

Use of Dublin Core metadata for describing and retrieving digital journals

Assumpció Estivill	Ernest Abadal	Jorge Franganillo	Jesús Gascón	J. M. Rodríguez Gairín
Universitat de Barcelona	Universitat de Barcelona	Universitat de Barcelona	Universitat de Barcelona	Universitat de Barcelona
Tel: +34 93 4035773	Tel: +34 93 4035887	Tel: +34 93 4035784	Tel: +34 93 4035788	Tel: +34 93 4035777
Fax: +34 93 1035772	Fax: +34 93 4035772	Fax: +34 93 4035772	Fax: +34 93 4035772	Fax: +34 93 4035772
estivill@ub.edu	abadal@ub.edu	Franganillo@ub.edu	gascon@ub.edu	rodriguez.gairin@ub.edu

Abstract:

This article summarises the experience of applying Dublin Core metadata to digital journal articles. Some of the problems encountered during this process are discussed.

Keywords:

Metadata, Dublin Core, Journal articles.

1. Introduction

The journal *BiD: textos universitaris de Biblioteconomia i Documentació*¹ was probably among the first Spanish journals in digital format to include Dublin Core (DC) metadata in the headers of the journal, of each number, and of each article. It was most certainly the first digital Spanish journal in the field of library and information science to follow this practice comprehensively.² Little by little, the use of Dublin Core has increased on the part of some Spanish journals. For example, *Cuadernos de documentación multimedia* includes DC metadata in the articles of its latest number, and work is progressing, at the Temaria portal,³ for introducing metadata into articles from Spanish digital journals specialising in library and information science. Likewise, the project, e-revist@s, from the Tecnociencia portal of the Fundación Española para la Ciencia y la Tecnología (FECYT) that gathers together scientific e-journals from Spain and Latin America, states that one of the recommended selection criteria —although not a *sine qua non* requisite— is the inclusion of DC metatags.⁴ Nevertheless, the use of these metadata is still not a standard practice for the majority of titles

published on the Web.

There are certain publications —such as *D-Lib magazine* or *Ariadne*— with long digital traditions, oriented towards topics dealing with information technologies and organisation of digital information and that have served as platforms for promoting the DC standard, that have not adopted this practice themselves.⁵ The present work analyses the experience of the use of DC metadata in *BiD* and in the Temaria portal of Spanish digital journals in library and information science, and reflects upon the advantages and disadvantages of this practice.

2. Assigning DC metadata to *BiD*, and to the Temaria portal

Beginning with number 4 in June 2000, *BiD* initially started including DC metadata in the header of each element of the journal. The principal objective was to experiment with the use of these metatags. Above all, the purpose was to make a commitment to standards, rather than merely being a matter of efficient retrieval of the journal's articles: after all, the majority of search engines did not recognise DC tags⁶ and neither did the journal's own internal search software (AtomZ)⁷. As such, in order to make article retrieval more efficient, specific metatags also had to be introduced for search software. Subsequently a database arranged according to DC elements was set up. As a result of this application, metadata was no longer entered into the headers of the journal's units. Later this practice was reinstated and now the metadata are integrated in the database as well as into the header of each of *BiD*'s 255 articles published to date.

Starting with this experience, in 2004 work began on the Temaria portal, whose objective was to facilitate searching for articles in Spanish library and information science journals through the use of DC elements. Currently this portal provides comprehensive indexing of the following journals: *Anales de documentación*, *BiD: textos universitarios de biblioteconomía i documentació*, *Cuadernos de documentación audiovisual*, *Cuadernos de documentación multimedia*, *Hipertext.net* and *Revista general de información y documentación*. Other titles are expected to be added in the near future. Currently 700 articles are searchable.

Temaria contains a guide, *Guía de procedimientos para la aplicación de metadatos*, explaining how to apply the DC standard and offering detailed instructions for supplying contents to the elements. The aim of this guide is to assure that metadata is assigned properly and uniformly.⁸

Thirteen of the 15 DC elements have been designated as obligatory and repeatable and they are applied to the description of the article as shown in Table 1. In the Temaria project the use of the "Coverage" element is optional and the "Source" element has been removed.

Element	Content
title	Title of the article.
creator	Author or authors of the article.
subject	Subject of the article.
description	Description of the article.
publisher	Publisher of the journal.
contributor	Contributors (editors and translators).
date	Date of publication of the journal issue that contains the article.
type	Type of resource (this is always "text").
format	Format of the article (for example, html, pdf).
identifier	Identifier of the article (generally its url, but it could also be another identifier, such as a DOI or a SICI). Bibliographic reference: title of the journal and numeric and chronological designation of the issue containing the article.
language	Language of the article.

Element	Content
relation	Relation of the journal article with other resources (translations, versions, etc.). Relation of the article with the journal.
rights	Information on copyright, intellectual property and judicial protection of the article.

Table 1. Definition of the content of DC elements in the Temaria project

Temaria has chosen to qualify the DC elements. Whenever possible, the qualifiers defined in *DCMI metadata terms*⁹ are employed. But local qualifiers that are more attuned to the portal's environment are also used, such as "personalName", "corporateName" and "address" applied to "Creator", "Publisher" and "Contributor". Table 2 summarises the qualifiers used in this project.

DC Elements	Element refinements	Coding schemes
creator publisher contributor	personalName* corporateName* address*	—
subject	—	TBD ¹⁰ CDU2000
description	abstract tableOfContents	—
date	Issued	W3C-DTF
type	—	DCMI Type Vocabulary
identifier	—	URI (URL, DOI, SICI...)
	bibliographicCitation	—
language		ISO 639-2
relation	isVersionOf hasVersion isPartOf isFormatOf hasFormat	URI (URL, DOI, SICI, ISBN, ISSN, etc.)
coverage	Spatial	ISO 3166 MARC 21 (043)
	temporal	W3C-DTF

Table 2. Summary of the Temaria project's qualifiers

* Local refinements

Of the DC's fifteen elements, the difference

between “Creator” and “Contributor” is unlikely to be operative for the majority of projects. If the intention was to move the standard closer to library practice, it has not succeeded since there is no equivalent between these elements and the MARC tags 1XX and 7XX. As for retrieval, it does not make much sense to distinguish between the two, especially when the very definition of “Contributor” is not very precise and can lead to ambiguity and confusion.

Throughout the development of the DC metadata set, the “Source” element has been interpreted in quite distinct ways. The different versions of the current guide *Using Dublin Core*, have not been capable of specifying and exemplifying the use of this element that, most surely, should form part of the “Relation” element.

Finally, the “Coverage” element does not make much sense either, within a set intended to be very general and for broad usage. In fact, this same concept is included in the “Subject” element, since it deals with the spatial and/or chronological content of the resource.

As to the qualifiers, the project has attempted to use only those that are essential in its context. For example, with the “Date” and “Relation” elements only those qualifiers that are useful for the project have been defined. The use of qualifiers reflects decisions that have been taken in relation to the content of the elements, such as the relatively recent adoption of the *Tesouro de biblioteconomía y documentación*. The usefulness of some of the qualifiers developed for the project has not been demonstrated —such as the refinements “personalName” and “corporateName” that complement the “Creator”, “Publisher”, and “Contributor” elements. On the other hand, being that Temaria is a portal that brings together journals from the academic sector, it is possible that a mechanism could be developed in the future for annotating the affiliation of the authors and linking it to the corresponding author.

One of the main problems in applying DC metadata to the description and retrieval of journal articles is the slow progress of the formula for including the bibliographic citation of the journal that contains it, especially considering how essential it is to apply DC to this type of resource. The Dublin Core Metadata Initiative’s Citation Working Group, in operation since 1998, has as its objective to develop a mechanism for including information about the journal that contains the

article (title of the journal, numeric and chronological designations and pagination). After, several proposals that eventually were rejected, the *Guidelines for encoding bibliographic citation information in Dublin Core metadata* has finally been approved as a DCMI recommendation in June 2005.¹¹

Assigning metadata to *BiD* and to the other journals on Temaria’s portal has reflected—as far as the bibliographic citation is concerned—solutions found in the professional literature as well as those provided by the Citation Working Group over the years of its activity. Various solutions have been tried for annotating this very necessary element for the description and retrieval of journal articles, including the following:

- Use of the “Source” element.
- Development of a local field that did not appear in the public metadata of the article.
- Use of the “bibliographicCitation” qualifier in the “Identifier” element developed according to the guidelines given in *DCMICite: a bibliographic citation Dublin Core structured value (DCSV) encoding scheme*¹² but ultimately rejected by the DCMI’s Usage Board.
- Use of the “bibliographicCitation” qualifier in the “Identifier” element, according to the latest proposal of the Citation Working Group.¹³ At present the option for annotating the bibliographic reference according to the Open URL standard is not being used, but instead it is being annotated textually.

The “Relation” element with the “IsPartOf” qualifier has always been used in the project to relate the article to the journal’s title, by means of its ISSN.

3. Conclusions

The DC metadata set had an initial moment of acceptance in developing some subject portals, most of which grew out of library-related projects. Nevertheless, ten years after its creation and after its approval as an ISO standard (ISO 15836:2003), it does not appear to enjoy a generalised acceptance. The DC elements were designated as the common metadata format by the Open Archives Initiative (OAI), but a recent study on the use of DC by 100 data providers, registered under the OAI, reveals that only 82 of them make metadata available for analysis. The study also

shows that only five of the fifteen DC elements —“Creator”, “Identifier”, “Title”, “Date” and “Type”— are present in 71% of the cases, and 44 of the 82 providers only use the “Creator” and “Identifier” elements in half of their products.¹⁴ In addition, the five previously mentioned elements are so obvious that their use cannot easily be attributed to the existence of the DC standard.

Some authors are very critical of this metadata set and with the slow development that it has suffered.¹⁵ Certainly, if one compares the acceptance of DC ten years after its creation and the level of development reached, with the acceptance and development of a format as complex as MARC in the same length of time, the conclusion is too obvious to need commentary.

As to the application of DC metadata to *BiD* and to the Temaria portal, the conclusion is that its use has provided a standardised framework for the project. It has also permitted the journal's metadata to be included automatically in the e-revist@s database of the Tecnociencia portal after having first been adapted to the Protocol for Metadata Harvesting of the Open Archives Initiative. The most negative aspect is that the entire process has progressed haltingly as a result of the lack of specifics for some elements and of the absence of clear guidelines for including bibliographic citations.

References

1. *BiD: textos universitaris de biblioteconomia i documentació*. <<http://www.ub.es/bid/>>.
2. Carme Cantos Puig. Les metadades a les revistes electròniques espanyoles de biblioteconomia i documentació. *BiD: textos universitaris de Biblioteconomia i Documentació*. desembre, núm. 11. <http://www2.ub.es/bid/consulta_articulos.php?fihero=11cantos.htm> [Available: 09-04-2005].
3. *Temaria: revistas digitales de biblioteconomia y documentació*. Facultat de Biblioteconomia i Documentació, Universitat de Barcelona, Grup de Recerca Organització i Recuperació de Continguts Digitals, 2005. <<http://bidoc.ub.es/temaria/>>. [Available: 10-04-2005].
4. *E-revist@s: selección y criterios de calidad de las revistas que se incorporen al proyecto*. FECYT: CSIC. <<http://www.tecnociencia.es/e-revistas/proyecto/proyecto4.jsp>>. [Available: 10-04-2005].
5. A search on “Dublin core” in the magazine *Ariadne* (<http://www.ariadne.ac.uk/>) retrieved 385 records; the same search performed in *D-lib magazine* (<http://www.dlib.org/dlib.html>), retrieved 316. [Available: 06-04-2005].
6. Robin Henshaw, Edward Valauskas, “Metadata as a catalyst: experiments with metadata and search engines in the Internet journal, First monday”. *Libri*, vol. 51, no. 2 (2001), p. 86–101.
7. Atomz. <www.Atomz.com>.
8. *Guia de procediments per a l'assignació de metadades Dublin Core en el marc del projecte Temària, revistes digitals de biblioteconomia i documentació*. <<http://bidoc.ub.es/temaria/librestil.php>>. [Available: 10-04-2005]
9. *DCMI metadata terms*. DCMI, 10-01-2005. <<http://dublincore.org/documents/dcmi-terms/>>. [Available: 10-04-2005].
10. *Tesauro de biblioteconomia y documentació* (Madrid: CINDOC, 2002).
11. *Guidelines for encoding bibliographic citation information in Dublin Core metadata*. DCMI, 29-03-05. <<http://epub.mimas.ac.uk/DC/dc-citation-guidelines/>>. [Available: 10-04-2005].
12. Ann Apps, *DCMICite: a bibliographic citation Dublin Core structured value (DCSV) encoding scheme*. 21 Aug. 2003. <<http://epub.mimas.ac.uk/DC/citdcsv.html>>. [Available: 10-04-2005].
13. *Guidelines for encoding bibliographic citation information in Dublin Core metadata*. DCMI, 29-03-05. <<http://epub.mimas.ac.uk/DC/dc-citation-guidelines/>>. [Available: 10-04-2005].
14. Jewel Ward, “Unqualified Dublin Core usage in OAI-PMD data providers”. *OCLC systems & services*, vol. 20, no. 1, (2004), p. 40-47.
15. See, for example, Jeffrey Beall (2004), “Dublin Core: an obituary”. *Library hi-tech news*, no.8, p. 40-41.