

The Dublin Core Metadata Initiative: None of us is as smart as all of us

Stuart Weibel, Ph.D.
Senior Research Scientist
OCLC Office of Research
weibel@oclc.org

The Dublin Core Metadata Initiative [DC-HOME] is nearing the fifth anniversary of what has become the broadest international, interdisciplinary effort in resource description on the Internet. It is the leading initiative for improving resource discovery on the Web. How this has come to pass is a combination of good timing and asking questions that were on the minds of many people in many places.

The luck and the good intentions that got the Dublin Core started would not have been enough to sustain the initiative through a turbulent 5 years of discussion, halting starts, revisions, and passionate debate. The thread that stitches the initiative together is *trust*:

- *trust* that our differences in language, vocabulary, and culture are less important than our common goals
- *trust* that emerges from openness to the perspectives of many disciplines and traditions in resource description
- *trust* that none of us is as smart as all of us: our collective understanding is greater than that of any individual or isolated group

It is helpful to review the progress made over the past year and examine how the year's work shapes the agenda for the upcoming 7th Dublin Core Workshop [DC-7], and informs our expectations for the next year of our work.

Scope of the 1999 Workplan

The 6th Dublin Core Workshop [DC-6], held in November of 1999 in Washington, DC, resulted in a workplan with four major components:

- formalizing the Dublin Core Process
- Dublin Core 1.1 refinements
- Dublin Core qualification semantics
- promoting syntactic convergence for encoding Dublin Core metadata, including HTML and RDF

The Dublin Core Directorate: Changes in Process

The Dublin Core Metadata Initiative began as a single workshop, evolved into a series of workshops, and eventually into an international collective. The evolution has not ended, and it has become clear that this loose organization of stakeholders requires a formal process to manage its activities and to assure representation from the many communities involved. The year now ending has witnessed a maturation of Dublin Core governance that has improved communication and distributed responsibilities for accomplishing the goals of the initiative.

The changes that have been instituted are a result of the development of formal guidelines for managing Dublin Core efforts and structural changes in the committees that oversee the Dublin Core initiative [DC-PROCESS].

The Dublin Core process has been refined by a subcommittee of the Dublin Core Advisory Committee (DC-AC), and reflects our best effort at providing a formal means of conducting our business. The major change in our governance structure has been the formation of the Dublin Core Executive Committee (DC-EC). This committee, comprised of 5 individuals, is charged with the week-to-week management of Dublin Core activities. This is achieved through weekly teleconference calls, the maintenance of a task list and an open-issues list to help make explicit the tasks necessary to accomplish our goals.

The DC-EC is not a decision-making committee, but rather an operational committee. The DC-AC is the primary decision making body of the Dublin Core Metadata Initiative. The DC-AC is a consolidation of two previous committees, the Technical Advisory Committee and the Policy Advisory Committee. It was found that as a practical matter, these two committees were acting largely as a single committee, so

combining them seemed a sensible means of streamlining the governance of the Dublin Core. The membership is selected much the same as it was prior to consolidation: working group chairs are automatically members and a broad cross section of International policy leaders have been selected to represent stakeholder communities.

DC 1.1 Refinements

Experience with the Dublin Core elements in a broad spectrum of applications suggested that several of the DC element definitions would benefit from revisions to clarify their meaning. None of the refinements are intended to change the basic semantics of the elements. These semantics have been stable for three years, the result of long standing consensus and deployment in a broad array of applications. Rather, the revisions should make it easier to apply the definitions consistently.

Working groups completed proposed revisions in June, and following a public comment period in July, the revisions were ratified by the DC Advisory committee and now constitute a formal Dublin Core Recommendation [DC-1.1]

Standardization Efforts

The semantics of the basic 15 elements (what has become known as DC 1.0) has been captured in the Internet Engineering Task Force (IETF) RFC 2413. The Dublin Core community used the IETF as a standardization path because it provides a formally maintained repository of standards documents with relatively simple requirements for ratification.

An Internet Draft (a precursor to an RFC) has been developed for the expression of DC metadata in HTML [DC-HTML]. A related document describing best practice for including DC qualifiers in HTML, has been submitted as a DC Note as well [DCQ-HTML]. As these works mature, they will establish formal standards for encoding DC metadata in HTML, a goal that should help to promote interoperability among different DC applications.

As of August, the revisions to DC 1.0 have been completed and ratified by the Dublin Core Advisory Committee. This revised document will comprise the primary input to the standardization processes of both CEN (the European standardization organization) and NISO (the North American standards organization)[CEN, NISO]. The intention is to

have these processes proceed in parallel, working on the same document, so as to minimize duplication of effort and resulting in a single standards document that has global scope.

Another area of standardization of immediate concern to the Dublin Core community has to do with the progress of the Resource Description Framework (RDF) and Extensible Markup Language (XML). The DC community has played a major role in the development of RDF, a technology that Tim Berners-Lee describes as the key to achieving a semantic web of inference. The rapid development of XML as the primary markup standard for advanced publishing on the Web has introduced overlap in the specification of schemas (an important part of the infrastructure necessary to support machine processable metadata systems). At this writing there are ongoing negotiations between the RDF and XML communities to resolve these conflicts. This important issue should be resolved soon after these meetings.

DC-Qualifiers

Qualifiers have been recognized since DC-4 as integral to effective use of DC and as a tool for extensibility and refinement. The use of qualifiers will inevitably reduce interoperability, however, and hence the DC community has an interest in promoting a common set of qualifiers to reduce the diffusion of semantics. The identification of a standard set of DC qualifiers emerged as one of the primary objectives of the 1999 workplan.

Working Groups have been gathering input on current use of qualifiers among Dublin Core applications and are currently engaged in discussions that will culminate in recommendations about a standard set of qualifiers that will enjoy broad adoption in the DC community. It is expected that these discussions will be an important part of the agenda for DC-7 in Frankfurt.

It is not expected that these qualifiers will meet the needs of every community, but rather, that they will serve applications which do not have well defined alternative qualification schemes. Communities having their own qualification vocabularies and schemes will be able to invoke these schemes as well, enabling such communities to tailor their metadata to the needs of their constituents while remaining

within the generalized metadata architecture of the Dublin Core initiative.

Relationships to Other Metadata Standards

MARC/AARC2

The influence of the library cataloging standards MARC and AACR2 on Dublin Core is well known. Librarians have comprised about a third of active participants in the Dublin Core Metadata Initiative from its inception, and there is perhaps no community that has a stronger tradition of resource description standards than the library community. It is natural that the philosophy of MARC cataloging has had an influence, and as the largest legacy resource description system, it is important that these systems operate as seamlessly as possible.

The integration of MARC/AACR2 and DC is among the specific objectives of the CORC project at OCLC [CORC]. The CORC project has demonstrated that a single database will support the needs of both communities. Any record in the CORC database can be created or displayed in either format. The CORC team has worked closely with the Office of MARC Standards in the Library of Congress and with the Dublin Core Directorate to promote interoperability between these formats.

Other DC Projects have invested significant effort in cross walks between DC and MARC/AACR2 as well, including the Library of Congress, the Nordic Metadata Project, the National Library of Australia, and the United Kingdom Office of Library and Information Networking to name a few. The close attention to these issues on the part of researchers and implementers from many projects will help assure that the large investment in MARC/AACR2 cataloging will not be isolated from more broadly based resource description on the Internet. It also helps to bring to bear the long-standing expertise of the library world on the problems of resource description on the Internet.

Z39.50 and Dublin Core

Z39.50 is an important part of infrastructure supporting libraries and database producers. While it has not had a major impact in the Web environment, it nonetheless has a prominent role in providing search interoperability across

heterogeneous databases, which are increasingly available via the Web.

The Dublin Core is the proposed cross-domain attribute set for Z39.50 [LEVAN] and a recent proposal for improving interoperability among Z39.50 applications also includes the Dublin Core as a central feature [BATH].

DC and INDECS

The Dublin Core community has been engaged in discussions with the INDECS community since July of 1998. It is evident that there is substantial overlap between the functional requirements of our respective groups, however, differences in our underlying approaches have caused us to focus on different means of expressing and encoding metadata.

As of the middle of 1999 there are informal agreements between our groups to use the DC namespace for top level categories where semantic overlap exists. Applications of electronic commerce and intellectual property rights management will require additional functionality and authority control which can be provided via the use of INDECS-specific qualifiers and additional metadata elements beyond the DC 15.

DC, IMS, and Ariadne

The educational community is another example of a constituency that has particular requirements for metadata to promote instructional and curriculum management goals. The Dublin Core community has had educational projects in a number of countries from its inception but there are other initiatives that focus on these needs as well, including the IMS Project [IMS] in the United States and the Ariadne project [ARIADNE] in Europe.

These efforts have substantial semantic overlap among their metadata sets but differences in approach have made it difficult to promote interoperability. The DC-Education Working Group has been established to address the functional requirements of the educational materials sector within the DC framework. As a result of the establishment of this group, closer discussions have been initiated between the DC world on the one hand and the IMS/Ariadne coalition on the other.

It is not clear at this time what will develop from these discussions, though certainly the international educational community will be better served to the extent that interoperability

among these standards is improved. Further, these standards should be, as far as possible, consistent with the resource description infrastructure for Web resources in general. The Dublin Core has pursued a modular metadata approach of this type since the Warwick Workshop, and we continue to promote this notion of a general resource description core that can be enhanced and refined to support the needs of particular content domains.

The DC-7 Agenda

DC-7 will convene at Die Deutsche Bibliothek in Frankfurt at the end of October to bring together long time DC participants and many new participants [DC-7]. Nominees from 28 countries have been invited to bring their experience and perspectives to bear on active discussions of current problems and the formulation of the workplan for the coming year.

Consolidating qualifier semantics will be the primary theme of the meeting, but longstanding issues (internationalization and standardization)

as well as recent developments (the formulation of domain-specific working groups such as DC-Education) will play a significant role in the agenda.

The international diversity of the Dublin Core community has always been an important part of the vitality of the group. The translation of Dublin Core standards into 25 languages [DC-INT] testifies to the global importance of these issues, and also highlights the difficulty of maintaining semantic consistency in a multilingual, multicultural environment. The broad array of nationalities in attendance at DC-7 will help develop this aspect of the initiative and promote the means to advance this important dimension of global resource discovery.

Finally, we will continue refining the governance and management process of the Dublin Core Metadata Initiative in our attempt to consolidate the trust of the largest interdisciplinary aggregation of resource description experts in the Internet community.

References

[ARIADNE] ARIADNE Project (home page)
<http://ariadne.unil.ch/>

[BATH] The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery.
<http://www.mailbase.ac.uk/lists/dc-general/1999-08/0018.html>

[CEN] European Committee for Standardization: Information Society Standardization System (home page).
<http://www.cenorm.be/iss/>

[CORC] The CORC Project (home page)
<http://purl.oclc.org/corc>

[DC-6] dc:DC The Sixth Dublin Core Metadata Workshop
<http://purl.oclc.org/dc/workshops/dc6conference/index.htm>

[DC-7] The 7th Dublin Core Metadata Workshop (home page)
<http://www.ddb.de/partner/dc7conference/index.htm>

[DC-HOME] The Dublin Core Metadata Initiative (home page)
<http://purl.org/DC>

[DC-INT] Dublin Core Metadata Initiative: Multiple Languages Working Group
<http://purl.org/DC/groups/languages.htm>

[DC-PROCESS] The Dublin Core Metadata Initiative: About the Organization and Process
<http://purl.oclc.org/dc/about/index.htm>

[DCQ-HTML] Cox, Simon. Recording qualified Dublin Core metadata in HTML. 1999-08-16
<http://purl.oclc.org/dc/documents/notes-cox-19990816.htm>

[IMS] IMS (home page)
<http://imsproject.org/>

[INDECS] INDECS: Interoperability of Data in E-commerce Systems (home page)
<http://www.indecs.org/>

[LEVAN] Levan, Ralph, The Z39.50 Cross-Domain Attribute Set. Version 1.4, 1999-08-04.
<http://www.oclc.org/~levan/docs/crossdomainattributeset.html>

[NISO] National Information Standards Organization (home page)
<http://www.niso.org/>

[RFC-2314] Weibel, S.; Kunze, J.; Lagoze, C.; Wolf, M. Dublin Core Metadata for Resource Discovery.
IETF #2413. The Internet Society, September 1998.
<http://www.ietf.org/rfc/rfc2413.txt>