

A Model for Copyright and Licensing: Elsevier's Copyright Model

Anna Tordai
Elsevier, Amsterdam, The Netherlands
a.tordai@elsevier.com

ABSTRACT

With the rise of digital publishing and open access it has become increasingly important for publishers to store information about ownership and licensing of published works in a robust way. In a fast moving environment Elsevier, a leading science publisher, recognizes the importance of sound models underlying its data. In this paper, we describe a data model for copyright and licensing used by Elsevier for capturing elements of copyright. We explain some of the rationale behind the model and provide examples of frequently occurring cases in terms of the model.

Categories and Subject Descriptors

J.7 [Computers in Other Systems]: Publishing; K.4.1 [Computers and Society]: Public Policy—*Intellectual Property Rights*

Keywords

copyright, data model, publishing

1. INTRODUCTION

In the past decade with the rise of digital publishing there have been significant changes in the perception of copyright in the publishing industry [4] [7] [8]. For Elsevier as a major academic publisher it has become increasingly important to keep a record of the copyright status and licenses connected to each piece of content. Journal articles form the vast majority of Elsevier's published content and with the increase of various open access options¹ given to authors this copyright and license landscape has become more and more complex. Licensing in particular plays a role in open access publishing and Elsevier's policy in adherence to the Creative Commons² (CC) licensing approach is that all open access articles must have end-user licenses. The open access options that Elsevier currently provides have not only introduced a variety of licenses for end-users which may allow copying and reuse, but also mean that an article's copyright

and license state may change even after publication. For example, an article published in a subscription-based journal may end up in an issue that is funded by an external third party to become open access, or an author may later opt for his article to become open access after it has been published. As a result, it has become vital for Elsevier to record and update the copyright and license state of every article it publishes.

In this paper we describe the data model that captures the copyright and licensing information of a journal article. This data model is in the process of being implemented at Elsevier and was mainly driven by Elsevier requirements in that existing systems involved in the publishing process must be able to retrieve and supply copyright information automatically even after the article has been published. The model allows us to generate copyright and license notices on the fly so that information presented in products is always accurate. In an increasingly dynamic digital publishing environment we recognize the importance of sound underlying data models that may be extended and updated quickly. The draft model went through internal reviewing process by publishing and legal domain experts. It is important to note that the model was created from the publisher perspective.

The paper is organized as follows: in Section 2 we provide context for the data model in terms of Elsevier's policy around the storage and access of its own content. Next, in Section 3 we describe the copyright model and provide some example cases in Section 4. In Section 5 we describe some of the literature around copyright and compare it to our model. Finally we end with a conclusion.

2. ELSEVIER CONTENT MODEL

The copyright and license data model is a small part of a larger model of Elsevier's content which is composed of scientific publications, enhancements of publications, as well as data from products in specific domains that support scientists in knowledge discovery such as Reaxys³ and Embase⁴. In the past this content was stored in dedicated data stores for each product but as part of Elsevier's current content strategy, our aim is to present these repositories as a virtual whole and describe the content using existing (Linked Data) standards. The content and metadata are made accessible (internally) through APIs with the ability to add, retrieve, modify and delete content and metadata in what we call the Virtual Total Warehouse (VTW) [2]. This content is composed of digital assets of various types such as the

¹<http://www.elsevier.com/about/open-access/open-access-options>

²<http://creativecommons.org/>

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.

ACM 978-1-4503-3473-0/15/05.

<http://dx.doi.org/10.1145/2740908.2742029>.

³www.elsevier.com/Reaxys

⁴www.elsevier.com/online-tools/embase

XML and PDF versions of journal articles, books, figures, as well as various enrichments of content that form the basis of Elsevier products such as ScienceDirect⁵. The metadata of the content exposes its intrinsic properties necessary for retrieval and filtering, such as the type of content, the title, creator(s), and also various characteristics that may change over time even though the content itself does not change.

The metadata describing the content is stored as JSON-LD [9] a W3C recommendation and all Elsevier content stored in the VTW adheres to Linked Data principles [3]. Each piece of content is identified with a unique resolvable URI and is described using a set of properties and values. The context in each JSON-LD file provides mappings from JSON to RDF. We use external metadata standards such as Dublin Core Terms⁶ and PRISM⁷ for capturing content characteristics and PROV⁸ to describe the provenance of digital assets. The values of properties are often resources from vocabularies or lists of values in SKOS⁹ format.

3. COPYRIGHT MODEL DESCRIPTION

The model described in this paper is part of the larger Elsevier content model. It describes copyright-related properties of two classes of objects: Articles and Journals. The model may be adapted to other classes of published objects such as books by expanding the possible values of properties. Although articles are aggregated in journal volumes and issues, they have the same copyright owner as the journal and are not included in the model. In terms of copyright only the relationship between an article and the journal is important which is modelled using the *dct:isPartOf* property. We have defined dedicated copyright namespaces within the Elsevier domain:

- The copyright namespace *cp*: <http://vtw.elsevier.com/data/ns/properties/Copyright-1/>
- The open access namespace for open access licensing related properties *oa*: <http://vtw.elsevier.com/data/ns/properties/OpenAccess-1/>

Next we describe the copyright properties of Journal and Article class objects.

3.1 Journal Properties

Elsevier publishes a variety of journals that are either owned by Elsevier, owned jointly by Elsevier and a third party, or owned by a third party. The owner(s) of the journal, or publication to use a more generic term, is also the copyright holder of the publication. Elsevier is an international company and is the parent company of operating companies registered in different countries. For tax purposes each journal is registered under one of the Elsevier operating companies which is in essence the publisher of the journal. If the journal is owned by Elsevier, then the copyright holder of the journal is the same as the publishing operating company.

A journal has the following properties related to ownership and copyright:

- *cp:copyrightHolder*: The copyright holder(s) of the journal. There may be one or more values for this property, which should be the identifier of the publication owner(s).
- *cp:copyrightHolderType*: The type of copyright holder(s) of the journal of which there may be one or more. Currently we have two possible resource values: Elsevier and “Other publisher”.
- *cp:operatingCompany*: The Elsevier operating company the journal is registered under. We have a small list of values representing the Elsevier companies with URI’s and labels. Example of the label is Elsevier Ltd.

Note that normally the *cp:copyrightHolderType* ought to be attached the Copyright Holder class but we needed to keep the metadata as flat as possible. Currently the data cannot be queried as a graph by internal systems that make use of the copyright information. Instead, the retrieval is limited to key value pairs associated with journals and articles. In the future we hope to remedy this issue and improve the model.

3.2 Article Properties

The copyright and license characteristics of an article are determined by the journal publishing (license) agreement signed by the author that stipulate the rights of the author, the publisher and the end users. Generally speaking there are two types of agreements: ones where authors do not pay a publishing fee but are asked to transfer copyright to the journal owner; and “author-pays” open access where authors sign a licensing agreement granting the journal owner a license to publish the article. Elsevier’s policy is that all open access articles must have an end-user license so that the rights of the readers with respect to use and copying are clear.

A complicating factor in copyright is that in some cases the authors cannot transfer copyright because it is not theirs to transfer. There are cases where the British Crown or the EU claims copyright and grants the publisher a license to publish the work. The US government having funded the work described in the article may claim the article is not copyrightable, which is not the same as the British Crown claiming copyright, and granting a license to the publisher.

Given this landscape the copyright and license model must capture the rights of the author, the publication owner and the end users and cater for exceptions where the copyright holder is someone other than the author or the publication owner.

An article has the following properties:

- *cp:copyrightHolder*: Holds the identifier of the copyright holder of the article who may be the same as the journal copyright holder. The article may have zero or more copyright holders.
- *cp:copyrightHolderType*: Holds the type of the copyright holder of the article of which there may be zero or more. The values are limited to a small set including the values used for the journal copyright holder type: Elsevier, Other publisher, Authors, Crown, US-Government and Other. This list extensible with new types. Strictly speaking the US Government is not a valid copyright holder type as they claim the work is not copyrightable. However, for analytics purposes it is useful to distinguish between this case from cases with no copyright holder.

⁵www.sciencedirect.com

⁶<http://dublincore.org/documents/dcmi-terms/>

⁷www.prismstandard.org/

⁸<http://www.w3.org/TR/prov-overview/>

⁹<http://www.w3.org/2004/02/skos/>

- *cp:copyrightYear*: Holds the year the article was copyrighted. Note that the journal has no copyright year as it is an abstract entity composed of articles organized into issues.
- *cp:copyrightTransfer*: Describes the extent to which the copyright is held by the creator of the work vs the publisher, i.e., the degree of transfer to the publication owner. One of the following values represented by resources may be chosen: “Full transfer to publication owner”, “No transfer to publication owner” and “Transfer not applicable”. The last value is used when the work is in the public domain or when no agreement is sought to be signed by the author.
- *cp:licenseGrantedToPublicationOwner*: Describes the type of license granted to the publication owner which is particularly pertinent if the copyright was not transferred, as without such a license the work could not be published. One of the following values represented by resources may be chosen: “Exclusive license to publish”, “Non-exclusive license to publish”, “Custom license to publish”, “Implied non-exclusive license to publish”, “No license to publish”, and “License not applicable”. Most values represent some type of license granted to the publication owner with two exceptions; first, if copyright was not transferred and the publication owner has no license to publish, then the work may not be published; second if the publication owner has copyright over the work a license to publish does not apply.
- *oa:userLicense*: Represents end-user rights and may have zero or one value. If the article has no user license the Elsevier Terms and Conditions apply¹⁰. If there is a license the values are restricted to the various Creative Commons licenses and an Elsevier license¹¹. This property comes from our internal Open Access model that allows Elsevier to register the open access status of articles and various types of open access funding connected to journals and journal issues. A description of the Open Access model is out of scope for this paper.
- *cp:authorRights*: Describes additional author rights applicable when the author has transferred copyright to the publication owner. The property is not used for default cases but only when rights go beyond the default author rights such as the right to use the article for teaching without needing to obtain special permission.
- *cp:copyrightAgreement*: Stores the type of agreement signed by the article author. Currently we are unable to fill this property as there is direct source of this information; however this property represents an important aspect of copyright pointing to original legal document the details of which cannot be wholly captured by the model.

In addition, we have some properties that are not described in this paper used as flags necessary for triggering processes in systems that use the properties and values.

Given the choice of values the space of all possible combinations appears to be rather large. However, in practice there are some combinations that make no sense and would never occur. These business rules are not formalized cur-

rently in the model, but systems making use of it do apply some checks to validate that the copyright and license properties of articles follow these rules. An example of such a rule is that if the copyright is transferred to the publication owner a license granted to the publication owner is not applicable as the publication owner may publish the article without a need for such a license.

4. EXAMPLE COPYRIGHT CASES

In this section we describe some frequently occurring copyright scenarios and how they are represented by the model.

4.1 Copyright transferred to publication owner

In this scenario the author signs a journal publishing agreement transferring copyright to the journal owner in 2013. The journal in this example is co-owned by Elsevier and a third party owner called the “Society of Things”, and the journal is registered with the UK branch of Elsevier: Elsevier Ltd. The article is not open access, therefore the Elsevier terms and conditions apply for end users. The copyright holder of the article is Elsevier Ltd and the “Society of Things” and there is no need for a publishing license.

The article has the same copyright holder as the journal and in terms of the model inherits the value of the copyright holder from the journal.

The journal properties and values in human readable terms are:

- *cp:copyrightHolder*: Elsevier Ltd, Society of Things
- *cp:copyrightHolderType*: Elsevier, Other Publisher
- *cp:operatingCompany*: Elsevier Ltd

The article properties in human readable terms are:

- *cp:copyrightHolder*: Elsevier Ltd, Society of Things
- *cp:copyrightHolderType*: Elsevier, Other Publisher
- *cp:copyrightYear*: 2013
- *cp:copyrightTransfer*: Full transfer to publication owner
- *cp:licenseGrantedToPublicationOwner*: Not applicable
- *cp:copyrightAgreement*: JPA v1.3

The *oa:userLicense* and *cp:authorRights* properties would have no values.

4.2 Author pays for Open Access

In this scenario the author pays a fee in an open access journal retaining copyright over the article and signing a journal publishing licensing agreement granting the right to publish the article to the journal owner in 2014. The article has an associated CC-BY license in this example.

The article has the following properties:

- *cp:copyrightHolder*: John Doe
- *cp:copyrightHolderType*: Authors
- *cp:copyrightYear*: 2014
- *cp:copyrightTransfer*: No transfer to publication owner
- *cp:licenseGrantedToPublicationOwner*: Exclusive license to publish
- *oa:userLicense*: CC BY license
- *cp:copyrightAgreement*: JPLA v2.1

In this case the author retains copyright captured by the copyright transfer property and the *cp:authorRights* property has no value.

4.3 Crown claims copyright

Here the author signs the journal publishing agreement in 2012 and claims that the British Crown holds the copyright. The journal owner is granted an exclusive license publish

¹⁰<http://www.elsevier.com/legal/standard-terms-and-conditions-of-supply>

¹¹<http://www.elsevier.com/open-access/userlicense/1.0/s>

the article. The article is not open access and therefore the Elsevier terms and conditions apply to end users.

The article has the following properties:

- *cp:copyrightHolder*: Crown
- *cp:copyrightHolderType*: Crown
- *cp:copyrightYear*: 2012
- *cp:copyrightTransfer*: No transfer to publication owner
- *cp:licenseGrantedToPublicationOwner*: Exclusive license to publish
- *cp:copyrightAgreement*: JPA v1.3

5. RELATED WORK

Several ontologies describe the Intellectual Property (IP) domain which covers copyright. In [10] Zhang et al. present ontologies for IP protection. They are process oriented and concern the lifecycle of an IP work including its usage and focus on modeling aspects of IP violations such as piracy. ALIS IP [5] is an IP ontology based on French copyright law centered on three concepts: “work of mind”(1) created by the “author(s)”(2) who are entitled to “IP rights”(3). The ontology covers various rights aspects, such as the distinction between moral rights (eg: attribution rights) and economic rights (eg: copying-, performance rights) and allows categorizations of the works according to type and authorship (eg: collaborative work). The Copyright Ontology by García et al. [6] is very similar to ALIS IP in the concepts it covers, such as the distinction between a work, authorship and rights and similarly to Zhang et.al. has a strong focus on Digital Rights Management and e-commerce.

These ontologies are designed to cover IP law and processes revolving around IP but did not cover our requirements for a copyright model. First, they have a very broad coverage and include many process oriented concepts we do not need. Second, we would have had to create custom properties in order to model use cases such as governments retaining copyright, and capture specific publisher rights. Instead, we developed a light-weight model that covers all our requirements and will create mappings to external ontologies in the future.

Creative Commons provides a legal framework and expression language for communicating the allowed degree of sharing and reuse of content (on the web) thus capturing the rights of end-users. The CC licenses appear in three formats: legal code describing the license in legal terms, an explanation of the license and in Creative Commons Rights Expression Language (CC Rel) [1] a machine readable representation that search engines may recognize. CC Rel is a set of properties needed to license the work including but not limited to its title, the attribution name, the license URI, permissions and prohibitions pertaining to the work. The language is light weight but does not include elements we need in our copyright model such as the copyright holder (attribution name and URL are weak pointers) and the copyright year. We do use CC licenses to indicate end-user rights in open access articles.

6. CONCLUSION

In this paper we presented the copyright and license model that is being implemented at Elsevier for journal articles. Our model describes the rights of the authors, the publisher (publication owner), and the end-user community. The aim is to always identify the copyright holder, the degree of copy-

right transfer from the perspective of the publication owner, the publishing license granted to the publication owner if applicable and the end-user license. In addition, the model allows for some exceptional cases such as those where external agents such as governmental bodies keep copyright or do not allow the work to be copyrightable. We also provided a number of examples of the scenarios that occur in publishing.

The ontologies in IP domain described the previous section have too broad a coverage and we would have had to extend them in order to adequately capture copyright information we need. In the future we intend to map our properties to external ontologies.

The current model is centered on journal publishing but could be extended to apply to book publishing by extending the values of properties and capturing editor rights.

We hope that this model will prompt a dialogue with other (science) publishers on the best way of sharing and storing copyright information. Since journals are routinely bought and sold, and open access content may be reused to some degree, simple standard ways of capturing copyright information would benefit the publishing industry making the navigation of a complex legal landscape easier.

7. REFERENCES

- [1] Abelson, H., Adida, B., Linksvayer, M., and Yergler, N. 2012, CC REL: The Creative Commons Rights Expression Language. The Digital Public Domain. Foundations for an Open Culture
- [2] Allen, B.P. 2011. Linked Data Standards and Infrastructure for Scientific Publishing. W3C Workshop on Linked Data Enterprise Data Patterns (Dec. 2011). <http://www.slideshare.net/bpa777/ledp2011-v3>
- [3] Berners Lee, T. July 27 2006. Linked Data - Design Issues. (Retrieved March 7 2015). <http://www.w3.org/DesignIssues/LinkedData.html>
- [4] Camp., L.J. 2002 DRM: Doesn't Really Mean Digital Copyright Management. Proceedings of the 9th ACM Conference on Computer and Communications Security 78-87.
- [5] Cevenini, C., Contissa, G., Laukyte, M., and Riveret, R. 2009. Ontologies and Law: A Practical Case of the Creation of Ontology for Copyright Law Domain. Handbook of Research on the Social Dimensions of Semantic Technologies and WebServices. 819-37
- [6] García, R., Castella, D., and Rosa, G. 2013 Semantic Copyright Management of Media Fragments. In Proc. 2nd International Conference on Data Management Technologies and Applications (DATA).
- [7] Ryan, B. 2013. 1 - Copyright basics. Optimizing Academic Library Services in the Digital Milieu. Chandos Publishing 3-12.
- [8] Seeley, M., and Wasoff, L. 2012. 15 - Legal Aspects and Copyright. Academic and Professional Publishing. Chandos Publishing 355-383.
- [9] Sporny, M., Longley, D., Kellog, G., Lanthaler, M., and Lindström M. 2014. JSON-LD 1.0 A JSON based serialization for Linked Data W3C Recommendation. <http://www.w3.org/TR/json-ld/>
- [10] Zhang, X.M., Liu, Q., and Wang, H.Q. 2012. Ontologies for intellectual property rights protection. Expert Systems with Applications