

# Towards an Expression Language for Licensing Content in the Connected Semantic Web

Victor Rodriguez-Doncel<sup>1</sup>, Jaime Delgado<sup>1</sup>

<sup>1</sup> DMAG, Universitat Politècnica de Catalunya, Barcelona, Spain  
{victorr, jaime.delgado}@ac.upc.edu

**Abstract.** Rights Expression Languages (RELs) defined several years ago like MPEG-21 REL or ODRL were conceived as XML documents to govern the content distribution in DRM platforms, relying on centralized content repositories, proprietary author registries, and authorization servers to control the access. As of year 2011, the panorama in the Internet and its uses has changed and a new generation of Rights Expression Languages is needed, with a broader scope, meant for open systems as opposed to closed platforms, making use of the Linked Open Data information instead of referencing organizational repositories and suitable to be applied to user generated content in Social Networks. This paper foresees and describes a new generation of RELs, where information is upgraded from the plain XML language to a semantic level, where its scope of use is broadened from closed environments to the open Internet, where permissions granted in a license are put in the context of all the data already present and interconnected in the Semantic Web, and where a rigorous authorization can be done based on the deontic logic premises. As an intermediate solution, this paper also proposes ODRL as a good starting point, describing the modifications needed to accommodate the current language definition to the new challenges described in the paper.

**Keywords:** Rights Expression Languages, Semantic Web, ODRL, Linking Open Data

## 1 Introduction

In the last years technology has boosted the presence of multimedia content on the Internet, being Internet video alone the main source of Internet traffic at present. Also, high value material is being dumped by public initiatives (e.g. the Europeana project) and private ones (e.g. Google libraries digitalization) complementing the user generated content which floods the Internet (in Youtube, videos are being uploaded at a pace of 24 hours of new content per minute, and this figure is growing). Other sources like Internet radios, IPTV and in general broadcasting services offer massive multimedia content heterogeneously accessed and catalogued. This constitutes a huge amount of multimedia content in the Internet which could be used as prime material

---

for the creation of other audiovisual works or which simply could be aggregated in better organized channels by third parties.

However, this explosive growth in multimedia availability is not being paralleled by technologies to effectively search and access material with a clear rights statement. There is no Google search for “*images to be acquired for less than 10€ and which can be used in my country for commercial purposes*”, because authorship information is usually not well attributed and the terms and conditions under which it could be used are still less clear.

And yet, reusing existing material (adapting, integrating and synchronizing) is a crucial activity for creating content. If after a casual search, a content creator sees some material in the Internet and he is unsure whether its use is allowed or not, he will miss an opportunity; because even if he was willing to pay for the material, he does not have the channels to do it. Both the creator of the original work and the creator of the derivative work suffer from a technology hurdle. There are indeed platforms to market images and video (in the fashion of iPhotoStock<sup>2</sup>, etc.), and platforms where Creative Commons pictures can be found (like Flickr<sup>3</sup>) but these and other closed stock systems exclude the vast majority of available material in the Internet. Multimedia material is in practice massively used and recycled but there is much uncertainty about the consequences of its use which sometimes hinders the creation process.

Annotations on multimedia material should include authorship information, in order to attribute a work to its creator, and, if possible, to other agents involved in its value chain, like contributors, publishers or rights holders. In addition, annotations should include licensing information, i.e., the terms and conditions under which different uses of the content are allowed. Most of the metadata formats include elements to describe this (for example, DublinCore<sup>4</sup> has elements for “*creator*” and for “*rights*”) but they go no further in their definition (how to use them, etc.) nor match between the different metadata formats. The authorship and licensing information should be able to express complex conditions, possibly made up of the logical combination of atomic ones, and liable of being evaluated by an authorizer.

## 2 Limits of existing technologies

### 2.1 Technologies for expressing authorship information

Intellectual Property rights are considered in a similar way throughout the different law systems in the world and this makes possible a common global treatment for the Internet audiovisual assets. These quasi-universally accepted rules and practices are very well summed up in [1] and [2] (the latter given by the WIPO (World Intellectual

---

<sup>2</sup> iPhotostock. <http://www.iphotostock.com>

<sup>3</sup> Flickr, <http://www.flickr.com>

<sup>4</sup> The Dublin Core Metadata Initiative, <http://dublincore.org>

Property Organization), which administers the international treaties which grant this harmonious vision of the Intellectual Property).

There is no consensus, however, on how the authoring information is represented and stored. There are publicly managed registries (like the "*Registro de la Propiedad Intelectual* (Intellectual Property Registry)" in Spain) and those handled by the Collective Author Societies (like SGAE in Spain) but they are very seldom put in connection to metadata in multimedia archives or used for the exploitation in distribution platforms. There are a few identification systems which are universally accepted like DOI (Digital Object Identifier) for the identification of intellectual property objects, ISBN (International Standard Book Number) for books or ISWC (International Standard Musical Work Code) for musical works, but none of them is free of charge nor agile to be used.

Actually, these identifiers are widely used in official, controlled distribution platforms, but mostly ignored in the Internet. In fact, authorship of a given content published in the Internet is often missing at all. Even though most of the multimedia formats reserve space for metadata, and there is usually at least one metadata field for expressing who is the author, this information is largely missing in common contents published in the Internet. Not only authoring tools do not emphasize enough the need for authoring the content (i.e. web composers should remind when finding undeclared "creator" fields, or automatically fill them), but also content authors do not feel the need of adding this information if it is not with a licensing purpose.

However, the introduction of new licensing expressions attached to the content may encourage content creators to state more clearly the authorship.

Note that also there are no control mechanisms that prevent attributing himself a work of a different person. Again, the trends are that DRM are not about encrypting and enforcing, but about providing fair channels of communication between the end user and the authors.

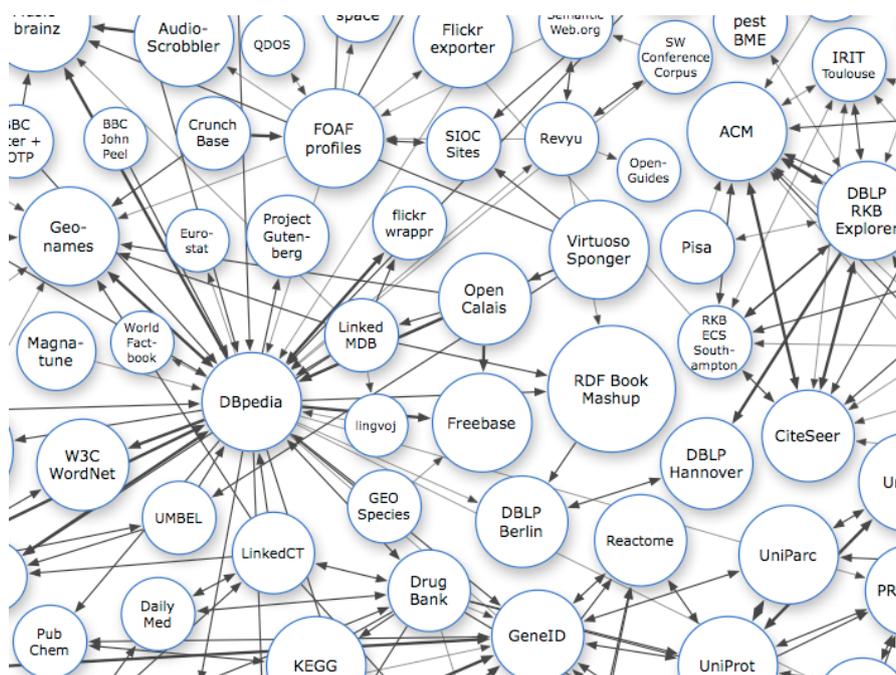
## **2.2 Technologies for expressing licensing information**

Representation of the licensing information is even more fragmented. Several Rights Expression Languages (REL) have been defined by different standardization organisms, but none of them has achieved universal acceptance or use. A good overview of the elements in a REL is done in [3].

The representation of rights and licensing information in computer languages covered by RELs appeared from the late nineties, and these have been XML languages among which the most spread ones are MPEG-21 REL [4] and ODRL [5], the first being restricted to MPEG-21 environment and the second with a successful variant in the mobile playground known as OMA DRM [6]. Besides these languages with a public specification, other closed specifications have had success in different private initiatives and have played an important role in the Digital Rights Management systems (e.g. that of Microsoft, or Apple's iTunes). None of them, however, has foreseen a generalized used, and its expressivity level has remained to that of XML (missing the advantages of a semantic representation with RDF or even

backed by an OWL ontology). Only timid attempts to make a semantic expression of rights have to be mentioned, like IPRonto [7], OREL [8], or the reflections in [9], but they did not pass from the academic level nor pretended to act in an open context.

All of the existing rights expression languages express conditions related to context information provided by the system itself, and they are not open to possible external sources of context information such as an Internet search from semantic information sources (e.g. Linked Data<sup>5</sup> from the web, Fig.1).



**Fig. 1 Part of the Linking Open (LOD) Data Project Cloud Diagram, taken from <http://linkeddata.org>**

However, the current generation of RELs is not satisfactory for labeling published material in the Internet. Traditional RELs are intended and used in the context of a Digital Rights Management System (DRM), and very often they require of authorship registries, centralized repositories and authorization servers. The process of registration in these servers sometimes requires off-line steps, formal requirements and additional payments, all of this obstructing its use for the always dynamic user generated content.

This need for managed repositories in order to reference and identify the content creates a dependency on non-free entities, in opposition to user generated content published in Social Networks –easily identifiable by a URL–, even more sophisticated repositories like Freebase<sup>6</sup> or any other in the Linking Data Cloud. The same applies

<sup>5</sup> Linked Data, <http://linkeddata.org>

<sup>6</sup> Freebase, <http://www.freebase.com>

regarding with the provision of one identifier to users; which can be as simple as a profile in a Social Network.

An additional problem that RELs are suffering is their inability to grow in a universally accepted manner. The rigid set of conditions which can be expressed with current REL is reduced to at most a dozen of verifiable assertions (e.g. a date time stamp or the number of times a user has reproduced a content and so on), and this pre-defined set of conditions is fixed, soon becoming too short. The foreseen extension mechanisms for this vocabulary do not foster reuse. For example, there is a MPEG-21 Rights Data Dictionary Registration Authority<sup>7</sup>, where terms can be registered from £20 each word, but there are no published extended terms.

The context information to take the authorization decision is supposed to come from reliable providers trusted by the authorization system, what is satisfactory for a number of cases, but this policy restricts the number of sources of information. If conditions in licenses were put in terms of information to be found in the Web of Linked Data, an almost unlimited set of assertions could be used for licensing – any RDF triple present in the web (for a detailed review on Linked Data and multimedia see [10]). The lack of trustworthiness of the sources can be compensated if its reputation is accepted by both parties. Datasets from public institutions or *dbpedia* might be fallible but good enough for non-critical transactions, i.e. where one mistake in the authorisation among many is tolerable.

To sum up, there are zettabytes of images and video online, but poorly labeled, with authorship seldom attributed and hardly marketed. Creative Commons constitutes a notable exception: material licensed as Creative Commons is well annotated (although not always in a computer readable way), authorship is clearly stated and use conditions are clear. There is not, though, a commercial equivalent universally acknowledged, and the existing rights expression languages are limited to the DRM system for which they were thought.

Regarding open licensing, the initiative of Creative Commons has spread since its inception 10 years ago, and nowadays tons of authored content is public in the Internet freely but subject to certain conditions. Creative Commons licenses are text documents, whose exact form is referenced by using some symbols now universally known. However, it is worthy to observe that works licensed as Creative Commons do not have this information accessible by a computer; text documents include the symbols usually somewhere around the second page, audio or video documents are labeled by having the symbols (images) in the webpage where they are available. Their RDF version [11], which would fix this problem, has not succeeded. Creative Commons suffers two major drawbacks: the expressivity of the language is reduced to a small set of pre-defined clauses (e.g. “non-commercial use”), and it never pays back directly to the author.

All these pending technological challenges are hindering the exploitation of audiovisual content on the Internet, and specially the development of new business models. It can be envisioned that as these technological problems become solved, new business models will appear; and that the appearance of new business models will

---

<sup>7</sup> <http://iso21000-6.net/>

increase the effectiveness in authoring, searching, licensing and interoperating multimedia content.

### **3. Features of the new generation licensing languages**

The relevance of content in the Internet has changed dramatically in the last few years, so much that a new paradigm of licensing languages is needed, and not only this, it is possible.

Although the Semantic Web had been envisioned long ago [12], it has not been until the recent years that the web has started to become populated with RDF related triples. Now, a vast collection of interconnected data is living in the Internet following the guidelines proposed by the Linked Data community, and there is plenty of information on musical works, movies, artists and many other data liable to be used for creating licenses. The first acknowledged REL, XrML (eXtensible Rights Markup Language), was born in 1998. Creative Commons and ODRL were born in 2001, XACML did in 2003, and finally MPEG-21 REL in 2004. By now, the time for declaring a new kind of authoring and licensing language has arrived, taking advantage of the Semantic Web mature technologies and the data availability.

Based on the former analysis, the features of the new generation of authorship and licensing languages can be profiled through the following ideas.

#### **Authorship Expression Languages in an Interlinked Web**

Existing Rights Expression Languages and policy languages described in the previous Section live in a non-interconnected world. With the advent of a cloud of interconnected RDF triples, where audiovisual content and its authorship are not the least important [13] (see the LinkedMDB for movies, the Music Brainz database for music etc.), a new generation of RELs should breed. This generation of licenses on multimedia content should be linked to the existing dataset repositories, with the advantages of having universal identification of content and creators, and enabling new decentralized business models. References to closed author registries should be avoided.

The authorship information should be contained in the metadata fields of the resource, and both content and users should be unequivocally related through a public URL – for example the user profile in a Social Network–, in one of the above mentioned data stores, or similar.

#### **Authorship Expression Languages aware of the Value Chain**

The achieved expression for the rights and licensing information should be able to link to the different content assets a work is based upon, and the previous artists and contributors that added value to the asset. An ideal technology to use would be that of the Media Value Chain Ontology [14], an international standard (Part 19 of MPEG-21) aimed at describing the media value chain from the intellectual property point of

view but which has never been taken outside a closed system.

### Using Semantic Expression Languages

New expression languages should be moving from XML to RDF, and they should allow the expression of more complex terms of use subject to be authorized by semantic authorizers, i.e., the authorization decision being performed by a logic reasoner on OWL and SWRL statements. At the price of an increased complexity, there is a number of advantages in this change:

- Transparency, as the reasoning logic does not remain in a authoriser application whose code is unknown, but it is publicly stated within the licence, i.e. everybody can read the rule. Moreover, the rule can be evaluated with many different existing reasoners.
- Clear semantics, as the XML elements need an external interpretation to understand their meaning. In comparison, the RDF triples are clear in relation to the authorisation –just by looking at the rule.
- Logical elements are an intrinsic part of the language

The existing work already present in the literature gathers significance if put in the context of the existing knowledge in the Internet under the form of connected RDF statements. Current rights expression languages relate to contextual information in a closed manner. For example, a certain condition may constrain the use of a material to a current location or the current time, and the actual values are acquired by the system itself. Thus, this information is taken by the authorizer itself in a closed manner. However, it might be of the highest interest to pose constraints and restrictions based on information outside the system (like an external web service), and acquired intelligently from different sources with a semantic analysis. Joining the information provided by *myspace*, the *dbpedia*, or the BBC Programmes, imaginative licensing can be made. Licensing of a song can be dynamically based on the music ranking lists (i.e. make it free until it reaches the top 40), on its broadcasting frequency at BBC or the current state of the artist. As an example, the following simple SPARQL query can help at verifying if the artist is dead long ago enough as the property rights to have expired.

```
SELECT ?object WHERE {
  <http://dbpedia.org/resource/Bob_Marley>
  <http://dbpedia.org/property/died> ?object .}
```

### Expressions allocable in the multimedia metadata fields

On despite of the disparate definitions of the different metadata formats, as well as their vagueness (for example DublinCore includes elements for *creator*, *contributor*, *publisher* and *rights* but it does not go much further than their enumeration), a more homogenous placeholder could be found.

During 2011, the W3C Ontology for Media Resource will become a W3C Recommendation [15]. This pivotal ontology, conceived for the interoperability of the existing metadata formats for multimedia defines a model around which the rest of 17 formats (like MPEG-7, EXIF, TV Anytime or Youtube) are articulated and can be interoperated through an accompanying API. These formats describe the vast majority

of systems in use, and after this effort, it will be easier to find a uniform placeholder for the authorship and licensing and information.

**Content is not protected but licensing may be not free**

Having no protection mechanisms (encryption etc.) the user of the new generation of licensing languages does not need to be reduced to the non-profit sector only. Paid use of the content should be expressible, and the licensing information should point out to the methods to acquire more rights if so intended by the rights holder. This makes a difference with the existing Creative Commons.

All the previous changes could lead to the following advantages:

- Searches made based on the licensing information will be made faster and more reliable
- User generated content shall be exploitable. For example, if a television channel uses user generated content (like the videos of the recent revolts in North Africa), the original author will be able to be compensated because he will be reachable. And even if compensations for professional use of user generated content don't come, the author may have the forensic proof to claim for such compensation in front of a judge.
- New business models will be enabled after the spread of this technology.

## **5. An intermediate step: adapting ODRL to some of the requirements**

Reaching the expression of licenses as RDFs in new languages starting from scratch might be a cold start. However, from the existing REL languages and for its flexibility, ODRL would be a good candidate to implement some of the features listed in the previous section.

In the following, we describe how ODRL might come close to the requirements described before:

### **Authorship Expression Languages in an Interlinked Web.**

ODRL is in its conception a language to represent rights (its foundation model is intended for offers and agreements), but not property claims, i.e. positive assertions to state a property claim. These property claims are left out of the scope of current RELs, but giving an integrated approach would benefit language integrity. In addition to the `permission`, `prohibition` and `duty` elements, a simple claim element would be highly useful *per se*.

Regarding the *interlinkability* feature, ODRL neither prevents representing users with any linkable URL, nor provides support mechanisms for the same. In the interlinked space, specific terms should be given for the links, and for their characterization. Actually, in the context of using ODRL to manage privacy, a vocabulary extension of ODRL was proposed in [15] including terms for “all friends”, “groups”, etc.

Not the least, another feature of the Interlinked Web is the ability to communicate bidirectionally. In this context, the author of a published content may wish to be reachable to further negotiate a permission, etc. while this idea was pointed out in [17], no further action was taken.

#### **Authorship Expression Languages aware of the Value Chain.**

Managing downstream rights through the value chain, i.e. being able to grant rights and impose conditions at different levels in the value chain is explicitly expressed in the requirements of ODRL 2.0, and the `o:nextpolicy` element has been added, but such construct was missing in ODRL 1.1. For more advanced value chain expressions, parties and content might be represented with instances of the MPEG Media Value Chain Ontology, linked with ODRL expressions, but an explicit work should be made to align the concepts.

#### **Using Semantic Expression Languages.**

Even though ODRL is specified in English, a formal semantics was taken in [18] and a derived ontology described in [19]. However, no step has been given in the last 6 years to promote these attempts to fully accepted versions, until very recently when a XML/RDF encoding has been planned. To fully interoperate with the semantic cloud of triples, ODRL should reinforce these efforts and finish the RDF version of its language.

#### **Expressions allocable in the multimedia metadata fields**

The Open Digital Rights Initiative and the Dublin Core Metadata Initiative formed in 2005 a working group to develop a joint ODRL/DCMI profile (something shown possible in [20]), but they failed to gather attention and the group was closed down soon after.

This study shows that even though ODRL does not currently fulfill the ideal requirements of Section 4, it is not incompatible with many of them. Actually, it has been seen that many of the individual recommendations have already been explored independently but for different reasons, abandoned. We strongly believe that the vision of having these features together would fit the current trends of a Web semantically interlinked, highly connected, and with a large load of user-generated content within social networks or not. By adapting to these features, ODRL might gain a momentum and achieve massive spread.

## **6. Conclusion**

This paper has tried to envision a new generation of licensing expression languages to appear in the Internet for the published content other than Creative Commons and other than the rather restricting current Rights Expression Languages.

To reach this goal, several research problems will have to be attacked: how to represent deontic logic propositions in terms of OWL (Web Ontology Language)

constructs, how to relate the content identification and user identification with the data already present in the Internet (for example with the authors databases) or how to embed these rights and licensing language expressions in the existing metadata formats. Not the least important, an insight on several side problems will have to be made, such as how to make searches on this authoring and licensing information (e.g. “*which works of this author can I find at a cost fewer than X?*”), etc., which relationships will there be with the content dumped through social networks and which new business models may be favored by these technologies.

Finally, in order to make a softer approach, the suitability of ODRL or modifications thereof has also been studied. We have found that some of the ideas pointed out in this paper had already been analyzed and discarded individually. However, given the new Internet trends and the new content creation practices, the joint application of these changes in ODRL might boost its use as a lingua franca in the open exchange of content.

## References

1. Helberger N, Cabrera F.J, Nikoltchev, S.: Copyright and related rights in the audiovisual sector. Technical report, IRIS. Institute for Information Law. European Audiovisual Observatory (2000)
2. World Intellectual Property Organization: Understanding Copyright and Related Rights. Technical Report 909e, World Intellectual Property Organization. ISBN 92-805-1265-4 (2008)
3. Guth S.: Rights Expression Languages, Lecture Notes in Computer Science, vol. 2770/2003, pp. 101--112 (2003)
4. ISO/IEC 21000-5:2004, Information technology — Multimedia framework (MPEG-21) — Part 5: Rights Expression Language, 2004.
5. The ODRL Initiative, Open Digital Rights Language (ODRL), <http://odrl.net>
6. Open Mobile Alliance, OMA DRM 2.0 specification, [www.openmobilealliance.org](http://www.openmobilealliance.org)
7. Delgado, J., Gallego, I., Llorente, S., García, R.: IPRonto: An Ontology for Digital Rights Management, in Proceedings for the 16th Annual Conference on Legal Knowledge and Information Systems. (2003)
8. Huiying Li Yuzhong Qu, Xiang Zhang: An ontology-based rights expression language. In Proc. of the 13th international World Wide Web conference on Alternate track papers & posters. (2004)
9. Nadah, N., Dulong de Rosnay, M., Bachimont, M.: Licensing digital content with a generic ontology: escaping from the jungle of rights expression languages. In Proceedings of the 11th international conference on Artificial intelligence and law (ICAIL '07). Ed. ACM, New York, NY, USA, pp. 65--69 (2007)
10. Schandl, B., Haslhofer B., Bürger, T., Langegger, A. and Halb, W.: Linked Data and multimedia: the state of affairs, Multimedia Tools and Applications, February, 1-34. (2011)
11. Abelson, H., Adida, B., Linksvayer, M., Yergler, N.: ccREL: The Creative Commons Rights Expression Language. White Paper, March (2008)
12. Berners-Lee, T.: Weaving the Web. Ed. Harper, San Francisco. ISBN 0062515861 (1999)

13. Hausenblas, M., Halb, W.: Interlinking Multimedia Data, in 'Linking Open Data Triplification Challenge at the International Conference on Semantic Systems (I-Semantics08)'. (2008).
14. Rodriguez-Doncel, V., Delgado, J.: A Media Value Chain Ontology for MPEG-21, IEEE Multimedia, Oct-Dec, (2009)
15. WRC Recommendation, Ontology for Media Resources 1.0, March 2011
16. Ianella, R., Findem, A.: Privacy Awareness: Icons and Expression for Social Networks, 8th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods, Namur, (2010)
17. Arnab, A., Hutchison, A.: Extending ODRL to Enable Bi-Directional Communication, In Proceedings of the Second International ODRL Workshop, ADETTI / ISCTE, Lisbon, Portugal (2005)
18. Pucella, R., Weissman, V.: A Formal Foundation for ODRL, in Workshop on Issues in the Theory of Security (WITS), (2004)
19. Garcia, R., Gil, R.; Gallego, I. and Delgado, J.: Formalising ODRL Semantics using Web Ontologies. Open Digital Rights Language Workshop, ODRL'2005 Porto, Portugal. (2005)
20. Peig, E., Delgado, J.: Embedding ODRL statements in Dublin Core, 2nd International ODRL Workshop Lisboa, 7-8 July (2005)