

Semantic Digital Rights Management for Controlled P2P RDF Metadata Diffusion

Roberto García, Giovanni Tummarello

GRIHO – Human-Computer Interaction Research Group
Universitat de Lleida, Spain - roberto@griho.net
SEMEDIA – Semantic Web and Multimedia Group
<http://semidia.deit.univpm.it> - g.tummarello@gmail.com

Since the early works in the W3C Semantic Web initiative, RDF has been generically indicated as a potential basis for legally binding exchange of semantically structured information. In this paper we introduce and detail a procedural framework that could support such legally binding exchange. The proposed methodology is based on a Copyright Ontology, a copyright conceptualisation which includes concrete rights expression languages like MPEG-21 REL, and RDF model decomposition based on the Minimum Self Contained Graph theory. The procedure seems particularly useful when applied to P2P semantic web scenarios.

1. Introduction

The knowledge representation capabilities of RDF are agnostic with respect to the content and the purpose for which it is used. Since the early works in the W3C Semantic Web initiative, however, a few use cases stood out and among these there was the idea that RDF might have been potential basis for legally binding exchange of semantically structured information [1].

In this paper we address a scenario which is becoming more and more common on both the Semantic Web and in “Web 2.0” websites; information does not simply go directly from the source to the intended destination. Instead, information is *mashed up, aggregated, filtered, republished, annotated*, etc. This happens notably with RSS feeds but more on the “Semantic Web”, with frameworks such as DBin [2] where peers collect bits of RDF (related to resources of common interest) which can then be redistributed either to other peers or web republished.

Clearly however, not all data sources would in any case agree on uncontrolled use and redistribution of their produced content. For example, a stock price web service might be willing to provide real time information to a subscriber as long as “it is not publicly redistributed before 10 minutes”. Similarly, in a DBin P2P RDF group, a user might want to give information to other peers “as long as it is redistributed only to those who have a verified @deit.univpm.it address”.

In such scenarios, simple access control to the information sources (e.g. password protected) does not suffice and a non machine readable licence (e.g. a fixed licence that one has to agree with a “I understand the terms and condition” checkbox at sign

up time) would not allow any automatic and dynamic handling of such information distribution scenarios.

The procedure we discuss in this paper addresses such needs and enables a source peer (from here on *source*) to provide a piece of RDF to a receiving peer (*receiver*) in a manner which could provide the technical basis for legal protection.

2. The proposed exchange procedure: outline

In this section we describe the procedure by which the source provides RDF to the receiver along with a licence which specifies how such information may be used. The procedure involves multiple steps requires trust of the identity of the remote party, i.e. the parties must know or have a way to track the legal identity of the creator of the public key that will verify the signing of the licences. There are many ways by which this can be achieved (e.g. via a certifying third party like for example *Verisign*) so the discussion of these is outside the scope of this paper.

For the rest of the discussion we will use the term *cite* to indicate a pointer to the information, e.g. an URL. A *non dereferenciable citation* is a citation by the way of, for instance, a digital hash: a receiver can check that it refers to the information just when it has the information itself or via a third party. With the term *quote* we indicate providing the information itself along with additional control information.

In time steps, the exchange proceeds as follows:

- 1) **R makes a request to S.** As a result of such request R expects S to give information expressed in RDF. *Optionally:* The request is digitally signed so to provide R with a way to make a “personalized” licence offer
- 2) **S receives the request,** creates the RDF for the answer and uses the minimum self contained graph (MSG) decomposition as highlighted in the next chapter to obtain a set of digital hashes which enable to *cite* in a non dereferenciable way the information it is willing to give. Uses the hashes in a licence created with the methodology described in section 3 and sends the result, from here on called *proposal*, to R. *Optionally:* signs the proposal so to provide S with the guarantee that if agreed, the answer will actually be provided within the specified terms
- 3) **R receives the proposal** and, if it decides that the terms are agreeable, signs it and returns it to S. *Optionally:* thanks to the properties of MSGs, R can check if the answer correspond to information which is already locally known. In this case R could drop the request as not interesting, or proceed, e.g., in case it is important for R to prove that the information was in fact legally acquired.
- 4) **S receives the signed proposal,** stores it and replies with the answer computed in 2). *Optionally:* the signed proposal might be countersigned to allow R to prove that the information was obtained by legal means.

2.1. An introduction to the Minimal Self Contained Graph theory

In this section we will illustrate the Minimum Self Contained Graph (MSG) theory. The discussion will deepen that first illustrated in [3] and will provide the bases for the understanding precisely the procedure.

Let's first define what is the minimum "standalone" fragment of an RDF model. As blank nodes are not addressable from outside a graph, they must always be considered together with all surrounding statements, i.e. stored and transferred together with these. MSG are the smallest components of a lossless decomposition of a graph which does not take into account inference such as provided by OWL, as concepts such as RDF-Molecules show [4]. We will here give a formal definition of MSG (minimum Self-contained Graph) and will cite some important properties (for proofs, see [3]).

Def 1. An RDF statement involves a name if it has that name as subject or object.

Def 2. An RDF graph involves a name, if any of its statements involves that name.

Def 3. Given an RDF statement s , the Minimum Self-contained Graph (MSG) containing that statement, written $MSG(s)$, is the set of RDF statements comprised of the statement in question and, recursively, for all the blank nodes involved by statements included in the description so far, the MSG of all the statements involving such blank nodes;

It is possible to show however that the choice of the starting statement is arbitrary and this leads to a unique decomposition of the RDF graph into MSGs.

It is also possible to prove that:

Theorem 1. If s and t are distinct statements and t belong to $MSG(s)$, then $MSG(t) = MSG(s)$.

Theorem 2. Each statement belongs to one and only one MSG.

Corollary 1. An RDF model has a unique decomposition in MSGs.

This is a consequence of theorem 2 and of the determinism of the procedure.

As a consequence of the Corollary 1, a graph can be incrementally transferred between parties by decomposition into MSGs and transfers with granularity down to one MSG at a time. Such transfer would be, as consequence of theorem 2, maximally network efficient as statements would never be repeated.

Definition 4. The RDF Neighbourhood (RDFN) of a resource is the graph composed by all the MSGs involving the resource itself.

Content based identifiers for MSGs

MSGs are standalone RDF graphs. As such they can be processed with algorithms such as canonical serialization. We use an implementation of the algorithm described in [5], which is part of the RDFContextTools Java library [6], to obtain a canonical string representing the MSG and then we hash it to an appropriate number of bits to reasonably avoid collisions. This hash acts as a unique identifier for the MSG with the fundamental property of being *content based*, which implies that two remote peers would derive the same ID for the same MSG in their DB. Sets of such IDs are used to identify the information covered in the licences.

3. Semantic Digital Rights Management

Lately, there have been great works and debate surrounding Digital Rights Management, or DRM. A DRM system (DRMS) is composed of IT components and services along with corresponding law, policies and business models which strive to enable controlled distribution of content and associated usage rights.

It is important for different DRMSs to interoperate. One of the main initiatives for DRM interoperability is the ISO/IEC MPEG-21 standardisation effort. The main interoperability facilitation components are the Rights Expression Language (REL), which is based on a XML grammar and so syntax-based, and the MPEG-21 Rights Data Dictionary (RDD) which captures the semantics of the terms employed in the REL [7]. This one, however, does so without defining a formal semantics [8].

The limitations of a purely syntactic approach and the lack of formal semantics can be overcome using a semantics based approach based on ontologies [9]. Web ontologies are used in order to benefit from the Semantic Web initiative efforts and facilitate its integration in the Web context. The Copyright Ontology [10], of which we give here an overview, is a conceptualisation effort based on OWL.

The copyright domain is a very complex one and its conceptualization is a very challenging task. In order to facilitate this, the Copyright Ontology conceptualisation task has been divided in three parts. Each part concentrates on a portion of the problem. The conceptualisation starts from building a model for the more primitive part, the Creation Model. Then, the following step is to build the Rights Model, and, finally, the Action Model on the roots of the two previous ones. This section just sketches the main points of these three models. For more details, see [11].

The Creation Model defines the different forms a creation can take. These can be classified on the three top categories common in many upper ontologies: **Abstract**, a mental concept, **Object**, a continuant or enduring and **Process**, an occurrent or perdurant. [12].

The Rights Model follows the World Intellectual Property Organisation (WIPO) recommendations in order to define the rights hierarchy. There are the economic rights plus the moral rights, as promoted by the WIPO and adopted by all the countries adhered to the Berne Convention [13].

The more relevant rights in the DRM context are the economic rights as they are related to productive and commercial aspects of copyright. The Action Model corresponds to the primitive actions that can be performed on the concepts defined in the Creation Model and which are regulated by the rights in the Right Model.

For instance, for the economic rights, these are the actions governed by them:

- **Reproduction Right:** *reproduce*, commonly speaking *copy*.
- **Distribution Right:** *distribute*. More specifically *sell*, *rent* and *lend*.
- **Public Performance Right:** *perform*; it is regulated by copyright 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*.

The action concepts are complemented with a set of relations that link them to the action participants. The relations are adopted from the linguistics field and they are based on case roles [14].

The previously introduced pool of primitive actions can be combined in order to build different value chains in the copyright domain. It is complemented with a set of axioms that restrict the ways actions, rights and creation types are related.

The P2P RDF metadata diffusion scenario is governed by the Reproduction and Communication Rights. The Reproduction Right governs the *Copy* action that reproduces a piece of metadata from Peer A, where the piece resides originally, to Peer B, where the piece also resides when the copy is completed.

The Communication Right governs the generic action *Communicate*. This action corresponds, among others, to the situation where the agent responsible for a peer makes content available to others from the place and time individually chosen by them. Therefore, in the context of P2P diffusion, this is the right required by a peer in order to make a piece of metadata available for others to copy.

In order to complete the action model, there are also the licensing actions: *Agree* and *Disagree*, the building blocks for any license, as the one shown in. Fig. 1.

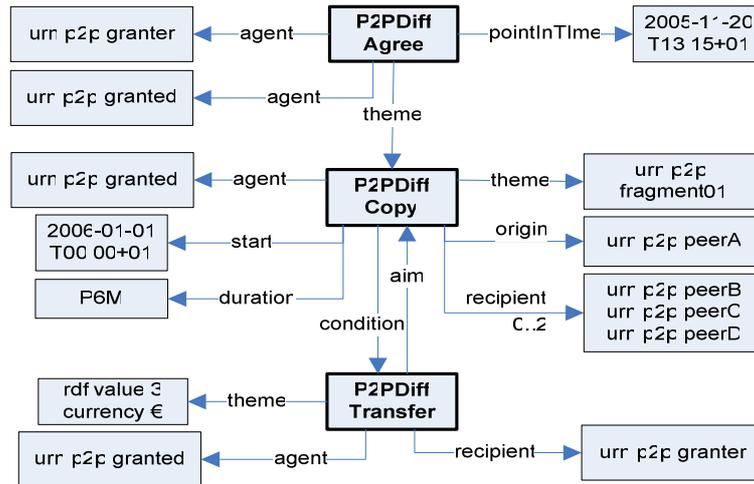


Fig. 1. Model for an agreement on a copy action pattern plus a condition

The deontic operators are implicit in the agreement 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 object. Finally, it is also possible to model prohibitions. This can be done in two ways, by agreeing on a negated pattern or by using the *Disagree* action.

3.1. License Checking, an example

The main objective has been to provide a straightforward and efficient implementation geared towards an extensive use of DL (Description Logic) reasoners.

Licenses are modelled as OWL Classes and copyrighted content intended uses are modelled as instances. In order to check if a usage (instance) is authorised by a set of licenses (classes) a DL reasoner is used to classify the instance in the available classes. If the instance is classified into a class that models an agreement, the *Agree* class as specified in the Copyright Ontology, the usage is authorised.

Suppose, for example, that we want to model a license that allows the agent "granted" to copy the metadata "fragment01" from "peerA" to either "peerB", "peerC" or "peerD". Additional restrictions are that at most it can be simultaneously copied to 2 peers (as a result of an individual copy action) and that the copy can be performed from January 1st 2006 to June 30th 2006.

Table 1 shows the class pattern for the *theme* values of the license *Agree*. The pattern is for *Copy* actions, so it is a subclass of *Copy*, and it is equivalent to the class resulting from the intersection of four OWL restrictions, which constitute the necessary and sufficient conditions that would trigger the classification of authorised usage instances.

Table 1. Class pattern for the actions authorised by the example license

| | | |
|-----------------------|--|-----|
| Pattern \sqsubseteq | Copy | (1) |
| Pattern \equiv | $\forall \text{pointInTime}.\geq 2006-01-01T00:00:00, \leq 2006-06-30T23:59:59 \sqcap$ | (2) |
| | $\exists \text{agent}\{granted\} \sqcap \exists \text{origin}\{peerA\} \sqcap \exists \text{theme}\{fragment0001\} \sqcap$ | (3) |
| | $(\leq 2 \text{ recipient}) \sqcap$ | (4) |
| | $\forall \text{recipient}\{peerC, peerD, peerB\}$ | (5) |

3.2. Implementation

The Semantic DRMS is implemented at two levels. The ground level is about OWL-DL and can be implemented with a common Description Logic reasoner. Pellet¹ has been selected because it can reason over custom data types and this has been very useful to check licensing time ranges.

This however must be complemented with a metalevel that implements the deontic aspects that are implicit in the conceptual model. This metalevel guides the DL checks that have to be performed in order to capture the semantics of the implicit obligations, permissions and prohibitions. The metalevel has been also implemented programmatically.

MSG theory and tools has been implemented in [2] based on the Jena and the Sesame toolkits. The entire procedure as described in this paper is covered in the implementation of an upcoming version of the DBin platform [2] but it will be made available as a standalone library to be used embedded in other applications which exchange RDF.

¹ Pellet OWL Reasoner, <http://www.mindswap.org/2003/pellet>

4. Conclusions

The copyright ontology constitutes a complete framework for representing copyright value chains and the associated flow of rights situations, agreements, offers, etc. This general framework can be specialised and used in conjunction with the Minimum Self Contained RDF graph theory to implement a P2P RDF diffusion mechanism which could form a base for legally binding agreements.

The proposed methodology works based on typical semantic web tools. Licences are implemented as an OWL-DL ontology so an implementation only needs a Description Logic classifier to determine if an action is permitted.

One could say that the proposed approach would be limited to the case of protection against “verbatim” redistribution of information. While this is the case technically (MSG IDs would change with any simple modification, e.g., the insertion of a meaningless triple attached to any blank node), this does not change the validity and applicability of the procedure. It is in fact long established that copyright laws protect not only the exact representation of the protected work but also derived representations. The case is similar to one licensing a photo from a collection, changing a single pixel and wanting to redistribute it as one’s own production outside fair use limits.

We believe this work can have wide applicability and cover real world requirements. The development of this idea was in fact motivated by the need to support much requested use cases in the Semantic Web P2P framework of DBin. As per DBin version 0.4, information is in fact exchanged just based on a URI based request. Under this condition, all that is known by a peer which involves that URI (at MSG level) is shipped to the requesting peer. Thanks to the procedure we propose in this paper it will be now possible to support important use cases involving information which should be exchanged but just in controlled conditions.

4.1. Related Work

While we consider DRM a natural approach for the purpose of this paper, there exist several general policy system which have been applied to SW scenarios. Ontology-based approaches rely on the expressive capabilities of Description Logic languages, such as OWL. DL reasoners can be then used to classify policies and contexts and enable deductive inferences for policy checking.

This is the approach for the Copyright Ontology implementation presented in this paper. A generic policy language also following this approach is KAoS [15] which can reason about licenses by ontological subsumption. KAoS requires however OWL-Full reasoning capabilities and its implementation is based on a theorem prover.

In contrast, rule-based approaches take the perspective of Logic Programming to encode policies as rules with variables. Rei is a policy framework based on rules [16]. Rules are expressed as triples following a pattern that is typical of logical languages like Prolog. In fact, Rei is developed using the XSB Prolog engine. Rei overcomes the variables limitation and enables the definition of policies that refer to dynamically determined values. However, this prevents it from exploiting the full potential of the

OWL language. In fact, Rei rules knowledge is treated separately from OWL ontology knowledge due to its different syntactical form.

To overcome the limitations of this trade-off between ontology and rule-based policies, some have proposed a hybrid solutions [17]. This is also the choice for the Copyright Ontology implementation, as in fact SWRL is used for some axioms and for metalevel reasoning.

References

1. Resource Description Framework (RDF): Concepts and Abstract Data Model. W3C Working Draft 2002. RDF <http://www.w3.org/TR/2002/WD-rdf-concepts-20020829>
2. G. Tummarello, C. Morbidoni, P. Puliti, F. Piazza, "The DBin Semantic Web platform: an overview", WWW2005 Workshop on The Semantic Computing Initiative (SeC 2005)
3. Tummarello G.; Morbidoni C.; Puliti P; Piazza F. "Signing individual fragments of an RDF graph" , 2005, World Wide Web Conference 2005 Poster Track
4. Ding L.; Finin, T; Peng, Y; Pinheiro da Silva, P; , McGuinness, D , "Tracking RDF Graph Provenance using RDF Molecules" , 2005, Proceedings of the Fourth International Semantic Web Conference, November 2005
5. Carroll, J "Signing RDF Graphs", 2003, International Semantic Web Conference 2003
6. Tummarello, G.; Morbidoni C.; "RDFContext Tools 0.2", <http://semedia.deit.univpm.it/tiki-index.php?page=RdfContextTools>
7. Wang, X.; DeMartini, T.; Wragg, B.; Paramasivam, M.; Barlas, C.: "The MPEG-21 rights expression language and rights data dictionary". IEEE Transactions on Multimedia, Vol. 7, No. 3, pp. 408-417, 2005
8. García, R.; Delgado, J.: "An Ontological Approach for the Management of Rights Data Dictionaries". In Moens, M. & Spyns, P. (ed.): "Legal Knowledge and Information Systems". IOS Press, Frontiers in Artificial Intelligence and Applications Vol. 134, 2005
9. García, R.; Gil, R.; Delgado, J.: "A Web Ontologies Framework for Digital Rights Management". In press, Journal of Artificial Intelligence and Law, Springer, 2006
10. Copyright Ontology, <http://rhizomik.net/ontologies/copyrightonto>
11. García, R.: "A Semantic Web Approach to Digital Rights Management". PhD Thesis, Technologies Department, Universitat Pompeu Fabra, Barcelona, ES, 2006. <http://rhizomik.net/~roberto/thesis>
12. Niles, I.; Pease, A.: "Towards a Standard Upper Ontology". In Welty, C.; Smith, B. (eds.): Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS), Maine, USA, 2001
13. Berne Convention, <http://www.wipo.int/treaties/en/ip/berne>
14. Sowa, J.F.: "Knowledge Representation. Logical, philosophical and computational foundations". Brooks Cole Publishing Co., 2000
15. Uszok, A., et al.: "KAoS policy management for semantic web services". IEEE Intelligent Systems, Vol. 19, Num. 4, pp. 32-41, 2004
16. Kagal, L.: "A Policy Based Approach to Governing Autonomous Behavior in Distributed Environments". PhD Thesis, University of Maryland, Baltimore County, USA, 2004
17. Bradshaw, J.; Kagal, L.; Montanari, R.; Toninelli, A.: "Rulebased and ontology-based policies: Toward a hybrid approach to control agents in pervasive environments". In Proceedings of the ISWC2005 Semantic Web and Policy Workshop, 2005