

Semantic Lenses as Exploration Method for Scholarly Articles

Silvio Peroni¹, Francesca Tomasi², Fabio Vitali¹, and Jacopo Zingoni¹

¹ Department of Computer Science and Engineering, University of Bologna, Italy
{essepuntato,fabio}@cs.unibo.it, jacopo.zingoni@gmail.com

² Department of Classical Philology and Italian Studies, University of Bologna, Italy
francesca.tomasi@unibo.it

Abstract. In a move towards an enrichment of the metadata models that are used in the electronic publication of scholarly literature, modern publishers are making steps towards *semantic publishing*. The possibility to explore a collection of scientific papers (a digital library, a repository or an archive of data) using different and multiple facets, i.e., different and multiple points of view on the digital collection, increases on the one hand the success of information retrieval and on the other hand the availability of richer data sets. Multiple facets are the natural navigation method made possible by an adequate ontological representation of a class of homogeneous documents. Context and content of published journal articles are thus components that in the representation of information at the metadata level constitute a fundamental approach to semantic enhancement. In this paper we introduced a test in using a particular semantic publishing model, called *semantic lenses*, to semantically enhance published journal articles.

Keywords: context and content, document semantics, semantic publishing.

1 Introduction

It is a truism to assert that the richness of the metadata model used in digital collections is instrumental in expanding and enhancing the uses made possible by them on the collection, and that models that are too simple may well result in widespread adoption, but on the other hand provide a weak representation of the information contained in the collection, and may induce conceptual errors and misrepresentations, as we discussed (among many) in [1].

Nowhere this is more visible than in the *publishing domain*. Publishers started to use the Web as distribution channel since its early origins [2], but their market exploded with the advent of XML-based languages (e.g. (X)HTML and DocBook), ebook formats (e.g. EPUB and PDF), online vendors (e.g. Apple's iTunes bookstore and Amazon's bookstore), and tablet reading devices (e.g. iPad and Kindle). Similarly, metadata associated to electronic publications while inheriting the results of a multiseccular discipline, library studies, have managed to

coalesce into a number of very simple, minimal, models, such as Dublin Core [3], that although pretty successful are inevitably crippled by their own simplicity.

Simultaneously to the evolution of the Web into the Semantic Web, modern publishers (and in particular scholarly publishers) have taken steps to enhance their digital publications with semantics, an approach that is known as *semantic publishing* [4]. In brief, semantic publishing is the use of Web and Semantic Web technologies to represent formally the meaning of a published document, by specifying a large quantity of information about it as metadata and to publish them as Open Linked Data. As a confirmation of this trend, recently the Nature Publishing Group (publisher of *Nature*), the American Association for the Advancement of Science (publisher of *Science*) and the Oxford University Press have all announced initiatives to open their articles' reference lists and to publish them as Open Linked Data^{1,2}. The open archive movement³ is increasing in the field scientific papers publishing and big commercial companies, like i.e. Springer, support the idea of "institutional repositories" and the concept of "open access publishing" as a solution that "makes your work immediately and permanently available online for everyone, everywhere"⁴. A significant increase of open access journals reveal the impact of new methods of digital publishing⁵.

Open archives as repositories for the dissemination, the interchange and the preservation of scholarly articles and related metadata but also open access as method of publication are becoming a strategy and a paradigm in the field of publishing. Digital libraries of scientific papers use these techniques, theories and methods in order not only to speed up the access given to publications but also to increase the amount of digital data i.e. research articles, they can associate to such articles. Even editors and publishers that did not marry into the open access philosophy are creating digital collections of scientific papers under the guise of freemium platforms for accessing for free the metadata related to their publication, and pay for the full-text of the articles. Even many aggregation platforms (e.g. Elsevier Science and Emerald⁶) found in the possibility to give access to big collections of scientific papers a new way of exploring knowledge.

Many of these platforms are defining semantic models to enhance the digital representation of their articles. However, this enhancement is not a straightforward operation, since it involves much more than simply making semantically

¹ Science joins Nature in opening reference lists:

<http://opencitations.wordpress.com/2012/06/16/science-joins-nature-in-opening-reference-citations>

² Oxford University Press to support Open Citations:

<http://opencitations.wordpress.com/2012/06/22/oxford-university-press-to-support-open-citations>

³ <http://www.openarchives.org>

⁴ <http://www.springer.com/open+access?SGWID=0-169302-0-0-0>

⁵ See the Directory of Open Access Journals:

<http://www.doaj.org/doaj?func=newTitles&fromDate=2012-11-14+00%3A00%3A00&untilDate=2012-12-14+19%3A15%3A31>

⁶ Elsevier Science (<http://www.sciencedirect.com>), Emerald (<http://www.emeraldinsight.com>).

precise statements about named entities within the text. For instance, the sentence “Christopher Marlowe was the real author of many Shakespeare’s plays” has possibly one formal representation, but its use in a scholarly document may be characterized in many different ways, as a claim, an hypothesis, a rebutted concept, or even as an example of an English sentence in a paper not discussing Shakespeare’s plays at all (as this one).

In [5], we showed how several relevant interpretation layers exist beyond the bare words of a scientific paper – such as the context of the publication, its structural components, its rhetorical structures (e.g. Introduction, Results, Discussion), or the network of citations that connects the publication to its wider context of scholarly works. These points of view are usually combined together to create an effective unit of scholarly communication so well integrated into the paper as a whole and into the rhetorical flow of the natural language of the text, so as to be scarcely discernible as separate entities by the reader.

In this paper we use a well-known scholarly paper, *DelosDLMS – The Integrated DELOS Digital Library Management System* by Agosti *et al.* [6], to investigate the feasibility and the usefulness of separating these aspects into eight different sets of machine-readable semantic assertions (called *semantic lenses*), where each set describes one of them, from the most contextual to the most document-specific: research context, authors’ contributions and roles, publication context, document structure, rhetoric organization of discourse, citation network, argumentative characterisation of text, and textual semantics.

The rest of the paper is organised as follows. In Section 2 we introduce some significant works related to semantic publishing experiences and models. In Section 3 we show an application of semantic lenses onto a particular scholarly article. Finally (Section 4) we conclude the paper sketching out some future works and briefly present a prototype named TAL (*Through A Lens*), an HTML interface for scholarly papers.

2 Related Works

Much current literature concerns both the proofs of concepts for semantic publishing applications and the models for the description of digital publishing from different perspective. Because of this richness, here we present just some of the most important and significant works on these topics.

In [4], Shotton *et al.* describe their experience in enriching and providing appropriate Web interfaces for scholarly papers enhanced with provenance informations, scientific data, bibliographic references, interactive maps and tables, with the intention to highlights the advantages of semantic publishing to a broader audience. Along the same lines, in their work [7] Pettifer *et al.* introduce pros and cons of the various formats for the publication of scholarly articles and propose an application for the semantic enhancement of PDF documents according to established ontologies.

A number of vocabularies for the description of research projects and related entities have been developed, e.g. the VIVO Ontology⁷ – researched for

⁷ VIVO Ontology: <http://vivoweb.org/ontology/core>

describing the social networks of academics, their research and teaching activities, their expertise, and their relationships to information resources – and DOAP, the *Description Of A Project*⁸ – an ontology with multi-lingual definitions that contains terms specific for software development projects.

One of the most widely used ontology for describing bibliographic entities and their aggregations is BIBO, the *Bibliographic Ontology* [8]. FRBR, *Functional Requirements for Bibliographic Records* [9], is yet another more structured model for describing documents and their evolution in time. One of the most important aspects of FRBR is the fact that it is not tied to a particular metadata schema or implementation.

Several works have been proposed in the past to model the rhetoric and argumentation of papers. For instance, the SALT application [10] permits someone such as the author “to enrich the document with formal descriptions of claims, supports and rhetorical relation as part of their writing process”. There are other works, based on [11], that offer an application of Toulmin’s model within specific scholarly domains, for instance the legal and legislative domain [12]. A good review of all the others Semantic Web models for the description of arguments can be found in [13]. A comprehensive analysis of the application of Semantic Web ideas and techniques in digital repositories can be read in [14].

3 Context and Content through Semantic Lenses

In [5] we introduced the idea that the semantics of a document is definable from different perspectives, where each perspective is represented as a *semantic lens* that is *applied* to a document to reveal a particular semantic facet.

A faceted classification system [15] in the field of library science is a bottom-up scheme that divides a subject into concepts and gives rules to use these concepts in constructing a structured subject. This approach makes it possible use a kind of poly-hierarchical relationship between the elements of the description [16].

But facets have to be transformed in an ontology in order to give access to the deep meaning of the documents. An ontology has been defined⁹ to formally define these lenses so as to allow the annotation of resources such as scholarly papers. In addition, since the application of the semantic lenses to a document is an *authorial activity*, i.e. the action of a person (the original author as well as anyone else) taking responsibility for a semantic interpretation of the document, we also need to record the provenance of the semantic statements according to the *PROV Ontology* (*PROV-O*) [17].

In the following subsections we introduce the lenses using the well-known paper *DelosDLMS – The Integrated DELOS Digital Library Management System* by Agosti et al. [6] as the scholarly article on which the small snippets of semantic lenses are based.

⁸ DOAP: <http://usefulinc.com/ns/doap>

⁹ Lens Application Ontology (LAO): <http://www.essepuntato.it/2011/03/lens>

3.1 Describing the Context

Writing a scientific paper is usually the final stage of an often complex collaborative and multi-domain activity of undertaking the research investigation from which the paper arises. The organizations involved, the people affiliated to these organizations and their roles and contributions, the grants provided by funding agencies, the research projects funded by such grants, the social context in which a scientific paper is written, the venue within which a paper appears: all these provide the research *context* that leads, directly or indirectly, to the genesis of the paper, and awareness of these may have a strong impact on the credibility and authoritativeness of its scientific content.

The concept of context is a polysemic textual situation because it runs across a variety of different disciplines. In general “the broad notion of context [is] constituted by the interactions and relationships between a TE [target entity] and its environment” [18]. In particular in the archival domain this concept regards the need to separate the description of document from the description of people that create the documents. The EAC-CPF (Encoded Archival Context-Corporate Bodies, Persons and Families) is a DTD¹⁰, an XML Schema¹¹ and now an ontology [19] for translate the ISAAR (CPF), the International Standard for Archival Authority Records [20], in a formal language.

Daniel Pitti states that “relations between records, creators, and functions and activities are dynamic and complex, and not fixed and simple. Creators are related to other creators. Records are related to other records. Functions and activities are related to other functions and activities. And each of these is inter-related with the others. [...] By developing dedicated semantics and structures for describing each descriptive component and its complex interrelations, we can build descriptive systems that are far more efficient and effective than those we have realized in print” [21]. So the context reflects the need to separate the object (the paper) from the information surrounding it, and in fact the context reflects the relationships between data and structured metadata, but is also an interpretation key of the document as a complex entity whose information emerges only when analysing the elements of the document in their specific context.

Given these assumptions, we need to point out that semantic lenses have to be used as a complex system, in a network perspective of interconnected scopes, rather than as a hierarchical model of independent layers.

Three lenses are designed to cover the contextual aspects of a scholarly text:

- *Research context*: the background from which the paper emerged (the research described, the institutions involved, the sources of funding, etc.).
- *Contributions and roles*: the individuals claiming authorship on the paper and what specific contributions each of them provided.
- *Publication context*: any information about the event (e.g. conference or workshop) and publication venue of the paper (such as the proceedings or the journal), as well as connections to the other papers sharing the same event or venue.

¹⁰ ISAAR(CPF) DTD: <http://www3.iath.virginia.edu/eac/>

¹¹ ISAAR(CPF) Schema: <http://eac.staatsbibliothek-berlin.de>

Using [6] as the basis for the annotations example, we describe the *contextual environment*, that is the *research context*, that made possible writing this paper by using¹² *FRAPO*, the *Funding, Research Administration and Projects Ontology*¹³, as shown in the following excerpt. The excerpt specifies that the European Commission, as a funding agency (#1), funded the network of excellence DELOS (#2) that led to the aforementioned paper (#3)¹⁴:

```
:research-context {
:ec a frapo:FundingAgency ; foaf:name "EU Commission" ; #1
  frapo:funds [ a frapo:Endeavour ; #2
    foaf:name "A Network of Excellence on Digital Libraries";
    frapo:hasOutput :delosdlms ] . } #3
```

Then we use *SCoRO* (the *Scholarly Contributions and Roles Ontology*¹⁵) and its imported ontology *PRO* (the *Publishing Roles Ontology*¹⁶) [22] to identify the *roles and contributions*. Once again, in order to be concise, only the code for one of the many contributors will be shown, in this case for the first one, Maristella Agosti. We can identify her *role* (e.g. being affiliate with the University of Padua during the realization of the paper – #4) and her *contribution* (#5) within the context of this paper.

```
:contributions-and-roles {
:agosti a foaf:Person ; foaf:name "Maristella Agosti" ;
  pro:holdsRoleInTime [ a scor:OrganizationalRole ; #4
    pro:withRole scor:affiliate ;
    pro:relatesToOrganization [ a frapo:University ;
      foaf:name "University of Padua" ] ;
    pro:relatesToDocument :delosdlms ],
  scor:makesContribution [a scor:ContributionSituation ; #5
    scor:withContribution scor:writes-manuscript-draft ;
    scor:withContributionEffort scor:major-effort ;
    scor:relatesToEntity :delosdlms ] }
```

We then describe the *publication context* of the paper using FaBiO, the *FRBR-aligned Bibliographic Ontology* [23] and BiRO, the *Bibliographic Reference Ontology*¹⁷, specifying the conference proceedings in which the paper was published

¹² Note that all the ontologies used or suggested in this paper to describe “lenses” statements have been chosen as an appropriate and convincing example of an ontology that fulfils the requirements for the lens, since they allow us to fully describe all the document aspects we are interested in. However, their choice is not unique, and many other ontologies may exist to fulfil the same role, so as to allow the use of other models (such as those described in Section 2) instead of them.

¹³ *FRAPO*: <http://purl.org/cerif/frapo>

¹⁴ This and the following RDF examples are written in Turtle (<http://www.w3.org/TeamSubmission/turtle/>), with namespace definitions defined at <http://www.essepuntato.it/2013/tal/prefixes>.

¹⁵ *SCoRO*: <http://purl.org/spar/scoro>

¹⁶ *PRO*: <http://purl.org/spar/pro>

¹⁷ *FaBiO*: <http://purl.org/spar/fabio>; *BiRO*: <http://purl.org/spar/biro>.

(#6) and the list of its references to other related documents (#7) – which is crucial for semantic publishing:

```
:publication-context {
# The textual realisation of the paper
:version-of-record a fabio:ConferencePaper ; #6
  frbr:realisationOf :delosdlms ;
  dcterms:title "DelosDLMS - The Integrated DELOS Digital
    Library Management System" ;
  prism:doi "10.1007/978-3-540-77088-6_4" ;
  frbr:partOf [ a fabio:ConferenceProceedings ;
    dcterms:title "Proc. 1st International DELOS Conference";
    fabio:hasPublicationYear "2007"^^xsd:gYear ]
  frbr:part [ a biro:ReferenceList ; #7
    co:element [ biro:references
      <http://dx.doi.org/10.1109/ICCV.1998.710779> ] ... ]}
```

3.2 Describing the Content

The semantics of *the content* of a document, i.e. the semantics that is implicitly defined in and inferable from the text, can be described from different points of view. For example, the overall *structure* of the text – i.e. the organization of the text of the document into structured containers, blocks of text, inline elements – is often expressed by means of markup languages such as XML and LaTeX, that have constructs for describing content hierarchically.

In the field of textual editing, the TEI schema [24] represents a standard model for the encoding of humanistic texts using an embedded markup. The *Guidelines* elaborated in the TEI project reflect on different aspects of the interpretative intervention of the editor in describing textual entities. A big effort is now devoted towards the translation of this XML Schema into an ontology in the domain of cultural heritage, mapping TEI onto CIDOC-CRM [25] a conceptual model for describing entities used in cultural heritage documentation.

In an Semantic Web context, we would rather use an ontology that describes the markup structures in OWL. For this we need a way to separate the document from its interpretation, i.e., a way to apply a meta-syntax for stand-off annotations of textual content with fully W3C-compliant technologies. For this reason, we use *EARMARK* [26], an ontology¹⁸ of a markup metalanguage, to describe the structure of the document as a set of OWL assertions to associate formal and explicit semantics [27]. Through the *Pattern Ontology (PO)*¹⁹ [28] in combination with EARMARK we can associate a particular structural semantics to markup elements, such an element *h3* expressing the concept of being a block of text (#9), or the *div* element containing it being a container with an header (#8), as shown in the following:

¹⁸ EARMARK: <http://www.essepuntato.it/2008/12/earmark>

¹⁹ PO: <http://www.essepuntato.it/2008/12/pattern>

```

:structure { :div a earmark:Element ;
  la:expresses pattern:HeadedContainer ; #8
  earmark:hasGeneralIdentifier "div" ;
  c:firstItem [ c:itemContent ... ; c:nextItem [
  c:itemContent :h-sec-2 ; ... c:nextItem [ ...
  c:itemContent :p4-sec-2 ... ] ] ] .
:h-sec-3-1 a earmark:Element #9
  la:expresses pattern:Block ;
  earmark:hasGeneralIdentifier "h3" ;
  c:firstItem [ c:itemContent :r-h-sec-3-1 ] .
# Text node within :h-sec-3
:r-h-sec-3-1 a earmark:PointerRange ...
:p1-sec-3-1 a earmark:Element # Sec 3.1, Par 1
  la:expresses pattern:Block ;
  earmark:hasGeneralIdentifier "p" ... }

```

Just a little above a purely structural perspective, we place the identification and organization of the *rhetorical components* of the text, such as a section being an *Introduction*, some paragraphs describing the *Methods* of the research, or the presented *Results* or the paper's *Conclusion*), in order to label explicitly all the meaningful aspects of the scientific discourse.

Such rhetoric characterization of markup structures can be specified through *DoCO*, the *Document Components Ontology*²⁰, and *DEO*, the *Discourse Elements Ontology*²¹. The following excerpt expresses that the elements *div*, *h3* and *p*, introduced in the previous excerpt, represent, respectively, a *section* of the paper (#10), a *section title* (#11), and a *paragraph* (#12) introducing some *background* assets (#13):

```

:rhetoric { :div la:expresses doco:Section . #10
:h-sec-3-1 la:expresses doco:SectionTitle . #11
:p1-sec-3-1 la:expresses doco:Paragraph , #12
  deo:Background } #13

```

Besides its structural and rhetorical characterisation, a document takes also part to a *citation network* with its cited documents, in particular taking into account the *reasons* for particular citations – e.g. to express qualification of or disagreement with the ideas presented in the cited paper – which may significantly effect the evaluation of a citation network itself.

For instance, analysing the content of the paper, for instance the aforementioned 1st paragraph of the 3rd section (1st subsection) of the paper (i.e. *p1-sec-3-1*), we encounter several citations to other works that are introduced for a particular reason, e.g. to express qualification of or disagreement with the ideas presented in the cited papers. Using *CiTO*, the *Citation Typing Ontology*²² [23], we provide descriptions of the nature of the citations, as shown in the following example, where paper “[5]” is cited as a source of background information (#14), and paper “[7]” is also cited as evidence supporting a statement (#15):

²⁰ DoCO: <http://purl.org/spar/doco>

²¹ DEO: <http://purl.org/spar/deo>

²² CiTO: <http://purl.org/spar/cito>


```

:citation { :delosdlms
  # citation to [5] in Sec 3.1, Par 1
  cito:obtainsBackgroundFrom #14
    <http://doi.ieeecomputersociety.org/10.1109/ICME
      .2005.1521528> ;
  # citation to [7] in Sec 3.1, Par 1
  cito:citesAsEvidence #15
    <http://dx.doi.org/10.1109/ICCV.1998.710779> }

```

In addition, strictly correlated with the citational aspects of a document, we can detail the organization of the claims and the arguments of the paper (providing evidences to a claim). The argumentative organisation of discourse is described using *AMO*, the *Argument Model Ontology*²³, that implements Toulmin's model of argumentation [11]²⁴ in OWL, as shown in Fig. 1 and introduced in the following excerpt:

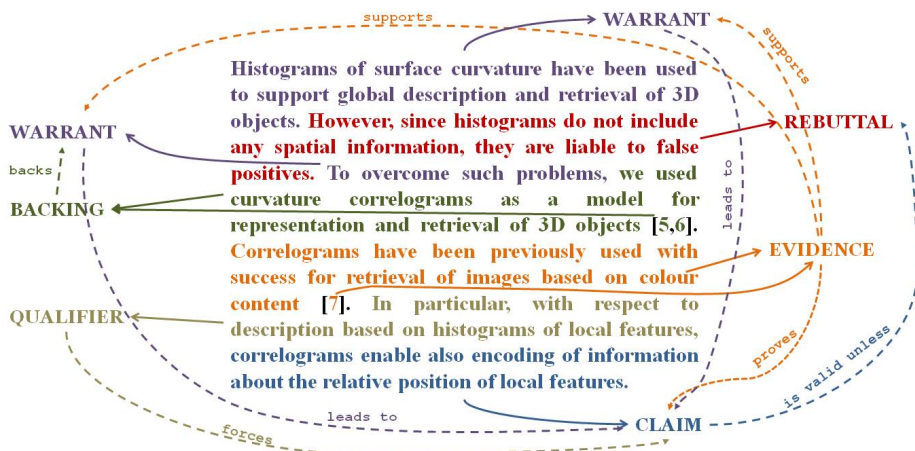


Fig. 1. Argument organisation of the 1st paragraph of Sec. 3.1 in the DelosDMS paper

```

:argumentation { :argument a amo:Argument ;
  amo:hasClaim :r-cl-p1 ; # correlograms...about
  amo:hasQualifier :r-qual-p1 ; # In...features
  amo:hasRebuttal :r-reb-p1 ; # However...false positive
  amo:hasEvidence :r-ev-1-p1 ... # Correlograms...content

```

²³ AMO: <http://www.essepuntato.it/2011/02/argumentmodel>

²⁴ Toulmin proposed that arguments are composed of statements having specific argumentative roles: the *claim* (a fact that must be asserted), the *evidence* (a foundation for the claim), the *warrant* (a statement bridging from the evidence to the claim), the *backing* (credentials that certifies the warrant), the *qualifier* (words or phrases expressing the degree of certainty of the claim) and the *rebuttal* (restrictions that may be applied to the claim).

```

amo:hasWarrant :r-war-2-p1 ... # To...problems
amo:hasBacking :r-back-1-p1 ... # we used...3D objects
:r-qual-p1 amo:forces :r-cl-p1 .
:r-cl-p1 amo:isValidUnless :r-reb-p1
:r-ev-1-p1 amo:proves :r-cl-p1 ; amo:supports :r-war-1-p1 ...
:r-war-2-p1 amo:leadsTo :r-cl-p1 .
:r-back-1-p1 amo:backs :r-war-2-p1 . ...

```

Finally, the *textual semantics*, i.e. the very message contained in a piece of text, is the final step in the definition of the semantics of a piece of text. For instance, the formal description of a claim needs to be expressed in such a way as to represent as faithfully as possible the meaning of the claim itself. Since each document expresses content in domains that are specific of the topic of the paper, we do not seek to provide an encompassing ontology to express claims. In some cases, the claim of an argument can be encoded through using a simple model, e.g. DBpedia [29], as shown in the following excerpt. In other cases, an appropriate specific ontology for the domain might be chosen freely.

```

:semantics {dbpedia:Correlogram a dbpedia:Mathematical_model}

```

4 Conclusions

The evolution of modern digital collections implies that the metadata we associate to their content are enhanced and enriched with more and more information. Simple metadata model may increase the likelihood of their adoption, but eventually result in simple annotations and possibly in errors and misrepresentations of the associated documents. Modern publishers are now approaching digital publishing from a semantic perspective (aka *semantic publishing* [4]).

In this paper we verified our *semantic lenses* [5] to semantically enhance a published scholarly article with direct, explicit, and hopefully correct annotations about the context, structure and argumentation of the paper as well as its actual content. Since one of the criteria for evaluating digital libraries as complex systems is the performance, which “depends strongly on the formats, structure and representations of the content” [30], we strongly believe that the use of semantic lenses as ontological keys could markedly improve usefulness of a library of scholarly articles. We are now working on *Through A Lens (TAL)*, a prototypical application²⁵ we developed as proof of concept of the use of semantic lenses in a real-case scenario, that enables the navigation and understanding of a scholarly document through these semantic lenses. We are now analysing the outcomes of a user testing session we undertook to demonstrate the efficacy of TAL when addressing tasks requiring deeper understanding and fact-finding on a document. Finally, along the lines of our previous work [28], we plan to develop automatic and semi-automatic approaches – based on ML and NLP techniques – for the enrichment of documents according to semantic lenses.

²⁵ Available at <http://www.essepuntato.it/2013/tal/LensedMika.html>

References

1. Peroni, S., Tomasi, F., Vitali, F.: Reflecting on the Europeana Data Model. In: Proceedings of the 8th Italian Research Conference on Digital Libraries. Revised Selected Papers, pp. 228–240 (2012)
2. Whalley, W.B., MacNeil, J., Munroe, G., Landy, S., Power, S.: Developing a flexible structure for a pure e-journal. In: Rowland, F., Meadows, J. (eds.) Proceedings of the 1st ICC/IFIP Conference on Electronic Publishing: New Models and Opportunities (1997), <http://elpub.scix.net/data/works/att/97112.content.pdf>
3. Dublin Core Metadata Initiative. DCMI Metadata Terms. DCMI Recommendation (June 14, 2012), <http://dublincore.org/documents/dcmi-terms/>
4. Shotton, D., Portwin, K., Klyne, G., Miles, A.: Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Computational Biology* 5(4), e1000361 (2009), doi:10.1371/journal.pcbi.1000361
5. Peroni, S., Shotton, D., Vitali, F.: Faceted documents: describing document characteristics using semantic lenses. In: Proceedings of the ACM Symposium on Document Engineering, pp. 191–194. ACM, New York (2012), doi:10.1145/2361354.2361396
6. Agosti, M., Berretti, S., Brettlecker, G., Del Bimbo, A., Ferro, N., Fuhr, N., Keim, D.A., Klas, C., Lidy, T., Milano, D., Norrie, M.C., Ranaldi, P., Rauber, A., Schek, H., Schreck, T., Schuldt, H., Signer, B., Springmann, M.: DelosDLMS - The Integrated DELOS Digital Library Management System. In: Proceedings of the 1st International DELOS Conference, pp. 36–45 (2007), doi:10.1007/978-3-540-77088-6_4
7. Pettifer, S., McDermott, P., Marsh, J., Thorne, D., Villeger, A., Attwood, T.K.: Ceci nest pas un hamburger: modelling and representing the scholarly article. *Learned Publishing* 24, 207–220 (2011), doi:10.1087/20110309
8. Darcus, B., Giasson, F.: Bibliographic Ontology Specification. Specification Document (November 4, 2009), <http://bibliontology.com/specification>
9. IFLA Study Group on the Functional Requirements for Bibliographic Records: Functional Requirements for Bibliographic Records (FRBR). Final Report (1998), http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm
10. Groza, T., Möller, K., Handschuh, S., Trif, D., Decker, S.: SALT: Weaving the claim web. In: Aberer, K., et al. (eds.) ISWC/ASWC 2007. LNCS, vol. 4825, pp. 197–210. Springer, Heidelberg (2007)
11. Toulmin, S.: The uses of argument. Cambridge University Press, Cambridge (1959)
12. Lauritsen, M., Gordon, T.F.: Toward a general theory of document modeling. In: Proceedings of the 12th International Conference on Artificial Intelligence and Law, pp. 202–211 (2009), doi:10.1145/1568234.1568257
13. Schneider, J., Groza, T., Passant, A.: A review of argumentation for the Social Semantic Web. Semantic Web -Interoperability, Usability, Applicability (2012), Pre-press available at: <http://iospress.metapress.com/content/016x47v66347462v/fulltext.pdf> (last visited July 25, 2012) (in press)
14. Koutsomitropoulos, D.A., Solomou, G.D., Alexopoulos, A.D., Papatheodorou, T.S.: Semantic Web enabled digital repositories. *International Journal on Digital Libraries* 10, 179–199 (2009), doi:10.1007/s00799-010-0059-z
15. Ranganathan, S.R.: Theory of Library Catalogue. Madras Library Association, Chennai (1938)

16. Quintarelli, E.: Folksonomies: power to the people. Presented at the ISKO Italy-UniMIB meeting (2005), <http://www.iskoi.org/doc/folksonomies.htm>
17. Lebo, T., Sahoo, S., McGuinness, D.: PROV-O: The PROV Ontology. W3C Working Draft (03 May 2012). World Wide Web Consortium (2012), <http://www.w3.org/TR/prov-o>
18. Lee, C.A.: A framework for contextual information in digital collections. *Journal of Documentation* 67(1), 95–143 (2011), doi:10.1108/00220411111105470
19. Mazzini, S., Ricci, F.: EAC-CPF Ontology and Linked Archival Data. In: *Semantic Digital Archives*. In: *Proceedings of the 1st International Workshop on Semantic Digital Archives* (2011), <http://ceur-ws.org/Vol-801/>
20. International Council on Archives: ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, 2nd edn. (2003), [http://www.icacds.org.uk/eng/ISAAR\(CPF\)2ed.pdf](http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf)
21. Pitti, D.: Creator Description: Encoded Archival Context. In: Taylor, A.G., Tillett, B.B., Baca, M., Guerrini, M. (eds.) *Authority Control in Organizing and Accessing Information: Definition and International Experience*, pp. 201–226. Haworth Information Press, Binghamton (2004)
22. Peroni, S., Shotton, D., Vitali, F.: Scholarly publishing and the Linked Data: describing roles, statuses, temporal and contextual extents. In: *Proceedings of the 8th International Conference on Semantic Systems*, pp. 9–16. ACM, New York (2012), doi:10.1145/2362499.2362502
23. Peroni, S., Shotton, D.: FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* (2012), doi:10.1016/j.websem.2012.08.001
24. Text Encoding Initiative Consortium: TEI P5: Guidelines for Electronic Text Encoding and Interchange (2005), <http://www.tei-c.org/Guidelines/P5>
25. Ore, C.E., Eide, O.: TEI and cultural heritage ontologies: Exchange of information? *Literary and Linguistic Computing* 24(2), 161–172 (2009), doi:10.1093/ll-c/fqp010
26. Di Iorio, A., Peroni, S., Vitali, F.: A Semantic Web Approach To Everyday Overlapping Markup. *Journal of the American Society for Information Science and Technology* 62(9), 1696–1716 (2011), doi:10.1002/asi.21591
27. Peroni, S., Gangemi, A., Vitali, F.: Dealing with Markup Semantics. In: *Proceedings the 7th International Conference on Semantic Systems*, pp. 111–118. ACM, New York (2011), doi:10.1145/2063518.2063533
28. Di Iorio, A., Peroni, S., Poggi, F., Vitali, F.: A first approach to the automatic recognition of structural patterns in XML documents. In: *Proceedings of the 2012 ACM Symposium on Document Engineering*, pp. 85–94. ACM, New York (2012), doi:10.1145/2361354.2361374
29. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia – A Crystallization Point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* 7(3), 154–165 (2009), doi:10.1016/j.websem.2009.07.002
30. Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C., Kovacs, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, C., Solvberg, I.: Evaluation of digital libraries. *International Journal on Digital Libraries* 8(1), 21–28 (2007), doi:10.1007/s00799-007-0011-z