

Published in Technicalities, Jan./Feb. 2007, v. 27, no. 1

Falling off the Metadata Edge

By Diane Hillmann

I spent a few days in England the beginning of December, attending a small workshop with the title “Contextual Metadata and the Teaching & Learning Context” organized as part of the University of Southampton-led MURLLO Project. The workshop was held in a small, older, very British hotel (they said Cromwell slept there, but didn’t mention whether he recommended it!). I was one of only two non-UK people there, which meant that a host of the unfamiliar acronyms had to be explained to me, but other than that there were few language barriers.

The participants were an intelligent, interesting lot, with varied experiences to share, primarily in what they term the “higher education sector,” in and around funded projects aimed towards the re-use of the ubiquitous “learning objects.” The problems they struggled with are not dissimilar to those encountered in U.S. projects, or anywhere. Most pressingly, the resources proliferate but the funding stream rarely takes the provision of metadata seriously, much less its sustained management over time. In addition, conflict persists between those who advocate highly complex solutions requiring intensive analysis and application of multiple pedagogical terms and the pragmatic minimalists who would be happy if there were more projects identifying intended audiences. Thus, even if there was agreement (which there decidedly is not) about what combination of attributes constitutes a good metadata record, these projects tend to operate in an environment where much is posited and little known.

What there seems to be agreement on, to a large extent, is that they want to be able to use the LOM schemas, based on working decisions of several years standing. They are also interested in Dublin Core, but far less so, given the widespread feeling amongst many embedded educational experts that DC is far too general for their needs. I tend to disagree, but I’ve stopped arguing about it; in the end, there will be crosswalks—good ones, and bad ones—and we’ll all survive quite nicely, I’m sure.

The topics of discussion were interesting ones: What is context? Are all educational attributes contextual by default? What do users really need (and who are they anyway)? What is coming up in the future that might save us from forcing people at gunpoint to fill out forms (and will it come soon enough)? As usual I was far more interested in talking about some concrete things that could be done sooner rather than later, but it was hard to resist the blue sky discussions that kept taking over.

So, why the heck am I spending good column inches on this workshop, when educational metadata is very much a niche interest in most libraries? Good question. My answer is based on what I think is coming to the libraries near you, and I contend that the issues that plague educational metadata are the stalking horses for the ones that will bedevil us, very soon now. Best to be forewarned, lest we be taken by surprise.

Particularly in academic libraries we are increasingly seeing calls for re-examination of the library missions, particularly as it relates to the mission of the larger institution. As part of this there is a new focus on library collections, based on the impact of new technologies on patterns of use. In a recent article for LRTS, Mark Sandler writes:

"Fast forward now to a world where a single digital copy of an article or book can be delivered to multiple users, anytime, anywhere. This is a world in which publishers can deliver in real time the books or articles as needed by users--electronically or in print--rather than libraries or retail booksellers stockpiling the content on consignment; a world in which a user can locate and buy a print copy of almost any known book--new or used--and expect delivery the next day; a world in which a single catalog of books (and non-books) can be searched at the word level, leading users to library holdings and purchase opportunities. This is the world today, or the world that we know to be close at hand. It is potentially a world of disintermediation for libraries of all types, but especially for those research libraries that have historically defined themselves in terms of the extent of holdings rather than the relevance of services." [1]

Sandler, and others looking at the future of library collections, see the current focus on the published products of scholarship, where libraries have traditionally put most of their effort, making way for a new focus on primary research materials. These primary materials are not the product of the scholarly enterprise, but instead the precursor to those products. As more of the secondary products—the books and journal articles—find their way to the web in digital form and are fully indexed by Google, Google Scholar, and other non-library search applications, libraries will be forced to re-examine whether they are spending their “cataloging” dollars effectively. As libraries shift their collection development effort to acquire, manage and preserve these primary materials, they will also need to shift their cataloging dollars from approaches used now on the published products collected redundantly by libraries, to others more appropriate for the wide variety of resources that are the fodder for most academic research.

Most of you reading the points on “primary material” noted above will recognize that what I’m talking about sounds suspiciously like what archivists already do, and that’s one of the interesting things about these visions of library futures. Ironically, archivists went down their own separate path many years ago—they were never very happy with how AACR2 related to their materials, and in response they developed their own cataloging guidelines and formats. While it is certainly true that the strategies developed by archivists to manage their physical assets have been quite effective and efficient, they may not scale well to digital materials, nor be a good basis for the kinds of record sharing likely to be a required part of this new world, where primary materials are the main focus of libraries rather than the interesting margins.

Let me be clear that although I find Sandler’s vision compelling, and have seen strong indications that his vision is shared by other collection developers, I make no claim to a functioning crystal ball. There are certainly other, equally interesting visions of what the future will look like for libraries, but what stays the same is the need for reusable data to drive flexible services. As in the current environment, some of this data must come from others, since we suspect that users will wish to combine "virtually" materials originating from different sources for delivery to users. For this vision to work, we know we will need better infrastructure (servers rather than shelving) and a sustainable and efficient way to describe, manage and preserve these materials.

So where does this fit in to the topic I started with: educational metadata and context? Consider that one of the ways the current archival approach is likely to be problematic for future applications is that it seeks, like traditional cataloging, to describe materials outside a specific context. So, for example, in our current environment, a historic image of the Statue of Liberty would be described in an “objective” manner, and whatever use is made of it in a course, or a book or poster is not considered relevant to that

description, and is not captured in any way by the system. Because it is not considered relevant, the context of use of the image is not available to associate the image to other materials, or provide an indication to a user that the image was considered significant enough to be chosen for a number of purposes (and therefore might be ranked higher in subsequent searches).

In the educational metadata world, it has long been an item of faith that to promote re-use, materials must be described without a context, but the workshop participants had clearly moved beyond that position to one diametrically opposite—context is now considered essential. The interesting question is: why the 180 degree change in viewpoint?

I suspect that one reason is the exponential growth of services like Amazon, whose incredible success can be ascribed in part to its exploitation of usage information, albeit for commercial purposes. When the results of a search for a book on Amazon are displayed, the user sees things like “people who viewed X also viewed Y,” lists of other materials on the same topic supplied by other users, and recommendations for possible additional items of interest. Amazon also makes freely available feeds of information that can be used by other services to provide links to books on Amazon from other pages. All this drives buyers to Amazon, ready to purchase something either new or used from the main site or their vast network of re-sellers.

Consider for a moment how such an approach might work in a library world. The Statue of Liberty image would be linked to the course where it was used (which would provide the context for the image developed by the professor giving the course, and links to his/her other courses), and to the book wherein it was published (if the publication was digital there would be potential for links to text around the image and other text and images). Many of these links would provide far more information about the image than would be possible even in the best hand-crafted metadata. A system that allowed users to take advantage of the additional context available from usage data, might very well be more likely to meet user needs than even the best traditional catalog. Our users are voting with their feet, and we need to start paying more attention to what their behavior means to our mission and priorities.

[1] Sandler, Mark. *Collection Development in the Age Day of Google*, LRTS, Oct. 2006, 50(4) p. 239-243.