CONCEPTUALISATION OF RIGHTS AND META-RULE OF LAW FOR THE WEB OF DATA

Pompeu Casanovas*

Abstract

This article deals with some regulatory and legal problems of the Web of Data. Data and metadata are defined. Digital Rights Management (DRM) and Rights Expression Languages (REL) are introduced. Open Digital Rights Language (ODRL), Licensed Linked Data Resources (LLDR) and Creative Commons Licenses are referred. The development of REL by means of Ontology Design Patterns such as LLDR, or Open Licenses sustained by Policy Models such as ODRL, situates the discussion on metadata at the regulatory level. With the development of the Web of Data the Rule of Law needs to evolve to a Meta-Rule of Law, incorporating tools to regulate and monitor the semantic layer of the Web. This means reflecting on the construction of a new public dimension space for the exercise of rights.

Keywords: Web of Data, Rights Expression Languages (REL), Rule of Law, Open Digital Rights Language (ODRL), Linked Licensed Data Resources (LLDR)

* *Royal Melbourne Institute of Technology (RMIT). Centre for Applied Social Research (CASR), Melbourne, Victoria, Australia*
** *This paper originates at the RMIT Seminar Series "The New Rule of Law: Open Rights, Social Intelligence", held at RMIT, in Melbourne, on May 29th, 2015. I especially thank Victor Rodriguez-Doncel and Renato Iannella for their invaluable comments and suggestions.*
*** *This paper was previously published by Democracia Digital e Governo Eletrônico (Brazil).*

1. Introduction

The Web of Data, the so-called Internet of Things and the emergence of Smart Cities are changing the whole regulatory framework of the Rule of Law.[3] Traditionally, the Rule of Law conceptualizes the principle that tyranny and totalitarian forms of government should be excluded from ruling a social body. In the legal tradition, this expression is defined as the set of practices, norms, rules and principles that allow the functioning of the market and social bonds —the civil society— while securing justice. This is what Hans Kelsen, Herbert Hart, Alf Ross or, to mention the American side of the Rule of Law, Roscoe Pound or Karl Llewellyn intended to do: *The restriction of the arbitrary exercise ofpower by subordinating it to the scope of well-defined laws.* And contrary to what is believed, after World War I and II, jurists and legal philosophers were very aware of what they were carrying out *collectively* within the notions of *primary* and *secondary* norms or rules, and *legal systems.*[64] But fifty years later, as already noticed by several legal scholars — e.g. Lawrence Lessig (2001, 2006), when we think of the substance of "laws" in the digital world, we have to think not only of legal systems but of standards, protocols and technical languages as well, what LESSIG terms "code". As we will see later on, we should add *ethics* as well to this list.

This paper intends to come to terms with the new ways of regulating the contemporary societies that have emerged through the Web 2.0 (the Social Web) and the Web 3.0 (the Semantic Web).[65] It points at seeking a new conceptualisation for rights and the Rule of Law, as well, based on the social and political transformation that has occurred in the past ten years. These are preliminary thoughts, still to be fleshed out with a deeper insight. But, still, pouring new wine into old wineskins has some risks. In law, you never start from scratch.

The paper singles out four topics relevant for this cultural change: (i) data and metadata to structure the flow of information, (ii) social intelligence and crowdsourcing, taking into account the collective properties both of human and computational cognition; (iii) formalisation of languages of law, making norms and rights manageable, (iv) security

[64] Put it broadly, it was clear for them that the powers of the state had to be controlled through norms or rules that brought about competences, constituted institutions, and could monitor and control the dynamics of normative changes.

[65] The Web 2.0 includes services, platforms and applications, end-users, prosumers (both producers and consumers of information), citizens, and social networks that constitute the grassroots of the new digital neighbourhood. The Web 3.0 includes the methods, languages and computer devices that allow turning content - the information spread over the web - into structured information, that is, into shareable and reusable knowledge.

and privacy to protect individuals and communities from ancient and new threats (to prevent violence across the Web).

2. Big Data

We may understand big data: (i) as a magnitude, (ii) as an attitude, (iii) as a cultural and organizational shift. What are we talking about? Every second 30.000 gigabytes of data (1 gigabyte= 10^9, 1.000 000.000 bytes) are shed on the web (Marz; Warren, 2014). Twitter daily generates half a billion of tweets. We are able to process and analyze daily about 50 millions of tweets, extracting patterns and trends using schemas and "memes" of information (Asbagh et al., 2014). The last Gartner *Hype Cycle for Emerging Technologies,* delivered in August 2014, does not locate *big data* at the peak of emerging expectations, but on the edge of already known and yet non-mature technologies. The Internet of Things substitutes big data at the peak position, but *Data Sciences* are coming up as new emergent sciences.[66] Huge amounts of data are produced daily through the sensors of smartphones, automatically sending information regardless the will of their owners.[67] The speed of mobile technology, taking off and outnumbering personal computer regarding users in 2008, is one of the trend application topics. Emerging new political notions such as *digital neighbourhood*, *crisis mapping*, and *political crowdsourcing* would not have been possible without it (Heinzelman; Brown and Meier, 2011; Poblet, 2011; Poblet and Casanovas, 2012; Poblet, 2013; Poblet; Noriega; Plaza, 2014).

Accordingly, regulations are switching forms and manners. The difference lies on the regulation of data (actions, intentions, results ...). The past way of ruling assumed a simple ontology, where human knowledge could be treated as separated knowledge *about* human behavior (be understood as experience or as external behaviour). Now, the structuring of data by means of metadata incardinate action *and* knowledge at the same time in a more complex dynamic flow (action, knowledge, shared knowledge, meta-knowledge) in real time: i.e. it is endowed through an *intelligent* flow. This is called *Open Source Intelligence*, *Crowd intelligence* or *Social Intelligence* (Poblet et al., 2014). I will not go through the differences now (Casanovas, 2014). While crawling the web, this flow can be spotted and situated according to its granularity: a single individual, a group, a community, and the interrelations among them. The important point is turning information into *knowledge*, and to managing and monitoring this knowledge we should make some distinctions.

1. We should distinguish at least between three types of languages expressing knowledge:
2. Natural language, (ii) technical (expert) language, (iii) formal language. Expert language is most needed, as rules and norms are usually formulated in natural languages (English, Spanish, French...). Formal language is the only one that machines can understand. Sometimes, all three kinds of language are put together to convey content. E.g. Creative Commons licenses incorporate a "three layer design" to make them more comprehensible and ease their usage - legal code, human readable, machine readable.[68]
3. We should distinguish *semantic* metadata (human or automated annotations added to the content) from *structural* metadata. The latter adds information about *creation, purpose, origin, time, author, location, network, language and data standards.* Metadata is data that refer and describes data. As they are defined by the W3C, they have the feature of being automatable: for the Web, metadata is machine-understandable information, expressible into a programming language.[69]
4. We should also distinguish scientific and technological achievements, from their usages, functions and roles. I am not saying that big data is neutral. What I am contending is that we should calibrate that tools are used and situated in very different contexts and courses of action. They might foster participation and *digital awareness* (empowerment of people), and at the same time they might bring about more control (power over people, monitoring their flow of information).

3. The "Giant Global Graph"

Two examples come to my mind: (i) military uses of metadata for security issues (ii) DBpedia.

David Cole at *New York Review of Books* (2014) a reliable source, refers to it crudely:

Of course knowing the content of a call can be crucial to establishing a particular threat. But metadata alone can provide an extremely detailed picture of a person's most intimate associations and interests, and it's actually much easier as a technological matter to search huge amounts of metadata than to listen to millions of phone calls. As NSA General Counsel Stewart Baker has said, "*metadata absolutely tells you everything about somebody's life. If you have enough metadata, you don't really need content"* When I quoted Baker at a recent debate at John Hopkins University, my opponent, general Michael Hayden, former director

[66] http://www.gartner.com/newsroom/id/2819918

[67] At least: Light, Proximity, Two cameras, Three microphones (ultrasound), Touch, Position (GPS, Wifi - fingerprint, Cellular -trilateration-, NFC, Bluetooth - beacons-), Accelerometer, Magnetometer, Gyroscope, Pressure, Temperature, Humidity.

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

[69] http://www.w3.org/Metadata/ Metadata is machine understandable information for the web.

of the NSA and the CIA, called Baker's comment "absolutey correct", and raised him one, asserting, "^e *kill people based on metadata* (Cole, 2014, *online).*

Thus, metadata triggers action, as data does. This means that it has semantic content, after all.

A well-known non-military example stems from Wikipedia. Big data acquire much more sense when it comes to massive publication of linked data. It is the so-called Linked Open Data [LOD] project. Today, everybody uses Wikipedia to find information. Wikipedia is the seventh most popular website in the world.[70]

Since 2007, there is a DBpedia project linking databases according to the best practices and guidelines of the W3C, and building a large-scale, multilingual knowledge base by extracting structured data from Wikipedia editions in 111 languages. The largest DBpedia knowledge base which is extracted from the English edition consists of over 400 million facts that describe 3.7 million things. The knowledge bases that are extracted from the other 110 editions in other languages consist of 1.46 billion facts and describe 10 million additional things.[71]

References are tied using Semantic Web languages, especially *Resource Descriptive Framework* [RDF]. The search language is SPARQL, *Protocol and RDF Query Language* [SPARQL], currently being drawn 3,000 million triples — subject / object / relation in all natural languages — describing some four and a half million objects.

In 2011, another sister project was put in place, *Wikidata*, "a free linked database that can be read and edited by both humans and machines"[72] containing more than 14.000.000 data editable items (June 2015) in all Wikimedia languages. Wikidata aims at provide statements given in a particular context.[73]

The second Semantic Web generation is already known by the promotion of the Semantic Web Linked Data Project to achieve the objective of "a single global data graph", or what Tim Berners-Lee describes as the "Giant Global Graph".[74 75]

This idea is still far from real. But a new visualization of the State of the LOD cloud was published on April 24th 2014. There is an increase of 271% compared to 2011; significantly, the field that has experienced a larger growth (306% corresponding to 199 sets of large databases) corresponds to data made public by governments that are following a policy of transparency *(Linked Government Open Data)}*[6] The size of the circles indicates their valency, the degree or number of edges incident to the vertex, as shown in the graph plotted by Schmachtenberg et al. (2014) (Figure 1).

In this last version of the graph, they had to add the category of social networking to the previous topical categories (media, government, publications, life sciences, geographic, cross-domain, user-generated content). Social networking is by far the largest category (520 datasets, 48% of all datasets) (Schmachtenberg; Bizer; Paulheim, 2014).

4. Rule and Meta-Rule of Law

The functioning of social intelligence and rights — what agents or humans can, might, must, or must not do with regard to each other — are connected. Rights matter, and should not be bartered with other kind of interests. Languages to express, manage and operate rights through the Semantic Web, Big Data and the Internet of Things are key to understand the normative side of the web, and how it can evolve, for the good or the bad.[1] Likewise, rights can be modelled and designed into electronic institutions, which are able to buy, sell or auction goods, and enter into disputes or mediate between contenders.[76 77]

70 After Google, Facebook, Youtube, Yahoo, Baidu and Amazon. Vid. http://www.alexa.com/topsites (June 2015).

71 For a full explanation (raw-based Infobox extraction, ontologies, NLP, etc.) see Lehmann et al. (2012).

72 https://www.wikidata.org/wiki/Wikidata:Main Page

73 "Rather than stating that Berlin has a population of 3.5 million, Wikidata contains the statement about Berlin's population being 3.5 million as of 2011 according to the German statistical office." (Lehmann et al., 2012, p. 23).

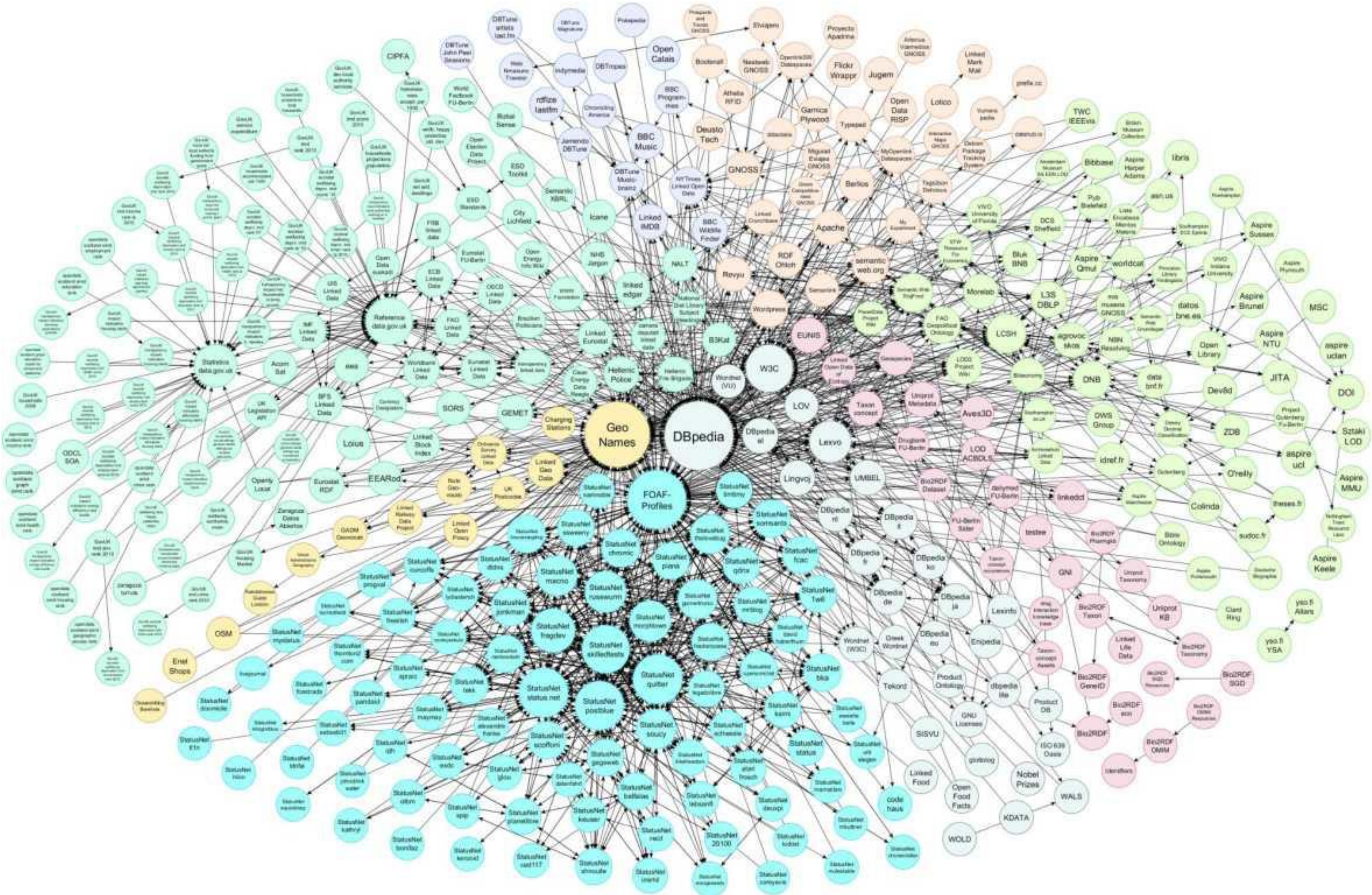
74 http://dig.csail.mit.edu/breadcrumbs/node/215

75 http://lod-cloud.net/

76 Cfr. e.g. about licenses (Governatori et al., 2013).

77 Cfr. Rodriguez-Doncel et al. (2013a). Linked data rights ontology was released on 1st September 2014: http://oeg-dev.dia.fi.upm.es/licensius/static/ldr/. For electronic institutions and social intelligence, see the results of the EU coordination action SINTELNET: https:/ /ec.europa.eu/digital-agenda/en/news/sintelnet-european- network-social-intelligence.

Figure 1. Linking Open Data cloud diagram 2014.



Media (Geographic Publications User-Generated Content Government Cross-Domain Life Sciences Social Networking Crawlable Linked Datasets as of April 2014)

Source: Schmachtenberg et al. (2014).

All of this has fostered new ways of looking at the patterns and rules on digital content. Creative Commons (CC), Linked Open Data projects (LOD), Linked Open Government Data (LOGD), Open Science (OS), blockchain technologies *(smart contracts*, self-enforcing digital contracts)[78], Free Access to Law (FAL)[79], MetaLex[80], AkomaNtoso[81], OASIS standardization efforts[82], among other trends, try to counterbalance the pervasive pressure that the uncontrolled management of increasingly structured information is putting on our lives. Open Rights is certainly a metaphor. But it points at the *global ethical dimension* of transparency, understandability and shareable values that could be added to the norms and regulations that operate on the web through data and metadata:

1. Dialogue, and not justpower, is emerging as a source of law across technology. People, *we the people*, have a new opportunity to take the floor.
2. Information principles can be embedded into the making of this new digital society.
3. *Privacy by design, data protection by design, security by design* are other terms used to express the construction of a new Rule of Law, or Meta-Rule of Law, comprising humans and programs, rights and languages, alike.
4. We have to face in the next years the management of a new self, a personal identity which is complex, plural, multidimensional and durable on the Web.[83]

5. Rights Expression Languages (RELs) for the Web of Data

The idea of Open Rights is not just a metaphor, for rights can be structured into conceptual automatable schemes. "The topic of rights expression is coming up nearly everywhere that metadata is used to describe digital resources. [...] RELs themselves do not act on digital content, they need to be used in systems that implement the rights management that they express" — Karen Coyle wrote in 2004 for the Library of Congress. There is with no surprise that second generation RELs is being developed ten years later, in parallel with the Web of Data.

Rights Expression Languages (REL) are technical languages that have specific syntax (grammar) and semantic vocabulary rules for expressing what kind of uses are permitted, forbidden or obligatory. Reuses and transfers of the content, actions such as copy, print or play, type of users, time and space are the subject matter of RELs. They can be viewed as expression of copyright, contract or license agreements, control over access and/or use. RELs have been developed within Digital Rights Management (DRM) technologies to express machine readable relationships at different level of depth, although DRM deployment is closer to patents and turf battles of rights holders companies (Coyle, 2007).

This is not a new idea. But, like computational ontologies, REL are not neutral and emerged from real market needs, reflecting market constrictions, conflicts and potentialities at the same time. These practical origins should be taken as they are, because each big company tried to develop in business its own way of modelling rights.

As ontologies, 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).[84] Permissions and restrictions can be modelled according to Creative Commons principles (ccREL), Open Digital Rights Language (ODRL)[85], MPEG-21[86], 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 (Rodriguez-Doncel Et Al., 2013; Rodriguez-Doncel; Gomez- Perez; Mihindukulasooriya, 2013).

6. REL story so far

There is an interesting story to be told here, because companies developed RELs to implement and enforce *their own policies:* what purchasers of digital goods (entities or services) could, could not, ought or ought not do with what they were paying for. In North-American law, e.g., especially after the *Digital Millennium Copyright Act* (1998), the extension of the first sale doctrine[87] and the exhaustion principle to

[78] A blockchain is simply a chronological database of transactions recorded by a network of computers.

[79] http://www.worldlii.org/ For an explanation, see Greenleaf, Chung And Mowbray (2015).

[80] http://doc.metalex.eu/

[81] http://www.akomantoso.org/

[82] https://www.oasis-open.org/. see esp. http://www.legalxml.org/governance/

[83] See a brief description in Casanovas et al. (2014).

[84] For the whole story of RELs in the nineties (Katz, Stefik, Iannella, etc.), see Jamkhedkar, Heileman (2009).

[85] https://www.w3.org/community/odrl/

[86] The MPEG-21 standard has been accepted as ISO 21000. Part 5 of ISO 21000/MPEG-21 contains the Rights Expression Language. ISO 21000/MPEG-21 Part 6 provides a structure for a data dictionary for the REL.

[87] The first sale doctrine starts the distribution chain of purchased products: it entails one exception to the copyright owner's distribution right. Once the good is legally sold, the buyer can dispose of it at his own ease. He can take care or destroy it because the copyright owner's has already satisfied his right, and cannot prevent the buyer to behave normally in the market. On the contrary, what continues is the copyright owner's reproduction right. It is forbidden to make copies. The four normative rationales for the first sale doctrine are access, preservation, privacy, and transactional clarity.

digital goods was alleged to make a balance between copyright and consumer's protection, via restrictive licensing. But actually if every possible digital product incorporates a self-executable license to protect it, then Digital Rights Management (DRM) technologies - the reaction of industry against piracy - enables content publishers to enforce *their own access policies on content* (not only restrictions on copying or viewing, but executing, printing, altering of works or devices, etc.).

This would foster the emergence of normative scenarios in which, contrary to the Kelsenian legal "closing rule" - all what is not expressly forbidden is permitted - the assumption could be expressed as follows: all what is not expressly permitted is forbidden (Moscon, 2011). Some authors have labelled the situation as monopolistic, and even "feudal".[88]

There are several interesting interrelated features in this broad regulatory landscape, embracing *all* digital goods, and reaching *all* micro-situations (relations between the enduser, the subject of the legal act of selling - or renting, leasing, etc., the copyright owner, the ISP, and eventually the company).

1. The first one, early noticed by Benkler (2001), is the competition for the creation of a new institutional ecosystem, in which rights have been expanding their scope to practically all interactions inside and outside the market.[89]

2. The second is the pragmatic structure of this normative world implementing intellectual property rights through and within licenses: rights can be enforced in a close secure closed environment ("trusted platform") or on the open Web, and companies prefer the former "controllable" scenario.

3. Thus, tight competition among companies to impose their own solutions as de facto standards increased the lack of interoperability between licenses.

Interoperability is not the whole, but a part of the story. What is really at stake is the consumer liberty of choice, and the right of making decisions about personal rights. As stated by Hiram Melendez-Juarbe (2009, p. 194) *"while flexibility does not necessarily follow interoperability, interoperability may follow flexibility"*. The personal use of licenses might entail the need for interoperable and eligible licenses, not the other way around. Interoperability, per se, does not constitute a solution for the social effects of the expansion of rights. Actually, nothing prevents that interoperability alone would enhance more control than freedom.

Moreover, after DMCA enactment in 1998, the judicial interpretation of consumer laws entered into play.[90] According to some observers, Courts are well-equipped to limit copyright exclusivity, enabling copy owners to make traditionally lawful uses of their copies, including resale through secondary markets (Perzanowski; Schultz, 2011). But a clear definition of digital first sale doctrine is still pending. Others defend that tangible and digital goods are radically different, and therefore propose dropping the first sale doctrine because licensing framework provides an alternative to a digital first sale (Tobin, 2011).

This is not an ideal situation. Consumers' reaction count as well. For instance, as Iwahashi (2011) illustrates, REL can be used to monitor and control consumers' affordances. Let's reproduce his example of consumers' resilience.

The Moving Picture Expert Group Rights Expression Language (MPEG REL) works by associating an XML header (extra metadata) with each file that is controlled by MPEG REL. The header contains a standardized definition of the rights associated with the file for the user. Each copyrighted file is stored as data on the user's computer, but with a MPEG REL header attached. Thus, since its launching in the market, Apple iTunes managed to make that a rented movie is automatically deleted thirty days after it is downloaded, or twenty-four hours after the user started watching it. But people invented methods to face this limitation.

Mechanism for online movie rental protection can be circumvented using a few different methods. An early circumvention technique to extend the length of movie rentals has since been fixed, but it makes an interesting circumvention example. Before renting a movie, the circumventor would set his computer clock ahead by about twenty years. He would subsequently rent the movie and start viewing it and then set his clock back to today's date. This

[88] "The current system of digital property transfer disenfranchises consumers and inevitably creates a monopoly on the distribution of digital materials, since only the original distributer retains the right to sell. (…).Today's digital property transactions resemble a feudal system in which the digital copyright owner is able to dictate the terms and overall use of the property to the end user through use of contracts of adhesion" (Richardson, 2014, p. 196).

[89] According to Benkler (2001, p. 85-86): "[…] the effort to define the new parameters has meant a struggle over intellectual property rights. In the U.S., we have seen a vast expansion of rights in multiple dimensions. The term of copyright was lengthened. Patent rights were extended to cover business methods. Trademarks were extended by the Federal Anti-Dilution Act of 1995 to cover entirely new values, becoming the basis for liability in the early domain-name trademark disputes." […].."Only companies whose business models depend on licensing rights reap the benefits of strong rights. Everyone else simply has to pay higher prices for input" (ibid.).

[90] Under 17 U.S.C. 5 1201(a)(1)(A), "No person shall circumvent a technological measure that effectively controls access to a work protected under [the Copyright Act]" (Iwahashi, 2011, p. 491).

made the rental period last for twenty years instead of the typical thirty days (Iwahashit, 2011, p. 507).

Apple dropped DRM from all iTunes music files in 2009, letting consumers transfer tracks between computers or onto their mobile phones. Strong property rights defence proved to backfire on copyright owners. Economic analysis of law confirms this assertion. Control itself might become a valuable good. In some situations, copyright owners might prefer liability-rule protection for users to liability-rule protection for owners. "*The parties preferences across rules depend on the value they place on control itself* (DiCola, 2015, p. 666).

7. Rights in the public domain

Thus, people reacted.[91] Against the private framework created by company-driven DRM, other successful initiatives followed, proposing to redefine the public digital space through the empowerment of end-users. The most popular initiative is Creative Commons, with millions of people using its licensing system.[92] When it comes to the computational structure of rights, Open Digital Rights Language has been equally successful. Fifteen years ago, it represented a shift and a tipping point as well. In a seminal W3C position paper, Renato Iannella wrote in 2001.

Traditional DRM (even though it is still a new discipline) has predominately taken a closed approach to solving problems. That is, DRM has primarily focused on the content protection issues more than the rights management issues. Some argue that this skew in emphasis towards content protection diminish the rights of the end users, as well as content creators. Hence, we see a movement towards 'Open Digital Rights Management' (ODRM) with clear principles focused on interoperability across multiple sectors and support for fair-use doctrines (Iannella, 2001, p. 1).

Within the Web of Data, the idea that each digital resource may be accompanied by a description of ruling metadata (description of the rules governing its use) gains strength. There is an active W3C ODRL Community Group at W3C, with the aim of developing and promoting an open international specification for Policy Language expressions.[93] ODRL was adopted by the Open Mobile Alliance as the standard REL for mobile content in 2004, but it has not become a W3C standard yet.

ODRL is flexible, based on the extensible Markup Language (XML), and therefore modifiable, without any specific DRM software to use it. *It is a language to express rights.* It does not control access, but usage, compatible with other systems to control identification and validation of users. ODRL has a data dictionary with rights (expressed through permissions) and their limits (expressed through context, constraints, and requirements). This data dictionary can be easily modified to satisfy new situations and consumers needs.

REL have received some criticisms too, some of them involving their complexity and applicability in DRM systems,[3] while others referring to the fragmentation of multiple scenarios and the problem of reusability of ontologies (Nadah; Dulong De Rosnay; Bachimont, 2007) or its extension to Web services (Gangadharan; Weiss, 2007; Gangadharan, 2009). It is a common *motto* among jurists that formal languages — such as REL— are not able to express the subtleties of legal language. But I don't think they have to. Mimicking the pragmatic use of natural language in specific cases and settings is not the main objective of making tools for the Web o Data.

Both ccLL and ODRL have tackled the problem of complexity and reusability in a different way. ccLL are user-centred, the user is asked to chose between different types of pre-established licenses. This is an *external* point of view, in which end users are defined as participating in an outer context, taking control of the content through CC licensing. On the contrary, ODRL situates itself in an *internal* point of view, in which the inner context is created by incorporating users into a simplified controlled and abstract structure of formalised rights.

Given the abstract nature of language, several solutions have been proposed to improve interoperability and consumers' empowerment. In a way, the Web of Data has helped to sharpen the ideas of scalable abstraction, composition, and more abstract and simple ontology design patterns (ODP)[94] [95] to ease reusability. There are several works and ongoing research strategies to composite, evaluate or improve the compatibility of multiple datasets licenses from different fields (argumentation theory, NLP, deontic logic, service licensing, etc.).

---

[91] See Edwards et al. (2013) for the copyright story in UK and Europe, claiming for the public interest. "A democratic copyright policymaking process must accommodate the modes of justification offered by users to allow copyright law to reconnect with the public interest goals at its foundation." (ibid. 2013, p. 9).

[92] As it is well-known, Creative Commons was founded in 2001 by Lawrence Lessig, Hal Abelson, and Eric Eldred. It is a non-profit organization, with more than 800 million people using its licensing system.

[93] https:// www.w3.org/community/odrl/

[94] "[...] current RELs are too complex, and lack a manageable standard partitioning of their functionality that would allow them to be more easily incorporated into DRM applications" Jamkhedkar; Heileman; Martinez-Ortiz, 2006, p. 60; also, Jamkhedkar; Heileman 2009).

[95] ODP are based on "knowledge patterns": "small, well connected units of meaning which are 1) task-based, 2) well-grounded, and 3) cognitively sound" (Nuzzolese et al., 2011, p. 520). An "Ontology Design Pattern" is a "reusable successful solution to a recurrent modeling problem" (Presutti, 2012, slide 7). Cfr. Gangemi (2007).

I will just mention two of them, (i) a single License structure connected with a general vocabulary repository based on the idea of *normative compliance* (Governatori et al., 2013),[96] and (ii) a ODRL ontology, connected with a general ODRL model of governance and the proposal of a general Licensing Ontology Design Pattern. Both proposals are focusing on Linked Open Data.

The first initiative is based on a deontic logic solution to reconcile a set of licenses associated to heterogeneous datasets. A composite license can rely on a deontic logic semantics (permission, prohibitions and obligations) which is compliant with the normative semantics of each single license composing it. This can be called *compliance by design,* as a policy-driven strategy to make policies and personal decisions compliant with the particular law governing each license.

8. Licensed Linked Data Resources (LLDR)

The second initiative is centred on the ODRL ontology, released on March 2015, within the general framework of ODRL governance, and the related attempt to produce a Licensing Ontology Design Pattern (a conceptual scheme to be reused as the kernel or template to Licensing further modelling). This vocabulary defines 24 classes,[97] 5 6 properties, one concept scheme *(actions),* 61 concepts[98] and 18 named individuals (Mcroberts, Rodriguez-Doncel, 2015).

This is a selected vocabulary, intended to cover all areas and processes in which licensing is involved. But the REL structure that lies at the centre of it to facilitate the rights m anagement of linked database is quite simple (Figure 2). Rodriguez-Doncel et al. (2013) built up a content (non-logical) design pattern out of it[40].

The intent of the content pattern *Licence Linked Data Resources* (LLDR) is to represent the relation that exists among a rights expression, an action, an agent, a LD resource and a condition. In particular, the core idea of the pattern is to model: *a rights expression which allows/prohibits/obliges to make an Action (Right) to an Agent over a LD resource under a condition*" (Rodriguez-Doncel et al. 2013, p. 1).

Some properties should be highlighted: (i) the pattern combines the commonalities of the most six used licenses (among them ccLL, MPEG-21 REL and ODRL), (ii) it expresses nary relations, (iii) the *LinkedDataRights* is a superclass representing the applicable rights to Linked Data resources, (iv) rights expressions appear naturally in groups and not separately (licenses and typical authorizations are actually aggregations of atomic rights expressions); (v) the aggregation relationship can be represented in OWL using a *partOf-whole* relation pattern.

Still, this has to be populated with an extended vocabulary (e.g. massive repositories such as LIMO[41]), and there is room for improvement. LLDR is under review.[99] But it is clear that publishing REL along with the digital asset (database) allows (i) an easier handling and monitoring of the rights involved in the relation between data and end users, (ii) a faster and cleaner implementation of policies, (iii) a consistent compliance with rights to be performed and duties to be respected.

We should distinguish (i) the performative act of *informing about rights* that is accomplished by conveying the content of rights as metadata; (ii) from the performative act of *qualifying a relation as legal* by selecting a specific license that qualifies the social bond between the end user, the database, and the data owner.

As we will contend in the next section, the possibility of carrying out extended and traceable (provenance) sequences of legal acts through metadata is not only a technological application, but it challenges the relationship between law and policies. Transparency and public accountability are key (Raines, 2012-2013). The network of legal relations is made explicit and transparent, redefining the public space, and by the same move, requiring a

---

[96] Those are the research questions: " i) How to express the deontic component of the licensing terms in a machine-readable format?, and ii) How to compose in a compliant and automated way the licensing terms associated to a set of heterogeneous data to produce a single composite license?" (Governatori et al., 2013, p. 152).

[97] As example: Action | Agreement | All | All Connections | All Groups | AJU2ndConnections | Asset | ConflictTerm| Constraint | Duty | Group | Individual | Offer | Operator | Party | Permission | Policy | Privacy | Prohibition | Request | Rule | Set | Ticket | Undefined Term. Cfr. http://www.w3.org/ns/odrl/2/ODRL21

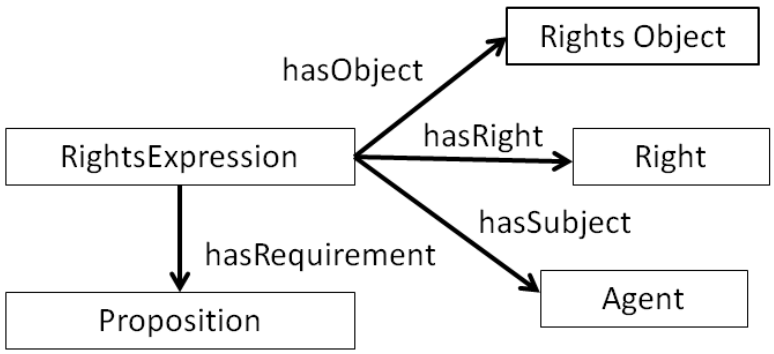
[98] acceptTracking | adHocShare | aggregate | annotate | anonymize | append | appendTo | archive | attachPolicy |attachSource | attribute | commercialize | compensate | concurrentUse | copy | delete | derive | digitize | display | distribute | ensureExclusivity | execute | export | extract | extractChar | extractPage | extractWord | give | grantUse | include | index | inform | install | lease | lend | license | modify | move | nextPolicy | obtainConsent | pay | play | present | preview | print | read | reproduce | reviewPolicy | secondaryUse | sell | share | shareAlike | textToSpeech | transfer | transform | translate | uninstall | use | watermark | write | writeTo Cfr. http://www.w3.org/ns/odrl/2/ODRL21

[99] Cfr. especially Timothy Lebo's review (2013), pointing at: (i) the "core three" properties hasObject, hasRight, hasSubject, (ii) the use of OWL to allow reasoning; (iii) the use of Directed Qualification Pattern (instead of n-ary relations) to facilitate the tracking (provenance) between agents and the database; (iv) the suggestion of incorporating factual violations of rights (and not only rights) to the pattern. About semantic provenance and best practices, see Moreau et al. (2015).

redefinition of what we understand by the Rule of Law.

Metadata centres the Rule of Law on the expression, representation, and performance of individual and collective rights. Standards (or adapters), protocols, recommendations, and behavioural patterns do not relate primarily to power-grounded norms released by the authority of the State, but to pluralistic, company or community-based conceptual models and rules created by means of computer languages. This adds complexity to legal systems. I will call this new regulatory semantic layer *Meta-Rule of Law*. But, as stated above, this is not happening without tensions affecting rights holders - between DRM and REL, between private and public licensing, between companies and consumers, between states and other kind of organisations.

Figure 2. Relationship expressed in Rights Expression Languages



Source: Rodnguez-Doncel et al. (2013); Rodnguez-Doncel; Gomez-Pérez; Mihindukulasooriya, 2013).

9. ODRL Governance

Linked Data Best Practices push for posting licensing information as a criterion for quality: [42]

Do you provide licensing metadata? Web data should be self-descriptive concerning any restrictions that apply to its usage. All Linked Data published on the Web should include explicit license or waiver statements. A common way to express such restrictions is to attach a data license to published data. Doing so is essential to enable applications to use Web data on a secure legal basis (Mendes et al., 2012, p. 2).

There is still a long way to go before these nuclear structures can be successfully and broadly implemented into linked databases, web-services, and digital assets. All surveys conducted so far indicate that few databases make use of licenses.[100] Schmachtenberg, Bizer And Paulheim (2014) found recently that only 9,96% of all reported datasets (DBpedia) provided licensing information in RDF. The provision of information varies widely across topical domains. More than a third of all government datasets provide licensing information, while none of the geographical ones does.

But in spite of this slow taking-off, the legal-apex of licensing (or related instruments) being systematically added to digital assets will be the dominant practice in the next future.

The W3C ODRL Community Group, lead by Renato Iannella, aims at developing and promoting "an open international specification for Policy Language expressions". Figure 3 shows the underlying ODRL Policy governance model (do notice that it targets the business models of companies, organisations, administrations, and state agencies.

For the sake of clarity, I am not reproducing here the last version of the model, but a simpler one.[101]

Policy holds ODRL policy together. "In its encoded form, e.g. in an XML document, it makes the policy addressable from the outside world via its unique identifier (uid) attribute". It is worth noting that Permission, Prohibition and Duty are introduced as subclasses of *Rule,* a superclass, to avoid the "redundancy of having very similar, but separately developed classes in an application's source code".

[100] "Conclusions? Few documents provide licencing information directly as part of the document meta-data. Further still, there is a palpable need for (i) an agreed-upon licencing property, and (ii) an agreed set of common licence URIs; to avoid consumers again having to hard-code support for all alternatives used by publishers. The most complete proposal along these lines is provided by the Creative Commons vocabulary" (Hogan et al., 2012, p. 32 and ff.)

[101] The reader can find the last version (2.1) at https://www.w3.org/community/odrl/model/2.1/. I thank Renato Iannella for letting me know this recent updating of his work. The Global Standards Body of News Media (ITPC) has recently adopted ODRL as standard: https://iptc.org/standards/rightsml/

R:ule implements the ODRL core model. What kind of reasons lay behind this solution?

This is a computational engineering diagram that is not stemming from deontic logic nor legal philosophy. Jones' and Sergot's deontic spaces, or Hohfeld's fundamental concepts are far from it. Permission, Prohibition and Duty work as simple linguistic operators or computational triggers for action, not as logical functors, to avoid undesired effects and make the licensing system feasible, the payments effective, and governance flexible. All kind of policies might be introduced to extend the model under the super or upperclass *Rule.*

But nothing guarantees that *Rule* work in a democratic way, or according to ethical principles, or according to constitutional values. This is where the notion of Meta-Rule of Law is needed and comes into play to shape *institutionally* what can be achieved through the implementation of REL.

10. Meta-Rule of Law

The development of REL by means of Ontology Design Patterns such as LLDR, or Open Licenses sustained by Policy Models such as ODRL, situates the discussion on metadata at the regulatory level. Self-executive metadata consisting of automated rights and duties as particular actions constrained by defined conditions might behave as Lego or Minecraft building blocks. How these blocks ought to be organized to became *legal*?

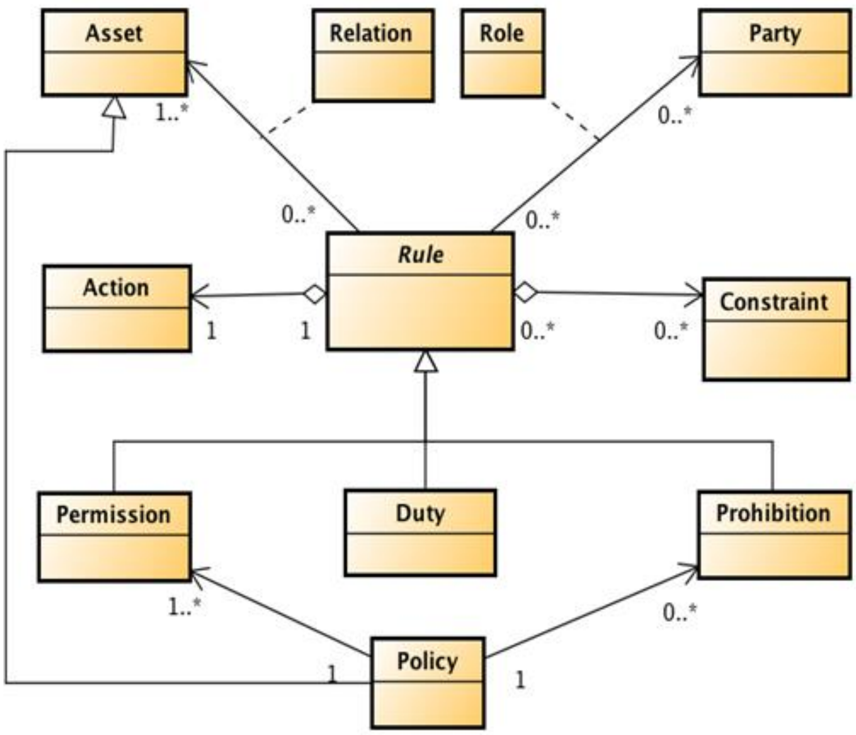
Validity or *legality* of policies, contracts, norms or standards is a complex propriety that cannot be taken for granted. *Rule* or *LinkedDataRights* (in LLDR) do not convey legality *per se.* Compliance by design populating systems by means of massive vocabularies do not guarantee either that actions can be considered *legal*.

This is the domain of Meta-Rule of Law that should be worked out. Pointing at unique identifiers (uid) to connect REL with the outside world is not enough. There is no point at replicating meaning in an ostensive way.

I have introduced elsewhere the notions of *normative* Semantic Web Regulatory Models (nSWRM), and *institutional* Semantic Web Regulatory Models (iSWRM) to perform this kind of connection (Casanovas, 2015). Global ethics, hard and soft law, policies, and standards, should be addressed in a theoretical way as components of *intermediate* institutions to build the public dimension of a more consistent society.

The Rule of Law, then, *the restriction of the arbitrary exercise of power by subordinating it to the scope of well-defined laws,* could be internally fleshed out. There is no need for change the fundamental notion. But we should take care of the restriction of the arbitrary exercise of power through data by subordinating it to the scope of well-defined metadata.

Figure 3. ODRL Abstract Policy Model.



Source: Iannella (2012).

Acknowledgments

This research has been partially funded by the following related projects. RTC- 2014-29467: Crowds eAssessment, *Ministerio de Economia y Competitividad (MEC), Gobierno de Espana; RTC-2014-2666-7:* Info_COR Modulo de servicios para la implementation del equilibrio entre empresa y consumidor en la resolution online de reclamaciones de consumo; *MEC; DER2012- 39492-C02-01* Crowdsourcing: instrumentos semanticos para el desarrollo de la participation y la mediation online, *MEC.*

References

1. Asbagh, M. J. et al. Clustering memes in social media streams. Social Network Analysis and Mining, v. 4, n. 1, p. 237, 2014.
2. Benkler, Y. The battle over the institutional ecosystem in the digital environment. Communications of the ACM, v. 44, n. 2, p. 84-90, february/2001.
3. Casanovas, P. et al. Semantic Web: a Legal Case Study. In J. Davis, R. Studer, P. Warren (Ed.) Semantic Web Technologies: trends and research, Chichester: Ed. John Wiley & Sons, p. 259-280, 2006.
4. Casanovas, P. et al. OPJK and DILIGENT: Ontology Modelling in a Distributed Environment, Artificial Intelligence and Law. Vol. 15, Issue 2 (June 2007), p. 171-186., 2007.
5. Casanovas, P. Open Source Intelligence, Open Social Intelligence and Privacy by Design. In: European Conference on Social Intelligence, 2014, Barcelona. Proceedings of the European Conference on Social Intelligence. Barcelona: CEUR-WS vol. 1283. 2014, p. 174-185. Available at <http://ceur-ws.org/Vol-1283/paper_24.pdf>. Accessed on June 25th 2015.
6. Casanovas, P. Semantic Web Regulatory Models: Why Ethics Matter. Philosophy and Technology. Special Issue on Information Society and Ethical Inquiries, v. 28, n. 1, p. 33-35, 2015.
7. Casanovas, P.; Moreso, J. J. (Eds.). El ámbito de lo jurídico. Barcelona: Ed. Crítica, 1994, 2000.
8. Casanovas, P. et al. Law, Social Intelligence, nMAS and the Semantic Web: an Overview. In: Casanovas, P. et al. (Eds.). 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. LNAI 8929. Heidelberg: Springer, 2014, p. 1-10.
9. Casellas, N. Legal ontology engineering: methodologies, modelling trends, and the ontology of professional judicial knowledge. Vol. 3. LGT Series, Heidelberg: Springer Science & Business Media, 2011.
10. Coyle, K. Rights Expression Languages. A Report for the Library of Congress. February, 2004. Available at <http://xml.coverpages.org/LOC-CoyleReportREL2004.pdf> Accessed on June 25th 2015.
11. ______. DRM and patents. Coyle's InFormation, February 19th 2007. Available at: <http://kcoyle.blogspot.com.au/search/label/DRM> Accessed on June 25th 2015.
12. Cole, D. We Kill People Based on Metadata. New York Review of Books, Blog NYRB, May 10th, 2014. Available at <http://www.nybooks.com/blogs/nyrblog/2014/may/10/we-kill-people-based-metadata/?insrc=wbll>. Accessed on June 25th 2015.
13. Dicola, P. Valuing control. Michigan Law Review, v. 113, n. 5, p. 663-710, 2015.
14. Edwards, L. et al. Framing the consumer. Convergence: The International Journal of Research into New Media Technologies, v. 19, n. 1, p. 9-24, 2013.
15. Gangemi, A. Design Patterns for Legal Ontology Construction. In: P. Casanovas et al. (Eds.): Trends in Legal Knowledge: The Semantic Web and the Regulation of Electronic Social Systems. Florence: European Press Academic Publ., 2007, p. 180-201.
16. Gangadharan, G. R. Licensing of Open Apis. Technology Innovation Management Review, April/2009. Available at < http://timreview.ca/article/243>. Accessed on June 25th 2015.
17. ______.; Weiss, M. An Introduction to Rights Expression Languages. Technology Innovation Management Review, December/2007. Available at <http://timreview.ca/article/105>, Accessed on June 25th 2015.
18. Governatori, G. et al. One License to Compose Them All. A Deontic Logic Approach to Data Licensing on the Web of Data. In: International Semantic Web Conference, 12, 2013, Sydney. The Semantic Web–ISWC 2013, LNCS, Berlin, Heildelberg: Springer, 2013, p. 151-166.
19. Greenleaf, G.; Chung, P.; Mowbray, A. Supporting and influencing data privacy practice: The free access International Privacy Law Library. Computer Law & Security Review, v. 31, n. 2, p. 221-233, 2015.
20. Heinzelman, J.; Brown, R.; Meier, P. Technology, Crowdsourcing and Peace Mapping: New Theory and Applications for Conflict Management. In: POBLET, M. (Ed.). Mobile Technologies for Conflict Management. Law, Governance and Technology Series, v. 2. Dordrecht, Heidelberg: Springer, 2011, p. 39-53.
21. Hogan, A. et al. An empirical survey of Linked Data conformance. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, v. 14, p. 14–44, 2012.
22. Iannella, R. Open Digital Rights Management. A Position Paper for the W3C DRM Workshop. In: Workshop on Digital Rights Management for the Web. World Wide Web Consortium, 2001, Sophia-Antipolis, France. Available at: <http://www.w3.org/2000/12/drm-ws/pp/iprsystems-iannella.html>. Accessed on June 25th 2015.
23. ______. ODRL Version 2.0 Core Model. Final Specification: 24 April 2012. Available at <https://www.w3.org/community/odrl/model/2.0/> Accessed on June 25th 2015.
24. Iwahashi, R. How to circumvent technological protection measures without violating the DMCA: an examination of technological protection measures under current legal standards. Berkeley Technology Law Journal, v. 26, n. 1, p. 491-526, 2011.
25. Jamkhedkar, P. A.; Heileman, G. Rights Expression Languages. In: Shiguo, L. (Ed.). Handbook of Research on Secure Multimedia Distribution. New York: IGI Books, 2009, p. 1-21.
26. Jamkhedkar, P. A.; Heileman, G.; Martínez-Ortiz, I.

The problem of Rights Expression Languages. In: DRM’06, 2006, Alexandria, Virginia, USA. Proceedings of the ACM Workshop on Digital Rights Management. NY: ACM, 2006, p. 59-68.

27. Lebo, T. Reviews: Tim Lebo about LicenseLinkedDataResources, Ontology Design Patterns, 2013/8/2. Available at: <http://ontologydesignpatterns.org/wiki/Reviews:Tim Lebo_about_LicenseLinkedDataResources>. Accessed on June 25th 2015.
28. Lehmann, J. et al. DBpedia: A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web, v. 1, p. 1–29, 2012.
29. Lessig, L. 2001. Code and Other Laws of Cyberspace. New York: Basic Books, 2001.
30. ______. Code. V.2. New York Basic Books, 2006. Available at <http://codev2.cc/download+remix/ Lessig-Codev2.pdf>. Accessed on June 25th 2015.
31. Marz, N.; Warren, J. Big data: Principles and best practices of scalable real time data systems. N.Y.: Manning Ed., MEAP, 2014.
32. Mcroberts, M.; RODRÍGUEZ-DONCEL, V. ODRL Version 2.1 Ontology, W3C, 2015-03-05. Available at: <http://www.w3.org/ns/odrl/2/ODRL21> Accessed on June 25th 2015.
33. Meléndez-Juarbe, H. DRM Interoperability. Boston University Journal of Science and Technology Law, v. 15, p. 181-219, 2009.
34. Mendes, P. N. et al. D2.1 Conceptual model and best practices for high-quality metadata publishing. 2012. Planet Data. Network of Excellence. EU Project FP7 – 257641. Available at <http://planet-data.eu/sites/default/files/D2.1.pdf>. Accessed on June 25th 2015.
35. Moreau, L. et al. The rationale of PROV. Web Semantics: Science, Services and Agents on the World Wide Web, p. 1-23, 2015 (in press). Available at: <http://dx.doi.org/10.1016/j.websem.2015.04.001> Accessed on June 25th 2015.
36. Moscon, V. Rights Expression Languages: DRM vs. Creative Commons. Italian Journal of Library and Information Science, v. 2, n. 1, p. 4593-4619. Giugno/June/2011.
37. Nadah, N.; Dulong De Rosnay, M.; Bachimont, B. Licensing Digital Content with a Generic Ontology: Escaping From the Rights Expression Language Jungle. In: International Conference on Artificial Intelligence and Law, 11. 2007, Stanford, United States. Proceedings of the 11th International Conference on Artifcial Intelligence and Law, NY: ACM, 2007, p.65-69.
38. Nuzzolese, A. et al. Encyclopedic Knowledge Patterns from Wikipedia Links, L. Aroyo et al. (Eds.), International Semantic Web Conference 2011, Part I, LNCS 7031, Heidelberg: Springer Verlag, p. 520–536.
39. Perzanowski, A.; Schultz, J. Digital Exhaustion. University of California Los Angeles Law Review, v. 58, n.4, p. 889-946, 2011.
40. Poblet, M. (Ed.) 2011. Mobile Technologies for Conflict Management. Online Dispute Resolution, Governance, Participation. LGTS n. 2, Dordrecht, Heidelberg: Springer Verlag.
41. Poblet, 2013. Visualising the Law: Crisis Mapping as an Open Tool for Legal Practice. Journal of Open Access to Law, v 1 n. 1. Available at:< https://ojs.law.cornell.edu/index.php/joal/article/view/ 12/13> Accessed on June 25th 2015
42. Poblet, M.; Casanovas, P. Crowdsourced crisis mapping: how it works and why it matters. The Conversation, May 16th, 2012. Available at: <https://theconversation.com/crowdsourced-crisis-mapping-how-it-works-and-why-it-matters-7014> Accessed on June 25th 2015.
43. Poblet, M.; Noriega, P.; Plaza, E. (Eds.) Crowd 2014: Crowd intelligence: Foundations, Methods and Practices. Proceedings of the Sintelnet WG5 Workshop on Crowd Intelligence: Foundations, Methods, and Practices. Barcelona: CEUR-WP, vol. 1148, 2014, p. 16-23, available at <http://ceur-ws.org/Vol-1148/> Accessed on June, 25th 2015.
44. Poblet, M.; et al. Open-Access Grant Data: Towards Meta-Research Innovation In: International Conference On Legal Knowledge And Information Systems, 27, 2014. Legal Knowledge and Information Systems. Foundations on Artificial Intelligence. Amsterdam: IOS Press, 2014, p. 151-156.
45. Presutti, V. Ontology design patterns: an Introduction. 2012. Available at <http://www.slideshare.net/ sssw2012/valentina-presutti-ontology-design-patterns-an-introduction>. Accessed on 25 june 2015.
46. Raines, J. The Digital Accountability and Transparency Act of 2011 (DATA): Using Open Data Principles to Revamp Spending Transparency Legislation, New York Law School Law Review , vol. 57, 2012-2013, p. 313-360.
47. Richardson, M. S. The Monopoly on Digital Distribution, Pacific McGeorge Global Business & Development Law Journal, v. 27, 2014, p. 157-171.
48. Rodríguez-Doncel, V. et al. License Linked Data Resources Pattern. In: Workshop on Ontology and Semantic Web Patterns, 4, 2013, Sydney. Proceedings of the 4th Workshop on Ontology and Semantic Web Patterns. Sydney: 2013. CEUR-WS n. 1188. Available at <http://ceur-ws.org/Vol-1188/>. Accessed on June 25th 2015.
49. Rodríguez-Doncel, V.; Gómez-Pérez, A.; Mihindukulasooriya, N. Rights declaration in linked data. Available at: <http://oa.upm.es/21512/1/ RodriguezDoncelEtAl_COLD2013.pdf> Accessed on June 25th, 2015.
50. Schmachtenberg, M.; Bizer, C.; Paulheim, H. Adoption of the Linked Data Best Practices in Different Topical Domains. In: International Semantic Web Conference, Riva di Garda, 2014. IISWC-13th Part I, LNCS, 8796, Dordrecht, Heidelberg: Springer, 2014, p. 245-260.
51. Schmachtenberg, M. et al. Linking Open Data cloud diagram 2014. Available at: http://lod-cloud.net/ Accessed on June 25th, 2015.
52. Tobin, J. Licensing As a Means of Providing Affordability and Accessibility in Digital Markets: Alternatives to a Digital First Sale Doctrine. Journal of the Patent and Trademark Office Society, v. 93, p. 167-206, 2011.