linguistic LD for all linguistic datasets that are published following the LD principles (Vilasuero et al. 2014). Three essential issues need to be addressed for such data to be easily exploitable by language technologies.

First of all, datasets need to be enriched with machine-readable licensing information so that applications can reason about conditions under which it is legitimate to use a particular resource. While in the general case open licenses are preferable and compatible with the Web-style publishing used in LD, some use cases and datasets might require more restricted licenses or even datasets to be closed while being at the same time linked. Second, linked linguistic datasets need to have a minimum quality in order to build trust by end applications. Finally, we need shared and agree-upon vocabularies and guidelines to foster standardization and thus easier exploitation of resources. We have briefly discussed these issues in the first sections of this chapter, and presented some preliminary investigations on the distribution of licenses in the Linguistic LD Cloud.

Different Licensing models can co-exist in Licensed Linguistic LD, from totally open to more restrictive licenses through to completely closed datasets. However, these models and LLLD should be adopted and adapted within the ethical, political, and legal constraints that apply in different countries and legal cultures. This is a big problem. Even CC licenses can be turned down and redefined by Courts in national jurisdictions, as it already happened in Germany.

Therefore, we contend that intermediate institutions such as *Semantic Web Regulatory Models* are needed in order to bridge the different kinds of regulations, and coordinate at different levels the exercise of rights with law and policy compliance. Regulations, and rights on and through the Web of Data do not operate in the same way that the law used to do it in a non-virtual world.

Regulatory systems need to become more complex, in order to be simpler. This is not a paradox. If we want to introduce some order and solve the "Frankenstein" puzzle (Corcho, 2015) in our already legally hyper-regulated global world, we should devote some effort to build up the necessary tools that social ecosystems require to work in favour of everybody. The pragmatic perspective, a hybrid and symbiotic approach, can help to achieve this goal.

Acknowledgments

This work has been carried out under the projects: DER2012-39492-C02-01 CROWDSOURCING, IPT-2012-0968-390000 CROWDCRISISCONTROL. We thank Pablo Noriega and Marta Poblet for previous reading and useful comments.

References

Andrighetto, G., G. Governatori, P. Noriega, and LWN van der Torre, 2013. *Normative multi-agent systems*. Saarbrücken/Wadern: Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, http://www.dagstuhl.de/dagpub/978-3-939897-51-4.

Antoniou, G., and F. van Harmelen, 2004. *A Semantic Web Premier*. The MIT Press: Cambridge, Third ed., with P. Groth and R. Hoekstra, 2012.

Athan, T., G. Governatori, M. Palmirani, A. Paschke, and A. Wyner, 2015. *LegalRuleML: Design principles and foundations*, in: *Reasoning Web. Web Logic Rules*, 151-188. Springer: Dordrecht, Heidelberg, 151-188.

Berners-Lee, T., 1999. *Weaving the Web: The Past, Present, and Future of the World Wide Web.* Orion Books: London.

Berners-Lee, T. *Linked Data*, 27 July 2006 (last change 18 June 2009) http://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., T. Heath, and T. Berners-Lee, 2009. *Linked data-the story so far*, in Semantic Services, Interoperability and Web Applications: Emerging Concepts, "IJSWIS", 205-227, http://linkeddata.org/docs/ijswis-special-issue

Bolander, T., P. Casanovas, C. Castefranchi, R. Craven, M. Franssen, M. d'Inverno, A. Herzig, A. Jones, P. McBurney, W. Nash, P. Noriega, N. Osman, J.Padget, J. Pitt, D. Pearce, C. Sierra, L. Tummolini, H. Verhagen, 2014. *Issues in Social Intelligence: Problems and Challenges*, SINTELNET FP7-ICT-2009-C Project No. 286380, version: 02.2015 b).

Bonacin, R., J.C. Dos Reis, H. Hornung, and M.C. Baranauskas, 2013. An ontological model for supporting intention-based information sharing on collaborative problem solving, in "*International Journal of Collaborative Enterprise"* 3, 2/3: 130-150.

Breuker, J., 2013. *A cognitive science perspective on knowledge acquisition*, in "International Jurnal of Human-Computer Studies", 71: 177–183.

Breuker, J., P. Casanovas, M.C.A Klein, and E. Francesconi (ed.), 2009. *Law, Ontologies and the Semantic Web*, IOS Press: Amsterdam.

Cardoso, J., 2006. *Approaches to developing semantic web services*, in "International Journal of Computer Science" 1,1: 8-21.

Casanovas, P., 2012. *Legal Crowsourcing and Relational Law. What the Semantic Web Can Do for Legal Education*, in "Journal of Australian Law Teachers Association" 5 , 1/2: 159-176, http://www.alta.edu.au/resources/PDFs/JALTA/Full%20Volumes/2012_vol5_no1&2.pdf

Casanovas, P., 2015a. *Semantic Web Regulatory Models. Why Ethics Matter?*, in "Philosophy & Technology": 28, 1: 33-55.

Casanovas, P. 2015b. "Conceptualisation of Rights and Meta-rule of Law for the Web of Data", in "Democracia Digital e Governo Eletrônico" 1, 12: 18-41, http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/34399 , reprinted in "Journal of Governance and Regulation" 4, 4: 118-129.

Casanovas, P. and Zeleznikow, J. (2014). *Online Dispute Resolution and Models of Relational Law and Justice: A Table of Ethical Principles*, in: P.Casanovas et al. (eds), *AI Approaches to the Complexity of Legal Systems IV-V*, LNAI 8929, Springer: Dordrecht, Heidelberg, 55-69

Casanovas, P., M. Poblet, N. Casellas, J. Vallbé, V. R. Benjamins, J. Contreras, R. Peña, M. Blázquez. (2006). Semantic Web: a legal case study, in: J. Davis, R. Studer, P. Warren (ed.), *Semantic Web: trends and research*, John Wiley & Sons: London, 259-280

Casanovas, P., Díaz, L., Magre, J., Poblet, M. (ed.), 2009. *Materiales del Llibro Blanco de la Mediación en Cataluña. Vol. I. La mediación: conceptos, ámbitos, perfiles, indicadores.* Generalitat de Catalunya, Departament de Justícia CEJFE: Barcelona.

Casanovas, P., U. Pagallo, G. Sartor, and G. Ajani (2010). "Introduction: Complex Systems and Six Challenges for the Development of Law and the Semantic Web", in: P. Casanovas et al. (Eds.) *Artificial Intelligence and Complex Legal Systems*, LNAI 6237, Springer, Heidelberg, Berlin, 1–11.

Casanovas, P., N. Casellas, and J.J. Vallbé, 2011. *Empirically Grounded Developments of Legal Ontologies: A Socio-Legal Perspective*, in: G.Sartor, P. Casanovas, M.A. Biasiotti, and M. Fernández-Barrera (ed.), *Approaches to Legal Ontologies: Theories, Domains and Methodologies*, Springer-Verlag: Dordrecht, Heidelberg, 49–68.

Casanovas, P., J. Arraiza, F. Melero, J. González-Conejero, G. Molcho, and M. Cuadros., 2014a. *Fighting Organized Crime Through Open Source Intelligence: Regulatory Strategies of the CAPER Project*, in: R. Hoekstra (ed.), *Legal Knowledge and Information Systems: JURIX 2014: The Twenty-Seventh Annual Conference*, 271, Amsterdam: IOS Press, 189.

Casanovas, P., U. Pagallo, M. Palmirani, and G. Sartor (ed.), 2014b. *AI Approaches to the Complexity of Legal Systems IV. Social Intelligence, Models and Applications for Law and Justice Systems in the Semantic Web and Legal Reasoning*; AICOL IV-V, LNAI 8929, Springer: Dordrecht, Heidelberg.

Casellas, N., 2011. *Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge.* Springer: Dordrecht, Heildelberg.

Chiarcos, C., S. Hellmann and S. Nordhoff, 2012. *Linking linguistic resources: Examples from the Open Linguistics Working Group*, in: C. Chiarcos, S.Nordhoff and S. Hellmann (ed.), *Linked Data in Linguistics. Representing Language Data and Metadata*, Springer: Dordrecht, Heidelberg, 201-216.

Clancey, W.J., P. Sachs, M. Sierhuis, and R. van Hoof, 1998. *Brahms: Simulating Practice for Work Systems Design*, in "International Journal of Human-Computer Studies" 49: 831–65.

Compton, P., 2013. "Situated cognition and knowledge acquisition research", in *International Journal of Human-Computer Studies* 71: 184–190.

Corcho, O. , M. Poveda-Villalón, and A. Gómez-Pérez, 2015. *Ontology engineering in the era of linked data*, in "Bulletin of the American Society for Information Science and Technology", 41, 4: 13-17.

Davies, J., M. Potter, M. Richardson, S. Stinčić, J. Domingue, and G. González-Cabero, 2009. *Towards the Open Service Web*, in "BT Technology Journal" 26: 2.

d'Aquin, M., E. Motta, M. Sabou, S. Angeletou, L. Gridinoc, V. Lopez, and D. Guidi, 2008. *Towards a new generation of semantic web applications*, in "IEEE Intelligent Systems", 24: 20–28.

d'Aquin and A. Gangemi, 2011. *Is there beauty in ontologies?*, in "Applied Ontology", 6, 3: 165-175.

d'Inverno, M., M. Luck, P. Noriega, J.A. Rodriguez-Aguilar, C. Sierra, 2012. *Communicating open systems*, in "Artificial Intelligence" 186: 38–94.

Domingue, J., D. Fensel, D., J.A. Hendler (ed), 2011. *Handbook of Web Semantics Technologies*, Heidelberg, Dordrecht: Springer.

Feigenbaum, E.A. , 1984. *Knowledge engineering: the applied side of artificial intelligence*, in "*Annals of the New York Academy of Sciences*" 426: 91–107. Original publication in 1980 as Report of the Stanford Department of Computer Science.

Feigenbaum, E.A., 1992. *A Personal View of Experts Systems: Looking Back and Looking Ahead*, in “Knowledge System Laboratory”, Report n. 92-41 KSL, Stanford.

Francesconi, E., M.W. Küster, P. Gratz, and S. Thelen, 2015. *The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services*, in: *Electronic Government and the Information Systems Perspective*, Springer: Dordrecht, Heidelberg, 29-39.

Francopoulo, G., G. Monte, N. Calzolari, M. Monachini, N. Bel, M. Pet, and C. Soria, 2006. *Lexical markup framework (LMF)*, in: *International Conference on Language Resources and Evaluation-LREC* 2006- LREC 2006, 2006, G^enes/Italie, 2006. <inria-00121468>

Gaines, B.R. 2013. *Knowledge acquisition: Past, present and future*, in "International Journal of Human-Computer Studies" 71: 135–156.

Gómez-Pérez A.. Mariano Fernández-López, and Oscar Corcho, 2004. *Ontological Engineering With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer: Dordrecht, Heidelberg.

González-Conejero, J. , R. Varela-Figueroa, J. Muñoz-Gómez, and E. Teodoro, 2014. *Organized Crime Structure Modelling for European Law Enforcement Agencies Interoperability through Ontologies*. In: P. Casanovas et al. (ed.) *AI Approaches to the Complexity of Legal Systems*. AICOL IV-V, Springer: Dordrecht, Heidelberg, 217-231.

Guarino, N., and P. Giaretta, 1995. *Ontologies and knowledge bases towards a terminological clarification*, in: J.N.I. Mars (ed.) Towards very large knowledge bases: knowledge building & knowledge sharing , IOS Press: Amsterdam, 25-32.

Gruber, T.R., 1993. *A translation approach to portable ontology specifications*, in *Knowledge acquisition* 5, 2: 199-220.

Gruber, T. R. 2000. *It is what it does: The pragmatics of ontology for knowledge sharing*. Presentation available at https://www.yumpu.com/en/document/view/31975231/it-is-what-it-does-the-pragmatics-of-ontology-as-tom-gruber

Hellmann, S., J. Lehmann, and S. Auer, 2012. *Linked-data aware uri schemes for referencing text fragments*, in: *Knowledge Engineering and Knowledge Management*, Springer: Dordrecht, Heidelberg, 175-184.

Hendler, J., and T. Berners-Lee, T., 2010. *From the Semantic Web to Social Machines: A research challenge for AI on the World Wide Web*, "Artificial Intelligence" 174: 156-161.

Hitzler, P. and F. van Harmelen, 2010. *A Reasonable Semantic Web*, "Semantic Web Journal" 1, 1: 39-44. http://corescholar.libraries.wright.edu/cse/25

Hua, J. 2008. *Study on Knowledge Acquisition Techniques*, in "*Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium IEEE"* 1: 181-185.

Iannella, R. (2003) *Digital Rights Management Technology*, available at NICTA http://www.nicta.com.au/pub?doc=764

Kasper,G., and K.R. Rose, 1999. *Pragmatics and SLA*, in "Annual Review of Applied Linguistics" 19: 81–104.

Levinson, S., 2003. *Space in Language and Cognition*. Cambridge University Press: Cambridge.

Lewis, R. (ed.), 2007. *Dereferencing HTTP URIs*, (31 May, 2007), W3C, http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14

Liu, K., 2009. *Pragmatic Computing – A Semiotic Perspective to Web Services*, in J. Filipe and M.S. Obaidat (ed.) "C*ommunications in Computer and Information Science*" 23: 3–15.

McCrae, J., G. Aguado-de-Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gómez-Pérez, J. *Gracia et al*., 2012*. Interchanging lexical resources on the Semantic Web,* in "Language Resources and Evaluation" 46, 4: 701-719.

Markoff, J., 2006. *Entrepeneurs See a Web Guided by Common Sense* (11/12/2006). http://www.nytimes.com/2006/11/12/business/12web.html?_r=1&pagewanted=1&ei=5088&en=254d697964cedc62&ex=1320987600

Marz, N., and J. Warren, J., 2014. *Big data. Principles and best practices of scalable real time data systems*, Manning Ed.: MEAP.

Mazak, A., and B. Wally, 2014. *Pragmatic-based Ontology Design and Alignment*. In: *Proceedings of the 8th International Workshop on Value Modeling and Business Ontology (VMBO 2014),* 8 p.

Musen, M. A., 2013. *The knowledge acquisition workshops: A remarkable convergence of ideas*, in "International Journal of Human-Computer Studies" 71: 195–199.

Noriega, P. 1999. *Agent mediated auctions: the fishmarket metaphor*. Barcelona: IIIA.

Noriega, P., and C. López del Toro, 2009. *Towards a platform for online mediation*, in: M.Poblet, U. Schield, J. Zeleznikow (ed.), *Legal* and *Negotiation Decision Support Systems* (*LDSS 2009*), a *Post-conference Workshop* of the *12th International Conference* on *Artificial Intelligence* and *Law*, *Proceedings*, IDT Series, *vol*. *4*, *2009*, pp. 69-76.

Noriega, P., and C. Lopez del Toro, 2010. Annex of the *Catalan White Book on Mediation*, cfr. , http://www.llibreblancmediacio.com/

Noriega, P., and M. d'Inverno, 2014. *Crowd-Based Socio-Cognitive Systems,* in: M. Poblet, P. Noriega, and E. Plaza, 2014. *Sintelnet WG5 Workshop on Crowd Intelligence: Foundations, Methods and Practices*, 1-15. http://ceur-ws.org/Vol-1148/CROWD2014_Full.pd

Noriega, P., J. Padget, H. Verhagen, and M. d'Inverno, 2014. *The challenge of artificial socio-cognitive systems*, in The XIX International Workshop on Coordination, Organizations, Institutions and Norms in Multiagent Systems, COIN@AAMAS2014, Paris.

Noriega, P., and M. Sensoy, (ed.), 2015. *Agent Teams for Design Problems*, The XVIII International Workshop on Coordination, Organizations, Institutions and Norms in Multiagent Systems, COIN@ AAMAS2015, Istambul.

O'Reilly, T. What is Web 2.0. *Design Patterns and Business Models for the Next Generation of Software* (09/30/2005), at http://oreilly.com/web2/archive/what-is-web-20.html .

Øhrstrøm, P., H. Schärfe, and S. Uckelman, 2008. *Jacob Lorhard's Ontology: A 17th Century Hypertext on the Reality and Temporality of the World of Intelligibles*, in: P. Eklund and O. Haemmerlé (Eds.), *ICCS 2008*, LNAI 5113, Springer: Dordrecht, Heidelberg: Springer, 74–87.

Pedrinaci, C. and J. Domingue, John, 2010. *Toward the next wave of services: linked services for the web of data*, in "Journal of Universal Computer Science" 16, 3, 1694–1719.

Pedrinaci, C., J. Domingue, and A.P. Sheth. 2011, *Semantic Web Services*, in: J. Domingue et al. (ed.) *Handbook of semantic web technologies*, Springer: Berlin, Heidelberg, 977-1035.

Poblet, M., (ed.), 2011. *Mobile Technologies for Conflict Management: Online Dispute Resolution, Governance, Participation*. Springer: Dordrecht, Heidelberg.

Poblet, M., 2013. *Visualizing the law: crisis mapping as an open tool for legal practice*, "Journal of Open Access to Law" 1: 1. https://ojs.law.cornell.edu/index.php/joal/article/view/12

Poblet, M., P. Noriega, C. Pérez del Toro, J. Suquet and Josep Redorta, 2010. *Capítulo 16. Tecnologías para la mediación en línea: estado del arte, usos y propuestas*, in: Casanovas, P., Magre, J., Lauroba, E. (ed.) 2010, *Libro Blanco de la Mediación en Cataluña* Departament de Justícia, Generalitat de Catalunya, Huygens Editorial: Barcelona, 896-1008. http://www.llibreblancmediacio.com/

Poblet, M., P. Noriega, and E. Plaza, 2014. *Sintelnet WG5 Workshop on Crowd Intelligence: Foundations, Methods and Practices.* http://ceur-ws.org/Vol-1148/CROWD2014_Full.pd

Poblet, M., E. Teodoro, J. Gonzalez-Conejero, R. Varela, and P. Casanovas, 2015. *A co-regulatory approach to stay safe online: reporting inappropriate content with the MediaKids mobile app.* "Journal of Family Studies", http://dx.doi.org/10.1080/13229400.2015.1106337.

Reidenberg, J.R., R. Gellman, J. Debelak, A. Elewa, and N. Liu, 2013. *Privacy and Missing Persons After Natural Disasters*. Washington, DC and New York, NY: Center on Law and Information Policy at Fordham Law School and Woodrow Wilson International Center for Scholars.

Rodríguez-Doncel, V., M.C. Suárez-Figueroa, A. Gómez-Pérez, M. Poveda, 2013a. *License linked data resources pattern*, in: In *Proc. of the 4th Int. Workshop on Ontology Patterns* , 2013, Sydney. *Proceedings of the 4th Workshop on Ontology and Semantic Web Patterns.* CEUR-WS n. 1188, http://ceur-ws.org/Vol-1188/.

Rodriguez-Doncel, V., A. Gómez-Pérez, and N. Mihindukulasooriya, 2013b. in: O. Hartig et al. (ed.) *Rights declaration in Linked Data*, CEUR 1034. http://ceur-ws.org/Vol-1034/RodriguezDoncelEtAl_COLD2013.pdf

Sartor, G., P. Casanovas, M.A. Biasiotti, and M. Fernández-Barrera (ed.), 2011. *Approaches to Legal Ontologies: Theories, Domains and Methodologies*, Springer: Dordrecht, Heidelberg.

Schindler, H.R., J. Cave, N. Robinson, V. Horvath, P.J. Hackett, S. Gunashekar, M. Botterman, S. Forge, and H. Graux, 2012. *Europe's policy options for a dynamic and trustworthy development of the Internet of Things*, RAND Europe. *Prepared for European Commission, DG Communications Networks, Content and Technology (CONNECT)*, RAND Corporation, SMART 2012/0053 http://www.rand.org/content/dam/rand/pubs/research_reports/RR300/RR356/RAND_RR356.pdf

Schmachtenberg, M., C. Bizer,, A. Jentzsch, and R. Cyganiak, R. 2014. *Adoption of the Linked Data Best Practices in Different Topical Domains*, in *The Semantic Web – ISWC 2014*, LNCS 8796, Springer: Dordrecht, Heidelberg, 245-26.

Searle, J.R., 1995. *The Construction of Social Reality*. Free Press: New York.

Searle, J.R., 2005. *What is an institution?*, in "Journal of Institutional Economics", 1(01):1–22.

Searle, J.R. 2010. *Making the Social World: The Structure of Human Civilization*, Oxford University Press: Oxford.

Singh, M.P., M. Arrott, T.Balke, A. Chopra, R. Christiaanse, S. Cranefield, F. Dignum, D. Eynard, E. Farcas, N. Fornara, F. Gandon, G. Governatori, H. Khanh Dam, J. Hulstijn, I. Krueger, H. Lam, M. Meisinger, P. Noriega, B.T.R. Savarimuthu, K. Tadanki, H. Verhagen, and S.Villata, *The uses of norms*, in: Andriguetto et al. (ed.) *Normative multi-agent systems*. Saarbrücken/Wadern: Schloss Dagstuhl, 191-229.

Sperber, D. and D. Wilson, 1995. *Relevance: Communication and Cognition*, Second Edition, Blackwell Publishers: Oxford, Cambridge.

Studer, R., V.R. Benjamins, D. Fensel, 1998. *Knowledge engineering: Principles and methods*. "Data and Knowledge Engineering" 25, 1-2: 161-197.

Suárez-Figueroa, M.C., A. Gómez-Pérez, F. Fernandez-López, 2012. *The NeOn methodology for Ontology Engineering*, in M.C. Suárez-Figueroa et al. (eds.), *Ontology Engineering in a Networked World*, Springer: Dordrecht, Heidelberg, 9-34.

Tuomela, R., 2011. *Searle's New Construction of Social Reality*, in "Analysis" 71: 706–2011.

Verschueren, J. 1999. *Understanding Pragmatics*, Arnold: London.

Vila-Suero, D., V. Rodríguez-Doncel, A. Gómez-Pérez, P. Cimiano, J.P. McCrae, and G. Aguado de Cea, 2014. *3LD: towards high quality, industry-ready linguistic Linked Licensed Data*. European Data Forum-2014, 1-29.

Wagner, C., 2006. *Breaking the Knowledge Acquisition Bottleneck Through Conversational Knowledge Management*, in "Information Resources Management Journal" 19 (1): 70–83.