ARTICLE EXCERPTED FROM:

INFORMATION STANDARDS QUARTERLY

WINTER 2010 | VOL 22 | ISSUE 1 | ISSN 1041-0031

SPECIAL ANNIVERSARY EDITION: PART FOUR

PLANNED OBSOLESCENCE: A NEW MODEL FOR ACADEMIC PUBLISHING

DCMI: BEYOND THE ELEMENT SET

E-BOOKS: THE ETERNAL NEXT BIG THING

Z39.7 LIBRARY METRICS & STATISTICS DATA DICTIONARY

OPENURL KNOWLEDGE BASES RECOMMENDED PRACTICE

COREY A. HARPER

DUBLIN CORE METADATA INITIATIVE:

BEYOND THE ELEMENT SET

Dublin Core Metadata Initiative®

BACKGROUND »

The Dublin Core Metadata Initiative (DCMI) had its origins at a time when the World Wide Web was in its infancy. Over 15 years ago, in October of 1994, a hallway conversation took place at the 2nd Annual World Wide Web conference in Chicago. This discussion centered around the need for infrastructure to enable discovery of resources on the then nascent Web, despite the fact that it only included approximately "500,000 addressable objects" at the time. A few months later, a workshop was held to discuss a very basic metadata format for describing resources on the Web, and thus DCMI was born.

Between 1995 and 2001, DCMI held a series of workshops and meetings to discuss this need and to develop an extensible and broadly applicable standard. The perceived need was very specific, and focused on simple description for discovery purposes. By 1999, the set of 15 metadata elements was finalized and published as an RFC. *The Dublin Core Metadata Element Set* (DCMES) became a national standard in 2001 (ANSI/NISO Z39.85) and an international standard in 2003 (ISO 15386).

Shortly after the original publication of the element set, the DCMI broadened its scope to metadata practice and research, and added a peer-reviewed conference track and tutorials to its Workshop Series.

CONTINUED »

The linked data approach has since seen rapid uptake throughout the web community, from players including Google, Yahoo, Thompson Reuters, New York Times, BBC, and libraries. As library data increasingly enters into this giant global graph of linked data, their users—and the systems developed for them—can reap the benefits of the "follow your nose" principles Ed Summers wrote about in ISQ one year ago.

Criticisms of Dublin Core

DCMI has received a fair amount of criticism since the inception of the DCMES, particularly from the library community, mainly focused on the overly simple structure and format of the element set. It is true that early discussions within the Initiative were very focused on this element set, though parallel to those discussions the organization began to put a premium on more broadly applicable metadata research.

Much of the criticism posits DCMI's overly-simplistic and generalized set of elements as a central weakness, noting that the standard does not offer the richness and specificity required for resource description. Often, such criticism illustrates this drawback through comparisons with MARC and other standards in use in the library community.

While valid, these criticisms only apply to the DCMES metadata *format*. The DCMI's own initial focus on a simple set of elements led to misconceptions about the initiative's purpose and the nature of the Dublin Core. As a result, the element set's shortcomings sometimes lead to a misevaluation of the usefulness of the Initiative itself.

Despite these criticisms, the DCMES has been widely used in many communities and has formed the basis of more specialized metadata element standards, which was DCMI's original intent.

Indeed, the focus in the early years of 1995 to 1999 on the fifteen elements was justified by the assumption, articulated in 1996 by Carl Lagoze, Clifford Lynch, and Ron Daniel in the so-called Warwick Framework, that simple Dublin Core descriptions would provide just one among potentially several resource description "packages", of varying richness and granularity, associated with a given resource. The Warwick Framework idea was one source of inspiration for work on a new Resource Description Framework (RDF) at W3C in 1997—a parallel development which, as discussed below, redefined the scope of DCMI itself.

DCMI has grown far beyond the set of 15 elements bearing its name. Today the Initiative provides a framework and model, as well as a set of principles for designing metadata. It is also a diverse community bound by a common interest in developing the underpinnings of rich, interoperable metadata. The real value proposition of Dublin Core lies in its commitment to interoperability, as well as in applicability of the organization's guidelines and recommendations to any metadata scheme, element set, or implementation syntax.

RDF and the Semantic Web

In recent years, some information professionals—particularly those outside of the library community—have begun to change their conceptualization of metadata. Historically, records—and not the statements about resources that they aggregate and package—have been treated as the central components of metadata. This was necessary, and to an extent still is, due to the attention being paid to how these packages are transmitted from one system to another. The MARC format has been central to library metadata in large part because of its usefulness as a communication medium for transmitting metadata, usually through a Z39.50 service. The problem with this conceptualization of metadata is that it arbitrarily limits the edges of description to what can be effectively packaged and transmitted in a record.

Instead of focusing on the aggregation of individual pieces of metadata, DCMI and the Semantic Web community are advocating a focus on the smallest components of a resource's description. The *RDF Concepts and Abstract Syntax* document, one of a suite of specifications that collectively define RDF, defines the syntax of RDF as being made up of triples—statements composed of a subject, predicate, and object where properties serve as predicates (e.g., dc:title), the subjects are denoted by URIs defining the resources about which statements are made, and the objects can either be textual strings or additional resources. For example, as can be seen in *Figure I*, this article has a triple with the subject being the article, a predicate of dc:identifier and an object of the doi:10.3789/isqv22n1.201004. A second triple for the same subject has the predicate dc:title and the object of "DCMI: Beyond the Element Set".

The architecture of the World Wide Web allows statements to be linked together and woven into a rich tapestry of

descriptions, forming a *graph* that extends its reach across data from myriad sources. This terminology is significant. In the "graph" paradigm, it becomes easier to envision how library metadata interacts with other metadata on the open web. As the graph grows, systems interested in metadata packages and records have a more diverse selection of descriptive information to utilize when building these structures.

The value of RDF lies in its use of URIs to identify both resources and properties. Unique URIs provide "hooks" for linking statement data from multiple sources. However, the unfamiliar language of formal modeling, the complex RDF documentation, and the difficulty of its XML representation presented a hindrance to widespread adoption of RDF. In 2000, Roy Tennant included RDF in a list of "dead" technologies, stating that obscure concepts like "directed labeled graphs" would limit uptake.

Despite the lack of widespread deployment, the Semantic Web community continued to refine their thinking and further develop the specifications. By 2006, Tim Berners-Lee had published a design note in which he reframed the Semantic Web discussion in much more useful terms by succinctly articulating both the simplicity and elegance of *linked data*. This design note focused on assigning URIs to resources, providing useful descriptive information at those URIs, and including links to other URIs. The linked data approach has since seen rapid uptake throughout the web community, from players including Google, Yahoo, Thompson Reuters, New York Times, BBC, and libraries. As library data increasingly enters into this giant global graph of linked data, their users—and the systems developed for them—can reap the benefits of the "follow your nose" principles Ed Summers wrote about in ISQ one year ago. Tennant has since published a pair of follow-up articles re-evaluating his initial conclusions due to the appealing nature of linked data.

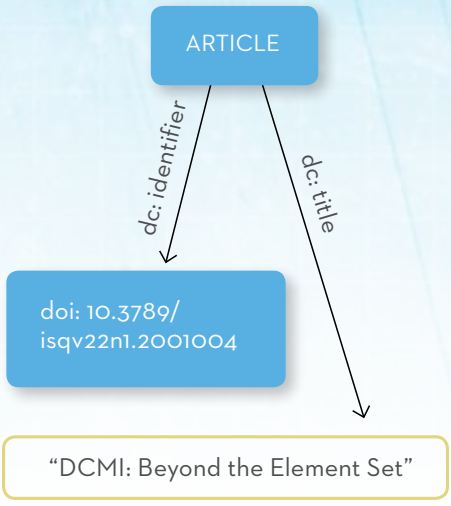


Figure 1. This article has a triple with the subject being the article, a predicate of dc:identifier and an object of the doi:10.3789/ isqv22n1.201004. A second triple for the same subject has the predicate dc:title and the object of "DCMI: Beyond the Element Set".

Metadata as Format vs. Metadata as Vocabulary – Qualified Dublin Core

In the early days of the DCMI, the connection of Dublin Core to RDF and the Semantic Web was not obvious, and many participants likely did assume that DCMES as a format was the end goal of their efforts. However, when early DCMI participants such as Eric Miller began working on RDF in 1997, some members of the community began to shift the focus of the conversation from a metadata *format* to a metadata *vocabulary*—a collection of carefully defined properties that could be used to make descriptive statements about resources. Subsequently, the DCMI and Semantic Web communities progressed on parallel tracks and influenced one another a great deal.

These changes in DCMI's own conception of its work began in the late 1990s, and are demonstrated by the notion of Qualified Dublin Core, which appeared on the DCMI website in July 2000. This introduction included both Element Refinement Qualifiers, which add specificity to the refined element, and Encoding Scheme Qualifiers, which provide constraints on the value space drawn on when populating the data of an element. The introduction of metadata element qualification marks DCMI's evolution into an organization with a broader scope.

In 2000 and 2001, as the DCMI began to discuss the implications of Qualified Dublin Core, the Initiative undertook efforts toward understanding how metadata practitioners would adjust and mold metadata schemas to meet particular application needs. In contrast, many in the library community saw Qualified Dublin Core as nothing more than a more detailed metadata format. As a result, libraries wanted a comprehensive schema defining how the format was to be used with record exchange protocols such as the Open Archives Initiative Protocol

CONTINUED »

When early DCMI participants began working on RDF in 1997, some members of the community began to shift the focus of the conversation from a metadata *format* to a metadata *vocabulary*—a collection of carefully defined properties that could be used to make descriptive statements about resources.

Take, for example, a digitization of a photograph of a famous painting. The 1:1 principle posits that a distinct record should be created for each manifestation of the object in question (i.e., the painting, the photograph of the painting, and the digitized version of the photograph), and that relation and/or source elements should be used to create linkages between each discrete record.

for Metadata Harvesting (OAI-PMH). This disparity further highlights the gulf between the record-based and vocabulary-based schools of thought concerning metadata interoperability. During these formative years, the DCMI prioritized its efforts to demonstrate that these viewpoints are not incompatible.

In a 2004 article, Jeffery Beall described the notion of qualifiers as local extensions that "defeat 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." While this is marginally true in the context of a record *format*, it misses the point of combining elements from an ever-growing pool. The elements form a vocabulary for resource description, which can be drawn upon to build more ad-hoc metadata formats according to the specific needs of a given application or community. This does *not* defeat the purpose of DC as a common exchange mechanism, but rather makes the concept more powerful by moving the definitions and specifications to a level that is more granular than a specification defined at the level of the aggregation. This notion of mixing and matching is familiar to users of XML specifications, who have a long history of defining XML elements per namespace and allowing a document to draw on elements from a variety of namespaces. The idea of mixing namespaces is generalized to the construction of metadata for all contexts in the Dublin Core notion of *application profiles*.

Dublin Core Application Profiles

In 2000, Rachel Heery and Manjula Patel introduced in an *Ariadne* article the concept of "application profiles as a type of metadata schema." This was the first published discussion of how to make well-modeled statement-based metadata in the context of record-based systems. With an application profile, the metadata *record* becomes an application-specific aggregation of statements that draw on the properties defined by Dublin Core *and* on properties that are defined elsewhere. This concept does not conflict with the need for many systems to exchange metadata records, and also allows external descriptive information to be linked on the basis of any particular resource's identifier. It enables management of metadata at a granular level while taking advantage of the web's open infrastructure. This increased specificity allows for customization of the vocabularies used in a description. Additionally, graphs can be merged to combine pieces of description (statements) from other sources. Application profiles allow projects to specify constraints to how elements from a vocabulary are used. In the most generic DCMES-based format, for example, all properties are optional and repeatable, but an application profile might specify that the "title" element is required and non-repeatable.

The "1 to 1" Principle

Among the first indications that RDF-based thinking was entering the DCMI dialogue was a spirited debate centered on something called the 1:1 principle. The general problem addressed by the 1:1 discussion is that of how to describe complex objects with regard to various metadata elements. The debate emerged around the creator element (i.e., for describing the affiliation of the creator of a resource), but applies to a variety of other elements.

Take, for example, a digitization of a photograph of a famous painting. The 1:1 principle posits that a distinct record should be created for each manifestation of the object in question (i.e., the painting, the photograph of the painting, and the digitized version of the photograph), and that relation and/or source elements should be used to create linkages between each discrete record. This principle, though often challenging to encode in a metadata record in the "document" sense, can be seen as an important contribution to the theory and practice of describing resources, and fits very well with the "follow-your-nose" principles of linked data.

Unfortunately, many existing metadata encodings make it difficult to tease out exactly what is being described by any particular piece of information in a record. Take, for example,

the publication statement of a MARC record, which includes data about the publisher's location at the time a particular book was printed. When taken out of the MARC context, the statement becomes a free-text string that violates the 1:1 principle by describing more than one "resource" (i.e., the location of the publisher is a property of the publisher rather than of the primary resource described in the record). This presents challenges when trying to make MARC data interoperate with data that is structured according to more modern principles of database normalization and relational data modeling.

Dublin Core Abstract Model and Ongoing DCMI Development

One very significant value of the DCMI is its ongoing work to make tools and principles like those developed in the W3C relevant in more traditional metadata spaces, including libraries. The DCMI serves as a bridge between the linked data community and other areas of metadata practice. Additionally, the close ties that the DCMI has with the W3C and the Semantic Web Community continue to influence DCMI's work, and vice-versa. This cross-pollination can be seen in development of the Dublin Core Abstract Model (DCAM) from 2003 through 2005. DCAM is designed to document the structure and to provide guidance in the use of Dublin Core compliant metadata and define how various pieces of metadata are combined to build descriptions of resources. A very significant feature of DCAM is that it is syntax independent.

The development of DCAM can be traced to efforts in the DCMI Architecture Forum to distill and make manageable the more challenging concepts in the suite of RDF specifications. The Architecture Forum felt that the central design principles of the Semantic Web could be applied to metadata practice without requiring RDF's obscure jargon and notoriously difficult XML syntax, so they attempted to craft a more accessible text to be used as a foundational data model for metadata. It is worth noting that this effort was finalized two years prior to Berners-Lee's note on linked data, a document with a similar purpose.

Some argue that DCAM tried to be too many things to too many people. To those who understood RDF, the additional value was hard to see. Why not just use the RDF data model as the data model? To those who were not already steeped in the terminology and concepts of the Semantic Web, it was a dense and impenetrable document. *Note: As this article goes to press, there is an ongoing discussion in DCMI about exactly this problem. Now that RDF language has become more familiar in the context of the Linked Data movement, it is argued that DCMI-specific terminology in DCAM should be further de-emphasized in favor of explicit alignment with RDF.*

If the DCMI revises DCAM to be more closely aligned with RDF and to still apply more broadly to other encodings and syntaxes, the current document's very useful constructs will continue to add value to the metadata conversation. One such construct that has particular value is the notion of the *description set*, which builds on the 1:1 principle by stating that a metadata description describes one, and only one, resource. At the same time, the DCAM authors acknowledge the complexity of applying this principle in practice, stating that,

> *"... real-world metadata applications tend to be based on loosely grouped sets of descriptions (where the described resources are typically related in some way), known here as description sets. For example, a description set might comprise descriptions of both a painting and the artist. Furthermore, it is often the case that a description set will also contain a description about the description set itself (sometimes referred to as 'admin metadata' or 'meta-metadata')."*

The concept of the description set provides a container to anchor a set of related descriptions around the description of one central resource in the context of a bounded entity—the record—further helping to bridge the chasm between the record-centric and property-centric approaches to metadata.

CONTINUED »

BRIDGING THE GAP

One very significant value of the DCMI is its ongoing work to make tools and principles like those developed in the W3C relevant in more traditional metadata spaces, including libraries. The DCMI serves as a bridge between the linked data community and other areas of metadata practice.

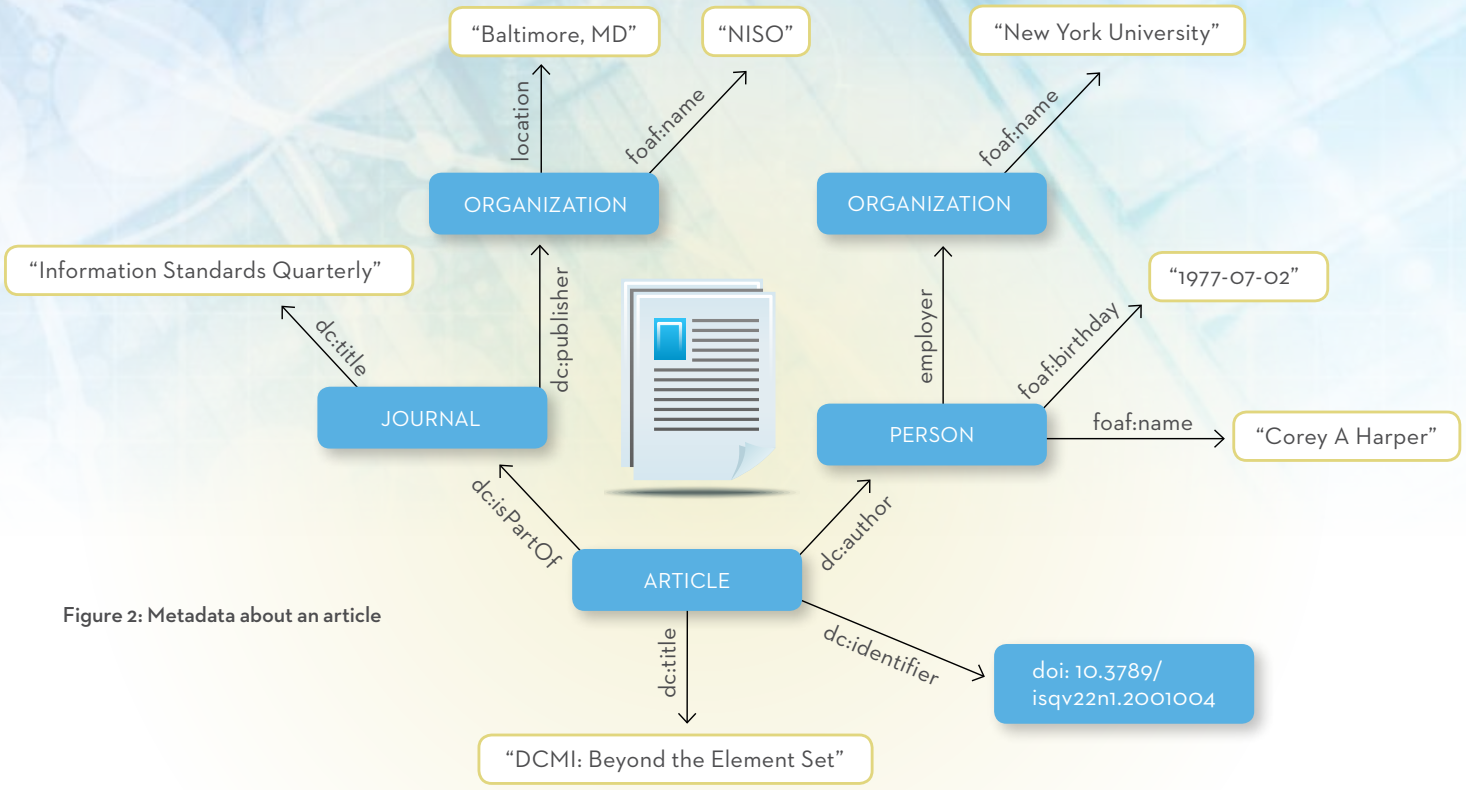


Figure 2: Metadata about an article

The promise of RDF is that, if different groups use the same identifiers for the same resources, the possible set of metadata to draw on in a resource description is theoretically boundless. However, this poses a problem in the context of applications that only need a subset of the metadata available. Having this center point to frame each conversation about a resource helps make the presence of so much metadata in the graph less problematic in those cases.

The DCAM helps the metadata retain its focus, forming a *description set* anchored around the central URI of the described resource. Figure 2 illustrates the concept using this article as an example.

This idea continues to shape DCMI's thinking, as can be seen in the *Guidelines for Dublin Core Application Profiles* and the currently under-development *Description Set Profiles (DSP)* documentation. According to DCMI's *Singapore Framework for Dublin Core Application Profiles*, a DSP "defines a set of metadata records that are valid instances of an application profile." The DSP provides rules for drawing the lines of demarcation around a portion of a graph, centered on the described resource, to facilitate the effective packaging of application specific metadata *records* describing that resource. While library data likely wants to represent the fact that NISO is located in Baltimore, another application may not care about this piece of information. It could apply its own rules to the same pieces of data to limit the triples included in its view of the description, then generate a record to represent that subset of data.

Similarly, it would be possible to hang additional triples off the identifier for "New York University." A library catalog application, if generating a MARC record from this data, would stop before processing information about the author's affiliation and about that organization, but likely *would* include selected information about NISO for inclusion in the publication statement.

At this time, however, the linked data uptake is new enough that rich vocabularies for describing entities like persons and organizations are limited and often very informal. This problem could potentially solve itself as libraries embrace the linked data movement. As noted earlier, a very large body of metadata specifications has focused on defining the metadata packages, and many of the necessary properties needed for describing related resources are already part of larger, XML-based standards. Rethinking the structure of these standards to support reuse as metadata vocabularies offers tremendous potential. For example, the elaborate record structures and rule sets governing library name authorities for both personal and corporate bodies provide a powerful

foundation upon which to build a robust vocabulary of properties for describing these entities. The resultant properties would offer a reputable solution for a set of challenges with which the semantic web community is struggling.

Additionally, basing these vocabularies on library authority records helps ensure backward compatibility with existing data, since it should be relatively easy to repackage subsets of these graphs using some sort of Library of Congress Name Authority File application profile. This legacy data could also be transformed into linked data in order to seed the graph with data converted from libraries' existing authority and bibliographic data. By including vocabulary-like components, recent efforts to update and revise the library community's bibliographic standards are helping to realize this transition.

RDA as RDF

There has been a great deal of discussion—and some controversy—around *Resource Description and Access* (RDA), the next generation of the library cataloging rules. However, until recently, much of this conversation has overlooked a very significant parallel effort that is happening between RDA and the DCMI community. A meeting between the developers of RDA and members of DCMI took place in 2007, at which time a DCMI Task Group was created to ensure that RDA could be treated as a Dublin Core Application Profile.

A recent article in D-Lib Magazine describes the challenges presented by the joint RDA/DCMI process and discusses the solutions that the task group and other participants have begun to put into place. Much of this work has involved systems and processes for defining these element sets as the types of data constructs that are used by RDF. The development of RDA includes the first attempt by the library community to implement *Functional Requirements for Bibliographic Records* (FRBR) in the context of a standard rather than after the fact, through algorithmic record "FRBR-ization." This is significant because it begins to define the various entities that the metadata is about, and allows the vocabularies being developed to adhere to the 1:1 principle, resulting in metadata that is both manageable and reusable.

Conclusion

The RDA work, while significant, is just one example of the possibility for various metadata communities to redesign their standards in order to ensure greater reuse and interoperability on the web. DCMI continues to engage in important work providing tools and guidelines to enable efforts like the RDA/DCMI collaboration.

Ongoing work to re-align the DCAM with the *RDF Model and Abstract Syntax* document will ensure that DCMI-compatible metadata of all stripes can interoperate well with other sources of linked data. Continued development of the *Description Set Profile* specification will refine the rules and guidelines for packaging statements into well-defined records for transmission and exchange. Additionally, this concept, when combined with the guidelines for application profile development, provides the tools needed to refine and augment these records for specific applications.

CONTINUED »

RELEVANT LINKS

Beall, Jeffrey. Dublin Core: An Obituary. Library Hi Tech News, 2004, 21 (8): 40-41
www.emeraldinsight.com/Insight/viewContentItem.do;jsessionid=514EE9D3AD14AAEC161E0A59221F85DB?contentType=Article&hdAction=lnkhtml&contentId=861505

Berners-Lee, Tim. Linked Data.
www.w3.org/DesignIssues/LinkedData.html

DCMI History
www.dublincore.org/about/history/

DCMI/RDA Task Group
dublincore.org/dcmirdataskgroup/

Description Set Profiles (DSP)
dublincore.org/documents/dc-dsp/

Dublin Core Abstract Model
www.dublincore.org/documents/abstract-model/

Dublin Core Metadata Element Set
dublincore.org/documents/dces/

The Dublin Core® Metadata Initiative
www.dublincore.org

Dublin Core Qualifiers
dublincore.org/documents/2000/07/11/dcmes-qualifiers/

Guidelines for Dublin Core Application Profiles
dublincore.org/documents/profile-guidelines/

Heery, Rachel and Manjula Patel. Application Profiles: Mixing and Matching Metadata Schemas. Ariadne, 2000, issue 25.
www.ariadne.ac.uk/issue25/app-profiles/

Hillmann, Diane, Karen Coyly, Jon Phipps, and Gordon Dunsire. RDA Vocabularies: Process, Outcome, Use. D-Lib, 2010, 16(1/2).
www.dlib.org/dlib/january10/hillmann/01hillmann.html

Interoperability Levels for Dublin Core Metadata
dublincore.org/documents/interoperability-levels/

MARC Standards
www.loc.gov/marc/

Open Archives Initiative
www.openarchives.org/

Resource Description Framework (RDF)
www.w3.org/TR/2004/REC-rdf-concepts-20040210/

RELEVANT LINKS

CONTINUED »

The Singapore Framework for Dublin Core Application Profiles
dublincore.org/documents/singapore-framework/

Summers, Ed. Following Your Nose to the Web of Data. ISQ, 20(44), 2008.
inkdroid.org/journal/2008/01/04/following-your-nose-to-the-web-of-data/

Tennant, Roy. Bringing Out the Dead (Technologies). Library Journal, 10/15/2000.
www.libraryjournal.com/article/CA156521.html

Tennant, Roy. The Semantic Web, Part I: Promises, Promises. Library Journal, January 6, 2009.
www.libraryjournal.com/blog/1090000309/post/1910038791.html

Tennant, Roy. The Semantic Web, Part II: Linked Data. Library Journal, January 7, 2009.
www.libraryjournal.com/blog/1090000309/post/1930038793.html

Warwick Framework
www.dlib.org/dlib/july96/lagoze/07lagoze.html

CONTINUED »

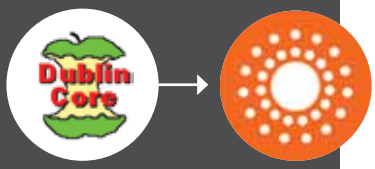
This process helps set the stage for reconciling the conceptual gap between standards for metadata packages and standards for metadata vocabularies. This will be of tremendous value to resources that have traditionally been on the margins of descriptive practice, such as special collections and audio-visual materials.

In addition to these valuable contributions, the DCMI has begun another effort to help harmonize metadata standards and ensure that as much metadata as possible will be compatible with the efforts described throughout this paper. The recent publication of the *Interoperability Levels for Dublin Core Metadata* document aims to guide a variety of audiences in evaluating the placement of their metadata along an interoperability continuum. The levels are meant to aid in decision making for communities that might wish to undertake efforts like the RDA work described above, by "specifying the choices, costs, and benefits involved in designing applications for increased levels of interoperability." DCMI recognizes the challenges of integrating myriad data formats into the linked data environment and is striving to be a central component in providing accessible and usable guidelines, specifications, and recommendations to support standards developers and metadata practitioners. | FE | doi: 10.3789/isqv22n1.201004

COREY A. HARPER <corey.harper@nyu.edu> is Metadata Services Librarian at New York University. Corey has been involved in DCMI since 2002 and works with the DCMI Registry and Library Application Profile Task Groups. He is also active in the DCMI Library and Registry Communities and sometimes the Architecture Forum, and serves on the Program Committee for DCMI Conferences.

Acknowledgements: *The author is indebted to the feedback and editorial input of both Tom Baker (DCMI) and Alexa Pearce (NYU).*

Dublin Core Celebrates its FIFTEENTH Anniversary: We've Come a Long Way



It started with a hall conversation at the second-ever Web conference in Chicago and led in March 1995 to a workshop in Dublin (Ohio) and the first draft of a "Dublin Core" metadata element set. The focus in those early years on core terms was reflected in the informal logo of an apple core. When the term set grew and a formal governance structure emerged, the Dublin Core Metadata Initiative (DCMI) developed the logo of a sunny orange core ringed with inner and outer circles of elements. Fifteen annual meetings later, held in almost as many countries, DCMI now has fifty advisory and oversight committees and an open-membership community of over two thousand people from fifty countries. Since their inception, DCMI vocabularies have remained among the most widely deployed metadata terms on the Web and continue to be maintained and developed using open review processes. The DCMI secretariat until 2009 was the Online Computer Library Center (OCLC) in Dublin. Now incorporated in Singapore and hosted by the National Library Board of Singapore, with Web servers hosted at the National Library of Korea, the international nature of the initiative is evident. One thing has not changed in fifteen years: the commitment to metadata standards and practices that will enhance the finding, sharing, and management of information.

Contributed by Tom Baker (DCMI)