

Semantic Copyright Management for Internet-wide Knowledge Sharing and Reuse

Research Paper

Abstract

Purpose In order to extract the full potential from Internet-wide knowledge sharing and reuse, the underlying copyright issues must be taken into account and managed using Digital Rights Management tools.

Design/methodology/approach Traditional Digital Rights Management and open licensing initiatives lack the required computerised support and flexibility to scale to Internet-wide copyright management. Our approach is based on a Semantic Web ontology that conceptualises the copyright domain.

Findings The Copyright Ontology facilitates interoperation while providing a rich framework that accommodates copyright law and copes with custom licensing schemes.

Research limitations/implications The ontology is based on the Description Logic variant of the Web Ontology Language. Despite its scalability, this variant has some expressive limitations that are going to be coped with the help of Semantic Web Rules in future versions of the ontology.

Practical implications The ontology provides the building blocks for flexible machine-understandable licenses and facilitates implementation because existing Semantic Web tools can be easily reused. Moreover, existing initiatives can be mapped to the ontology in order to make it an interoperability hub.

Originality/value The paper contributes a novel approach to Digital Rights Management, based on Semantic Web technologies, that takes into account the underlying copyright legal framework. This is possible thanks to the greater expressiveness of the Semantic Web knowledge representation tools.

Keywords. Knowledge sharing, Copyright law, Knowledge engineering, Semantics, Worldwide web

Introduction

Copyright management is a key issue for Internet-wide knowledge sharing and reuse because most of the artefacts used for knowledge storage and communication are governed by copyright rules.

Traditional Digital Rights Management (DRM) systems show their limitations when forced to interoperate in open environments like the Internet. Moreover, they are not expressive enough to easily accommodate the licensing schemes required by the new knowledge networks emerging in the digital space.

On the other hand, there are open licensing initiatives, like Creative Commons, which show really promising results. However, they lack the required computerised support and flexibility to scale to Internet-wide copyright management.

Our proposal facilitates both interoperation and automation, while providing a rich framework that accommodates copyright law and custom licensing schemes. It is based on a copyright ontology, which is implemented using the Web Ontology Language. This approach facilitates implementation because existing Semantic Web tools can be easily reused.

The rest of this article is organised as follows. First, we explore why we must consider copyright when we try to build an Internet-wide knowledge sharing and reuse network. Existing initiatives and their limitations are presented, from classical and standard DRM to open access proposals like Creative Commons. Then, our Semantic Web approach to copyright-aware DRM is presented, which is materialised in the Copyright Ontology and implemented using Semantic Web tools.

Knowledge Management and Copyright

The first thing to consider is why copyright must be taken into account when we talk about knowledge management. The issue is that, in order to share knowledge, it must be made explicit using some sort of support. This process is the entry point to copyright management, which automatically takes place in order to survey further actions over the resulting artefacts.

The support (a paper, a web page, a sound or video recording, a public presentation, a script, etc.) is used as the mechanism to recognise author's effort, the knowledge he has contributed and the knowledge from which he has derived his work.

Consequently, in order to foster knowledge sharing and reuse, copyright management is a key issue. Copyright law applies to any creation with a degree of originality, i.e. it contributes something new, and so it must be always taken into account, especially when knowledge crosses organisation

boundaries. It is one of the legal frameworks to be considered in order to avoid potential legal problems and on top of which content contracts and licenses are built (Litman, 2001).

Digital Rights Management

Traditionally, copyright management has been achieved through DRM systems. For instance, they have been used by record companies to protect music sold on the Internet and in enterprises in order to control content access.

DRM focuses on controlling content access, the last step in the copyright value chain, and pays little attention to the previous ones: creation, derivation, recording, communication, etc. This is enough in closed domains, like enterprise DRM or vertical content distribution channels.

However, traditional DRM is showing its limitations in Internet-wide scenarios or when it must accommodate new copyright schemes like open source or open access. For instance, a key scenario with these requirements is inter-organisational scientific and technological knowledge sharing and reuse among universities, research centres, etc.

Consequently, there are many efforts trying to solve these limitations. Some of them focus on DRM standardisation in order to achieve interoperability, while others look for innovative licensing schemes that empower the benefits of networked knowledge.

DRM Standardisation

The DRM Watch review on DRM standards (Rosenblatt, 2006) shows that interoperability is a key issue for DRM systems. It arises in the content distribution scenario, for instance when a user wants to consume content in any of the devices he owns, or in the organisational DRM scenario, when knowledge flows through organisations or external content is used in order to derive new knowledge.

The main response to DRM interoperability requirements has been the settlement of many standardisation efforts. One of the main ones is ISO/IEC MPEG-21 (de Walle, 2005), whose main interoperability facilitation component is the Rights Expression Language (REL) (Wang, 2005).

The REL is a XML schema that defines the grammar of a license building language, so it is based on a syntax formalisation approach. There is also the MPEG-21 Rights Data Dictionary (RDD) that captures the semantics of the terms employed in the REL, but it does so without defining formal semantics (García, 2005).

This syntax-based approach is also common to other DRM interoperability efforts and one of main causes of the lack of production

implementations also observed in the DRM Watch review (Rosenblatt, 2006). Despite the great efforts in place, the complexity of the DRM domain makes it very difficult to produce and maintain implementations based on this approach.

The implementers must build them from specifications that just formalise the grammar of the language and force the interpretation and manual implementation of the underlying semantics. This has been feasible for less complex domains but is hardly affordable for a complex domain like copyright, which also requires a great degree of flexibility.

Moreover, the limited expressivity of the technical solutions currently employed makes it very difficult to accommodate copyright law into DRM systems. Consequently, DRM standards follow the traditional access control approach. They concentrate their efforts in the last copyright value chain step, content consumption, and provide limited support for the other steps.

The limited support for copyright law is also a concern for users and has been criticised, for instance, by the Electronic Frontier Foundation (Doctorow, 2005). The consequence of this lack is basically that DRM systems fail to accommodate rights reserved to the public under national copyright regimes.

Consequently, the DRM world remains apart from the underlying copyright legal framework. As it has been noted, this is a risk because DRM systems might then incur then into confusing legal situations. Moreover, it is also a lost opportunity because, from our point of view, ignoring copyright law is also ignoring a mechanism to achieve interoperability.

It is true that copyright law diverges depending on local regimes but, as the World Intellectual Property Organisation [1] promotes, there is a common legal base and fruitful efforts towards a greater level of copyright law worldwide harmonisation.

Creative and Science Commons

As we have shown, the main DRM-world efforts are geared towards interoperability but ignore copyright (Camp, 2002; Samuelson, 2003). DRM keeps the traditional access control approach that is not prepared for the new challenges and opportunities offered by Internet-wide content sharing.

In fact, just Internet publishing risks are considered and the response is to look for more restrictive and secure mechanisms to avoid access control circumvention. This makes DRM even less flexible because it ties implementations to proprietary and closed hardware and software security mechanisms.

A new approach is necessary if we want to extract the full potential from the Internet as a knowledge sharing medium. Currently, many authors and rights holders are reluctant to putting their content online. The existence of this opportunity is clear when we observe the success of the Creative Commons initiative (Lessig, 2002), whose objective is to promote knowledge sharing and reuse through innovative copyright and licensing schemes.

Creative Commons [2] focus initially was on cultural, artistic and educational content. With this scope, there has been a great impact in Internet-wide knowledge sharing and reuse, especially due to open educational content licensing used in initiatives like MIT's Open Courseware [3].

An even greater impact is foreseen in the recently started Science Commons initiative (Wilbanks, 2006). The objective is to promote open access to scientific and technological knowledge, which might solve the contradiction that, in the era of networked content, this kind of knowledge is kept separated and dispersed in many repositories with the consequent loss of opportunities.

However, despite the success of Creative Commons licenses, who estimates more than 140 millions of works licensed under its terms, this initiative is not seen as an alternative to DRM. The main reason is the lack of flexibility of the available licensing terms. There are mainly six different Creative Commons licenses, all of them non-commercial, and just recently a rudimentary protocol [4] has been introduced for extending licenses with custom licensing schemes.

Moreover, although Creative Commons licenses are available in legal form for lawyers, readable "commons deed" for average users and metadata form for computers, there is a lack of formal representations. The Creative Commons metadata schema provides a reduced set of terms for building computer-oriented licenses.

There are three kinds of permissions (reproduction, distribution and derivative works), one prohibition (commercial use) and four requirements (attribution, notice, share alike and source code). For instance, this is not flexible and powerful enough to build the kinds of licenses required by Science Commons, as it is noted in the concept paper for this initiative (Wilbanks, 2006).

Consequently, although it is possible to provide computer support for simple services like content search, there are not mechanisms for customisation and advanced computerised support that enable an Internet-wide copyright-based alternative to DRM systems. And the recent license extension mechanism makes computerised support even harder because custom terms are based on user contributed unstructured text and links.

A Semantic Web Approach to DRM

Our proposal tries to solve the limitations observed in the current DRM and Creative Commons approaches. The underground reason for all of them is the lack of technological tools that allow building a flexible and expressive representation framework.

Such framework must deal with the underlying legal framework and, simultaneously, be automated in order to benefit from computerised support. This would make possible to extract all the potential from Internet-wide knowledge sharing and reuse with the support of accurate copyright management mechanisms.

The first objective is to overcome the limitations of purely syntactic approaches, like XML, and their lack of formal semantics. The best way to formalise semantics is to use knowledge representation technologies in order to build ontologies, which are the tool we propose for expressive and flexible computer-supported copyright management.

An ontology is a formal, explicit specification of a shared conceptualisation. Formal means that it is an abstract model of a portion of the world. It is an explicit specification because it is machine-readable and understandable. Shared implies that it is based on a consensus and it constitutes a conceptualisation because it is expressed in terms of concepts, properties, attributes, etc.

Moreover, as we want to operate through the Internet, the best choice is to use knowledge representation, and more specifically ontology languages, that can operate through this medium. The clear choice is Semantic Web ontologies based on the OWL standard [5], which provides a set of primitives that make possible to build web-sharable conceptualisations.

The increased expressivity of web ontologies allows us to include the underlying legal framework into the formalisation and to build the rest of the system on top of it. This is a key issue because, in order to build a generic framework that facilitates interoperability, the focus must be placed on the underlying legal, commercial and technical copyright aspects.

This is the approach for the Copyright Ontology [6], detailed in the following section. The expressiveness and generality of the resulting conceptualisations allows coping with the shortcoming of existing approaches and, additionally, it can be used as an interoperability facilitator for the main DRM standards, like MPEG-21, or Creative Commons licenses (García, 2007).

Finally, the ontology is implemented as an OWL Web ontology based on the Description Logic (DL) variant, OWL-DL. This implementation

facilitates DRM systems development as license checking is implemented using existing Semantic Web reasoners.

To the best of our knowledge, there is just one other ontological framework for DRM, OntologyX [7]. However, it is a commercial product for which there is little publicly available information. However, from the available information, it is clear that OntologyX concentrates on the kind of actions that can be performed on governed content and it does not take into account the underlying legal framework.

The Copyright Ontology

The copyright domain is quite complex so we face its conceptualisation in three phases. Each phase concentrates on a part of the whole domain. First, the objective is the more primitive part, the Creation Model.

Second, there is the model for the rights part, the Rights Model, and finally a model for the available actions, the Action Model, which is built on top of the two previous ones. This section describes all three models while full details are available from (García, 2006).

The Creation Model conceptualises the different forms a creation can take, which are classified depending on the three main ontological points of view (Niles, 2001):

- Abstract: something that cannot exist at a particular place and time without some physical encoding or embodiment.
 - **Work:** is a distinct intellectual or artistic creation. It includes literary and artistic works, music, pictures and motion pictures, but also computer programs or compilations, like databases.
- Object: it corresponds to the class of ordinary objects and also includes digital objects.
 - **Manifestation:** the materialisation of a work in a concrete medium, a tangible or digital object.
 - **Fixation:** the materialisation of a performance in a concrete medium, a tangible or digital object.
 - **Instance:** the reproduction, copy, of a manifestation, a fixation or another instance.
- Process: something that happens and has temporal parts or stages.
 - **Performance:** the expression in time of a work. Performers or technical methods might be involved in the process.
 - **Communication:** the transmission of a work among places at a given time. It is a process performed when the public is not present at the place and or time where the communication

originates. It includes broadcasts, i.e. one to many, but also communications from a place and at a time individually chosen.

The Rights Model follows the World Intellectual Property Organisation [1] recommendations. It includes economic plus moral rights, as promoted by WIPO, and copyright related rights, see Fig. 1. The most relevant rights in the DRM context are economic rights as they are related to the production and commercial aspects of copyright. Reproduction, Distribution, Public Performance, Fixation, Communication and Transformation Right are the economic rights.

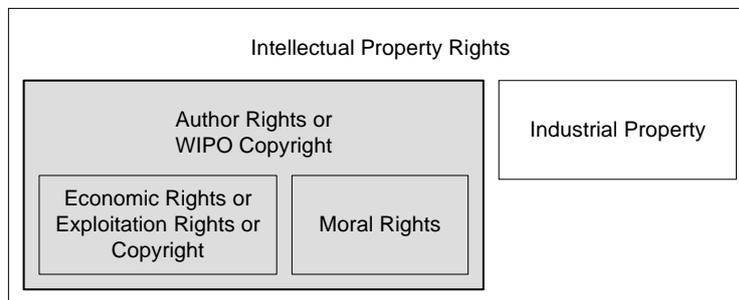


Fig. 1. The Rights Model in the Copyright Ontology

The last model, the Action Model, corresponds to the primitive actions that can be performed on the concepts defined in the Creation Model, as it is shown in Fig. 2. Actions are regulated by the rights in the Rights Model. For the economic rights, these are the governed actions:

- **Reproduction Right:** *reproduce*, commonly speaking *copy*.
- **Distribution Right:** *distribute*. More specifically *sell*, *rent* and *lend*.
- **Public Performance Right:** *perform*; it is regulated when it is a public performance and not a private one.
- **Fixation Right:** *fix*, or *record*.
- **Communication Right:** *communicate* when the subject is an object or *retransmit* when communicating a performance or previous communication, e.g. a re-broadcast. Other related actions, which depend on the intended audience, are *broadcast* or *make available*.
- **Transformation Right:** *derive*. Some specialisations are *adapt* or *translate*.

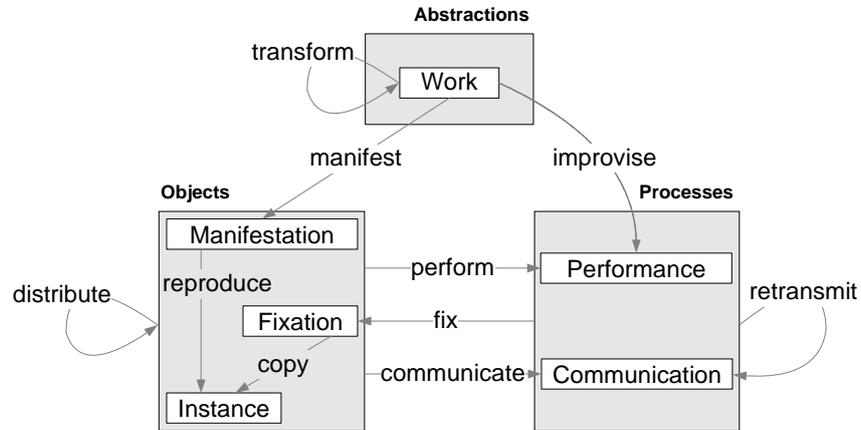


Fig. 2. Relations between the Action and Creation Models

The action concepts are complemented with a set of relations that link them to the action participants. This set is adopted from the linguistics field and it is based on case roles (Sowa, 2000). The case roles are shown in Table I.

Table I. Case roles

	initiator	resource	Goal	essence
Action	agent, effector	instrument	result, recipient	patient, theme
Process	agent, origin	matter	result, recipient	patient, theme
Transfer	agent, origin	instrument, medium	experiencer, recipient	theme
Spatial	origin	path	destination	location
Temporal	start	duration	completion	pointInTime
Ambient	reason	manner	aim, consequence	condition

The general case roles are shown at the top. *Initiator* corresponds to a participant that determines the direction of the process from the beginning. *Resource* is a participant that must be present at the beginning of the process, but not necessarily through it, and does not actively control what happens. On the other hand, *goal* determines the direction of the process from the end. Finally, *essence* is a participant that must be present at the end of the process, but not necessarily through it, and does not actively control what happens.

These generic case roles are specialised depending on the different kinds of facets of verbs. For instance, *patient* is an essential participant in an

action or process that undergoes some structural change as a result of the event, like in “The author revised [the paper]^{patient}”. On the contrary, *theme* is also an essential participant that may be moved, said, or experienced, but is not structurally changed.

The previously introduced pool of primitive copyright-related actions and case roles allows building expressions for many licensing schemes. This flexibility is possible because these building blocks are the more primitive ones, those coming from copyright domain concepts.

Fig. 3 shows how they can be used to model a license that combines commercial and open access licensing terms. It also presents a fundamental action, *Agree*, the primitive for any agreement. The objects of the agreement, connected through the *theme* relation, are two different patterns of authorised *Copy* actions.

The *Copy* pattern on the left grants Publisher Subscribers to copy some content identified by a DOI at any time point six months after “2007-01-01”. Any attempt to exercise this action pattern is subject to a commercial condition, a compensation of 3€. On the other hand, the *Copy* pattern on the right grants anyone to copy the same content, once the period of six months is surpassed, if the aim is non-commercial.

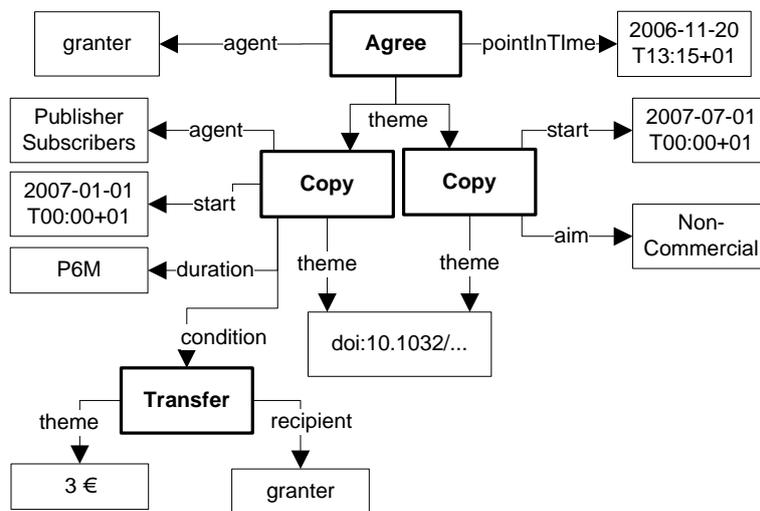


Fig. 3. Model for an agreement on a copy action pattern plus a condition

The deontic operators for permissions, prohibitions and obligations are implicit in the model. The agreement *theme* corresponds to an implicit permission, i.e. the theme of an agreement is permitted. The *condition* on the agreement theme corresponds to an obligation, i.e. in order to fulfil the theme

action it is necessary to satisfy the pattern defined by the *condition* property. Finally, it is also possible to model prohibitions using the *Disagree* action.

OWL Implementation

The previous conceptualisation is just an abstraction of the copyright domain. An implementation is required if we want to use it to build a computerised copyright management system. The Semantic Web approach is also productive in this respect because existing tools can be used to make the implementation quite straightforward.

The ontology has been implemented using the DL variant of the Web Ontology Language (OWL-DL), which is constrained in order to be managed by Description Logic (DL) reasoners. Such reasoners guarantee that OWL-DL ontologies can be put into practice, i.e. reasoned over, in an efficient way.

Existing DL reasoners are used to automatically check if actions on copyrighted content are authorised or not. Licenses are composed of *Agree* or *Disagree* actions, linked through a *theme* relation to patterns of actions that are correspondingly authorised or forbidden.

The pattern is implemented as an OWL class made up from the combination of a set of restrictions. Each restriction defines a constraint on how members of the class, the domain, are related through the specified property to other ones, the range class. The available restrictions in OWL are:

- **allValuesFrom:** all the values for the range of the restricted property must pertain to the given class. For instance, all values of the *agent* relation must pertain to the *Publisher Subscribers* class or, for the *pointInTime* relation, to the time range [2007/01/01–2007/06/30].
- **someValuesFrom:** there is at least one value that pertains to the given range class.
- **hasValue:** the range is limited to a specific individual, not a class of them. For instance, the *theme* of a *Copy* action must be the individual “doi:10.1032/...”.
- **cardinality:** this restriction limits the number of individuals that can be connected through the restricted property. A maximum, minimum or exact cardinality can be defined. For instance, the *recipients* of an action can be limited to just two individuals.

Restrictions are combined using the intersection, union and complement logical operators in order to compose the patterns of actions. For instance, Fig. 4 shows the pattern for the example presented in Fig. 3. For the set of all copy actions on “doi:10.1032/...”, the light grey area, to subsets are selected and their union constitutes the licensed actions pattern, the dark grey areas.

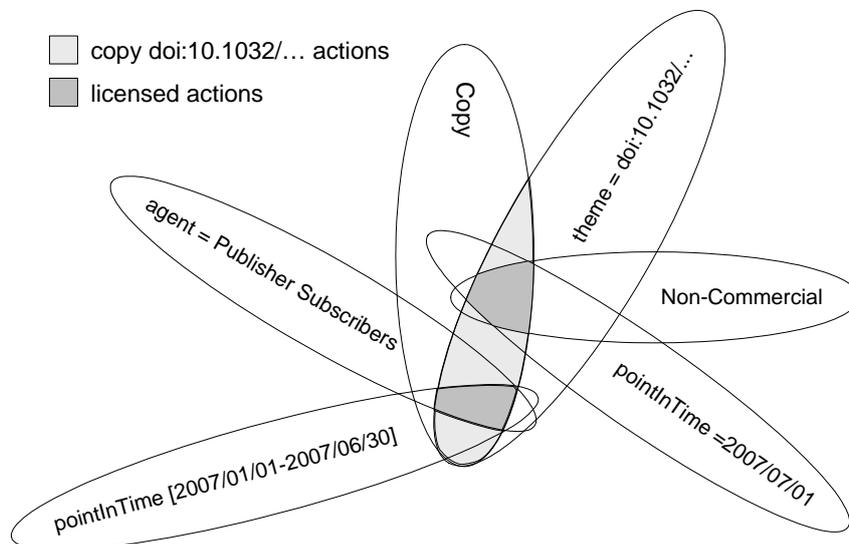


Fig. 4. Building an action pattern as an intersection of restrictions

As it can be seen in Fig. 4, each intersected restriction reduces the set of actions. For instance, the non-commercial pattern does not include any restriction on the agent of the action. Consequently, the licensed actions set includes any non-commercial copy action performed by anyone later than 2007-07-01.

DL reasoners are specially suited to classify individuals into classes. They can answer if an individual, considering its relations to other individuals and attribute values, satisfies all the restrictions of a class pattern and, thus, can be classified as an instance of that class.

In the context of the Copyright Ontology, this functionality is used to check if a particular action, modelled using the ontology as an individual, is allowed or not by a license. This corresponds to the fact that the action individual is classified into a class pattern that is the *theme* of an *Agree*. Another reading is that the license agrees on performing a set of actions that includes the requested one.

However, before the actions is authorised, it is also necessary to check that there is not any disagreement on the action. The DL reasoner checks if the action individual is classified into a class pattern that is the theme of a *Disagree*. Consequently, it is checked that there is an agreement on the action and no disagreement. This behaviour allows modelling complex licenses and revocation. There are more technical details about the Copyright Ontology OWL implementation in (García, 2006).

Conclusions

We are not profiting from the full potential of Internet-wide knowledge sharing and reuse because the underlying copyright issues, inherent to any knowledge expression medium, are not made explicit and dealt with. Instead, the reaction is to protect content using security mechanisms that limit the possibilities, especially for scientific and technological knowledge innovation.

A good example of the potential of a less restrictive approach is Creative Commons licensing schemes for open access and reuse of educational content, materialised in initiatives like MIT's OpenCourseware.

However, Creative Commons does not constitute an alternative to DRM. It lacks flexibility to incorporate alternative license terms, like commercial ones, and advanced computerised support. For instance, in the context of the Science Commons initiative, more expressive tools for computer-oriented licenses are being searched.

Our semantic web approach to copyright management constitutes an alternative. It provides an expressive conceptual framework, the Copyright Ontology, which provides the building blocks for flexible machine-understandable licenses. The ontology is rooted on copyright law and does take the underlying rights into account, even user rights like private copy or citation (Springer, 2007). On top of the ontology, it is possible to reuse existing logical reasoners in order to implement license checking and other services that enable sophisticated copyright management.

Altogether, it constitutes a tool that helps people state the copyright conditions for the knowledge they share and how it might be reused. A way to build an Internet-wide licensing network adapted to particular needs: commercial or non-commercial, open or closed access, reusable share-alike content, etc.

References

- de Walle, R.V. and Burnett, I. (2005), *The MPEG-21 Book*, John Wiley & Sons, London.
- Camp, L.J. (2002), "[DRM: doesn't really mean digital copyright management](#)", in *Proceedings of the 9th ACM conference on Computer and communications security*, Washington, DC, USA, pp. 78-87.
- Doctorow, C. (2005), "Critique of NAVSHP (FP6) DRM Requirements Report", Electronic Frontier Foundation. Available from <http://www.eff.org/IP/DRM/NAVSHP>

García, R. (2006), A Semantic Web Approach to Digital Rights Management, PhD Thesis, Universitat Pompeu Fabra, Barcelona. Available from <http://rhizomik.net/~roberto/thesis>

García, R. and Delgado, J. (2005) “An Ontological Approach for the Management of Rights Data Dictionaries”, in Moens, M. and Spyns, P. (Eds.), Legal Knowledge and Information Systems, IOS Press, Frontiers in Artificial Intelligence and Applications, Vol. 134, pp. 137-146.

García, R., Gil, R. and Delgado, J. (2007), “A Web Ontologies Framework for Digital Rights Management”, Artificial Intelligence and Law, Vol. 15 No. 2, pp. 137-154.

Lessig, L. (2002), The Future of Ideas: The Fate of the Commons in a Connected World, Vintage, New York.

Litman, J. (2001), Digital Copyright: Protecting Intellectual Property on the Internet, Prometheus Books, Buffalo.

Niles, I. and Pease, A. (2001), “Towards a Standard Upper Ontology”, in Welty, C. and Smith, B. (Eds.), Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS), ACM Press, New York, pp. 2-9.

Rosenblatt, B. (2006), “2005 Year in Review: DRM Standards”, DRM Watch, January 2. Available from <http://www.drmwatch.com/standards/article.php/3574511>

Samuelson, P. (2003), “DRM {and, or, vs.} the law”, Communications of the ACM, Vol. 46 No. 4, pp.41-45.

Sowa, J.F. (2000), Knowledge Representation. Logical, philosophical and computational foundations, Brooks Cole Publishing, New York.

Springer, M. and García, R. (2007), “Promoting Music Sampling by Semantic Web-enhanced DRM tools”, in Proceedings of the 5th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods, Virtual Goods'07, Koblenz, Germany.

Wang, X., DeMartini, T., Wragg, B., Paramasivam, M. and Barlas, C. (2005), “The MPEG-21 rights expression language and rights data dictionary”, IEEE Transactions on Multimedia, Vol. 7 No. 3, pp. 408-417.

Wilbanks, J. and Boyle, J. (2006), “An Introduction to Science Commons”, Science Commons concept paper. Available from http://sciencecommons.org/about/ScienceCommons_Concept_Paper.pdf

Footnotes

1. WIPO, <http://www.wipo.int>
2. Creative Commons, <http://creativecommons.org>
3. MIT's Open Courseware, <http://ocw.mit.edu>
4. Creative Commons Plus, <http://creativecommons.org/projects/ccplus>
5. Web Ontology Language (OWL), <http://www.w3.org/2004/OWL>
6. Copyright Ontology, <http://rhizomik.net/ontologies/copyrightonto>
7. OntologyX, <http://www.ontologyx.com>