Electronic Publishing '01 –
2001 in the Digital Publishing Odyssey
A. Hübler et al. (Eds.)
IOS Press, 2001

Metadata Usage in an Online Journal – An application Profile

Ana Alice Baptista
Altamiro Barbosa Machado[1]
analice@dsi.uminho.pt
Departamento de Sistemas de Informação
Escola de Engenharia
Universidade do Minho
Portugal

Abstract

In a chaotic environment like the Internet, data are not enough anymore. The description of resources is fundamental in order to keep some structure and make Internet services more efficient and more effective Metadata is, basically, data about data. However, metadata per se is also insufficient: with different kinds of services and software using different metadata and metadata structures, the problem persists. As it happens in other areas, standardization is a keystone to the metadata usage and implementation.

Dublin Core (DC) and RDF are two recommendations from two different initiatives: DCMI (Dublin Core Metadata Initiative) and W3C (World Wide Web Consortium).

In order to be widely used, the DCMI opted for broadly defining the DC semantics, while leaving the syntax issues open and undefined. This is the reason why RDF and DC match so well: RDF brings the syntax rules on which DC can be embedded.

The RDF schema, on its turn, makes it possible to design and implement, in a consistent way, project specific metadata vocabularies not covered by DC or other standard metadata vocabularies.

In this paper we will illustrate the use of DC, RDF and RDF Schema in the context of an online journal project: Informattica Online. An evaluation of this approach will also be presented.

0. Introduction

In a chaotic environment like the Internet, data are not enough anymore. Although indexing and information retrieval systems are more and more sophisticated, relevance and precision in user queries have not yet reached satisfactory levels.

In order to enhance information services' efficiency and effectiveness, semantic metadata needs to be used. However, with different kinds of services and software using different metadata and metadata structures, the problem persists. As in other areas, standardization is a keystone to the effective usage of metadata.

The Dublin Core Element Set (DCES) is a recommendation from Dublin Core Metadata Initiative (DCMI) and it consists of a core vocabulary of fifteen elements drawn specifically for the description of resources. Almost all of these elements have element refinement qualifiers, and/or encoding scheme qualifiers already approved by the Dublin Core Usage Committee (DCUC).

Element refinement qualifiers are, as the name suggests, other terms of the vocabulary that further specify the meaning of a given element. "A refined element shares the meaning of the unqualified element, but with a more restricted scope" (DCMI-DCQ, 2000). For instance, the element Date has as element refinement qualifiers the terms Issued, Created, Modified and Valid.

Element Encoding Scheme qualifiers, by their turn, "identify schemes that aid in the interpretation of an element value" (DCMI-DCQ, 2000). For instance, the same element Date, has as Encoding Scheme Qualifiers the W3C-DTF[1] and the DCMI[2] Period schemes.

Because DCMI is concerned with semantic issues, the DC needs a foreign syntax to be expressed. When DC is embedded in the Extensible Markup Language (XML) documents, the syntax chosen is, many times, the RDF/XML syntax. This is an XML application to express the Resource Description Framework (RDF) model (Lassila and Swick, 1999). The RDF Model is nothing but a set of triples (subject, predicate, object or resource, property, value). An RDF statement makes statements about resources using a property and having as a value the result of that property applied to the given resource. A statement can be seen as a triple composed of three elements: resource (subject), property (predicate) and value (object). A resource can be anything identifiable by a Uniform Resource Identifier (URI) (Harold 1999) pp.632.

The RDF Schema, in its turn, makes it possible to design and implement, in a consistent way, project specific metadata vocabularies not covered by DC or other standard metadata vocabularies. The RDF Schema Specification is a candidate recommendation since 27th March 2000 that "defines a schema specification language. More succinctly, the RDF Schema mechanism provides a basic type system for use in RDF models. It defines resources and properties such as rdfs:Class and rdfs:subClassOf that are used in specifying application-specific schemas" (Brickley and Guha, 2000).

In Informattica Online (IO) we used three vocabularies embedded in RDF syntax in order to fully describe our documents: DC[3], vCard[4] and IO[5]. To describe the vocabularies usage we created an RDF Profile (Heery and Patel, 2000) that will be presented in section 3.

In next section, we describe the contextualization in which we use the vocabularies. Then, we describe the general profile of Informattica Online and, afterwards, we draw some conclusion on the usage of these technologies, mainly RDF and RDF schema. Some glances over future work will follow.

1. Contextualization

Informattica Online (Baptista, Rodrigues et al., 1999) comprises the following document genres (Levy and Marshall, 1994): Scientific Article, Brief News, Comment, Anonymous Comment, Editorial, Revision, Revision Comment. Roughly, they are related as shown in Fig.1.

---

[1] http://www.w3.org/TR/NOTE-datetime
[2] http://dublincore.org/documents/dcmi-period/
[3] http://dublincore.org/2000/03/13/dces
[4] http://www.w3.org/2001/vcard-rdf/3.0 and ftp://ftp.isi.edu/in-notes/rfc2426.txt
[5] IO is a vocabulary namespace schema created by us using RDF schema. It can be found at http://www.dsi.uminho.pt/io/schemas/io-schema

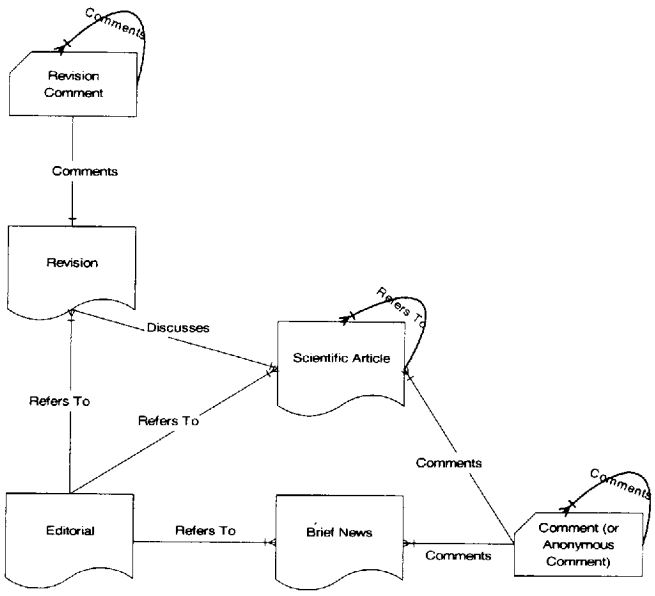

Fig. 1 – Relation between Informattica Online documents

For compatibility reasons, we decided to use widely used vocabularies as much as possible. Nevertheless, there were some specific semantics we needed that were not covered by them, leading us to create our own vocabulary and namespace schema using RDF schema. More specifically, we created the following terms:

- **Document Genre** - The Document Genre. This can be one of the following: Scientific Article, BriefNews, Comment, Anonymous Comment, Editorial, Revision and Revision Comment.
- **Document State** - The Document State within the system. This can be one of the following: underReview, underModification, forPublishing, forArchival, Archived, Published.
- **Approval** - Property to be used in the Revision Discussions Document. It can have four different values: Approved, Not Approved, Under Review, Under Modification
- **Version Number** - The version number of the document.
- **Document Code** - A code given to a document, independently of version, i.e., the document is considered to be always the same, whichever version it has.

2. IO General Profile

"Application profiles consist of data elements drawn from one or more namespace schemas combined together by implementors and optimised for a particular local application." (Heery and Patel, 2000). The Application Profile shows, in RDF, all terms and vocabularies and their specific meaning used in the scope of a specific application.

Application profiles are not the same as namespace schema. In fact, a namespace schema contains "all those elements defined by the managing body or registration authority (whatever that might be) for a particular namespace" (Heery and Patel, 2000) while an application profile is tailored for a particular implementation and

"will typically contain combinations of sub-sets of one or more namespace schemas" (Heery and Patel, 2000).

In order to build the general application profile of Informattica Online (see Fig. 2), we used the Schemas vocabulary (smes namespace[6]) being developed under the Schemas Project[7].

Each document genre needs a specific set of semantic metadata and therefore one profile, accordingly to its usage within the system. By making a union of these sets, we obtained the general set of vocabulary terms used in the overall system. Table 1 shows the terms and respective vocabularies used.

Table 1 – Vocabulary and respective terms used in Informattica Online

Vocabulary	Element	Element Refinement	Encoding Scheme
DC	Title	Alternative	
DC	Creator		
DC	Subject		
DC	Description	Table Of Contents Abstract	
DC	Publisher		
DC	Contributor		
DC	Date	Created Valid Issued	W3C-DTF
DC	Type		DCMI Type Vocabulary
DC	Format	Extent Medium	 IMT
DC	Identifier		URI
DC	Source		URI
DC	Language		RFC 1766
DC	Relation	Is Version Of Has Version Is Part Of Has Part Is Required By Requires Is Referenced By References Has Format	URI
DC	Rights		
vCard	FN		
VCard	EMAIL		
VCard	URL		
VCard	ORG		
IO	Document Genre		
IO	Document State		
IO	Approval		
IO	Version Number		
IO	Document Code		

[6] http://www.schemas-forum.org/registry/schemas/SCHEMAS/1.0/smes
[7] http://www.schemas-forum.org/

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://dublincore.org/2000/03/13/dces#" xmlns:dcq="http://dublincore.org/2000/03/13/dcq#"
xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#" xmlns:smes="http://www.schemas-
forum.org/registry/schemas/SCHEMAS/1.0/smes#">
    <smes:ApSchema rdf:about="http://www.dsi.uminho.pt/io/schemas/io-genProfile">
        <smes:isProfileOf rdf:resource="http://www.dsi.uminho.pt/io/semantics/general.html"/>
        <dc:title>The Informattica Online General Application Profile v1.0</dc:title>
        <dc:date>2001-04-12</dc:date>
        <dc:identifier>http://www.dsi.uminho.pt/io/schemas/io-genProfile</dc:identifier>
        <dc:creator rdf:parseType="Resource">
                <vCard:fn>Ana Alice Baptista</vCard:fn>
                <vCard:email>analice@dsi.uminho.pt</vCard:email>
                <vCard:URL rdf:resource=" http://www.dsi.uminho.pt/~analice"/>
                <rdfs:label>Ana Alice Baptista</rdfs:label>
        </dc:creator>
        <dc:publisher>Universidade do Minho</dc:publisher>
        <dc:type>Text</dc:type>
        <dc:language>en</dc:language>
        <dc:source rdf:resource="http://www.dsi.uminho.pt/io/semantics/general.htm"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#title"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#creator"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#subject"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#description"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#publisher"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#contributor"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#date"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#type"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#format"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#identifier"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#source"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#language"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#relation"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dces#rights"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#alternative"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#tableOfContents"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#abstract"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#created"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#valid"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#issued"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#extent"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#medium"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#isVersionOf"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#hasVersion"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#isPartOf"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#hasPart"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#isRequiredBy"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#requires"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#isReferencedBy"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#References"/>
        <smes:uses rdf:resource="http://dublincore.org/2000/03/13/dcq#hasFormat"/>
        <!--vCard-->
        <smes:uses rdf:resource="http://www.w3.org/2001/vcard-rdf/3.0#FN"/>
        <smes:uses rdf:resource="http://www.w3.org/2001/vcard-rdf/3.0#EMAIL"/>
        <smes:uses rdf:resource="http://www.w3.org/2001/vcard-rdf/3.0#URL"/>
        <smes:uses rdf:resource="http://www.w3.org/2001/vcard-rdf/3.0#ORG"/>
        <!-- io -->
        <smes:uses rdf:resource="http://www.dsi.uminho.pt/io/schemas/io-schema#documentGenre"/>
        <smes:uses rdf:resource="http://www.dsi.uminho.pt/io/schemas/io-schema#documentState"/>
        <smes:uses rdf:resource="http://www.dsi.uminho.pt/io/schemas/io-schema#approval"/>
        <smes:uses rdf:resource="http://www.dsi.uminho.pt/io/schemas/io-schema#versionNumber"/>
        <smes:uses rdf:resource="http://www.dsi.uminho.pt/io/schemas/io-schema#documentCode"/>
    </smes:ApSchema>
</rdf:RDF>
```

Fig. 2 - General Application Profile of Informattica Online

3. Conclusions

Although we tried to avoid creating a new vocabulary, we still needed some specific semantics not present in other vocabularies. Therefore, we created IO and defined a specific namespace using RDF Schema, which has shown to be very useful.

By their turn, DC, vCard and RDF syntax fit well and, although the characteristic verbosity of XML, the description of resources seems to be not only machine-understandable, but also human-understandable. The RDF profile of the application was also easy to develop.

Nevertheless, we have to say that the RDFMS Specification (Lassila and Swick, 1999) and the RDFS Specification (Brickley and Guha, 2000) are not easy to understand at first. However, this happens with almost all computer languages: after some experience their usage is straightforward.

Regarding RDF Namespace Schemas and Application Profiles, we expect that their usage and registry can bring some structure to the area.

4. Future Work

This project is an output of a PhD work in the field of Electronic Publishing. After having finished building the Application Profiles of all document genres, we intend to close the metadata definition part of our project.

Then some thoughts will have to take place to decide if this work should go on and implemented in a larger project with national and international partners.

5. References

Lassila O. and Swick R. R. (Eds). (1999). Resource Description Framework (RDF) Model and Syntax Specification. W3C Consortium.

Brickley D. and Guha R. V. (Eds). (2000). Resource Description Framework (RDF) Schema Specification. W3C Consortium.

DCMI-DCQ. (2000). Dublin Core Qualifiers, DCMI.

Baptista, A. A., E. Rodrigues, et al. (1999). Online Publishing as a Support for Scholarly Communication in Dynamic Knowledge Communities. ICCC / IFIP Conference on Electronic Publishing'99, Ronneby - Sweden, ICCC Press.

Harold, E. R. (1999). XML Bible. Foster City, IDG Books Worldwide, Inc.

Heery, R. and M. Patel (2000). "Application Profiles: Mixing and Matching Metadata Schemas." Ariadne(25).

Levy, D. M. and C. C. Marshall (1994). Washington's White Horse? A Look at Assumptions Underlying Digital Libraries. The First Annual Conference on the Theory and Practice of Digital Libraries, Texas - USA.

[1] Professor Altamiro Machado, my PhD supervisor and Head of the Information Systems Department at Universidade do Minho passed away on the 23rd March 2001 due to a heart attack. Because he was my supervisor and was aware of all my work, I decided to keep his name as an author in all documents we had in hand at the time he passed away.