The Death of Metadata

Jeffrey Beall

ABSTRACT. In the mid-1990s, information professionals and computer scientists began to pay much attention to the concept of metadata. The rise of the World Wide Web highlighted the need to enable discovery, or finding desired information, on the Internet. Metadata was the hot topic at library and information science conferences and in professional literature, and many librarians and others expended enormous efforts creating and implementing new metadata schemes. One of these schemes, the Dublin Core, emerged in 1995 and quickly became the metadata standard of choice for digital objects. Although abundant resources have been put into the development and implementation of Dublin Core, the schema has largely failed. In addition, many other metadata schemes and profiles have also been developed, most of them specific to a particular community. The result is a Tower of Babel of metadata schemes, and sharing metadata among professional communities is becoming increasingly difficult. Multiple schemes also make the ideal of successful federated searching all but impossible. Full-text searching, offered as a solution by some, is also a failure for most serious information-seeking needs. On the other hand, the implementation of the MARC format in libraries has been the most successful metadata implementation in history. MARC, along with various content standards, has worked successfully for decades as a metadata schema. The library community's implementation of the MARC metadata standard should be more widely adopted or used as a model for metadata schema design. A single, proven, comprehensive

Jeffrey Beall, MA, MSLS, is Catalog Librarian, Auraria Library, University of Colorado at Denver and Health Sciences Center, 1100 Lawrence St., Denver, CO 80204 (E-mail: jeffrey.beall@cudenver.edu).

This paper is a slightly revised version of the paper the author presented at the Canadian Metadata Forum in Ottawa, Ontario on September 27, 2005.

The Serials Librarian, Vol. 51(2) 2006
Available online at http://ser.haworthpress.com

doi:10.1300/J123v51n02_05

metadata standard will better enable discovery and control of information than a proliferation of minor schemes ever will. doi:10.1300/J123v51n02_05

KEYWORDS. Metadata, Dublin Core, MARC formats, information retrieval

Thank you and good morning. Although this session is entitled "Death of Metadata," I'll tell you right now that I do not believe that metadata is dead. It is abused and neglected, but it is definitely not dead. I used that title for my presentation mainly because that is the title the organizers of this forum gave it, and I decided just to stick with it. Probably a more accurate title for this presentation would be something like, "A critical analysis of the current state of metadata applications."

What I hope to do this morning is to provide a background on metadata, specifically the Dublin Core and MARC schemes, explain what a good metadata implementation needs to do, explain why full-text searching doesn't work very well for serious information-seeking needs, and explain why Dublin Core doesn't work very well. Also, I'll talk about the points of comparison for comparing metadata schemes, explain why MARC is successful, and describe MODS, the Metadata Object Description Schema. So, hopefully, after I am finished, we won't be talking about the death of metadata; instead we will be talking about the reinvigoration of metadata.

I'll begin by telling you a little about myself and about how I came to be invited to speak to you here today. I work as a catalog librarian at the University of Colorado at Denver and Health Sciences Center in downtown Denver, Colorado. I've been a cataloger for over fifteen years; my first ten years were spent at Harvard University in Massachusetts. In addition to cataloging, I am involved with database management and authority work for my library's online catalog. This means that I am responsible for the completeness, quality, and accuracy of the data in the online catalog. It also means that I am responsible for keeping the library's catalog current with all the national and international standards for cataloging data, which is another form of metadata.

I do some research and writing about topics related to my job. Specifically, I am interested in typographical and other data quality errors and how they limit searchers' access to library materials and to information available through the broader information landscape that is available through the Internet. Because my job and my research both deal with data quality, I am very critical whenever I observe a system or dataset that gets in the way of–or does not provide–effective, precise, and complete access to information.

So, on one day in the summer of 2004 I sat down and drafted an article about Dublin Core. The two things that motivated me to write the paper were what I saw as an unwarranted enthusiasm for the standard and the great ineffectiveness of it. I also thought it might be helpful for people to understand how the standard came to be created and what conflicts of interest there might have been in its creation. So, on that summer day, I sat down and I wrote my two-page article which I entitled "Dublin Core: an Obituary."

I thought I would likely have trouble finding someone who would publish such an article. Like I do for every article I write, I sat down and made a list of journals that I would send the article to. I made the list longer than a single journal in case the first ones rejected the article, then I would go down the list until, hopefully, one of them would be desperate enough for product and accept it. But much to my surprise, the first journal I submitted the article to accepted it, and "Dublin Core: an Obituary" was published in *Library Hi Tech News* in October, 2004.

The reaction to my article was quick. It was also negative and strong. I started getting e-mails from friends and colleagues advising me to make sure I saw all the negative things people were saying about me. Most of the reaction I observed was in two library-oriented e-mail lists, and I'll share some of that reaction with you in a minute. The first list that discussed my article was the "Metadata Librarians" list and the second was the list of the Library and Information Technology Association of the American Library Association.[1]

One thing I didn't realize prior to submitting my article was the reaction that the word "obituary" might cause in people, and especially when you associate the word "obituary" with individuals' jobs. I learned it can make them very defensive. So let me state right now that I am not trying to hurt anyone by saying negative things about Dublin Core or any other inadequate metadata scheme. I do not mean to threaten your job if your job depends on the continued existence of Dublin Core. I do think that a critical analysis of Dublin Core and metadata in general is warranted, however, and I hope to achieve that here this morning. So,

here's some of what was said about my article on the two e-mail lists. "While it's always nice to be cited, I'd hate for anyone to think I agree with Jeffrey Beall's article."[2]

This comment came from someone whom I had quoted in my article. It kind of went downhill from there. Someone else wrote: "Anyone who is familiar with the DCMI [that is, Dublin Core Metadata Initiative] process, or with the people who led the DCMI or the people who spent, and continue to spend, their time and talents defining and promulgating the standard, could not honestly write some of the comments you quote below."[3]

The "comments you quote below" were selected snippets from my article. And another person said: "I, too, read this article with raised eyebrows."[4]

In my article, I had said this: "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."[5]

And one person responded: "Ironically, most people deeply involved with Dublin Core are at this moment in 'far-flung' Shanghai for this year's meeting, but I do wonder how others see the DCMI and if this author expresses a common distrust/distaste of the DC activities."[6]

Actually, the fact that they were meeting in a remote place very nicely proved one of my points, that Dublin Core metadata standards development is pretty much limited to those organizations with the resources to send people to remote parts of the world every year.

Here are some more comments that were made: "I'm at the DC2004 conference in Shanghai, and as you can imagine, this article has created a great deal of irritation."[7]

> First, I'm surprised that the editors of Library Hi Tech News would even allow such an article to be published in the first place. The author is extremely ignorant about DC and DCMI. I could not begin to list the untrue statements that he asserts in the article, there are so many of them! The author fails to provide an [sic] empirical evidence to support any of his statements and he is hopelessly misinformed, misguided, and ignorant about the history, purpose, use, and utility of DC. In addition to the inaccuracies about DCMI, the author is clearly misinformed about Google, OCLC, the workings of search engines, CORC, and MODS. It is no wonder how he could produce a ridiculous article that is based on fiction, not facts.[8]

This last response was written by someone whose job title is "Metadata Librarian" and who has invested a lot of time into Dublin Core. I think my article was really the first one to openly criticize Dublin Core, and this caused a stir in the metadata orthodoxy.

HISTORY OF METADATA

So, to set the context, I think it would be a good idea now to look back on the brief history of metadata. I attend conferences of the American Library Association, and abruptly in the mid-1990s practically every presentation I attended at ALA began with these five words: "Metadata is data about data." I heard that phrase so many times that I began to get ill whenever I heard it. What was happening at this time was a change in terminology and a mixing of the library community and the technological community. According to Sherry Vellucci,

> The parallel world of library cataloging traditionally used the terms "bibliographic data" or "cataloging data" for this type of information. Catalogers used these terms when both the objects they cataloged and the bibliographic records were in a nonelectronic form. Catalogers continued to call this type of information cataloging data when the bibliographic record migrated to the Machine Readable Cataloging (MARC) format. They even called it bibliographic or cataloging data when they began to organize and describe local computer files. But when catalogers began to describe *networked* electronic resources using the same type of bibliographic data, the terminology changed. Suddenly the MARC record became metadata, and the cataloger's familiar world–circumscribed by the *Anglo-American Cataloguing Rules* (AACR) and the MARC formats–changed forever. This transformation was caused by a convergence with the broader world of information organization. The methods of organizing resources from the rather separate domains of library science, computer science, and information science all converged in this networked environment, and the term "metadata" became the commonly accepted term in all disciplines.[9]

For me as a librarian, one of the more frustrating aspects of metadata was the fact that the techies, that is, the technological folks, honestly believed that they had invented metadata in 1994, and we the librarians let them get away with thinking that. Librarians had, in fact, been dealing

with metadata for hundreds of years, in the form of book catalogs in monasteries right up to bibliographic records in online library catalogs. This trend is continuing today, for the hoi polloi are now employing folksonomies and acting as if they discovered the concept of the ontology. Next, they will probably "discover" authority control.

Right after Dublin Core was invented in 1995 it, for many people, became synonymous with metadata. Dublin Core was widely hailed as the successor to the MARC standard and as the final solution to discovery on the Internet. People boasted of the benefits of using Dublin Core before it had really been implemented anywhere. Countless librarians across North America and Western Europe joined the Dublin Core bandwagon and began implementing Dublin Core metadata applications. Interestingly, not much Dublin Core metadata was being created at that time (and relatively speaking, not much has been created since then), but systems librarians and information technologists were readying themselves for what they thought would be a voluminous amount of Dublin Core metadata that would facilitate discovery on the Internet. Remember, this was before the appearance of Google, and at that time Internet search engines were even less sophisticated than they are now.

I think a lot of Dublin Core's simplicity can be traced to a reaction against library cataloging standards, which were seen as too picky and excessively detailed. Dublin Core became sort of the anti-library standard. People in the technological community saw library metadata standards as too detailed and stigmatizing. Nevertheless, many librarians jumped on the DC bandwagon. Whereas the *Anglo-American Cataloguing Rules* were seen as overly prescriptive, Dublin Core was to be free-form and extensible. It was a mere frame upon which you could hang your own local implementation. You could use any content standard or none. It was sort of like a metadata scheme inspired by the spirit of freedom of the 1960s and combined with the doctrine of political correctness of the mid-1990s.

It was also sort of like a new product being released by Proctor & Gamble. Dublin Core was supposedly a new, fresh way of gaining access to information. Everybody was buying it! It's new and improved! Now, Dublin Core wasn't the first new metadata scheme after MARC, but it was surely the most talked-about one in the context of the early Internet.

Soon, though, numerous other schemes were under development, so many in fact, that some authors have referred to the set of schemes as a metadata ecology. Instead of sharing schemes, many different communities decided to develop their own, tailored, just, to their particular circumstances. It seems that the first step in metadata scheme creation is

the creation of an acronym for the scheme, and we have an abundance of these acronyms.

A result of this proliferation of metadata schemes–what I call the Tower of Babel of metadata schemes–is that it's become very difficult to crosswalk data from one scheme into another. Crosswalking is the process of mapping and exchanging data between two or more systems using two or more different metadata schemes. Because metadata schemes employ different content standards, such as different vocabularies for subject description, for example, crosswalking is hard to do well. Also, different schemes have such varying structures, a factor that also makes interoperability a challenge. That's why we hear so much about crosswalking and interoperability: they are nearly impossible to do well, and countless information professionals spend countless hours trying to perfect them. Dublin Core, especially simple or unqualified Dublin Core, is one of the most difficult metadata schemes in terms of interoperability because it is separated from content standards and because it uses so few elements of description. Another problem that a multiplicity of schemes creates is difficulty in federated searching. A federated search, theoretically, is a search initiated by a user done on a special search platform that goes out and searches multiple databases and then, supposedly, eliminates duplicates and presents a coherent and comprehensive set of results to the user. But this is really hard to do in an environment of multiple metadata schemes and multiple content standards. You can't program around the differences. So, creating so many metadata schemes is a colossal and probably needless duplication of effort. The proliferation of metadata schemes is tantamount to non-standardization. It's sort of like needing to have a different television set for each television channel. In fact, one might say that standards organizations approve so many standards that they are in fact promoting non-standardization.

WHAT METADATA OUGHT TO DO

Now I'd like to talk a little about why we need metadata. First, metadata adds value to information. This value-added process can be seen in several areas that I'll talk about in a minute. Second, it might be helpful if we divide metadata into three elements: structure, content, and value standard. Dublin Core is an example of metadata structure. Content generally refers to the separate elements of a metadata structure, such as subject, title, etc. Value standard refers to a standard system for a particular content

element, such as prescribing the use of the Library of Congress as the particular ontology (or vocabulary system) for the subject content.

Of course, metadata needs to work with some type of system for it to work well. In the library world, cataloging data is used by online library catalogs. On the Web, there are search systems that can exploit Dublin Core metadata, but none of them are very sophisticated. Also, Web search engines generally ignore metadata that is encoded in Web pages because of the tendency to misrepresent the information to cause a particular Web page to float to the top of search engine results pages. This tendency of search engines to ignore metadata has decreased the incentive for creator-produced metadata in Web pages.

Okay, one of the things that good metadata does well, especially when looked at in comparison with full-text searching, is to enable collocation. Collocation simply means displaying similar data elements together in a useful order, generally alphabetical or numerical. We know that collocation greatly improves information discovery. Figure 1 is an example of collocation I got from my library's online catalog. I performed a subject search on the topic "Colorado–History."

As you can see, the system presents a very neat and helpful display of the subject headings of works available to the information seeker. Particularly valuable here is the ability of the user to focus quickly in on a particular aspect of Colorado history. Collocation can be enabled by metadata for other elements of an information resource, including title, author, etc. By the way, the term "collocation" in this sense originated in the library community, and as a result, some technological people don't like to use it and instead use the term "resource aggregation" which basically means the same thing.

While we are looking at this example, let me point out another value that metadata adds, and that is what we call the "left-anchored index display." This is an example of that. It's closely related to collocation. Left-anchored index displays are valuable because they enable the user to quickly scan for what he is looking for.

Now let's look at the same search in Google.

There are a couple things to point out here. First, there's no real collocation. Entries are ranked by whatever system Google uses to rank works. And if anybody tries to tell you that this is an example of resource aggregation, then don't believe them!

Second, there are over 18 million hits, many of which I'm sure have nothing to do with Colorado history, and it's too many for anyone to sort through. This leads me to another thing that metadata needs to do, and that is enable search precision. Search precision is a measure of how

FIGURE 1. A Subject Search on "Colorado–History" in a Library Online Catalog

1	Colorado History --> Authority Record	1 entry
2	Colorado History	146 entries
3	Colorado History 1876 1950 --> Authority Record	1 entry
4	Colorado History 1876 1950 --> See also MILK CREEK, BATTLE OF	1 entry
5	Colorado History 1876 1950	20 entries
6	Colorado History 1876 1950 Biography	1 entry
7	Colorado History 1876 1950 Pictorial Works	1 entry
8	Colorado History 1951 --> Authority Record	1 entry
9	Colorado History 1951	2 entries
10	Colorado History Anecdotes	5 entries
11	Colorado History Anecdotes Juvenile Literature	1 entry
12	Colorado History Audiotape Catalogs	1 entry
13	Colorado History Bibliography	1 entry
14	Colorado History Bibliography Catalogs	1 entry
15	Colorado History Chronology	1 entry
16	Colorado History Civil War 1861 1865 --> Authority Record	1 entry
17	Colorado History Civil War 1861 1865	8 entries
18	Colorado History Curricula Standards	1 entry
19	Colorado History Exhibitions	1 entry
20	Colorado History Fiction	2 entries
21	Colorado History Juvenile Literature	11 entries
22	Colorado History Local	28 entries
23	Colorado History Local Bibliography	1 entry
24	Colorado History Local Exhibitions	1 entry

many of the hits in a retrieval are relevant compared with the total number of hits retrieved. Because there are so many hits in this Google search, and because we can be relatively sure that most of them have nothing to the history of Colorado, we can say that the Google search scores very low on search precision. But the search from my library's catalog–the one done through a system that searches the metadata, appears to be very precise, for all of the results listed look like they relate directly to Colorado history. Here, the metadata and the online catalog add value to the information in my library by efficiently telling me what information is available on the topic.

Metadata also ought to be easily sharable, to reduce duplication of effort. This, of course, is another way of saying that its data ought to be easily crosswalked and its structure ought to render easy interoperability. Sharing metadata is desirable because then it only has to be created once and then other organizations can copy it instead of taking time to create it themselves.

Next, metadata ought to provide consistency, another thing that helps searchers. This is achieved through what we in the cataloging commu-

nity call "authority control" and it basically means using the same heading for something every time within the metadata. For example, a single system without authority control might have the same author entered in various ways:

Montgomery, L. M. (Lucy Maud), 1874-1942
Montgomery, Lucy Maud
Montgomery, L. M.
Montgomery, L. M., b. 1874
Montgomery, L. M., 1874-
Montgomery, Lucille, d. 1942
Montgomery, Lucy, 1874-1942
Montgomery, L., n. 1942
MacDonald, Lucy Maud Montgomery, 1874-1942

This isn't a real example; I made it up. What I am trying to show is that without a system that insures consistency, an author's name could appear in many different ways, making it really difficult to collocate things by or about that author and therefore hindering research.

A good metadata system will be closely connected to an authority system that will foster the use of a single heading for any author, subject, etc., and will link, by means of cross references, from variant forms. In this case, the first heading listed is the one used by the Library of Congress and as well by Library and Archives Canada.

WHY FULL-TEXT SEARCHING DOESN'T WORK WELL

Perhaps in the talk after mine some of the problems with full-text searching will be discussed, but I'd be negligent if I didn't mention some of them here. I've already described a couple reasons why full-text searching is inadequate for serious information retrieval. First, there's no system that, without rich metadata, can search full text and create a neat, complete, left-anchored index display that collocates works by subject, title, or author. Also, I mentioned how full-text searching scores poorly on search precision. Let's look at some other reasons why this type of search fails:

False hits: If you want information about "mercury" and you do a full-text search using this term, you will get pages that contain the term *mercury*. But what exactly did you think when I said *mercury*?

Did you think about the planet? The element? The Roman god? The automobile? One of the great things that a controlled vocabulary and a good metadata scheme can do is to separate all these subjects so that they all do not come up in the same set of search results, creating noise. Here's how the Library of Congress does this:

- Mercury (Planet)
- Mercury
- Mercury (Roman deity)
- Mercury automobile

We've all had this problem when searching on Google, I'm sure, you enter a search and get pages that have nothing to do with what you want, but the terms you entered are by coincidence on other Web pages. We searching using words, but what we really want is to look for a concept.

- *Material in a different language or different spelling:* If you enter a term such as "labor movement" with labor spelled l-a-b-o-r, you may miss a lot of good information on that movement because many of the full-text resources use the spelling "l-a-b-o-u-r." Finally, if you enter a term in one language, it's likely that you will miss resources in every other language. A good metadata implementation will solve this problem because it will collocate resources on a given topic regardless of spelling or language.
- *Different term is used:* Now, if you want information on Timbits and include only that term in your full-text search, you will likely miss all the excellent resources on donut holes. Moreover, if you want information on roof gutters, you may miss all the excellent resources on eaves troughs. Here again, a good metadata implementation will solve this problem because it will collocate access to materials on the same topic regardless how they are presented in the resource.
- *Pages where the term under discussion is not used:* Sometimes full-text resources describe a topic and fail to explicitly state that topic in a way that matches a search. For example, Web pages on Dublin Core may refer to the standard simply as DC, and full-text searches for Dublin Core won't pull up these pages in the results.
- *Term occurs only in a graphic:* If a particular term occurs in an online resource only in a graphic, as often happens with names of corporate bodies, and that term does not occur elsewhere within the resource, it will likely not turn up on a search for that corporate body.

These are just some of the reasons why full-text searching alone is so ineffective for serious resource discovery. There are probably more. As the world of online resources grows bigger by the year, and as more resources are born digitally, these problems will intensify and become even more significant. We need high-quality metadata and a good system to exploit it in order to overcome the inefficiencies of full-text searching.

WHY DUBLIN CORE DOESN'T WORK WELL

I'd like to start out this section with a quote from Michael Gorman:

> The basic concept of metadata is that one can achieve a sufficiency of recall and precision . . . in searching databases without the time-consuming and expensive processes of standardized cataloguing. In other words, something between the free-text searching of search engines (which is quick, cheap, and ineffective) and full cataloguing (which is sometimes slow, labor-intensive, expensive, and highly effective). Like all such efforts to split the difference, metadata ends up being neither one thing nor the other and, consequently, has failed to show success on any scale, which is the touchstone by which all indexing and retrieval systems must be judged.[10]

I think that part of the reason we are all here today is that Dublin Core hasn't worked very well for the Government of Canada, just as Mr. Gorman predicted. The main reason Dublin Core doesn't work well is that it provides too few elements to describe complex resources and because of its separation from content standards.

> After many papers and numerous conferences (a process in which renegade librarians joined), a quasi-standard promoted by OCLC and called the Dublin Core emerged as the shining example of metadata and what it could achieve. The Dublin Core (DC) consists of 15 denotations, each of which has a more or less exact equivalent in the MARC record. As any true cataloguer knows, MARC contains far more than 15 fields and sub-fields, in addition to the information contained in coded fixed fields.[11]

He continues:

> Those who advocate metadata and, implicitly or explicitly, believe that the whole range of bibliographic data can be contained in 15 categories ignore the fact that the MARC formats are not the result of whimsy and the baroque impulses of cataloguers but have evolved to meet the real characteristics of complex documents of all kinds. What we have is a simplistic (in many ways na‹ve) short list of categories that is expected to substitute for cataloguing when put in the hands of non-cataloguers.[12]

What Mr. Gorman is talking about here is the flexibility of the MARC format that makes it suitable for the metadata needs of multiple communities. MARC has built-in extensibility. Most of Dublin Core's extensibility is not built in; rather, it is created at the local level, so the extensible elements are difficult to map to others. He also makes the point that resource description cannot be dumbed-down, like Dublin Core does it, to be effective. Effective resource description by necessity requires a certain amount of complexity and human touch, as well as a suitable carrier for this.

Creator-produced metadata doesn't work well and doesn't happen. Michael also wrote: "The fact is that the use of people without the skills and experience of cataloguers to complete metadata templates will lead, inevitably, to incoherent, unusable databases."[13]

I agree and think that metadata creation should be done by people dedicated exclusively to this task. I have a digital camera and I have some software on my computer at home that organizes the pictures. The software allows for adding metadata for each picture so that I can search or arrange the pictures according to things like date and subject, but the fact is I have never added any of the metadata. I'm too lazy to do it. When I need to find a picture I just scroll around until I find it, a process that takes longer and longer as I accumulate more pictures.

In his Web page entitled "Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia,"[14] Cory Doctorow says, "But info-civilians are remarkably cavalier about their information. Your clueless aunt sends you e-mail with no subject line, half the pages on Geocities are called 'Please title this page' and your boss stores all of his files on his desktop with helpful titles like 'UNTITLED.DOC'."

I think Dublin Core implementations are kind of like this. There is a place to enter the information, but not enough people do it. By the way, there is another thing that Doctorow mentions in his critique of metadata. He calls it "Mission Impossible–Know Thyself." He says, "People are dumb observers of their own behaviors. Entire religions are formed with the goal of helping people understand themselves better; therapists

rake in billions working for this very end."[15] In the context of my talk, this means that a particular government agency is probably not the best one to describe its own Web pages. It probably ought to be done by another government agency that specializes in resource description.

Finally on the subject of author-created metadata, let me share with you what Stuart Weibel, one of the creators of Dublin Core, wrote on this subject in the July/August 2005 issue of *D-Lib Magazine.* He said: "The answer is that almost nobody will spend the time, and probably the majority of those who do are in the business of creating metadata-spam. Creating good quality metadata is challenging, and users are unlikely to have the knowledge or patience to do it very well, let alone fit it into an appropriate context with related resources. Our expectations to the contrary seem touchingly naïve in retrospect."[16]

Another problem with Dublin Core is that there aren't really any good systems available commercially that exploit it, at least none that I am aware of. I know there are Web pages that have search screens that index Dublin Core metadata, but these often don't work well and often have results displays that are difficult to navigate or don't offer the helpful left-anchored index displays I described earlier. Moreover, some of these search platforms don't even exploit all of the Dublin Core fields. So you end up with an abbreviated set of fields in your schemes and then your search platform doesn't even index some of them–it's a wonder you can find anything at all.

In his article "Digital Libraries at a Crossroads," author Yannis Ioannidis, says, "The field is ready to design and build generic digital library management systems (DLMSs) that will have all the key features that appear fundamental in supporting digital library functionality as it arises in several possible contexts. All specialised functionality should then be developed on top of such systems."[17] I think this means that the digital library field realizes that it needs to create and exploit metadata in a more rigorous way than has been done up to now. Libraries have systems called integrated library systems that have functionality to acquire, catalog, and share information; digital libraries do indeed need similar systems, especially when it comes to metadata and searching.

Most big Web search engines ignore Dublin Core metadata that is encoded in Web pages. This is because of the already-mentioned tendency of Web page creators to misrepresent themselves to make their Web pages appear more often in search results or more towards the top of a particular result set.

Sometimes the only place that Dublin Core metadata exists is within the document that it itself describes, a situation that is often problematic

and contrary to the idea of discovery. In his article, "Is it Time for a Moratorium on Metadata?" Dick Bulterman says, "Where is it saved? Hopefully not within the data object itself, because then you can only see it once you've already found it."[18] The best metadata implementations have the metadata exist separately from the objects they describe. The metadata functions as a surrogate for and pointer to the digital object. By the way, sometimes metadata that rides along with the content it describes is called "intrinsic metadata," and metadata that is separate from the content it describes is called "extrinsic metadata."

COMPARING DIFFERENT METADATA SCHEMES

Now I'd like to shift a little bit and talk about comparing metadata schemes. Comparing schemes is valuable for several reasons, for example, before an institution implements a scheme, it will have to compare them as part of the selection process. A comparison might also help reveal weaknesses or strengths of one or more schemes. In order to compare different metadata schemes, I think it's essential to decide on the points of comparison one is going to use. And here are the 11 points of comparison I think are the most important:

1. *Level of description/specificity the scheme provides for; ability of the system to describe data in various formats [Web pages, books, etc.].* Here the *specificity* is also often referred to as *granularity*. In the Government of Canada context this may mean, for example, deciding between whether to describe an entire agency in one record or having a record for each of the agency's many Web pages.
2. *Connection to and compatibility with content standards and ability to encode data created according to available content standards, ontologies, etc., in the schemes.* Dublin Core was intentionally designed not to be closely connected to content standards, and this has proven in most cases to be a weakness, I think, because of the great variation among standards and the difficulty of sharing data.
3. *Availability of systems (such as search platforms) that can handle metadata created by the schemes; store the metadata; help create the metadata.* Earlier, I mentioned the DLMSs, the digital library management systems, that are needed to help manage digital libraries. As these systems are developed, they will probably favor some schemes over others. Because of its popularity, Dublin Core will surely be handled by these systems.

4. *Degree of community specificity of the schemes or built-in ability to handle metadata of multiple communities.* A scheme designed for Montgomery, e-commerce probably won't work well for intellectual information, such as that produced by Government of Canada agencies.
5. *Interoperability of the scheme.* Availability of systems for performing crosswalking of data created in the schemes. Also, availability of metadata in a scheme that is available for harvesting. This is related to how widely used a scheme is. The more widely used it is, the more sub-systems will have been developed to add-on to the functionality of the schemes.
6. *Proven success and reputation/popularity of the scheme*. Its future outlook.
7. *Amount of training required for individuals to gain proficiency in creating metadata in the scheme.* This is an area where MARC scores low, I must admit, for it requires a lot of training to be able to encode MARC metadata. Dublin Core, on the other hand, is easier to encode because of its fewer elements of description, so it doesn't require as much training.
8. *Viability of the organization behind the scheme/support network availability.*
9. *Ability of the scheme to accommodate a particular metadata function, e.g., rights management, preservation, discovery, etc.*
10. *Adaptability [extensibility] of the scheme to local needs.* This relates to "community specificity" that I mentioned earlier, but is different in that some metadata schemes can be changed at the local level, such as by adding certain new fields. Sometimes this is also called a particular "flavor" of a scheme.
11. *Scalability*. This refers to how big of a database the system can handle successfully. For example, a scheme with only a few elements of description is not as scalable as a system with many elements because when you have millions of records using a "few-element" scheme, it becomes harder to generate precise search results.

THE MARC MODEL (AND WHY MARC IS SUCCESSFUL)[19]

So now, taking into account these points of comparison, let's look at the MARC model and analyze why it's so successful. I think it's valuable to do this analysis not because I am trying to sell you MARC but because if we can model other schemes after the successful elements of

MARC, it might enhance the other schemes. So, here are 11 reasons why I think MARC is a strong metadata scheme.

1. *MARC has many elements of description*. First of all, MARC provides for a rich, detailed, discrete description. Instead of a single field for author, MARC offers multiple fields, such as personal author, corporate author, conference author, etc. This detail aids in resource discovery and description because it allows systems to separate out different elements according to their attributes and narrow search retrievals and to collocate like elements.
2. *MARC is connected to content standards that provide consistency.* You don't have to use the Library of Congress subject headings and the Library of Congress Classification and AACR2 in MARC, but the scheme is, admittedly, strongly connected to these content standards. It will accommodate others, quite well, though. For example, the 055 field in MARC 21 is entitled "CLASSIFICATION NUMBERS ASSIGNED IN CANADA" and the description says, "A classification or call number that is assigned by the Library and Archives Canada or a contributing Canadian library." So, MARC is designed to simultaneously accommodate multiple content standards and the scheme is always growing and changing to accommodate more.
3. *There are numerous vendors that sell systems that accommodate MARC data.* There is a healthy competitive environment of commercial vendors that make searching systems that are based on MARC data. I say healthy because there are many firms and lots of competition, and this in turn promotes innovation and improvement among the search engines. There are even some that are "open source" which means that the software is available for free.
4. *MARC is used by diverse communities of practice*. The MARC format is used to access intellectual material by a lot of different communities. Though chiefly implemented in libraries, MARC serves as a scheme for all types of materials, from books to online resources to videos, and works with material in virtually all disciplines.
5. *MARC is highly interoperable.* Metadata created according to the MARC standard is highly interoperable. This is because of the specificity of the scheme mainly. But practically speaking, there are many systems available that are able to share and convert MARC metadata to other applications. The scheme's ubiquity enhances its interoperability.

6. *MARC is popular and its future prognosis is strong.* The scheme is used worldwide. Several years ago there were variations of MARC according to country. There was US MARC, CANMARC, and UKMARC, but these have been united into MARC21, which is a reference to the 21st century. This unification of different flavors of MARC has further increased its standardization and made it even more interoperable than it was before.
7. *MARC metadata is generally created by professionals who are experts at metadata creation*. Most people see this as a strength, but others see it as a weakness. It's a strength because having professionals create metadata makes it more consistent and accurate. The people who see it as a weakness point to the higher cost of having professionals create the metadata. Libraries have made the decision that the higher cost is worth it and is better than having inferior or no metadata to describe the resources they make available. Let me say also that libraries do not collect and catalog every single book out there. One of the functions of a library is to select and describe with metadata the resources that the library believes are the ones its users want access to. I think the same ought to be true of metadata implementations elsewhere. You don't have to describe everything.
8. *MARC has a dynamic organization behind it.* The MARC homepage FAQ says: "The Network Development and MARC Standards Office at the Library of Congress and the Standards and the Support Office at the Library and Archives Canada maintain the MARC 21 formats. Input for development is provided by MARC 21 users from around the world, including libraries, library networks and utilities, and library system vendors."[20]
   I would also like to mention something about one of the content standards that is closely connected to MARC. It's called AACR2 and it is in the process of being revised to reflect the changing nature of how information is formatted. The name is being changed to *Resource Description and Access*, or RDA. Significantly, the change is being managed here in Ottawa by Tom Delsey. I know I speak for many of my cataloging colleagues when I say that we are excited about these revisions and look forward to their implementation.
9. *MARC accommodates multiple metadata functions.* Besides the more common discovery function of metadata, people are talking a lot these days about the rights management and preservation functions of metadata. MARC handles these functions very well.

10. *MARC is adaptable to local needs*. One example of the local adaptability of MARC is the call number field in OCLC. It's the 090 field, but this field isn't really defined in the international standard. OCLC made it up. Other bibliographic agencies make up fields as needed, sometimes as an interim measure before a new field is implemented in the standard or to express some local data field.
11. *MARC is scalable.* In terms of scalability, OCLC now has over 60 million records in its WorldCat database. The open source MARC- based systems that I mentioned earlier are often used by individuals to catalog their personal libraries, libraries that have only a few books, so scalability is not an issue with MARC.

So, those are the reasons why I believe MARC is an outstanding metadata scheme. MARC is the most successful metadata scheme ever implemented. I know that not everybody is going to adopt it, but I do see value in looking at its strengths and modeling after them.

MODS

Before I conclude I want to tell you about a relatively new metadata scheme. Its name is yet another acronym, MODS, and that stands for Metadata Object Description Schema. MODS is based on the MARC scheme, but it has three significant differences. First of all, unlike MARC, which uses numerical tags for data elements, such as the 090 field for the call number I mentioned a minute ago, it uses language-based tags. That is to say, instead of having the 090 field, MODS has a field called "local call number." Because it uses language-based tags, one might consider it more user-friendly. Also, MODS is expressed in XML. Moreover, MODS is a slightly simplified version of MARC.

One great advantage of having a scheme expressed in XML is that computer people love it. So, I predict that MODS will only gain in popularity, which is great, because it offers all the advantages that MARC does, only it's in XML.

CONCLUSION

So, in conclusion, I want to leave you with this. The two main reasons why Dublin Core has not been successful are the too few elements of description it offers and the fact that most implementations depend on creator-produced metadata, which just doesn't happen.

I believe the information you create on the Government of Canada Web sites is certainly worthy of a better metadata standard than Dublin Core. Rich metadata expressed in a well-developed scheme combined with a good search platform turns information into knowledge.

Metadata is surely not dead. It is waiting to be created in a quality way so that it will give life to the great information that the Government of Canada agencies present on their Web sites. Metadata is alive!

NOTES

1. The archive of the Metadata Librarians list is on the Web at: http://listserver.dreamhost.com/pipermail/metadatalibrarians-monarchos.com; The archive of the LITA-L list is on the Web at: http://lp-web.ala.org:8000/
2. Priscilla Caplan, e-mail to Lita-L mailing list, October 11, 2004.
3. David Dornan, e-mail to Lita-L mailing list, October 10, 2004.
4. Louise Ratliff, e-mail to Lita-L mailing list, October 11, 2004.
5. Jeffrey Beall, "Dublin Core, an Obituary," *Library Hi Tech News*, 21, no. 8 (2004).
6. Karen Coyle, e-mail to Lita-L mailing list, October 9, 2004.
7. Diane Hillman, e-mail to Lita-L mailing list, October 10, 2004.
8. Elaine Westbrooks, e-mail to Metadata Librarians list, October 26, 2004.
9. Sherry Vellucci, "Metadata and Authority Control," *Library Resources and Technical Services* 44, no. 1 (2000): 34.
10. Michael Gorman, "Authority Control in the Context of Bibliographic Control in the Electronic Environment," *Cataloging & Classification Quarterly* 38, no.3-4 (2004): 15.
11. Gorman, "Authority Control," 15.
12. Gorman, "Authority Control," 16.
13. Gorman, "Authority Control," 19.
14. Cory Doctorow, "Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia." http://www.well.com/doctorow/metacrap.htm
15. Cory Doctorow, "Metacrap."
16. Stuart Weibel, "Border Crossings: Reflections on a Decade of Metadata Consensus Building," *D-Lib Magazine* 11, no. 7/8 (2005), http://www.dlib.org/dlib/july05/weibel/07weibel.html
17. Yannis Ioannidis, "Digital Libraries at a Crossroads," *International Journal on Digital Libraries,* 5: (2005): 256.
18. Dick C.A. Bulterman, "Is it time for a Moratorium on Metadata?" *IEEE Multimedia*, v. 11, no. 4 (2004): 13.
19. For an opposing view of MARC, see: Tennant, Roy, "MARC Must Die," *Library Journal*, v. 127, no 17 (Oct. 15, 2002): 26 and Tennant, Roy, "MARC Exit Strategies," *Library Journal*, v. 127, no. 19 (Nov. 15, 2002):27-8.
20. Library of Congress, Network Development and MARC Standards Office, "Frequently Asked Questions (FAQ)," Library of Congress, http://www.loc.gov/marc/faq.html

doi:10.1300/J123v51n02_05