I Know Where I Am Going, Do You?

David Bade

Remarks at the ALCTS Serials Section, Continuing Resources Cataloging Committee, Update Forum "Continuing Resources Cataloging: Where in the World Are We Going?" ALA Annual Meeting, Washington, DC, June 25, 2007.

I must begin with the caveat that although I have many years experience creating, using and searching serials information in bibliographies, card catalogs, OPACs and online databases, I have never been involved in any activities involving serials cataloging and control other than the initial creation of a bibliographic record for a serial. In other words, in my cataloging work I have focussed entirely on title level access issues, never on article level access issues.

With that caveat, I gave these remarks the title "I know Where I Am Going, Do You?" in part as a humorous response to the title of this forum, but also for a serious reason: I really do know what I want from a bibliographic record, and I know that without any references to standards and metadata communities. I know where I am going when I create a record because of my engagement in searching various bibliographic information sources for both serials at article and title level, as well as by author and subject. I do that both as a librarian with an item in hand and as a researcher with a topic in mind or a citation in hand. My orientation towards cataloging continuing resources is rooted in that wider experience, and leads me first of all to comment on a few issues that underlie all of the questions we have been asked to address, issues that need to keep in our minds as we think about these questions.

First, serials are neither simple to describe nor to use. They have always been maddeningly complex and unpredictable objects, and online/Internet journals have compounded the problems rather than making them simpler. Barbara Tillett's desire to "Catalog it once and for all" cannot be extended to continuing resources. Attempts to simplify our treatment of such complex objects must not be undertaken on the assumption that simpler means easier or better. Recent work in joint cognitive systems theory has concluded that designing for simplicity is impractical, that instead we should design for whatever level of complexity exists. We need to develop transparent tools that give users control over their tasks rather than prosthetic devices that interpret the world for the user, simplifying that world but thereby depriving the user of control. Doing everything behind the scenes makes things much easier for the user so long as everything works, but it leaves the user powerless when anything fails for whatever reason.

Second, information technologies require not just any standards, but unambiguously defined data elements. For any task that one wishes to automate, all interpretation must be done by human beings at some point beforehand, because information technologies deal with information in only 2 ways: rigorously defined data elements, and algorithmic treatment of probabilities. Whenever accuracy is a goal, probabilities are not good enough.

Third, we cannot forget the reasons for and importance of citation, not only for the user of the library, but for the crucial scientific practice of reproducibility of results. What reproducibility of results is for the experimental sciences, checking the sources is for library based research. With the digitization, repackaging, reformatting, relabeling and revising that are rampant in the creation of electronic resources, these practices may eventually be rendered impossible *for electronic documents*. On the one hand, a digital object identifier is useless for citation purposes if the object lacks stability over time. On the other, a system which gives us a resource and a resource identifier that does not inform the reader/user of the provenance—original source and/or reproduction history—makes citation and research into context problematic. One example: Sarah Thomas' article "Quality in Bibliographic Control" is available in Cornell University's institutional repository. But the article as I found it (via a Google keyword search) had no indication of the author or publication details. I only knew what it was because I had read it previously. Someone who had not read that article previously would have no idea who wrote it when and published it where based on the item from the Cornell Repository document. That article is also available via findarticles.com, but without indication of volume, issue and pagination. I had to go find the publication details from a third search in an EBSCO database, get it again, and compare the three versions to make sure that what I had from EBSCO, the institutional repository and findarticles.com were in fact all the same item. Even then, I did not compare every word to determine if there were differences somewhere in the text. Because the first 2 versions found had incomplete metadata about who, when and where, it was impossible for me to determine whether those three versions were identical reproductions of a single paper or multiple versions without metadata distinguishing them and identifying them chronologically or otherwise. An article level service must have that metadata completely and accurately.

Fourth, any discussion of cooperation and interoperability must take seriously the diversity of purposes. Libraries cannot expect other organizations engaged in other practices in pursuit of other goals to do what we want them to do, even if we bend over backward to make it possible for them to fit into our plans. Technical systems tend to develop autonomously as possibilities reveal themselves; like markets, governments and criminals, technical developments do not follow what users of existing technologies imagine, hope for and plan for.

Fifth, implicit in all the questions is that the practices involved will all be automated processes. This means that the multiple imagined uses of cataloging/metadata are assumed to have no importance for the creation of the metadata. Given the proper packaging—whether Dublin Core, MARC, ONYX, XML or DHL—everyone can use the same data so it need only be created once. While it may seem evident enough that there is only one way to describe such things as pagination in a book, the key issue is rather that both what information one includes or excludes (e.g. pagination) and the form in which that information is recorded are always dependent upon the purposes for which it is recorded. In an environment of automated data exchange, either we have to be satisfied with whatever we do and do not get, no matter for whom or for what purposes it was

created, or we have to have a human being examine, evaluate and adapt that information to our particular purposes.

With the above assumptions and orientations, let me now tackle the three questions.

Question 1. Do libraries need to change their cataloging/metadata practice to accomodate the needs of other communities? I felt it necessary to answer this question by changing it. A better question, one which I feel comfortable answering, would be: How do libraries need to change their cataloging/metadata practices to accomodate their own needs? The reason I rephrase it is that it seems clear to me that we need to change practices, but not at all clear how far we must accomodate the purposes of other organizations doing other things for other purposes. The motivation for the original question is that we want organizations external to the library to do certain things for us, i.e. we want to outsource certain operations.

We can outsource certain operations either because we do not want to do them or because a service appears which could enable us to do things that we would not otherwise be able to do. Altering the ways we do things to accomodate newly available services or to take advantage of enhancements that may come along is something we should not shy away from, but we need to look carefully at what this will mean for all other aspects of our work—and our patrons practices—before inadvertantly crippling practices in one area in our eagerness to enable something else.

Our current systems, structures and standards have been developed to enable us to do what WE need to do, and they are therefore not necessarily suited to other users with other tasks and goals. For example, MARC was designed for libraries in order to enable users to discriminate between data elements so these could be searched individually or in combinations. The production of a complete MARC record therefore requires the creator of that record to interpret the document described in terms of those defined elements and to code them properly. It is that labor of interpretation and detailed encoding that gives databases such as OCLC their extraordinary powers of refinement in searching. That labor costs and costs dearly, but without it we would have records which humans could interpret but machines could only treat probabilistically. That difference is evident when we compare the results we get in Google versus the results in GoogleBooks, since the latter uses metadata, not simply probabilistic juggling of the universe of information available.

It is obvious that other metadata communities can and already do use MARC metadata, but it is just as obvious that there is no reason that they need to or will continue to do so if they find or develop some system that serves *their* needs more cheaply or better. And this will continue to be the case, no matter what cataloging/metadata systems we develop or adopt in the future. We need standards that serve our purposes and techniques and organizational structures to deal with other metadata communities with their purposes, practices, structures and standards, as they come and go, and insofar as we can efficiently and effectively benefit from that orientation of our efforts.

Finally, there are many pathways to failure. One sure way is to forget about what it is we are trying to do and try to satisfy the needs of someone or everyone else. Cooperation works when those cooperating have a clear understanding of a common goal. Hafter mentioned the observation made by two Canadian librarians in 1979 that the original objective of library automation was "to harness technology for the benefit of the user" but that this "was soon replaced by the substitute and less complex objective to apply technological innovation to the manipulation of bibliographical records." Our cataloging/metadata practices must serve our purposes, and whatever changes we make in our practices, they must continue to serve all OUR needs, for all our materials in all formats. If we need the discriminative power of a MARC record, then we should abandon it only if something more useful comes along, and in any case we should be very reluctant to alter our practices to incorporate metadata that lacks that power of discrimination. If we do not need that—and some libraries clearly do not—then simpler, cheaper systems will be adequate *for those libraries*.

Question 2. Should continuing resources cataloging standards change to accommodate greater use of non-library data? Again, the wrong question. Let us ask instead: What is the best way in which we can use non-library data to accomplish our goals? The problem with the original question is that the assumption is that non-library data fits the library's needs and purposes and it is the library that needs to change, not the data. There is a lot of non-library created metadata which I use daily. I regularly search online bookstores and databases (as well as print materials) for information on past or current issues of serials in attempts to interpret numbering (does 3 mean that it is the 3rd year of an annual and this one is for the year 1996, or is it the 3rd issue of the year 1996, following 1996 numbers 1 and 2?), dates, title changes and all sorts of other information. One thing I find is that the information is often not there, not encoded or not disambiguated, often it differs in different sources, and consequently is sometimes simply wrong. Those sources are important for me whether as confirmations or disconfirmations of my own interpretations, clear indications that insufficient information is available, or additions to my knowledge, but in every case I have to interpret what I find and evaluate it, as well as copy and paste or rekey it into my own MARC record in OCLC. In this respect the only thing that has changed from the days of print only information sources is the possibility of entering the information into the record in ways other than simply typing it all myself. To suppose that we could rely on metadata harvesting if only our metadata schemes were compatible is to assume that both interpretation and evaluation are unnecessary. That is only possible in an operation in which something, anything, is good enough, that is, in certain acquisitions, warehousing or circulation processes in which merely filling in the blanks permits the performance of the desired activity.

There certainly are and in the future will be even more ways in which we can use externally generated metadata for our purposes. But for any automated use of such metadata for our purposes, we must first guarantee that the content—not just the metadata scheme—is fit for our purposes.

Question 3. I am in a somewhat better position to comment on the CONSER Standard Record since I was a participant in the pilot project when it was called the Access Level

Record. Some of you may have read my brief but harsh criticism of that pilot in which I stated that I was very positively disposed to any standard that focussed on access, but that such an approach would necessarily focus on exactly those elements that require the greatest degree of interpretation and therefore the most costly activities. I find the standard viewed in itself, apart from all issues of implementation, to be laudable. In particular the insistance in the documentation that the cataloger is free at any point to intelligently adapt the guidelines to the specific resource being cataloged. There are not that many non-required fields identified, and the guidelines state that if the patrons would indeed benefit from the addition of a non-required field, catalogers should feel free to add that information. I myself intend to create and add more 130, 245 and 246 information than either the current or former standards suggested, for the simple reason that more information in those fields is not only helpful, but often crucial for me when searching and interpreting the records that I find.

One problem apparent in the standard itself is the conflict between the two goals indicated in this third question: the standard is intended to reduce redundancy and create less cluttered, easier to read public display records. There is actually very little redundancy in a MARC record. There are text based display fields that communicate to the human users, and there are coded and controlled fields for machine manipulation. That is not redundancy, they are in each case information created for different users. Most of the elements relegated to Not Required status in the standard are either there for machine manipulation and the patron display element is retained, or vice versa. Any time we delete the controlled, structured elements that the software reads, we are in for trouble in all future technical enhancements. Any time we delete the information available to the patron, we leave the patron in a state of ignorance. We can do either or both, and if indeed neither the machine nor the patron needs a particular element, let's forget it. But in every case let us be very clear about what we are doing and the consequences.

Unfortunately, standards can never be understood in and by themselves. They will be implemented as matters of policy. They will be implemented in an environment in which cost is a factor, and in the minds of many administrators, that is the factor which overrides all others. It was precisely in that environment and primarily for that reason that the CONSER Standard was proposed and implemented. From that origin, we can predict with a considerable degree of certainty the kinds of problems which the implementation of this standard will entail.

I doubt very much that in practice any standard focussing on subject and other access points will save time since these are the very elements of each record which require the most thought and take the most time to determine and create authorized headings for. If catalogers really do focus on access, then there will be an insignificant savings of time and money on the whole. Any reduction in the descriptive information will likely make the initial activities of searching for copy more time consuming in the long run anyway. My real fear, however, is that the CONSER Standard Record has been implemented solely in the interests of saving time and money with all the reference to access being simply propaganda. Very clear indications of that were evident in the report of the pilot

project. Abandoning the name "Access Level" also strongly suggested that access was emphasized in order to market the standard, not in order to guide its use.

One common theme in the literature of high reliability organizations is that quality assurance programs are often adopted and implemented for reasons other than a concern with quality. Usually those other reasons—prestige, following the trend, increasing managerial control over employees, public relations moves in the face of failure—guarantee that the program will fail in its ostensible purpose. If the CONSER Standard is implemented in institutions in the same manner as PCC was implemented in some places—as a managerial demand rather than an option available for the cataloger's discretionary use—then we can expect short term savings of time and money and long term reduction in access.

A policy implemented for reasons of cost will predictably enforce the least costly manner of implementation. In the creation of these records, the lowest allowable level of interpretation and information creation will become the maximum allowed. Not Required will become Not Allowed. At least one subject heading required will become Only one subject heading allowed. In the reuse of these records by institutions other than the creators, the policy will be "It is a CONSER Standard Record" so accept it as it is. Instead of being a record guaranteeing a certain minimum it will be treated as a record that must not be touched. The technical operations of exporting and importing will be matched organizationally by policies that insist that the cataloger see no evil and do no good.

If we maintain a dedication to serving the needs of the principal users of our bibliographic data, we can create metadata that others can harvest should they choose to do so, we can use metadata created elsewhere, adapting it to fit our purposes, and we can use the CONSER Standard Record, interpreted, modified, adapted and sometimes ignored, in our efforts to provide library users with the tools they need to perform the tasks in which they are engaged. If on the other hand, we get confused about what we are doing, for whom, and why, we may end up creating metadata that no one anywhere finds useful, and upon discovering that, we will once again be asking ourselves "Where in the world are we going?"