

# Semantic Web for Reliable Citation Analysis in Scholarly Publishing

Ruben Tous,  
Manel Guerrero,  
and Jaime Delgado

*Analysis of the impact of scholarly artifacts is constrained by current unreliable practices in cross-referencing, citation discovering, and citation indexing and analysis, which have not kept pace with the technological advances that are occurring in several areas like knowledge management and security. Because citation analysis has become the primary component in scholarly impact factor calculation, and considering the relevance of this metric within both the scholarly publishing value chain and (especially important) the professional curriculum evaluation of scholarly professionals, we defend that current practices need to be revised. This paper describes a reference architecture that aims to provide openness and reliability to the citation-tracking lifecycle. The solution relies on the use of digitally signed semantic metadata in the different stages of the scholarly publishing workflow in such a manner that authors, publishers, repositories, and citation-analysis systems will have access to independent reliable evidences that are resistant to forgery, impersonation, and repudiation. As far as we know, this is the first paper to combine Semantic Web technologies and public-key cryptography to achieve reliable citation analysis in scholarly publishing.*

In recent years, the amount of scholarly communication brought into the digital realm has exponentially increased.<sup>1</sup> This no-way-back process is fostering the exploitation of large-scale digitized scholarly repositories for analysis tasks, especially those related to impact factor calculation. The potential automation of the contribution-relevance calculation of scholarly artifacts and scholarly professionals has attracted the interest of several parties within the scholarly environment, and even outside of it. For example, one can find within articles of the Spanish law related to the scholarly personnel certification the requirement that the papers appearing in the curricula of candidates should appear in the Subject Category Listing of the Journal Citation Reports of the Science Citation Index.<sup>2</sup> This example shows the growing relevance of these systems today.

---

**Ruben Tous** (rtous@ac.upc.edu) is Associate Professor, **Manuel Guerrero** (guerrero@ac.upc.edu) is Associate Professor, and **Jaime Delgado** (jaime.delgado@ac.upc.edu) is Professor, all in the Departament d'Arquitectura de Computadors, Universitat Politècnica de Catalunya, Barcelona, Spain.

Nevertheless, current practices in citation analysis entail serious problems, including security flaws related to the publishing process (e.g., repudiation, impersonation, and privacy of paper contents) and defects related to citation analysis, such as the following:

- Nonidentical paper instances confusion
- Author naming conflicts
- Lack of machine-readable citation metadata
- Fake citing papers
- Impossibility for authors to control their related citation data
- Impossibility for citation-analysis systems to verify the provenance and trust of citation data, both in the short and long term

Besides the fact that they do not provide any security feature, the main shortcoming of current citation-analysis systems such as ISI Citation Index, CiteSeer (<http://citeseer.ist.psu.edu/>), and Google Scholar is the fact that they count multiple copies or versions of the same paper as many papers. In addition, they distribute citations of a paper between a number of copies or versions, thus decreasing the visibility of the specific work. Moreover, their use of different analysis databases leads to very different results because of differences in their indexing policies and in their collected papers.<sup>3</sup>

To remedy all these imperfections, this paper proposes a reference architecture for reliable citation analysis based on applying semantic trust mechanisms. It is important to note that a complete or partial adoption of the ideas defended in this paper will imply the effort to introduce changes within the publishing lifecycle. We believe that these changes are justified considering the serious flaws of the established solutions, and the relevance that citation-analysis systems are acquiring in our society.

## Reference Architecture

We have designed a reference architecture that aims to provide reliability to the citation and citation-tracking lifecycle. This architecture is based in the use of digitally signed semantic metadata in the different stages of the scholarly publishing workflow. As a trust scheme, we have chosen a public key infrastructure (PKI), in which certificates are signed by certification authorities belonging to one or more hierarchical certification chains.<sup>4</sup>

## Trust Scheme

The goal of the architecture is to allow citation-analysis systems to verify the provenance and trust of machine-readable metadata about citations before incorporating

---

them into their repositories. As a collateral effect, authors and publishers also will be able to store evidences (in the form of digitally signed metadata graphs) that demonstrate different facts related to the creating–editing–publishing process (e.g., paper submission, paper acceptance, and paper publication). To achieve these goals, our reference architecture requires each metadata graph carrying information about events to be digitally signed by the proper subject. Because our approach is based in a PKI trust scheme, each signing subject (author or publisher) will need a public key certificate (or identity certificate), which is an electronic document that incorporates a digital signature to bind a public key with an identity. All the certificates used in the architecture will include the public key information of the subject, a validity period, the URL of a revocation center, and the digital signature of the certificate produced by the certificate issuer’s private key.

Each author will have a certificate that will include as a subject-unique identifier the author’s Universal Resource Identifier (URI), which we explain in the next section, along with the author’s current information (such as name, e-mail, affiliation, and address) and previous information (list of former names, e-mails, and addresses), and a timestamp indicating when the certificate was generated. The certification authority (CA) of the author’s certificate will be the university, research center, or company with which the author is affiliated. The CA will manage changes in name, e-mail, and address by generating a new certificate in which the former certificate will move to the list of former information. Changes in affiliation will be managed by the new CA, which will generate a new certificate with the current information. Since the new certificate will have a new URI, the CA also will generate a signed link to the previous URI. Therefore the citation-analysis system will be able to recognize the contributions signed with both certificates as contributions made by the same author. It will be the responsibility of the new CA to verify that the author was indeed affiliated to the former organization (which we consider a very feasible requirement).

Every time an author (or group of authors) submits a paper to a conference, workshop, or journal, the corresponding author will digitally sign a metadata graph describing the paper submission event. Although the paper submission will only be signed by the corresponding author, it will include the URIs of all the authors.

Journals (and also conferences and workshops) will have a certificate that contains their related information. Their CA will be the organization or editorial board behind them (for instance, ACM, IEEE, Springer, LITA, etc.). If a paper is accepted, the journal will send a signed notification of acceptance, which will include the reviews, the comments from the editor, and the conditions for the paper to be accepted. If the paper is rejected, the journal

might send a signed notification of rejection. We feel that the notification of acceptance is necessary because in a certain kind of curriculum, evaluations for university professors conditionally accepted papers can be counted, and in other curriculums not. The camera-ready version will be signed by all the authors of the paper, not only the corresponding author like in the paper submission.

After the camera-ready version of the paper has been accepted, the journal will send a signed notification of future publication. This notification will include the date of acceptance and an estimate date of publication. Finally, once the paper has been published, the journal will send a signed notification of publication to the author. The reason for having both notification of future publication and notification of publication is that, again, some curriculum evaluations might be flexible enough to count papers that have been accepted for future publication, while stricter ones state explicitly that they only accept published papers.

Once this process has been completed, a citation-analysis system will only need to import the authors’ CA certificates (that is, the certificates of the universities, research centers, and companies) and the publishers’ CA certificates (like ACM, IEEE, Springer, LITA, etc.) to be able to verify all the signed information. A chain of CAs will be possible both with authors (for example, university, department, and research line) and with publications (for example, publisher and journal).

## Universal Resource Identifiers

To ensure that authors’ URIs are unique, they will have a tree structure similar to what URLs have. The first level element of the URI will be the authors’s organization (be it a university or a research center) ID. This organization id will be composed by the country code top-level domain (ccTLD) and the organization name, separated by an underscore.<sup>5</sup> The citation-analysis system will be responsible for assigning these identifiers and ensuring that all organizations have different identifiers.

Then, in the same manner, each organization will assign second-level elements (similar to departments) and so forth.

Author’s CA\_Id: <ccTLD>\_<A\_CA\_name>

Example: es\_upc

Author’s URI: author://<A\_CA\_Id1>/<A\_CA\_Id2> . . . <A\_CA\_Idn>/<author\_firstname>.<author\_familyname>

Example: author://es\_upc.dac/ruben.tous

(In this example “es” is the ccTDL for Spain, UPC (Universitat Politècnica de Catalunya) is the university, and DAC (Departament d’Arquitectura de Computadors) is the department.)

---

Creations' URIs are built in a similar manner to authors' URIs. But in this case, the use of the country code as part of the publisher's ID is optional. Because a creation and its metadata evolve through different stages (submission and camera-ready), we will use different URIs for each phase. We propose the use of this kind of URI instead of other possible schemes such as the Digital Object Identifier (DOI), because the ones proposed in this paper has the advantage of being human readable and contain the CAs chain.<sup>6</sup> Of course, that doesn't mean that once published a paper cannot obtain a DOI or another kind of identifier.

Publisher's CA\_Id: <P\_CA\_Id> or <country\_code>\_<P\_CA\_Id>

Examples: *lita* and *it\_ItalianJournalOfZoology*

Creation's URI: *creation://<P\_CA\_Id1> . . . <P\_CA\_Idn>/<creation\_Id>*

Example: *creation://lita.ital/vol27\_num1\_paper124*

## Confidentiality and Nonrepudiation

Nowadays, some conferences manage their paper submissions and notifications of acceptance (with their corresponding reviews) through e-mail, while others use a web-based application, such as EDAS (<http://edas.info/>).

The e-mail-based system has no means of providing any kind of confidentiality. Each router through which the e-mail travel can see their contents (paper submissions and paper reviews).

The web-based system can provide confidentiality through HTTP Secure (HTTPS), although some of the most popular applications (such as EDAS and MyReview) do not provide it; their developers may not have thought that it was an important feature. The following is a short list of some of the existing web-based systems:

- EDAS (<http://edas.info/>) is probably the most popular system. It can manage a large number of conferences and special issues of journals. It does not provide confidentiality.
- MyReview (<http://myreview.intelligence.eu/index.php>) is an open-source web application distributed under the GPL License for managing the paper submissions and paper reviews of a conference or journal. MyReview is implemented with PHP and MySQL. It does not provide confidentiality.
- ConfTool (<http://www.conftool.net>) is another web-based management system for conferences and workshops. A free license of the standard version is available for noncommercial conferences and events with fewer than 150 participants. It uses HTTPS to provide confidentiality.

- Microsoft's Conference Management Toolkit (CMT; <http://cmt.research.microsoft.com>) is a conference management service sponsored by Microsoft Research. It uses HTTPS to provide confidentiality, but it is a service for which you have to pay.

Although some of the web-based systems provide confidentiality through HTTPS, none of them provides nonrepudiation, which we feel is even more important. This is so because nonrepudiation allows authors to certify their publications to their curriculum evaluators.

Our proposed scheme always provides nonrepudiation because of its use of signatures. Curriculum evaluators don't need to search for the publisher's website to find the evaluated author's paper. In addition, our proposed scheme allows curriculum evaluations to be performed by computer programs. And confidentiality can easily be achieved by encrypting the messages with the public key of the destination of the message. It should not be difficult for authors to obtain the public key for the conference or journal (which could be included in its "call for papers" or included on its webpage). And, because the paper-submission message includes the author's public key, notifications of acceptance, rejection, and publication can be encrypted with that key.

## Modeling the Scholarly Communication Process

Citation analysis systems operate over metadata about the scholarly communication process. Currently, these metadata are usually automatically generated by the citation-analysis systems themselves, generally through a programmatic analysis of the scholarly artifacts unstructured textual contents. These techniques have several drawbacks, as enumerated already, but especially regarding the fact that there is metadata that cannot be inferred from the contents of a paper, like all the aspects of the publishing process. To allow citation-analysis systems accessing metadata about the entire scholarly artifacts lifecycle, we suggest a metadata model that captures a great part of the scholarly domain static and dynamic semantics. This model is based on knowledge representation techniques in Semantic Web, such as Resource Description Framework (RDF) graphs and Web Ontology Language (OWL) ontologies.

### Metadata and RDF

The term "metadata" typically refers to a certain data representation that describes the characteristics of an information-bearing entity (generally another data representation such as a physical book or a digital video file).

Metadata plays a privileged role in the scholarly

---

communication process by helping identify, discover, assess, and manage scholarly artifacts. Because metadata are data, they can be represented through any of the existing data representation models, such as the Relational Model or the XML Infoset. Though the represented information should be the same regardless of the formalism used, each model offers different capabilities of data manipulation and querying. Recently, a not-so-recent formalism has proliferated as a metadata representation model: RDF from the World Wide Web Consortium (W3C).<sup>7</sup>

We have chosen RDF for modeling the citation life-cycle because of its advantages with respect to other formalisms. RDF is modular; a subset of RDF triples from an RDF graph can be used separately, keeping a consistent RDF model. It therefore can be used with partial information, an essential feature in a distributed environment. The union of knowledge is mapped into the union of the corresponding RDF graphs (information can be gathered incrementally from multiple sources). RDF is the main building block of the Semantic Web initiative, together with a set of technologies for defining RDF vocabularies like RDF Schema (RDFS) and the OWL.<sup>8</sup>

RDF comprises several related elements, including a formal model and an XML serialization syntax. The basic building block of the RDF model is the triple subject-predicate-object. In a graph-theory sense, an RDF instance is a labeled directed graph consisting of vertices, which represent subjects or objects, and labeled edges, which represent predicates (semantic relations between subjects and objects).

Coming back to the scholarly domain, our proposal is to model static knowledge (e.g., authors and papers metadata) and dynamic knowledge (e.g., “the action of accepting a paper for publication,” or “the action of submitting a paper for publication”) using RDF predicates.

The example in figure 1 shows how the action of submitting a paper for publication could be modeled with an RDF graph. Figure 2 shows how the example in figure 1 would be serialized using the RDF XML syntax (the abbreviated mode).

So, in our approach, we model assertions as RDF graphs and subgraphs. To allow anybody (authors, publishers, citation-analysis systems, or others) to verify a chain of assertions, each involved RDF graph must be digitally signed by the proper principal. There are two approaches to signing RDF graphs (as also happens with XML instances). The first approach applies when the RDF graph is obtained from a digitally signed file. In this situation, one can simply verify the signature on the file. However, in certain situations the RDF graphs or subgraphs come from a more complex processing chain, and one could not have access to the original signed file. A second approach deals with this situation, and faces the problem of digitally signing the graphs themselves, that is, signing the information contained in them.<sup>9</sup> For

the purpose of the reference architecture described in this paper, we do not instruct which of the two described approaches for signing RDF graphs is to be used. The decision will depend on the implementation (i.e., on how the graphs will be interchanged and processed).

## OWL and an Ontology for the Scholarly Context

To allow modeling the scholarly communication process with RDF graphs, we have designed an OWL Description Logic (DL) ontology. OWL is a vocabulary for describing properties and classes of RDF resources, complementing RDFS’s capabilities for providing semantics for generalization hierarchies of such properties and classes. OWL enriches the RDFS vocabulary by adding, among others, relations between classes (e.g., disjointness), cardinality (e.g., “exactly one”), equality, richer typing of properties, characteristics of properties (e.g., symmetry), and enumerated classes. OWL has the influence of more than ten years of DL research. This knowledge allowed the set of constructors and axioms supported by OWL to be carefully chosen so as to balance the expressive requirements of typical applications with a requirement for reliable and efficient reasoning support. A suitable balance between these computational requirements and the expressive requirements was achieved by basing the design of OWL on the SH family of Description Logics.<sup>10</sup> The language has three increasingly expressive sublanguages designed for different uses: OWL Lite, OWL DL, and OWL Full.

We have chosen OWL DL to define the ontology for capturing the static and dynamic semantics of the scholarly communication process. With respect to the other versions of OWL, OWL DL offers the most expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). OWL DL is so named because of its correspondence with description logics.

Figure 3 shows a simplified graphical view of the OWL ontology we have defined for capturing static and dynamic semantics of the scholarly communication process.

Figure 4, figure 5, and figure 6 offer a (partial) tabular representation of the main classes and properties of the ontology. In OWL, properties are independent from classes, but we have chosen to depict them in an object-oriented manner to improve understanding. For the same reason we have represented some properties as arrows between classes, despite this information being already present in the tables. URIs do not appear as properties in the diagrams because each instance of a class will be an RDF resource, and any resource has a URI according to the RDF model. These URIs will follow the rules described in the above section, “Reference Architecture.” It’s worth mentioning that the selection of the included properties has been based in the study of several metadata formats and standards, such as Dublin

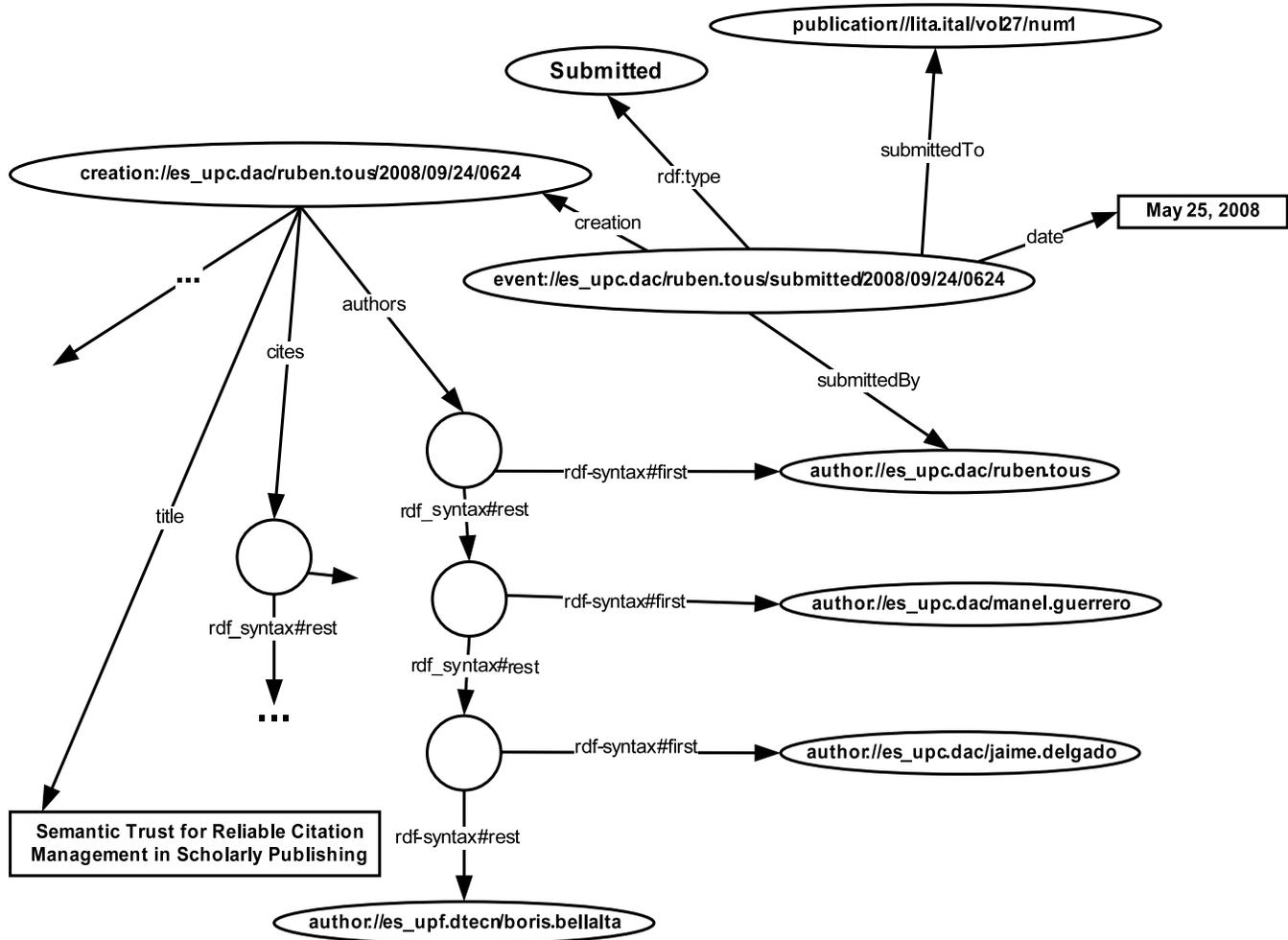


Figure 1. Example RDF Graph

Core (DC), DC's Scholarly Works Application Profile, vCard, and BibTEX.<sup>11</sup>

Figure 4 shows the class *Publication* and its subclasses, which represent the different kinds of publication. In the figure, we only show classes for journals, proceedings, and books. But it could obviously be extended to contain any kind of publication.

Figure 5 contains the classes for the agents of the ontology (i.e., the human beings that author papers and book chapters and the organizations to which human beings are affiliated or that edit publications). The figure also includes the *Creation* class (e.g., a paper or a book chapter).

Finally, figure 6 has the part of the ontology that describes the different events that occur in the process of publishing a paper (i.e., paper submission, paper acceptance, notification of future publication, and publication).

Note that instances of *Submitted* and *Accepted* event classes will point to the same creation instance because no modification of the creation is performed between these events. On the other hand, instances of *ToBePublished* and *Published* event classes will point to different creation instances (pointed by the *cameraReady* and *published-Creation* properties) because of the final editorial-side modifications to which a work can be subject.

## Advantages of the Proposed Trust Scheme

The following is a short list of security features provided by our proposed scheme and attacks against which our proposed scheme is resilient:

---

```

<rdf:RDF xmlns:rdf= "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:ex= "http://example.org/vocab#">
<ex:Submitted rdf:ID= "event://es_upc.dac/ruben.tous/submitted/2008/09/24/0624">
<ex:submittedTo rdf:resource= "publication://lita.ital/vol27/num1"/>
<ex:submittedBy rdf:resource= "author://es_upc.dac/RubenTous"/>
<ex:date rdf:datatype= "&xsd:date">2008-05-25</ex:title>
<ex:creation rdf:resource= "creation://es_upc.dac/ruben.tous/2008/09/24/0624"/>
</ex:Submitted>
<ex:Creation rdf:ID= "creation://es_upc.dac/ruben.tous/2008/09/24/0624">
<ex:title rdf:datatype= "&xsd:string"> Semantic web for Reliable Citation
Management in Scholarly Publishing
</ex:title>
<ex:authors rdf:parseType= "Collection">
<rdf:Description rdf:about= "author://es_upc.dac/ruben.tous"/>
<rdf:Description rdf:about= "author://es_upc.dac/manel.guerrero"/>
<rdf:Description rdf:about= "author://es_upc.dac/jaime.delgado"/>
<rdf:Description rdf:about= "author://es_upf.dtecn/boris.bellalta"/>
</ex:authors>
<ex:cites rdf:parseType= "Collection">
<rdf:Description rdf:about= "http://doi.acm.org/10.1145/511446.511532"/>
<rdf:Description rdf:about= "http://dx.doi.org/10.1137/S0036144502415960"/>
</ex:cites>
<ex:abstract> . . . </ex:abstract>
<ex:keywords> . . . </ex:keywords>
</ex:Creation>
</rdf:RDF>

```

---

**Figure 2.** Example RDF/XML Representation of Graph in Figure 1

- An author can certify to any evaluation entity that will evaluate his or her curriculum the publications that he or she has done.
- An evaluator entity can query the citation-analysis system and get all the publications that a certain author has done.
- An author cannot forge notifications of publication.
- A publisher cannot repudiate the fact that it has published an article once it has sent the certificate.
- Two or more authors cannot team up and make the system think that they are the same person to have more publications in their accounts (not even if they happen to have the same name).

## Implications

The adoption of the approach proposed in this paper has certain implications in terms of technological changes but also in terms of behavioral changes at some of the stages of the scholarly publishing workflow. Regarding the technological impact, the approach relies on the use of Semantic Web technologies and public-key

cryptography. The necessary changes do not apply only to the citation-management software, but also to all the involved parties in the publishing lifecycle (e.g., conference and journal management systems). Authors and publishers would be the originators of the digitally signed evidences, thus user-friendly tools for generating and signing the RDF metadata would be required. Plenty of RDF editors and digital signature toolkits exist, but we predict that conference and journal management systems such as EDAS could easily be extended to provide integrated functionalities for generating and processing digitally signed metadata graphs. This could be transparent to the users because the RDF documents would be automatically generated (and also signed in the case of the publishers) during the creating–editing–publishing process. Because our approach is based on a PKI trust scheme, we rely on a special setup assumption: the existence of CAs, which certify that the identity information and the public key contained within the public key certificates of authors and publishers belong together. To get a publication recognized by a reliable citation-analysis system, an author or a publisher would need a public-key certificate issued by a CA trusted by this citation-analysis system. The selection of trusted

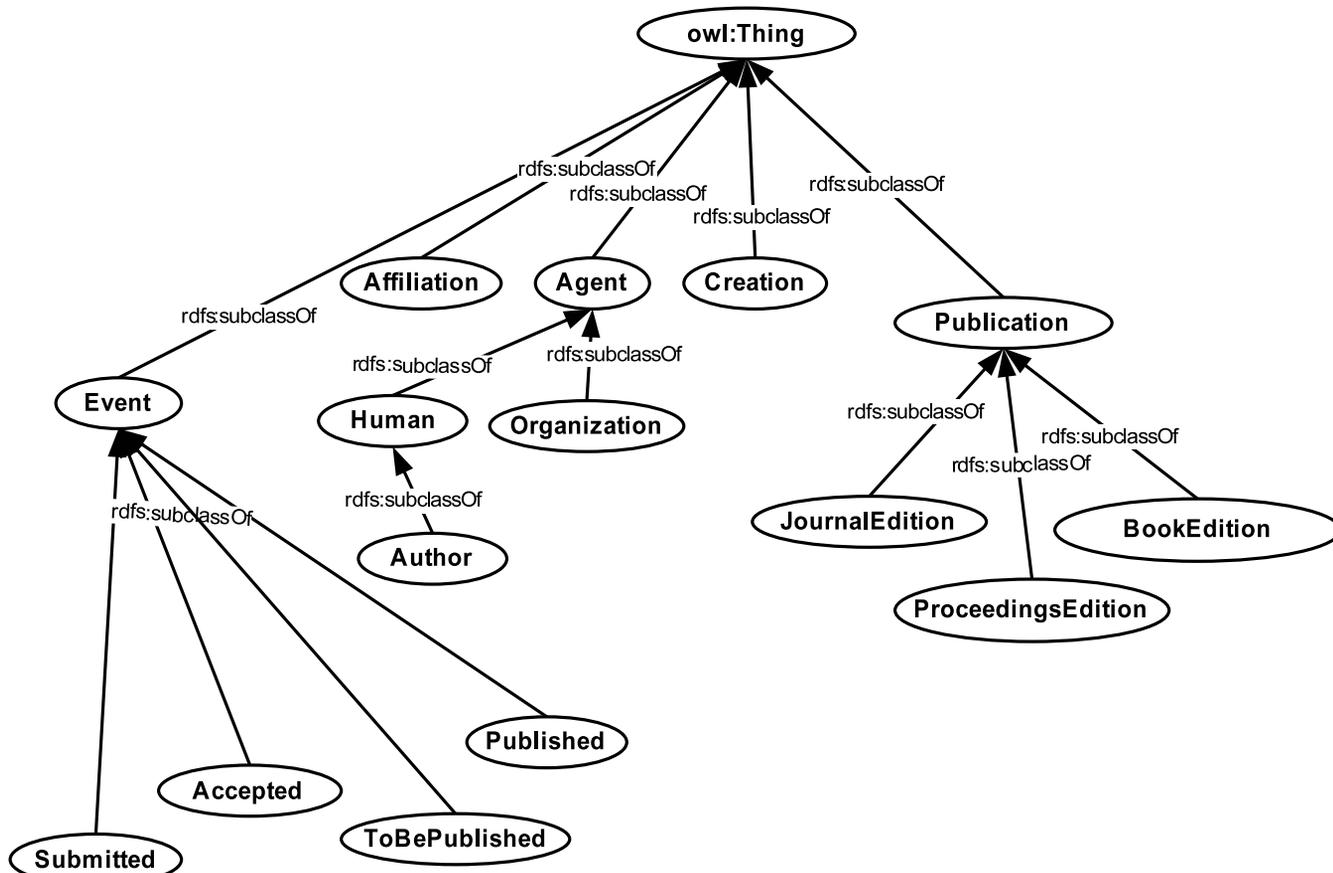


Figure 3. OWL Ontology for Capturing the Scholarly Communication Process

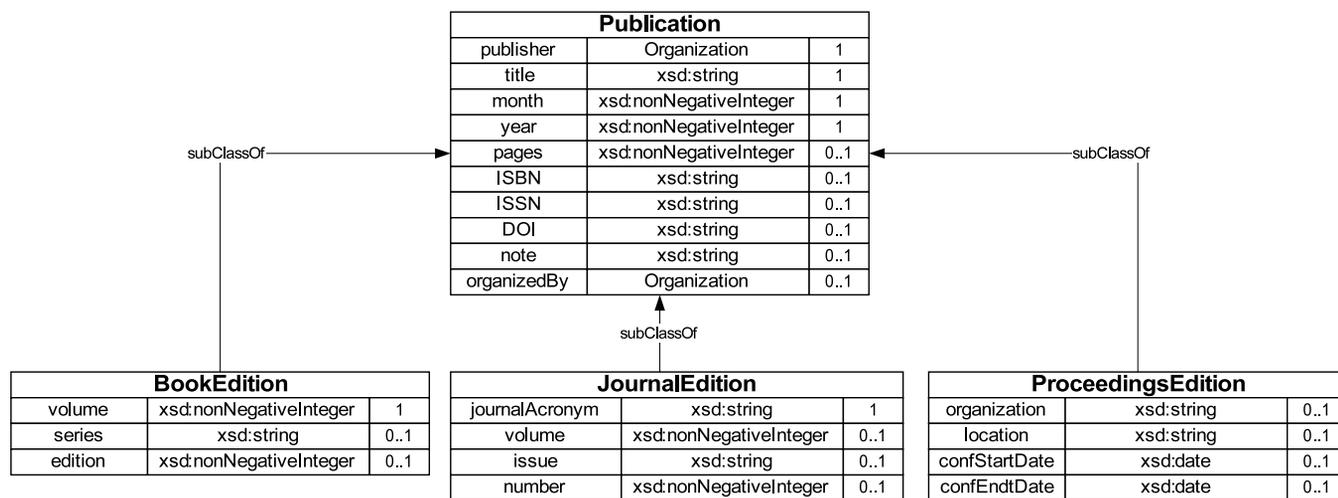
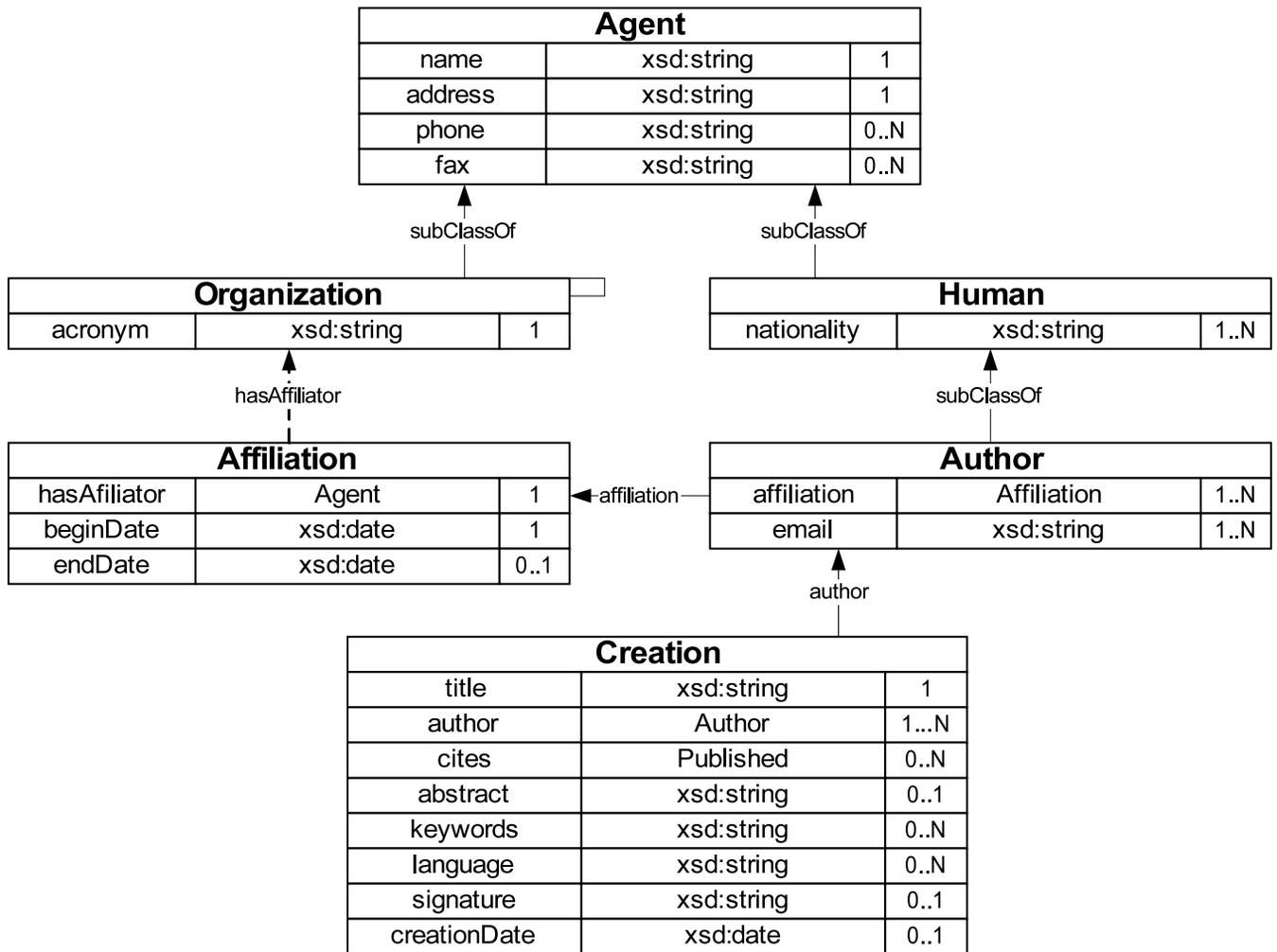


Figure 4. Part of the Ontology Describing Publications



**Figure 5.** Part of the Ontology Describing Agents and Creations

CAs by citation-analysis systems would require the deployment of the necessary mechanisms to allow an author or a publisher to ask for the inclusion of his or her institution in the list. However, this process would be eased if some institutional CAs belonged to trust hierarchies (e.g., national or regional), so including some higher-level CAs makes the inclusion of CAs of some small institutions easier.

Another technological implication is related to the interchange and storage of the metadata. Users and publishers should save the signed metadata coming from a publishing process digitally, and citation-analysis systems should harvest the digitally signed metadata. The metadata-harvesting process could be done in several different ways; but here raises an important benefit of the presented approach: the fact that it does not matter where

the citation-analysis system obtains the information or whether the information is duplicated. The proposed approach guarantees that the citation-analysis subsystem can always verify the provenance and trust of the metadata, and the use of unique identifiers ensures the detection of duplicates.

Our approach also implies minor behavioral changes for authors, mainly related to the management of public-key certificates, which is often required for many other tasks nowadays. A collateral benefit of the approach would be the automation of the copyright transfer procedure, which in most cases still relies on handwritten signatures. Authors would only be required to have their public-key certificate at hand (probably installed in the web browser), and the conference and journal management software would do all the work.

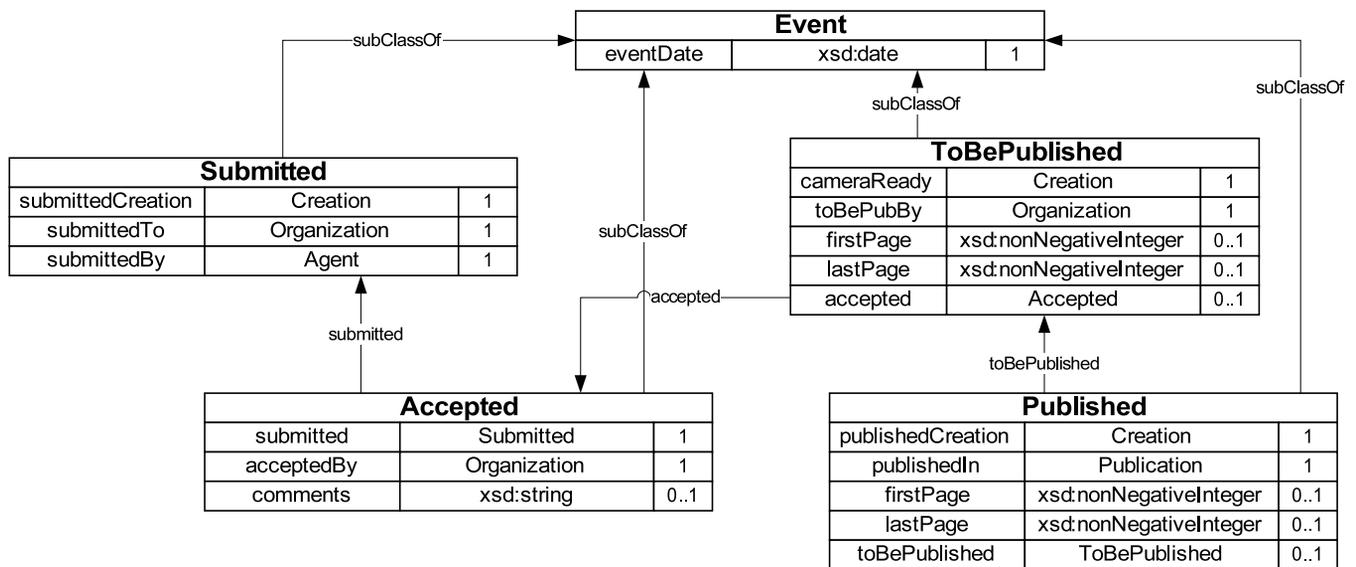


Figure 6. Part of the Ontology Describing Events

## Related Work

As far as we know, this is the first paper to combine Semantic Web technologies and public-key cryptography to achieve reliable citation analysis in scholarly publishing.

Regarding the use of ontologies and Semantic Web technologies for modeling the scholarly domain, we highlight the research by Rodriguez, Bollen, and Van de Sompel.<sup>12</sup> They define a semantic model for the scholarly communication process, which is used within an associated large-scale semantic store containing bibliographic, citation, and use data. This work is related to the MESUR (MEtrics from Scholarly Usage of Resources) project (<http://www.mesur.org>) from Los Alamos National Laboratory. The project's main goal is providing novel mechanisms for assessing the impact of scholarly communication items, and hence of scholars, with metrics derived from use data. As in our case, the approach by Rodriguez, Bollen, and Van de Sompel models static and dynamic aspects of the scholarly communication process using RDF and OWL. However, contrary to what happens in that approach, our work focuses on modeling the dynamic aspects of the creation–editing–publishing workflow, while the approach by Rodriguez, Bollen, and Van de Sompel focuses on modeling the use of already-published bibliographic resources.

Regarding the combination of Semantic Web technologies with security aspects and cryptography, there exist several works that do not specifically focus in the scholarly

domain, but which we have taken in consideration. In our approach, static and dynamic metadata cross many trust boundaries, so it is necessary to apply trust management techniques designed to protect open and decentralized systems. We have chosen a public-key infrastructure (PKI) design to cover such a requirement. However, other approaches exist, such as the one by Khare and Rifkin, which combines RDF with digital signatures in a manner related to what is known as the “Web of Trust.”<sup>13</sup> One aspect of any approach dealing with RDF and cryptography is how to digitally sign RDF graphs. As described above, in the section “Modeling the Scholarly Communication Process with Semantic Web Knowledge Representation Techniques,” there are two different approaches for such a task, signing the file from which the graph will be obtained (which is the one we have chosen) or digitally signing the graphs themselves (the information represented in them), as described by Carroll.<sup>14</sup>

## Conclusions

The work presented in this paper describes a reference architecture that aims to provide reliability to the citation and citation-tracking lifecycle. The paper defends that current practices in the analysis of impact of scholarly artifacts entail serious design and security flaws, including nonidentical instances confusion, author-naming conflicts, fake citing, repudiation, impersonation, etc.

The architecture presented in this work is based in the use of digitally signed RDF graphs in the different stages of the scholarly publishing workflow, in such a manner that authors, publishers, repositories, and citation-analysis systems could have access to independent reliable evidences. The architecture aims to allow the creation of a reliable information space that reflects not just static knowledge but also dynamic relationships, reflecting the full complexity of trust relationships between the different parties in the scholarly domain. To allow modeling the scholarly communication process with RDF graphs, we have designed an OWL DL ontology. RDF graphs carrying instances of classes and properties from the ontology will be digitally signed and interchanged between parties at the different stages of the creation–editing–publishing process. Citation-management systems will have access to these signed metadata graphs and will be able to verify their provenance and trust before incorporating them to their repositories.

Because citation analysis has become a critical component in scholarly impact factor calculation, and considering the relevance of this metric within the scholarly publishing value chain, we defend that the relevance of providing a reliable solution justifies the effort of introducing technological changes within the publishing lifecycle. We believe that these changes, which could be easily automated and incorporated to the modern conference and journal editorial systems, are justified considering the serious flaws of the established solutions and the relevance that citation-analysis systems are acquiring in our society

## Acknowledgment

This work has been partly supported by the Spanish administration (TEC2008-06692-C02-01 and TSI2007-66869-C02-01).

## References and Notes

1. Herbert Van de Sompel et al., “An Interoperable Fabric For Scholarly Value Chains,” *D-Lib Magazine* 12 no. 10 (2006), <http://www.dlib.org/dlib/october06/vandesompel/10vandesompel.html> (accessed Jan. 19, 2011).
2. Boletín Oficial del Estado (B.O.E.) 054 04/03/2005 sec 3 pag 7875 a 7887, <http://www.boe.es/boe/dias/2005/03/04/pdfs/A07875-07887.pdf> (accessed June 24, 2010). See also Thomson ISI Web of Knowledge, <http://www.isiwebofknowledge.com/> (accessed June 24, 2010); and Eugene Garfield, *Citation Indexing: Its Theory and Application in Science, Technology and Humanities* (New York: Wiley, 1979).
3. Judit Bar-Ilan, “An Ego-Centric Citation Analysis Of The Works Of Michael O. Rabin Based on Multiple Citation Indexes,” *Information Processing & Management: An International Journal* 42 no. 6 (2006): 1553–66.
4. Alfred Arsenuault and Sean Turner, “Internet X.509 Public Key Infrastructure: PKIX Roadmap,” draft, PKIX Working Group, Sept. 8, 1998, <http://tools.ietf.org/html/draft-ietf-pkix-roadmap-00> (accessed June 24, 2010).
5. Internet Assigned Numbers Authority (IANA), Root Zone Database, <http://www.iana.org/domains/root/db/> (accessed June 24, 2010).
6. For information on the DOI system, see Bill Rosenblatt, “The Digital Object Identifier: Solving The Dilemma of Copyright Protection Online,” *Journal of Electronic Publishing* 3, no. 2 (1997).
7. Resource Description Framework (RDF), World Wide Web Consortium, Feb. 10, 2004, <http://www.w3.org/RDF/> (accessed June 24, 2010).
8. “RDF Vocabulary Description Language 1.0: RDF Schema. W3C Working Draft 23 January 2003,” <http://www.w3.org/TR/2003/WD-rdf-schema-20030123/> (accessed June 24, 2010); “OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004,” <http://www.w3.org/TR/owl-features/> (accessed June 24, 2010).
9. Jeremy J. Carroll, “Signing RDF Graphs,” in *The Semantic Web—ISWC 2003*, vol. 2870, *Lecture Notes in Computer Science*, ed. Dieter Fensel, Katia Sycara, and John Mylopoulos (New York: Springer, 2003).
10. Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen, “From SHIQ and RDF to OWL: The Making of a Web Ontology Language” *Web Semantics: Science, Services and Agents on the World Wide Web* 1 (2003): 10–11.
11. See the Dublin Core Metadata Initiative (DCMI), <http://dublincore.org/> (accessed June 24, 2010); Julie Allinson, Pete Johnston, and Andy Powell, “A Dublin Core Application Profile for Scholarly Works,” *Ariadne* 50 (2007), [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Type\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme), <http://www.ariadne.ac.uk/issue50/allinson-et-al/> (accessed Dec. 27, 2010); World Wide Web Consortium, “Representing vCard Objects in RDF/XML: W3C Note 22 February 2001,” <http://www.w3.org/TR/2001/NOTE-vcard-rdf-20010222/> (accessed Dec. 3, 2010); and for BibTEX, see “Entry Types,” <http://nwalsh.com/tex/texhelp/bibtex-7.html> (accessed June 24, 2010).
12. Marko. A. Rodriguez, Johan Bollen, and Herbert Van de Sompel, “A Practical Ontology For The Large-Scale Modeling Of Scholarly Artifacts And Their Usage,” *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries* (2007): 278–87.
13. Rohit Khare and Adam Rifkin, “Weaving a Web of Trust,” *World Wide Web Journal* 2, no. 3 (1997): 77–112.
14. Carroll, “Signing RDF Graphs.”