Licensing Digital Content With A Generic Ontology: Escaping From The Rights Expression Language Jungle

Nadia Nadah, Melanie Dulong de Rosnay, Bruno Bachimont

Licensing Digital Content With A Generic Ontology: Escaping From The Jungle of Rights Expression Languages

Nadia Nadah
Heudiasyc CNRS/UMR 6599
Université de Technologie de Compiègne,
Centre de Recherche de Royallieu
BP 20529 60205 Compiègne
nadia.nadah@hds.utc.fr

Mélanie Dulong de Rosnay
CERSA CNRS
University Paris 2
10 rue Thénard
75005 Paris
melanie.dulong-de-rosnay@cersa.org

Bruno Bachimont
Direction de la Recherche et de l'Expérimentation,
Institut National de l'Audiovisuel
4, Avenue de l'Europe
94366 Bry sur Marne, Cedex
bruno.bachimont@ina.fr

ABSTRACT

Digital contents distributed over the internet are regulated by law and by technical management systems. The latter include a semantic component that describes licenses, i.e. rights of use which are granted to the user. These elements of Digital Rights Management (DRM) systems are called Rights Expression Languages (REL), they gather terms and relations needed to build licenses. Some are based on an ontology of online licenses, not necessarily related to applicable law and various legal systems, and cannot interoperate.
As a consequence, there is a need for a more generic way to express licenses. Here, generic means that rightholders should only need to express the license they need once, and semi-automatic tools should then translate this license so it can be browsed by any specific system. Hence it implies the necessity to be able to model concept semantics in order to translate a license expressed in generic terms into more specific terms that are compliant with the specific standards used by distribution systems. This work comes as part of larger studies on legal ontologies, legal systems and RELs.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; F.4.0 [**Mathematical Logic and Formal Languages**]: General

General Terms

Legal Aspects, Management

Keywords

Digital Right Management systems, Rights Expression Languages, semantic interoperability, legal ontologies, License

1. INTRODUCTION

Due to the pervasive use of DRM-based systems for the distribution of contents over networks, rightholders and distributors decribe the license associated to a particular content in language formats specific to the DRM system they use. For example, OMA ODRL profile is used for mobile phones, XrML for PC, Creative Commons for open content... They define as many licence languages as there are DRM systems formats.
Our approach is to model a generic level by means of an ontology describing the concepts and relations that are useful to define and declare licenses. Such an ontology would make it possible to declare generic licenses that reflect content holders intentions. The next step would then be the translation of this generic license - expressed with our ontology - into licences expressed with specific terms of the existing REL operational standards.
This "express once, translate many times" approach seems promising since it enables rightholders to express licenses in an assessable manner, validating through the ontology that the licensing conditions they write are really what they mean. Moreover, by mapping existing REL standards in this semantically modelled ontology, users may explicitly see whether the translation in, e.g, the ODRL format, are relevant or not, and whether there are discrepancies or inconsistencies between different formats. For example, does the concept labelled "Play" have the same meaning in ODRL or in XrML? Is it related to any national legal definition of reproduction rights?
Our aim is to create a REL based on an ontology, taking both the existing standards (ODRL, XrML, Creative Commons) and applicable legislations into consideration. We intend to seek the relevant level for a generic ontology between existing formats and legal systems. Interoperability is both a technical and a legal issue of standard definition and interpretation by the users, the machine and the lawyer.
This article introduces our work in the Medialex[1]project. Its purpose is to assist licensors' and licensee's work. They indeed have to express licenses associated to a content in each existing format of targetted distribution platforms because

[1]Medialex is an interdisciplinary research project supported by the RIAM Program (French Ministry of National Education and Research) between 2005-2008.

of the pervasive use of DRM-based systems to distribute contents over networks. As a consequence, there is a need for a more generic way to instantiate licenses, that is independent from specific formats. It is challenging to choose the correct terms and the relevant level to interpret them, both on an ontological and a legal point of view.
We will first outline the context of our work. We will then present the need of an ontology of licenses and the methodology used to construct this ontology. Finally, after describing the first results of our work, we will show our contributions and introduce perspectives.

2. DRM JUNGLE : THE PROBLEM

A Digital Right Management system is a consistent plan which control authorisations or restrictions of actions and uses. DRMs aim at considering all the possible usages for each type of digital content (text, video, picture, etc). Therefore the expression of licenses depends on the content's type. Indeed, possible actions on a video are not the same as possible actions on a text or a picture. They also depend of the rights held by the licensor and granted to the licensee.
The best known DRM systems constructors are Apple (FairPlay), Microsoft (WMDRM), OMA (ODRL), Sony (OpenMGX) and RealNetwork (Helix) [12]. DRMs represent a significant business. Despite the fact that there are not many DRM constructors, there is a monopolistic competition that leads to interoperability problems.
The following figure, show the basic mechanism of the major part of a DRM system.

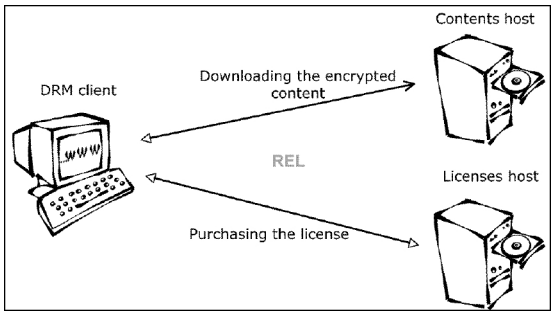

Figure 1: DRM Principle.

As shown in Figure 1, when a digital product is purshased, the license and the content are not sent at the same time. There are two steps : the content delivering, and the license delivering. Our work only focuses on license delivering.
The use of DRM systems generates an important problem. Indeed, it creates an inconsistency between some systems, softwares or hardwares. When a user receives a license, the key can only be used for reading with a specified device. In fact it almost depends on the DRM systems constructor. It is not possible to read on an iPod a music file which was packaged with a DRM (and REL) that is not one of Apple's. The pervasive use of proprietary DRM and REL-based systems in order to distribute content necessitates the definition of each license for each system. Figure 2 presents the "express once, translate many times" approach which enables content holders to express license in an assessable manner, validating through our ontology that what they write is really what they mean and how it can be expressed in several RELs.

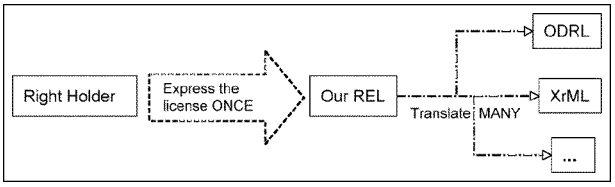

Figure 2: Our approach.

Our approach consists in modeling a generic way of expressing licenses, that would be both independent from and compatible with specific or dedicated formats and legal systems. Hence, we need an ontology which describes concepts and notionsthat may be useful to define and instantiate licenses. Such an ontology is first automatized through a language used to declare generic licenses that reflect rightholders intentions and legal requirements. The next step is the translation of this generic license expressed with the ontology into a series of licences expressed with specific terms of the existing standards. Most existing standards are not based on an ontology. Some are related to a Rights Data Dictionary providing definitions of REL terms. It implies that semantic relations between the words of those dictionaries words are not formalized. The words descriptions are given in a human- readable language (natural language). They are more or less precise, not automatable and not related to any legal system vocabulary or definitions.

3. APPROACH

the use of ontologies, seems to be the optimal way to approach the problem. Jaime Delgado, Isabel Gallego, Roberto García and Rosa Gil already considered this problem, and developed an Intellectual Property Rights Ontology : IPR0nto [8]. This ontology has been built by clearly defining the IPR domain.There are many initiatives around DRM standardisation, but few around REL interoperability (ODRL Creative Commons profile [11]). Our approach is quite similar to IPROnto, but our goal is not exactly the same. Indeed, we aim at construct an ontology of Licenses elements to define a Rights Expression Language in a generic way. It will thus be able to translate a license to others RELs (see Figure 2) and if possible to other legal systems. It is important to find a generic way to express licenses, that is both independent from and related to specific or proprietary formats and legal systems. It seems essential to specify that when we talk about licenses expressed in generic terms, we mean that is the language used to express the license which is generic. After the translation from our new language into one of the existing standards, there is only one license, which can be expressed in different formats. Generic means the ability to model concept semantics in order to translate the license expressed in generic terms into more specific terms compliant with the specific standards used by distribution systems, in order to interpret them according to applicable law.

4. BUILDING THE ONTOLOGY: METHODOLOGY AND PROBLEMS

4.1 What are ontologies?

Ontologies are conceptual resources for knowledge-based systems. They structure concepts, defined along with their semantics. Ontologies are knowledge representation tools.

Concepts and relations are then used to express knowledge-based content [1, 2] Consequently, our ontology of licenses should allow the expression of sentences such as :
"PARTIES lead ACTIONS on RESOURCES under CONDITIONS."

4.2 From existing standards to ontology

The first step is the identification of the concepts used by existing standards. To achieve this goal, we started by studying ODRL and MPEG Rights Data Dictionaries (RDD). There are a lot of iterative concepts, but those concepts are not classified in the same way and a given word can have many meanings according to the standard that is using it. This is why it is so important to define each concept, and also the context in which those concepts are used. This way, we can recognize redundant concepts and rename them in order to avoid redundancy. Consequently, we first selected the concepts we plan to use in our ontology, and classified them in our way. Then we completed the model by legal and contractual rules needed to build licenses.

4.3 From use cases to ontology.

The next step follows a bottom up methodology . Completing ODRL and MPEG scenarii, we defined a global use case to test our model [7] . The purpose was to be able to represent all licensing cases with our ontology. Indeed, with this case, it is easy to make sure that our ontology is generic and that we have not forgotten any possible event. This work leads us to face two problems linked to semantic representation.

4.3.1 Semantic continuity

Existing standards are not based on ontologies but on Rights Data Dictionaries. Therefore, semantics are not formalized. And descriptions of words are given in a human-readable natural language unrelated to any legal system. Moreover, there are discrepancies between different formats. For example, two terms labeled the same way in ODRL and in XrML (for example "Play") may not represent the same concept, while two terms labeled differently may represent the same concept. For example "Copy" in XrML and "Duplicate" in ODRL have the same meaning. The idea is to specify clearly all the concepts of existing standards. We also have to use the ontology to define all the semantic differences between standards in a homogeneous way .The issue of semantic interoperability and consistency is not only technical but also legal as legal terms do not match REL terms. The choice of the adequate level for the legal definition and interpretation of our ontology's terms still has not been fully developed yet.

4.3.2 Cognitive transparency

How can we make sure that a concept labeled in a specific way means what we want to say and will be interpreted by the judge in the meaning we intend? It does not seem possible to achieve this using RDD definitions as they are expressed in natural language. The legal domain is about interpretation of open-textured concepts [9], it is therefore essential to let the user validate the translation of his license. The licensor must be the one who chooses the terms to be used in the license expression. The ontology is the method and the solution at the same time. We must refer to the ontology of licenses to draw the translation steps and the modifications that it will imply at various technical and legal levels. Cognitive interfaces are defined as modes of questioning based on the dynamic shape of users requests rather than on the problem domain static knowledge, "taking into account a meta-knowledge implied in the situation of dialog with a view to adapting it to types of specific requests in a domain" [4] .

4.4 Building our ontology

As it would be possible to build several ontologies for the legal domain, our ontology is deduced from the task it is designed for [3] , i.e. copyright questions and context description and expression by all users and copyright balance maintenance within creative content electronic delivery.
In the chosen interdisciplinary approach, there are two types of components : the concepts and the relations of the ontology and the terms of the dictionaries. Our aim is to map an ontology with a legal dictionary (between a RDD and a thesaurus). That is to say link each terms of the dictionnaries to the ontology. During the ontology's construction we decided to increase the granularity . Indeed, there will be three different situations :

- a term from the legal dictionary corresponds to a concept from the ontology;
- a term from the legal dictionary corresponds to a group of concepts from the ontology or a set composed of a group of concepts linked by a relation described in the ontology ;
- a term from the ontology has no equivalent in the legal dictionary.

Some examples will be presented in the next section (Table 1) on the subject of translating the generic license expressed with the ontology into licenses expressed with specific terms of REL standards.

4.5 Linking our ontology to a Legal Dictionary

As we said, our goal is to translate the license expressed in generic terms into more specific terms compliant with the specific standards used by distribution systems. So, we introduce number of rules to apply the translation. Table 1 presents the mapping between the REL's terms and the concepts that we choose to represent each term. In the next section, we will introduce how we use this tool to translate the license by dealing with an example.

5. CURRENT RESULTS AND DISCUSSION

5.1 Preliminary results

For the moment our ontology is composed of about a hundred concepts. They are divided into five classes (Figure 3). These classes will be specified hereafter except the class called "Resource" which does not have any subclasses yet.

Defined Situation : This class concerns as much the resource as the license's partes.

- "Legal Situations" based on French legal rules dealing with situations related to local copyright law exceptions (to be compared to US fair use provision);
- "Destination" class , is used to allow access to a content in a restricted geographical area;
- "Source" concerns the way to get the content (streaming or downloading);
- "Device situation" means the device used to access the content (mobile, computer, DVD...).

Context Information : Context Information class merges concepts which allow the description of the resource (version, title, digital and physical location, ...)

License's Party : Like "Defined Situation class", "License's Party" class is what we call a legal class. It describes users, diffusers, and rightholders (author, performer, producer, etc).

Action : One of the most important class is "Action". The actions which can be performed on a content are divided into five classes.

- Render Actions (play, print...) - reproduction and performance rights in the legal dictionary
- Configuration Actions (install, uninstall ...)
- File management Actions (access folder informations, etc)
- Transport Actions (copy, transfer, ...)
- Derivative work Actions (extract, edit, etc) - adaptation rights in the legal dictionary

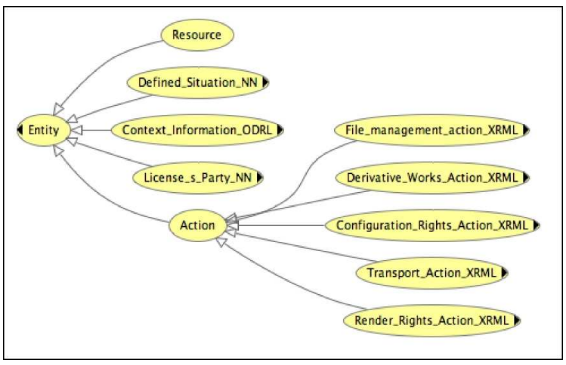

Figure 3: Ontology of licenses : an overview.

5.2 An example

As an illustration, we consider the example presented in [10], that is to say the case of "Alice" who has got the license to print 3 times an e-book titled "Why cats sleep and We don't.". The three following figures show the same example represented using ODRL, XrML, and our ontology's vocabulary. Before trying to translate any license, we need the xml schemas used by ODRL and XrML to validate our translation and to help us constructing this translation.
Due to the task's complexity we cannot translate the license directly. We must complete the translation step by step. The steps are as follows :

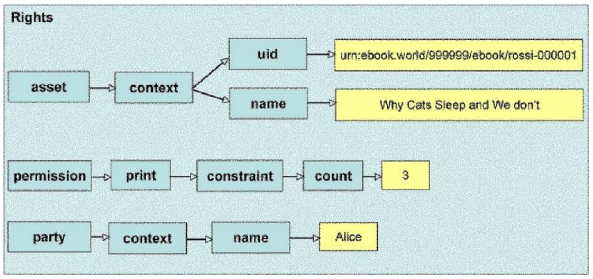

Figure 4: The example license's graph in ODRL.

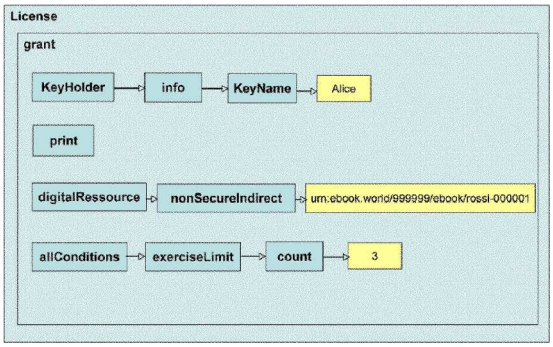

Figure 5: The example license's graph in XrML.

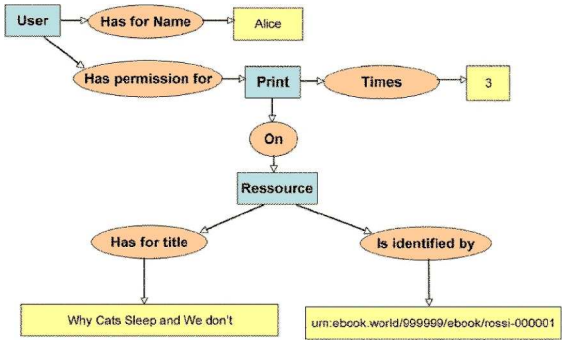

Figure 6: The example license's graph using our ontology.

- the application user (the licensor) chooses the concepts he/she wants to use to create his/her license;
- we construct the license in a generic way. This license will be used to perform the translation;
- the user choose the REL into which he/she wants to translate the license;
- the application creates an xml file in which we put the file's header and the namespaces that corresponds to the output REL;
- the application links each concept from the generically expressed license to the output REL's terms. If the translation needs more information, we ask the user to complete the generic expressed license.

As an example, in order to translate the use case presented previously into ODRL's REL, we start putting the marker <o-ex:rights>. Then, because of the concepts Resource, Action (Print is an Action in the Ontology corresponding to part of the Right of Reproduction of the legal dictionary) and License's Party(User is an License's Party in the Ontology), we respectively create the marker <o-ex:asset> , <o-ex:permission> and <o-ex:party>
If we consider the same sample in XrML, replacing the action "Print" by the action "Play". We need two concepts ("Play" **and** "Display"). So if the user has only selected the concept "Play" we must display a warning like : "In order to translate the notion Play in XrML you must select Play **and** Display". The following figure represents a view of our first prototype.

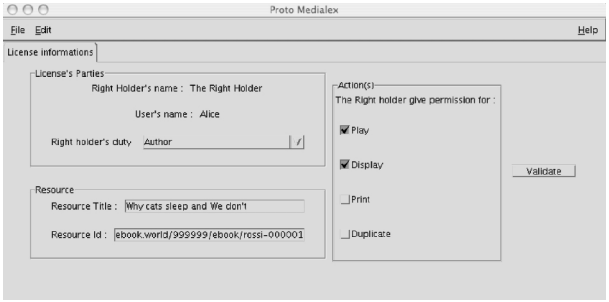

Figure 7: A Medialex's prototype view.

Standard	Term	Concepts
ODRL	Display	Display
ODRL	Play	Play
XrML	Play	Play + Display
ODRL	Print	Print
XrML	Print	Print
ODRL	Duplicate	Duplicate
XrML	Copy	Duplicate
ODRL	Display	Display
XrML	ExerciseLimit	Count
ODRL	Count	Count
XrML	KeyHolder	User
ODRL	Right Holder	Right Holder
XrML	Principal	License's Party
ODRL	Party	License's Party
XrML	Resource	Resource
ODRL	Asset	Resource

Table 1: Parallel between the ontology's concepts and the RDD's terms.

5.3 Discussion

The proposed model will be more complete than the existing standards. Indeed, we started by extracting concepts from those standards. Hence, if a concept is used by a standard and not by the other, we will be able to use it. Moreover, we completed our ontology using legal rules, so we can express licenses concerning French law exceptions. Additionally, we work at mapping ontology terms to a legal dictionary and further study the choice of the relevant level. As a concluding example, we specified rightholder class using legal classifications and legal rules, unlike XrML, which would only be able to calculate remuneration while we would be able to justify this with actors legal roles (the first person is the author, the second is the editor), making contractual sharing clearer for the end user.
We now have to complete our model using our complete use case. Then, we will develop the translation from our REL into the other standards and validate it. The definition of the adequate legal level for terms definition and interpretation deserves further work. We will develop a Legal Dictionary to be mapped to our ontology.

6. ACKNOWLEDGEMENTS

Danièle Bourcier, CERSA/CNRS - Centre Marc Bloch
RIAM Program (French Ministry of National Education and Research)

7. REFERENCES

[1] B. Bachimont. *Engagement Sémantique et Engagement Ontologique : Conception et Réalisation D'ontologies In Ingénierie Des Connaissances*, chapter 19, pages 305–324. Eyrolles, Février 2000.

[2] B. Bachimont, A. Isaac, and R. Troncy. Semantic commitment for designing ontologies: A proposal. In *Asuncion Gomez-Pérez and V. Richard Benjamins, editors, 13th International Conference on Knowledge Engineering and Knowledge Management, EKAW'2002*, pages 114–121, october 2002.

[3] T. Bench-Capon. Task neutral ontologies, common sense ontologies and legal information systems. In *Second International Workshop on Legal Ontologies, JURIX 2001*, 2001.

[4] F. Borges, D. Bourcier, E. Andreewsky, and R. Borges. A conception of cognitive interfaces for legal knowledge - evolution of the jurisque project on the risks of avalanches. In *ICAIL 2001 The Eighth International Conference on Artificial Intelligence and Law*, 2001.

[5] ContentGuard. Xrml 2.0. technical overview.

[6] K. Coyle. Rights expression languages - a report for the library of congress, February 2004.

[7] M. D. de Rosnay. An action-based legal model for dynamic digital rights expression. In *Legal Knowledge and Information Systems. JURIX 2006: The Nineteenth Annual Conference. Amsterdam: IOS Press*, pages 157–162. Tom van Engers (ed), December 2006.

[8] J. Delgado, I. Gallego, S. Llorente, and R. García. An ontology for digital rights management. In *16th Annual Conference on Legal Knowledge and Information Systems. JURIX 2003 Frontiers in Artificial Intelligence and Applications: IOS Press*, may 2003.

[9] H. Hart. *Le concept du droit*. Publication des Facultés Universitaires de Saint Louis, translation of the Concept of Law, Oxford University Press, 1976.

[10] R. Ianella. Open digital rights language (odrl) version 1.1. September 2002.

[11] R. Ianella. Odrl creative commons profile, july 2005.

[12] M. L. Smith. Digital rights management & protecting the digital media value chain. In *MUM '04: Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*, pages 187–191, New York, NY, USA, 2004. ACM Press.