

Dublin Core: An Obituary

Jeffrey Beall

Introduction

The Dublin Core metadata standard has rapidly become obsolete. The standard, which never really caught on, has outlived its usefulness. It is obsolete because it will soon be popularly replaced by the Metadata Object Description Schema (MODS), a new standard developed by the Library of Congress, and because it was designed as a tool for resource discovery before Google appeared and essentially put it out of business. Moreover, the standard failed because it was created with a profit motive in mind on the part of OCLC and because it was created mainly by managers rather than practitioners. Dublin Core also tried to be all things to all metadata users and lacked sufficient specificity to function as an operative description schema.

History of Dublin Core

The Dublin Core Metadata Initiative (DCMI) was named after Dublin, Ohio, where a meeting was held in 1995 among the staffs of OCLC and the National Center for Supercomputing Applications (Dublin Core Metadata Initiative, 2004). At that time, Web searching was inadequate; Google had not yet appeared, and the existing Internet search engines lacked sophistication and did not include many Web pages in their indexing, making search results incomplete and unsuitably organized. OCLC saw the emerging metadata standard as a potential revenue source. The organization theorized that it might sell metadata records much like it sells MARC records and wanted to reap the profits of this new revenue source. So the name "Dublin Core," which might strike some as cute at first, was a clever attempt by OCLC to associate OCLC with the emerging standard. The perspective of many librarians is that OCLC focuses chiefly on

entrepreneurship rather than librarianship. This view holds that the organization prioritizes creating new sources of revenue over serving library patrons. DMCI, at the time, held a lot of potential as the "MARC format" for the Web.

The idea behind Dublin Core was to facilitate resource discovery. That is to say, it was a system designed to enable Internet searchers to find Web pages that more precisely matched their interests, and to do so more precisely than the search engines available at the time of its invention. The metadata were to exist encoded in HTML tags within the Web page which it itself was describing. Thus it would "ride-along" with the Web page. Also, Dublin Core metadata could exist separately from the Web pages they describe, in much the same way that cataloging data is maintained in catalogs separately from books. The goal was to encourage all producers of Web content to include Dublin Core metadata within the Web pages they created, thereby facilitating the indexing of the Web. Also, the metadata could be used in Web catalogs – especially local catalogs – that users could search to access the Web pages they were looking for.

Soon after the Dublin Core standard emerged, OCLC created a database that was to serve as a prototype of a "metadata union catalog" through which OCLC could eventually sell the metadata. This database, now defunct, was called the Cooperative Online Research Catalog, or CORC. OCLC played up the name CORC, just as a business promotes its own brand names, and even distributed cork coasters emblazoned with the CORC logo at library conferences. The main reasons CORC failed parallel the reasons why Dublin Core itself has failed, but more importantly, OCLC eliminated CORC because it never generated revenue for the organization, for the emergence of superior Web discovery tools and the

accelerating demise of Dublin Core guaranteed that it never would.

The designers and operators of Internet search engines have never liked producer-created metadata. It is unreliable chiefly because it is abused so much. Web page producers want their pages to appear first in search engine results pages, and they often do anything necessary to make this happen, including misrepresenting the information contained within their Web page. So the idea of creator-supplied metadata made available over the World Wide Web was really dead from the start. Libraries were producing objective, DC metadata records, but the number of these records was so small compared to the total number of Web pages that their utilization and impact was minimal.

The weaknesses of Dublin Core

One of the things that drove the creation of a new metadata standard, that is, one different from MARC, was the high level of training required for information specialists to create valid MARC records. Moreover, MARC was seen as library-specific. The inventors of Dublin Core wanted a system with a broad application among different constituencies, and with broad flexibility. They also wanted a simple system. For example, in MARC, there are many different fields and subfields that can contain author information. In MARC, author information can be in the following fields, to name a few: 100, 110, 111, 700, 710, 711, etc. In Dublin Core, a single author field was prescribed, the "creator" field.

Dublin Core's simplicity is also one of its major weaknesses. However, as Hahn (2004) states:

Its fundamental flaw is that it is designed as a lowest common denominator system. Thus, converting from anything else (especially MARC) into DC results in a loss of specificity (and thus data) and

converting from DC into anything else (especially MARC) results in woefully substandard data.

Dublin Core attempted to solve the problem of this lack of specificity by incorporating into its design extensibility, that is, the ability to locally define or extend a particular field. For example, the general author field could be extended to a corporate author. But as Hahn (2004) explains:

DC proponents' recommended local fix for this – creating your own local fields and qualifiers that can be as specific as you want – defeats the purpose of using DC as a common language for data exchange, as your local customizations will likely be sufficiently different from everyone else's ...

Thus, one of the major flaws of DCMI is the near impossibility to “crosswalk” or convert data from Dublin Core into other schemes. Dublin Core scores very low on interoperability.

In her book *Metadata Fundamentals for All Librarians*, Priscilla Caplan (2003, p. 78-9) includes a chapter on Dublin Core, but after reading her description one gets the sense that she finds it an unworkable scheme. Describing the problems created by the simplicity designed into the schema, she writes:

Despite the simplicity of the Dublin Core scheme, certain problems have arisen repeatedly in applications. One issue concerns the overlap in meaning in the definition of some elements. Creator can be seen as a particular type of contributor, and source is a particular type of relation. This had led to confusion among implementers about when it is appropriate to use one element rather than another.

Also, in his article, “After MARC, what then?”, writer Andresen (2004) describes an implementation of the Dublin Core Metadata Standard in Denmark. Describing the results of the selection of DC as a metadata standard among Danish libraries, archives, and museums, he states:

It is not feasible to use Dublin Core internally in the sectors, as it is far too general and unable to cope with specific needs (Andresen, 2004, p. 47).

He continues:

The choice of Dublin Core is problematic because this rather simple format is basically focusing on Web resources (Andresen, 2004, p. 47).

Another problem with the simplicity approach promoted by DCMI is the lack of standardization of the data in the Dublin Core metadata. While in MARC the goal is to use standard headings, such as authorized forms of names, Library of Congress Subject Headings, and standard country and language codes, to name a few examples, no standard codes or data are required in Dublin Core. This lack of control creates inconsistency, and inconsistency is one of the greatest hurdles for resource discovery.

Simplicity was incorporated into the design of Dublin Core to make it attractive to a wide potential constituency of users. But this large constituency has failed to materialize, and as the weaknesses of Dublin Core become more known and understood, the standard is used less and less. The creation of MODS, in fact, is in part a reaction to the lack of specificity designed into Dublin Core. According to the Library of Congress (2004):

MODS is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format.

Problems in organization

Another major problem with Dublin Core is the initiative itself. DCMI is a bloated, European-style bureaucracy. The members of the initiative often bicker among themselves and often fail to achieve a consensus. According to Caplan (2003, p. 86):

The DCMI has been criticized for taking too long to produce basic guidance for Dublin Core implementers. There are still no approved recommendations for syntactical representation of qualified Dublin Core in XML and RDF. Guidelines for representing citations to journal articles in Dublin Core have been under development since 1998 and are still unfinished. Element refinement qualifiers for the creator, contributor, and publisher elements were omitted from the

Dublin Core Qualifiers because of lack of consensus within the usage board, and currently are still pending, despite a great need for these among implementers.

Perhaps one reason Dublin Core has failed can be traced to the makeup of its design group. The organization tried to be too inclusive and too international in its design and ended up pleasing very few. The designers were managers rather than practitioners and held organizational meetings in far-flung parts of the world, such as Finland and Australia, ensuring that day-to-day professionals who actually worked with metadata would be excluded.

Conclusion

The one good thing to come out of the failure and demise of the DCMI is MODS. MODS promises standard and specific data elements and wide interoperability. The lessons we have learned from Dublin Core's failure are these: metadata elements have to be well defined, with more universal agreement on the meaning of the elements. The metadata structure and content must have the ability to be easily crosswalked to other metadata schemes. The input of day-to-day metadata practitioners needs to be heeded when designing any metadata scheme. Significantly, the less-than-honorable intentions of OCLC to use the standard as a revenue generating source contributed to its failure.

REFERENCES

- Andresen, L. (2004), “After MARC – what then?”, *Library Hi Tech*, Vol. 22 No. 1, pp. 40-51.
- Caplan, P. (2003), *Metadata Fundamentals for All Librarians*, American Library Association, Chicago, IL.
- Dublin Core Metadata Initiative (2004), “History of the Dublin Core Metadata Initiative”, available at: <http://dublincore.org/about/history>
- Hahn, J. (2004), Personal communication.
- Library of Congress (2004), “Metadata object description schema”, available at: www.loc.gov/standards/mods

Jeffrey Beall (jeffrey.beall@cudenver.edu) is Catalog Librarian, Auraria Library, University of Colorado, Denver, Colorado, USA.