Rights declaration in Linked Data ⋆

Víctor Rodríguez-Doncel, Asunción Gómez-Pérez, and Nandana
Mihindukulasooriya

Ontology Engineering Group, Universidad Politécnica de Madrid, Spain

Abstract. Linked Data is not always published with a license. Sometimes a wrong license type is used, like a license for software, or it is not expressed in a standard, machine readable manner. Yet, Linked Data resources may be subject to intellectual property and database laws, may contain personal data subject to privacy restrictions or may even contain important trade secrets. The proper declaration of which rights are held, waived or licensed is a must for the lawful use of Linked Data at its different granularity levels, from the simple RDF statement to a dataset or a mapping. After comparing the current practice with the actual needs, six research questions are posed.

Keywords: Linked Data, licensing, intellectual property rights

1 Introduction

The term Linked Data (LD) is generally defined as *a set of best practices for publishing and connecting structured data on the Web* [1], and RDF is the preferred technology for representing this data. RDF information unit is the *triple* (a simple statement with a subject, a property and an object), being a RDF *graph* a set of triples, and a RDF *dataset* a collection of RDF graphs. Triples provide information about *resources* (identified by URIs) constituting pieces of data (or metadata —if the resource is data itself). The resources in a dataset are often linked to the resources in other datasets, through *RDF mappings*.

Linked Data is accessible to the public through the HTTP protocol, usually as RDF dumps in files or in SPARQL endpoints. However, being publicly available doesn't entitle the public to do any arbitrary action on the LD resources, and unless otherwise stated intellectual property (IP) rights and database rights will be in force if they exist. The most common practice with LD, however, is waiving some of the rights subject to certain conditions, using public notices called *licenses*. If these licenses are generous enough, they are called *open licenses*, and the Linked Data thus licensed is called Linked Open Data (LOD). The subset of LD that is not licensed as LOD has been termed *Linked Closed Data* (LCD) [2], Linking Enterprise Data [3] or simply proprietary data.

The lax and vague terms under which LD has been sometimes published has sufficed for all purposes, as the data publishers have been usually tolerant with its improper use not prosecuting it in the courts. However, misusing or disclosing high value data may suppose a real economic harm for those who have invested time and money in producing the dataset and it may cause legal trouble to the breaching user if sued. Further, it may discourage businesses entering the LD markets as they would fear having similar economic damage themselves. For the lawful use of Linked Data, a proper rights declaration understandable by humans and machines alike is a precondition. The following use case illustrates this need, which is not limited to IP-related laws but also to database laws, privacy laws and even trade secrecy laws.

> Alice, a data engineer, starts working today. She has been given a RDF dataset with valuable information, with no other indications, and she is unsure about what she can do. What is the risk of publishing it? Would she be breaking the IP law or, even worse, disclosing a trade secret? Can she edit the contents or even change the format of the dataset? What about distributing or selling the dataset?

The need for rights declaration is not limited to LCD, as lawfully using LOD also requires the satisfaction of the conditions in the license, nor is a different problem for either case, as the interplay of LOD and LCD in hybrid business models is likely to boost them. This paper describes the legal framework for publishing and consuming Linked Data which must be known by LD engineers (Section 2 and Section 3). It also makes an overview of the existing vocabularies for declaring rights and licenses in RDF (Section 4) to follow with an assessment of the actual use of these licensing terms (Section 5). Finally, the Section 6 opposes the legal requirements for Linked Data-centric business to appear with the existing of vocabularies and its actual use.

2   Legal Framework

Linked Data can become a high-value asset worthy to be protected. This protection can be achieved by means of secrecy –disclosing the data only to selected parties who have possibly paid, for example, as suggested in [4]. But law also offers a certain protection for data producers. Following the example before, Alice perhaps had received any of the datasets exemplified in Figure 1. The first contains literary works, the second a trade secret, the third personal data and the fourth mere data not subject to IP rights. Each case is different, existing rights associated to different parties (works' creator, the company holding the secret, the persons whose data is in a file), and all of the databases possibly subject to database rights. This section examines the different cases.

First, if data is not generally known to the public, if data confers some kind of economic benefit on its holder (and specifically by the fact that it is not generally known) and if it is subject of reasonable efforts, then it is object of protection by **trade secrecy** laws (in some jurisdictions named as *confidential*

Fig. 1. Examples of databases for which IP, trade secret and personal data law apply

information laws). Trade secrecy was included in the TRIPS agreements[1], and disclosing a trade secret is a prosecuted action everywhere, whose punishment is even harsher if the secret is of a military nature –if we are to have Military Linked Data. Disclosing a secret is an act which may be punishable under the criminal law. Other laws may preclude the communication of datasets if they express defamations, libels or other forbidden contents. This may have little interest for the data engineer, who yet should know that some datasets may not be spread out of a certain circle or even not communicated at all.

Second, data may be qualified to become object of protection by the **intellectual property laws**. Data, as a representation of a fact or an idea, is not necessarily the expression of an intellectual endeavour and in principle it does not get protection *per se*. But data can also represent an IP work (image, text...), in which case the IP law applies.

But also, if the selection and arrangement of others' literary and artistic works under the form of anthologies, databases etc. is the result of an intellectual creation, the mere collection is also under the umbrella of the intellectual property law (without prejudice of the rights of the original works' authors). This is universally acknowledged and can it be found in the Berne Convention[2].

Intellectual property rights comprise moral rights and exploitation rights. Moral rights are untransferable and unwaivable in some jurisdictions, and they include rights as the author being attributed, the work being respected or staying anonymous. Exploitation rights can be waived, licensed to the public, or traded in an economic exchange. Thus, the rightsholder of each of the exploitation rights may change along the time. These rights traditionally include the reproduction of the work (e.g., making copies), the distribution of copies (e.g., selling, renting etc.), the public performance, broadcasting or communication to the public and the transformation (including translation, adaptation etc.). Additionally, the so called *related rights* or *neighbouring rights* concern other categories of owners of rights different from the authors, namely, performers, producers, broadcasting organizations etc. A data curator or a dataset translator may also acquire related rights on the result of their work.

Third, in some jurisdictions, specific **database right laws** have been declared for the protection of databases which do not qualify to be intellectual property objects. This is the case of Europe, but not the United States, where no database right exists. This *sui generis* rights, as defined in Europe, protects

---

[1] Art. 39 in WTO Agreement on Trade Related Aspects of Int. Property Rights (1994)
[2] Art. 2.5 in Berne Convention for the Protection of Literary and Artistic Works (1886)

the "*qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents*"[3]. Extracting or re-utilizing the whole or a substantial part of the contents is prohibited unless permission is given. Naturally, exceptions usually exist for the case of educational purposes, injunction, public security etc., and in any case, after 15 years the database enters permanently into the public domain.

The combination of intellectual property rights and database rights (where applicable), generates a set of possible scenarios depending on (a) if the dataset contents are IP protectable assets or not (b) if the dataset creator has IP rights, database rights or none of them. Determining which of the scenarios corresponds to an actual case is not an evident task, as pointed out in [5]. For example, in the USA, Canada, Australia or Japan, a dataset with "the best of" a musician would be regarded as intellectual property object, for it would be a compilation involving an aesthetic judgment. But a "complete collection of works" of the same musician wouldn't, for it would be an automatable task. This collection would be protected, however, in Europe, if a verification work or any other similar effort was carried out.

Finally, if personal data is conveyed in the database, data protection laws have to be considered too[4]). These laws give no rights to the database creators, but rather impose certain obligations which have to be respected. These obligations include implementing security measures to be taken to physically and digitally protect the information, generating periodic security reports or keeping data access logs. In some jurisdictions, different levels of protection exist as a function of how sensitive the information is. As an example, the law in Spain defines three levels of confidentiality in a personal data file, ranging from the most trivial information but attributable to a person, to the most sensitive information like the sexual or religious preferences. Persons whose information is contained in a file have the right to access and rectify their records, which in any case can only be gathered for a declared purposed and can only live for a limited period of time.

To sum up, these laws ultimately protect the rights of (a) the authors who have created contents collected in a dataset (b) the dataset creators who have selected, curated and arranged the registers (c) the individuals whose personal information is in the dataset (d) third parties damaged if data is disclosed.

3   Rights Declaration for Linked Data

To precise the rights declaration for Linked Data, the following levels should be independently considered: (a) a single RDF triple, as the simplest unit of information, (b) RDF graphs or RDF datasets, as collections of data, (c) the

---

[3] Art. 7(1) in Directive 96/9/EC on the legal protection of databases (1996)

[4] For example, see the corresponding European Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data (1995)

RDF links, as mappings play the key role in the added value of Linked Data and (d) external resources referred by RDF.

Single RDF triples (or a reduced group of them) are not protected by intellectual property or database laws, which explicitly exclude individual data from the protection scope –unless a work or a full data collection is contained in a literal. However, they may be protected as trade secrets or its access by restricted for other reasons. As with copyrighted material, stamping a top secret or similar notice is merely informative and no additional protection is conferred by its addition.

In general, a **RDF dataset** matches the legal concept of database[5]. Its creator may claim database rights in certain countries, plus intellectual property rights if the dataset contains works creatively selected and arranged –a claim difficult to be justified in most of the cases. Database rights do not exist if, for example, the RDF dataset was only an effortless syntax transformation of what was included in another database. Other rights may exist over datasets if they contain trade secrets or personal data whose handling is subject to further restrictions.

RDF datasets aggregating data from different RDF sources require the specific authorization from the different dataset owners or the existence of a public license allowing to do so –in which case, possibly some conditions will have to be respected. **RDF mappings** are collections of triples relating resources in two or more different RDF datasets. Excepting for the case of automatic mappings, linking vocabularies or resources is a costly effort which almost immediately qualifies the work as a protectable asset: RDF mappings are a first class citizen in the Linked Data ecosystem.

Referring an external entity is an action always allowed: even if the resource is not yours, you can freely comment on it –or link it to your concept. But declaring a mapping (either an added-value mapping or merely re-using mappings already in the public domain) leads to opening the door to using information from different RDF resources with possibly a different legal character. The use of data obtained by following links in RDF mappings may be subject to rights whose declaration would ease the lawful use of Linked Data. This also applies to **resources referred by RDF** subject to protection, although most of users are possibly aware of this.

Finally, declaring if a Linked Data resource contains personal data or confidential information is merely informative, but it can ease its handling and strengthen the rightsholder in case of litigation. To sum up, for the lawful use of Linked Data, which may be created, acquired, transformed and published in a value chain where several parties intervene, a proper holistic rights declaration is a must.

---

[5] A database is: *a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible*, in European Directive 96/9/EC

4   Linked Data for rights declaration

If the declaration of rights for RDF data is needed, RDF itself can be the vehicle for its expression. The basic information to be given informs that a Linked Data resource (a triple, a graph, a dataset, a mapping,...) is subject to certain rights and if they are kept (e.g. a copyright statement), unconditionally released (e.g. a waiver notice) or given subject to certain conditions (e.g. licensed).

This allows us identifying three questions: which subjects can be attributed with rights expressions? Which predicates can be used for rights declaration? And which licenses can be used in the rights declaration? The rest of the section describes the existing choices to express these pieces of information.

4.1   Properties for Linked Data rights declaration

The predicate for rights declaration can be taken from Dublin Core (DC), perhaps the most used vocabulary in Linked Data after the language constructs (RDF, OWL, etc.): `rights` is one of the fifteen core properties defined in the Dublin Core Metadata Element Set[6] for use in resource description. Defined as *Information about rights held in and over the resource*, it has been generally used to include descriptions of the copyright information or references to rights management services. It is present in two different namespaces, usually prefixed as `dc`[7] and `dcterms`[8].

This predicate is generic enough as to be used to refer to any of the rights described in Section 2. Dublin Core specifies two properties refining the rights property: `accessRights` (*information about who can access the resource or an indication of its security status*) and `license` (*a legal document giving official permission to do something with the resource*). The former may be used to declare that a resource contains personal data (like a phone number), while the latter has been extensively used to declare the intellectual property license of a resource. The Creative Commons property `cc:license`[9], derived from `dc:license`, has also been used to point at a well-known license.

4.2   Subjects of Linked Data rights declaration

The subject of a rights declaration is the piece of information object of the rights, either a referred resource, an RDF triple, a dataset, or a mapping.

To declare rights of a **referred resource**, a simple property can be stated about the resource. The following example attributes a Creative Commons CC-BY license to an external resource.

```
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix ex: <http://www.example.org/ns#> .
ex:externalResource1
dcterms:license <http://creativecommons.org/licenses/by/>.
```

---

[6] http://dublincore.org/documents/dces/
[7] http://purl.org/dc/elements/1.1/
[8] http://purl.org/dc/terms/
[9] http://creativecommons.org/ns#

To declare access restrictions to a **RDF triple**, a reificated statement can be attributed with rights declaration. The following example attributes a privacy statement to a phone number.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
_:x rdf:type rdf:Statement ;
rdf:subject ex:Alice ;
rdf:predicate foaf:phone ;
rdf:object "654321987" ;
    dcterms:accessRights "PersonalData".}
```

Note the use of `dcterms:accessRights`, which according to Dublin Core can also be used to give *information regarding access or restrictions based on privacy, security, or other policies.* However, on despite of existing a vocabulary for privacy preferences ontology[10], no common term has been accepted to tag that a piece of information contains personal data and so a simple literal "PersonalData" has been given: an acknowledged URI is missing here.

A complete **RDF dataset** may be attributed within or outside the dataset it-self. Within the dataset, the most common practice is to attribute the URI of the dataset as in the example below.

```
<http://URI-OF-THE-DATASET />
   dcterms:license <http://creativecommons.org/publicdomain/zero/1.0/>.
```

The dataset can also be described in a separate RDF graph, possibly based on the VoID[11] or DCAT[12] vocabularies. In this case, the instance of `void:Dataset`, `dcat:Dataset` or even of its parent class `dctype:Dataset` would be attributed the corresponding rights declaration.

Finally, **RDF mappings** can receive the same treatment as RDF datasets, save that in VoID a dataset subclass is defined: `void:LinkSet` (a collection of RDF links between two datasets). This linkset can specify the referred dataset through the `void:target` property, which in turn can receive a rights declaration –for example a public domain license.

```
http://URI-OF-A-LINKSET> a void:Linkset ; # mapping
   void:target <http://URI-OF-A-DATASET>.
<http://URI-OF-A-DATASET>a void:Dataset . #external dataset
   dcterms:license <http://creativecommons.org/publicdomain/zero/1.0/>.
```

4.3   Rights declaration for Linked Data

The rights declaration should convey the information of which rights are held, waived or licensed. For data licensing, specific data licenses exist and can be identified by known URIs. It is the case of the Open Data Commons[13] (ODC) licenses, the Creative Commons license CC0 and licenses defined by some governments. This makes possible and easy the assignment of a license to a RDF dataset. The most common data licenses are:

---

[10] http://vocab.deri.ie/ppo#
[11] http://www.w3.org/TR/void/
[12] http://www.w3.org/TR/vocab-dcat/
[13] ttp://opendatacommons.org

– Public Domain Licenses. They waive all the possible intellectual property and neighbouring rights (database rights) of the dataset and its contents. There are two equivalent choices, the ODC-PDDL (Public Domain Dedication and License) and the CC0 public domain waiver.
– Attribution Licenses. They waive all the possible rights, requiring only the mere attribution. Example: ODC-By, attribution for data/databases.
– Share-alike Licenses. The rights are also waived requiring that derived or adapted databases keep the same license. Examples: ODC-ODBL (Open Database License), or the UK-OGL (UK Open Government License).

Some other licenses famous have been used, like the general Creative Commons licenses. These pre-defined licenses are also identifiable by URIs, but they are intended for general works and do not mention the database rights which might apply in places like Europe. These Creative Commons licenses always require attribution (BY), and they may require the share-alike (SA) condition, a non-commercial flag (NC) or the non-derivatives (ND) restriction. "Non-commercial" means that the work (nor derived versions thereof) can be use for profit, non-derivatives means that no transformations of the original work can be published. The combination of these conditions leads to having licenses known as CC-BY (only attribution), CC-BY-SA (with share alike), etc.

The imprecise use of licenses for datatasets is even more evident when licenses like the GFDL (GNU Free Documentation License) conceived for documents or even software licenses are used. Attending to their degree of restrictiveness, a categorization is shown in Table 1.

Kind of license	Licenses	Comment
Not specified	not specified	No license has been specified
Public Domain	cc-zero, odc-pddl	All the rights have been waived
Attribution	cc-by, odc-by	Attribution is required
Share alike	cc-sa, odc-odbl, uk-ogl, gfdl	Copyleft licenses
With restrictions	cc-nc, cc-nd, cc-nc-nd, etc.	More severe restrictions
Closed	all rights reserved, etc.	Closed licenses
Other	unknown, etc.	Not catalogued

Table 1. Classification of licenses by their restrictiveness

More complex ad-hoc licenses can be defined with one of the digital Rights Expression Languages like ODRL (Open Digital Rights Language) [6] or MPEG-21 REL [7], although they are XML based and do not intend to imbricate with the rest of the web of data as proposed in [8]. A new breed of vocabularies, interconnected and not intended for its use in specific Digital Rights Management systems is now appearing: vocabularies like LiMO[14], L4LOD[15] or ODRS[16], but

---

[14] http://data.opendataday.it/LiMo
[15] http://ns.inria.fr/l4lod/v2/l4lod_v2.html
[16] http://schema.theodi.org/odrs/

so far only the Creative Commons RDF ccREL [9] has been used by the Linked Data community.

5   Current practice in rights declaration in Linked Data

Quantitatively observing the current practice about rights declaration in Linked Data is a difficult task as RDF sources are multiple and embracing every piece of Linked Data in the web is not possible. Yet relevant or extensive parts of it can be analyzed.

For example, the LOD cloud is important for being a reference of high quality data, accounting 338 datasets, although biased regarding licensing: they are supposed to be openly licensed. A broader collection of datasets, easily accessible, is that listed in the CKAN archive (Comprehensive Knowledge Archive Network[17]), a registry of open data and content packages provided by the OKFN, excelling for its completeness at cataloguing existing datasets. A more selected collection of sources, periodically compiled and analyzed is that of the DyLDO[18], in a framework to monitor Linked Data over the time. It includes datasets from the LOD cloud and the Billion Triple Dataset challenge[19]. Finally, another source of study may be Sindice[20], a lookup index over resources crawled on the Semantic Web, which ingests RDF, RDFa and microformats.

Again recalling the study of Section 4, the questions to be answered by the experimental work can be formulated as *Which subjects* are actually attributed with rights expressions? *Which predicates* are actually used for rights declaration? And *which licenses* are actually used in the rights declaration?

5.1   Rights declaration for Linked Data in practice

In order to assess the use of licenses, a double test was made: determining which licenses were in use in the official LOD datasets, and which licenses were in use in the broader set of the LD datasets in CKAN. The set of LOD datasets could be obtained by using the REST API of CKAN (the LOD cloud diagram was formally managed through the CKAN repository). CKAN also records information about the license of each dataset, as declared at registering time. In a similar manner, the license in general LD datasets in CKAN was queried.

As of May 2013, 1,836 Linked Data datasets[21] were registered in the CKAN, belonging 338 of them to the LOD group. Each of the datasets had one or more resources (i.e. different data files, SPARQL endpoints etc.) but each of the datasets was homogenously licensed through the resources. The results of this observation are shown in Table 2, which has grouped the licenses with the

---

[17] http://datahub.io/
[18] http://swse.deri.org/dyldo
[19] http://km.aifb.kit.edu/projects/btc-2011/
[20] http://www.sindice.com/
[21] A dataset was considered to be LD if it had one resource marked with a type containing the following strings: rdf, rdfs, owl, ttl, turtle, nquads, ntriples, nt or sparql.

criteria of Table 1. This grouping hides, nonetheless, the fact that a 29% of the licenses were intended for works and not specifically for data.

-	*All Linked Data datasets in CKAN*	*LOD datasets in CKAN*
Kind of license	*Num. (%)*	*Num. (%)*
Not specified	469 (26%)	132 (39%)
Public Domain	291 (16%)	69 (21%)
Attribution	440 (24%)	66 (20%)
Share alike	322 (18%)	35 (10%)
With restrictions	143 (8%)	5 (2%)
Closed	43 (2%)	16 (5%)
Other	128 (7%)	3 (1%)
Total	1,836 (100%)	338 (100%)

Table 2. Licensing of Linked Data datasets

Disregarding the object where a license has been applied (a RDF dataset, external resources, etc.), an SPARQL query can be made to observe which kind of licenses are used in extensive pieces of Linked Data. Having made this query in Sindice, public domain and attribution licenses again gathered the largest percent of all the licenses: 63% against the 53% used for CKAN datasets. Share-alike licenses accounted for a 27%, against the 24% in CKAN datasets and licenses with restrictions (non-commercial, no derivatives) were 6% in Sindice against the 11% in CKAN datasets.

5.2   Properties for Linked Data rights declaration in practice

The goal of this observation is to assess which RDF elements are most used to specify a license. To achieve this, different SPARQL queries were made on Sindice, inquiring for each of the most common elements used for licensing. The results, shown in Table 3 shows the `dc:rights` as the champion. Yet, this element is used about one order of magnitude less than the `dc:title` element. The queries for Dublin Core included both namespaces as described in Section 4.1.

Vocabulary	Element	Usage	Usage (%)
Dublin Core	rights	5,905,519	59%
XHTML	license	3,825,939	38%
CreativeCommons	license	263,805	3%
Dublin Core	license	32,922	neglectable

Table 3. Relative use of licensing terms in Linked Data

More licensing elements proposed in other vocabularies were also tested, but their presence in Sindice was neglectable if not zero. These vocabularies included properties as the DOAP[22] `doap:license`, the PREMIS[23] `premis:licenseTerms`, the OMV[24] `omv:hasLicense`, the Music Ontology[25] `mo:License`, the VAEM[26] `vaem:hasLicenseType`, or more sofisticated classes in Dublin Core as `dcterms:RightsStatement` or `dcterms:LicenseDocument`.

5.3 Subjects of Linked Data rights declaration in practice

In the previous sections, the RDF triple, the RDF dataset and the RDF mapping had been identified as the key ingredients of Linked Data. In the following experiment, Sindice was queried to learn how often a licensing property had been applied to `rdf:Statement`s, `void:Dataset`s and `void:Linkset`s.

The experiment revealed that rights declaration had been expressed unevenly for these levels. Sindice included 48,968 reificated statemenets, of which 13,505 had rights information, but coming from exclusively a single dataset. RDF Datasets declared with `void:Dataset` accounted a total number of 4,549, of which 92 used a Dublin Core rights and 26 a Dublin Core license. Finally, none of the 1,163 mappings declared with `void:Linkset` and found by Sindice had rights information.

6   Conclusions

An ecosystem of entities (public bodies, academic institutions, enterprises, etc.) producing, transforming and consuming Linked (Open) Data in a marketplace is now starting to bloom [11], and it will presumable flourish more healthy if enough guarantees exist for all the parties in the value chain and their rights. However, the mismatch between the needs described in Sections 2 and 3 and the practices observed in Section 5 lead to formulating a series of pending challenges:

1. *Vocabularies for declaring rights information exist, but are not complete.* Terms like the Dublin Core `license` and `rights` have gained popularity (as shown in Section 4.1), but they fail to be precise. While it is vaguely assumed that the rights or licenses are IP related, other legal concerns as the privacy statements or confidentiality stamps (Section 2) are ignored.
2. *Vocabularies for licensing terms exist, but they need further development.* Some existing licenses are now well known and widely accepted. But specific terms of use are still referenced in natural text. The new vocabularies for licensing LD which are now sprouting should become more mature, better documented and accompanied of easy tools for producing rights expressions.

---

[22] http://usefulinc.com/ns/doap
[23] http://multimedialab.elis.ugent.be/users/samcoppe/ontologies/Premis/premis.owl
[24] http://omv.ontoware.org/2005/05/ontology
[25] http://musicontology.com/
[26] http://www.linkedmodel.org/schema/vaem

3. *Licensing information for Linked Data should be done at different granularity levels.* The description of RDF Datasets and mappings with VoID is adequate (see Section 4.2) but its use should be more spread to have a unified way of expressing rights declaration at different granularity levels, not forgetting RDF statements and RDF graphs.

4. *Many pieces of existing Linked Data lack a proper rights declaration or are incorrectly licensed.* A high percentage of the datasets in the LOD cloud (39%), champions of the LD, have no rights declaration specified, and 30% of the licensed datasets are using licenses for IP works unaccurate for database rights (see Section 5.1).

5. *Mappings have not received proper attention.* The fact that no mapping was licensed (Section 5.3) in the linksets with a VoID description reveals that LD creators may not have understood the value of a well-done mapping, and the importance of properly attributing rights for its use by third parties.

6. *There are no tools granting trust when aggregating data form different sources.* In order to easily build up richer datasets, reliable provenance information is needed, along with the precise knowledge of which licenses are possible when aggregating data from differently licensed sources, as described in [10].

As these leaks are not intrinsic of Linked Data, and they are technically solvable with appropriate vocabularies, standards and tools, it can be expected that the development of new LD business models will gradually bridge the gap.

References

1. Klyne G. and Carroll, J. J., eds. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. (2004)
2. Cobden, M. et al.: A research agenda for linked closed data, in Proc. of the 2nd Int. Workshop on Consuming Linked Data (2011)
3. Servant, F.P.: Linking enterprise data, in Proc. of WWW Workshop Linked Data on the Web, vol. 369 (2008)
4. Villata, S., Delaforge, N., Gandon, F., and Gyrard, A.: An Access Control Model for Linked Data. In On the Move to Meaningful Internet Sys., pp. 454—463 (2011)
5. Aliprandi, S.: Open licensing and databases. Int. Free and Open Source Software Law Review, North America, 4(1), pp. 5–18 (2012)
6. Ianella, R. and Guth, S.: ODRL Version 2.0 Common Vocabulary, W3C Community Group Final Specification (2012)
7. ISO/IEC 21000-5:2004, Information Technology– Multimedia Framework (MPEG-21)– Part 5: Rights Expression Language (2004)
8. Rodríguez-Doncel, V., Delgado, J.: Towards an Expression Language for Licensing Content in the Connected Semantic Web, in: Proc. of the 9th Int. Workshop on Virtual Goods (2011)
9. Abelson, H., Adida, B., Linksvayer, M., and Yergler, N.: ccREL: The Creative Commons Rights Expression Language. Technical report. (2008)
10. Villata, S. and Gandon, F.: Licenses Compatibility and Composition in the Web of Data. in Proc.of the 2nd Int. Workshop on Consuming Linked Data. (2012)
11. Harris K.: Selling and Building Linked Data: Drive Value and Gain Momentum, in Linking Enterprise Data, pp. 65–76, ed. Springer (2010)