BEYOND USMARC AMC: THE CONTEXT OF A DATA EXCHANGE FORMAT

JILL TATEM

ABSTRACT: Archivists' discussions about use of the USMARC AMC format so far have failed adequately to address goals and system design and implementation issues. The article focuses on one commonly articulated goal, improved end-user access to archival collections. It examines the issues of data quantity, data quality, and user-system interfaces and concludes that unresolved problems in all three areas present significant obstacles to end-user access to archival collections via current bibliographic descriptive networks.

The archival literature in recent years has been full of praise and promise for the USMARC AMC (Archival and Manuscripts Control) format. Archivists have been told that its use is inevitable[1] and that it is the best news for archivists since acid-free folders.[2] Many of these publications have done an excellent job of explaining what USMARC AMC is and how and why it came about. What has been missing is discussion of what USMARC AMC use can do (and, equally important, cannot do), what archivists want it to do, and how to ensure that the ways archivists use the format further professional goals.

Almost without exception, the reason for using the USMARC AMC format given in the archival literature to date has been the opportunity to become part of the larger information world.[3] Unfortunately, all too often the advantages of participation in this vast information community are assumed—as if they were so obvious as not to need enumeration. Conversely, the disadvantages are brushed aside as being too trivial for discussion among right-thinking and progressive archivists.

If archivists' discussion of goals has been superficial, exploration of system design and implementation issues has been virtually nonexistent. In the absence of discussion of other elements of information systems in which the format is used, the implication is that these undefined advantages will be ours as a necessary consequence of USMARC AMC use. If archivists are to exploit USMARC AMC fully and creatively they must begin a more critical and rigorous examination of their goals, and issues of system design and implementation.

The work of SAA's National Information Systems Task Force (NISTF) which led to the development of the USMARC AMC format was a remarkable achievement. With respect to professional goals one of NISTF's realizations was that "...a national information system probably would be justified on the

basis of its usefulness to archival organizations rather than to end users..."[4] The discussion in the archival literature since this statement appeared, however, has drifted away from organizational utility as justification for use of the format. USMARC proponents today consistently justify the format by referring to improved access to archival materials for end-users.[5] Since this has become the single most widely articulated goal, it is reasonable to ask if using USMARC AMC makes it easier for end-users to find archival collections when they ask, "Where can I find information about XYZ?"

Asking this question makes obvious the significance of archivists' failure to examine system components other than the format. There has been considerable confusion about the distinctions between the USMARC AMC format, the data recorded using the format, and manipulation of the data. However, it is the entire system (format, vocabulary, descriptive conventions, policies on descriptive levels, depth of subject analysis, the user interface, and other elements) that governs what end-users discover about our collections. Thus, benefits to be derived from use of the USMARC AMC format can not usefully be discussed without including the other components of information systems of which USMARC AMC is one part. The question becomes whether use of the USMARC AMC format as currently implemented within bibliographic networks and in-house systems makes it easier for end-users to find archival collections relevant to their information needs.

Three broad issues seem pertinent to that question: 1) Access: how much information about archival records is available in online catalogs and how readily available is it; 2) Data Quality: is it clear what entities are being described and are the analysis and vocabulary used to construct those descriptions understandable; 3) User Interface: how difficult is it for users to extract meaningful information from the system. Each of these issues, in turn, presents even more questions.

Access

In determining how much information exists, the most obvious measure is the number of series/collections that have been cataloged. A variety of published reports give both totals and an idea of trends. One early report claimed 47 repositories had entered 70,000 bibliographic records into the RLIN database as of 1 August 1986.[6] Another noted 45 repositories and 86,000 AMC records in RLIN in May 1987.[7] As of December 1988 the RLIN database included 160,000 AMC records.[8] OCLC as of August 1986 was cited as having 50,000 AMC records[9] and in July 1988, 80,000.[10] The Western Library Network, which implemented the AMC format less than a year ago, has around 500 records.[11] The rough total is nearly a quarter of a million records.

These figures only detail records in national databases. Micro-computer systems using software such as Cactus's Minaret and Michigan State's MicroMARC:amc and regional networks may boost the totals. In that regard, another interesting figure comes from the SAA automation survey of November 1987 that 70 repositories were currently using the USMARC AMC format and another 43 planned to do so within one year.[12] Assuming they all have done so, a questionable assumption, that represents 113 repositories out of 261 survey respondents.

For comparison, as of the 1985 issue, *The National Union Catalog of Manuscript Collections* (NUCMC) had described 56,435 collections in 1,321 repositories.[13] These figures are, at best, suggestive, but what they seem to indicate is that more records in fewer repositories are in online catalogs than in the manual equivalent of a national database. However, because NUCMC is now entering records into the RLIN database, this may be a distinction without a difference. What is encouraging is the speed with which this database is being constructed. Considering that the AMC format was not published until 1984 and that additional time was necessary to prepare for its use and actually create bibliographic records, this growth is a remarkable achievement.

A corollary question needs to be asked. At what level are those bibliographic records available—national, state, regional, institutional, repository? In practice, if end-users can only access the records cataloged in their home institutions it does them little good that many thousands more are out there somewhere. If access is possible, what practical obstacles stand between end-users and the records? Are the records, for example, available in an online public access catalog (OPAC) searchable at little or no cost? Can users search the catalog from their homes or offices or do they have to go to the repository? Or do they read a print out or catalog cards?

Even if end-users are affiliated with an institution that is a member of one of the national networks can they search that database? Many public catalogs are part of turnkey systems that allow access only to the bibliographic records of a single institution. If end-users want to search the union catalog of a bibliographic utility they must rely on a librarian intermediary. In any case, researchers almost always need some institutional affiliation because, as individuals, there is nothing available to them similar to BRS's AfterDark or Dialog's Knowledge Index.[14]

Because access is by institution, the membership figures of the national utilities might be instructive. As of October 1988, OCLC membership was 9,400, RLIN reported 99 members, the Western Library Network included 300 members[15].

Data Quality

Once end-users get access to archival bibliographic records, do the conventions and standards archivists use produce descriptions that make sense to them? Since descriptive cataloging data largely rest on application of the *Anglo-American Cataloguing Rules (AACR 2),* the effectiveness of those rules in producing descriptions intelligible to end-users is critical. Unfortunately, there is a substantial body of library literature devoted to attacking *AACR 2*'s effectiveness. As archivists become more experienced with *AACR 2* they will, undoubtedly, join whole-heartedly in pointing out its flaws. For example, one popular complaint revolves around rules 24.12 through 24.14 and what they do to institutional names. University archivists have discovered that the bizarre consequence of consistent adherence to the rules is that, within one university, some schools are entered directly under their own names while other schools are entered as subunits of the university. Those trained in cataloging can understand and sympathize with the reasons for these rules and the principles behind them. But most researchers have never even attended SAA's workshop on

library descriptive standards let alone semester-long graduate classes in cataloging. To them the results of following these rules must appear the cheapest and most self-indulgent form of self-aggrandizement.

Equally important are local descriptive policies. Are archivists describing material at a consistent level, i.e. record group, subgroup, series, etc.? If archivists are not consistent, are the principles guiding archival descriptive level choices intelligible? Do archivists know why cataloging is sometimes at the series level and at other times at the record group level or does it just "feel right"? If it is the latter does it also feel right to researchers? Alternatively, if archivists are consistent—for example, series-level and only series-level—was this policy reached because the records dictated this approach or is it a matter of administrative convenience?

Also important is the question of how much information archivists give users. Are end-users discovering five screens of biographical notes and, if so, should they? Do end-users ever want to know this much and, if so, at what point in their search? At the other extreme, are archivists mimicking the librarian's superficial subject analysis and assigning the equivalent of the library's average 1.4 subject entries per record? What fields and subfields are used? While it is a triumphant reaffirmation of the obvious it should be noted that if information is not part of a record researchers can not use it to find the record.

Is the vocabulary archivists use understandable to users? Among librarians, complaining about *AACR 2* is becoming second in popularity only to the tradition of Library of Congress Subject Headings-bashing. It is an understandable tradition. Who wants to explain to researchers why the catalog contains the heading *Military Education* but not engineering education because that is called *Engineering—Study And Teaching*?

User Interface

Assuming that bibliographic records are available through an OPAC, does the system's manipulation of the database help or hinder end-user retrieval? What fields and subfields are indexed and are they indexed in one file or separately? How many characters in the field are indexed? Can the user specify what data elements of the record to display? Are authority records available to users to guide them through the vocabulary? Are there suggestive error messages or are users simply faced with the discouraging "No records retrieved"? Can the user review the session's search history? Are printers available? Are truncation, stemming and soundex searches available? What about keyword, Boolean, phrase searching and adjacency? What are we doing to help the browser, who, to paraphrase Justice Potter Stewart, can't define it, but will know it when she sees it?

Answers to most of these questions of access, data quality, and interface suggest that there are significant barriers for end-users to overcome before they find archival collections. The current distribution system divides the universe of machine-readable catalog records among multiple bibliographic networks and local catalogs, effectively hiding many records from specific users' reach. Recent studies of how people use OPAC's have concluded that they like using them even though they do not perform effective searches. Frequently discussed obstacles to effective end-user searching include inability to formu-

late search strategies, misunderstanding of Boolean logic, and ignorance of the subject vocabulary used by librarians. Simply put, users do not understand either library descriptive practices or the existing capabilities of OPAC's. The current generation of catalogs does little to assist users in navigating even the limited capabilities available. Many users do not even understand what entities compose the universe described in online catalogs.[16]

All of which suggests that there is little reason to conclude that end-users are better off today than they were in 1982, when USMARC AMC was still a glimmer in the eye of the National Information Systems Task Force. There is a large mass of information to find. But archivists have not done enough to help users find it. Viewed in one way this is a continuation of the long-standing and often justifiable criticisms of archivists' idiosyncratic manual practices. The profession is simply using fancier tools—computers instead of card catalogs and registers—to achieve the same inadequate results. Archivists have not taken advantage of the opportunities presented by a structure for archival information exchange because still, eleven years after NISTF was formed, the basic questions have not been answered.

The first question is, what do archivists want these catalogs to do and who should they serve? Are they finding aids or "electronic pathfinders" to finding aids? Are they an outreach mechanism to alert potential users to the existence of collections or are they comprehensive management information systems? With whom do archivists intend to exchange information—other archival repositories, records creators, or researchers? As a beginning effort to answer these questions it would be instructive to survey institutions that are USMARC users to discover why they chose this approach, if their expectations have been met, and if not, what unmet needs they have identified. Information about how many and what types of institutions are AMC users would also be useful, as would reasons for nonusers' disinterest.

The second question is, what do specific types of users want these catalogs to do. Although this article discusses users as if they all had identical information needs, it is unlikely that they do. Calling for user studies is not an original thought. If, however, archivists are going to embark on the expensive process of developing online catalogs in order to assist end-users in discovering archival materials it is imperative to discover what users want to know. If, as seems probable, different users have different expectations then archivists will have to develop mechanisms to mediate conflicting needs.

The third question is, if there is no exact match between the expectations of catalog creators and catalog users, where and how is compromise reached. Contained within this question is the issue of how over-burdened archival staffs, many of whom lack experience and training in the theory and techniques of bibliographic description, are to add to or convert finding aids into the AMC format.[17] If archivists do not want to see use of the USMARC AMC format become a dividing line between the have's and have-not's of the archival community, then the profession must act to make online catalogs appealing and affordable to all institutions. Who will benefit from archival participation in the wider information community and who will bear the costs? Are the economic incentives of sharing resources such as cataloging data and holdings exchanged through interlibrary loan that were responsible for the growth of the bibliographic networks in the 1970s relevant to archives?[18]

Avra Michelson's article in the *American Archivist* has alerted archivists to a second facet of the issue of expectations: the problem of indexer inconsistency.[19] The most surprising aspect of her research is how surprised people have been at the results. Indexer inconsistency is not a new problem and there is no reason to expect that archivists would be immune to a disease that has plagued all other segments of the information business.

Prescriptive indexing literature, however, begins with the instruction that the nature of the collection and the information needs of the users must be the foundation of the index.[20] At the risk of being labelled an "idiosyncratist", this prescription suggests that describing different collections for different users will always entail certain differences in approach to analysis and vocabulary. Consistency, if it is the result of assigning descriptors regardless of user or collection context, may be neither possible nor desirable, particularly if it results in a smaller descriptive vocabulary. Rather than achieve profession-wide consistency, all archivists may be able to do is to explain what we are indexing and how. If, however, archivists can make local practices intelligible to end-users as well as to each other, they will have achieved something very helpful in bridging the gap between users and creators.

After lengthy deliberations, ample opportunity for comment by archivists, and substantial modification, archivists have adopted library structures for data exchange. As a result of this process, the USMARC AMC format is a significantly different animal from its predecessor manuscript format. The success of the current format is in large part a result of the process that produced it.

> The Task Force sanctioned the definition of existing practice with a consistent vocabulary, which it would then take as the "standard" and use in its design of national information systems....The process was open and open-ended and...actual data already in use was taken as a given.[21]

A similar process must be applied to the issues of descriptive practice, for analysis and vocabulary and for delivery systems. This would be less essential if existing library practice and systems met user expectations and could be adopted by archivists with minor modifications. Unfortunately, there is ample evidence that library practice and systems do not meet user expectations.

Despite the many questions to be answered, there is reason to be optimistic that archivists can address these issues successfully. In the United States, the Working Group on Archival Descriptive Standards[22] is examining the process by which the profession develops and adopts standards. The Government Records Project[23] continues to explore imaginative new uses of USMARC-based data as have several participants in the Bentley Library's Research Fellowship Program. The Canadian archival community also is exploring descriptive standards. *Archives, Personal Papers, and Manuscripts*[24], the embodiment of current archival bibliographic descriptive conventions, is being revised after only five years. Both the revision and the first edition incorporate substantial necessary modifications of the original *AACR 2* rules for manuscript cataloging. Beyond these projects, there is a vast body of library and information science literature on end-user searching of bibliographic databases and OPAC's from which to learn.

There remains, however, much more to do. Archivists must inform themselves about how allied professions have addressed similar questions and what

conclusions they have drawn. Those conclusions must be examined in light of archival experience and, if the conclusions do not hold up, archivists must persuasively discuss their needs and make original contributions. This need is even more important because of the recent integration of USMARC formats.

Archivists bring different, valuable, and heterogeneous perspectives to this process. The overwhelming portion of archival collection descriptions are not tied to LCSH, *AACR 2*, or LC and Dewey classification. Archivists do not carry the baggage of decades of copy cataloging. Administrative histories and biographical notes are a resource like nothing available in traditional bibliographic systems. Traditional hierarchical descriptions are a similar resource with enormous potential for aiding users in navigating catalogs.[25] It would be wrong to lose sight of the NISTF conclusion that "a national information system for archives and manuscripts collections derived from the intellectual resources of the archival profession...will, by the process used to create it, maximize its acceptance by the profession."[26]

Advice often repeated to systems analysts is the caution against "assuming the environment." It is a pitfall that NISTF avoided. The same cannot be said for subsequent developments. Presented with a vehicle for exchanging information, archivists have not adequately addressed the reasons to exchange information. Presented with available library descriptive standards, archivists have not adequately addressed archival needs for descriptive standards. Presented with a ready-made subject vocabulary, archivists have not adequately addressed archival needs for subject, function, form, and other vocabularies. Presented with ready-made delivery systems, archivists have not adequately addressed archival or user output requirements.

This article is not an attempt to argue against use of the USMARC AMC format. It is an attempt to argue for thoughtful use, for a rigorous analysis of goals and a creative reflection on possibilities, at both repository and professional levels. If these steps are neglected, the USMARC AMC format will be both an irrelevance and a failed opportunity.

ABOUT THE AUTHOR: Jill Tatem is presently assistant university archivist at Case Western Reserve University. She holds an MA in American history from the University of Dayton and an MSLS from Case Western Reserve University.

NOTES

1. Frederick L. Honhart and Richard L. Pifer, "Has the MARC AMC Format Really Changed the Archival Profession?" papers presented at the Midwest Archives Conference, Cincinnati, Ohio, 5 November 1988.
2. Steven L. Hensen, "The Use of Standards in the Application of the AMC Format," *American Archivist* 49 (Winter 1986): 32.
3. See Lisa Weber, "Archival Automation: the MARC AMC Format," *SAA Newsletter* (May 1987): 13; Katharine D. Morton, "The MARC Formats: An Overview," *American Archivist* 49 (Winter 1986): 28; Nancy A. Sahli, "Interpretation and Application of the AMC Format," *American Archivist* 49 (Winter 1986): 10, 12; Hensen,"The Use of Standards," 32. Notable exceptions are Margaret J. Kimball, "Workflow for Processing Manuscripts in Automated Systems," *Rare Books and Manuscripts Librarianship* 1 (Fall 1986): 117-126; David Bearman, "Why Adopt MARC?" *Archival Information Newsletter* 1 (Spring 1987): 3.

4. Richard H. Lytle, "An Analysis of the Work of the National Information Systems Task Force," *American Archivist* 47 (Fall 1984): 361.
5. David Bearman once again articulates a different approach in "Archives and Manuscript Control with Bibliographic Utilities: Challenges and Opportunities," *American Archivist* 52 (Winter 1989): 26-39; David Bearman and Richard Szary, "Beyond Authorized Headings: Authorities as Reference Files in a Multi-Disciplinary Setting," *Authority Control Symposium,* Occasional Papers of the Art Libraries Society of North America, no. 6 (1986): 69-78.
6. H. Thomas Hickerson, "Archival Information Exchange and the Role of Bibliographic Networks," *Library Trends* 36 (Winter 1988): 553.
7. Lisa Weber, "Automation Survey Results," SAA Newsletter (November 1987): 4.
8. Gregory Whitfield, RLIN, to Jill Tatem, personal communication, 15 December 1988.
9. Hickerson, "Archival Information Exchange," 553.
10. "OCLC Statistics by Format," *Bits and Pieces* (June/July 1988): 12.
11. Dave Wasser, WLN, to Jill Tatem, personal communication, 28 October 1988.
12. Weber, "Automation Survey," 5.
13. *National Union Catalog of Manuscript Collections* (Washington, D.C.: Library of Congress, 1988), iii.
14. These are two of the many subscriber services which offer access to online databases geared to non-professionals. Evening service, economical costs, and a relatively simple user interface facilitate independent, direct, end-user searching.
15. *Furthering Access to the World's Information: 1987/88 OCLC Annual Report* (Dublin, Ohio: OCLC, 1988); Whitfield; Wasser. Membership figures can be misleading both for comparative purposes and as indicators of the availability of union catalogs since most utilities have varying levels of membership and distribute products beyond their membership.
16. There have been dozens of OPAC user studies reported in the library literature, among them several summaries of the Council on Library Resources 1981-82 Online Catalog Evaluation Projects such as Joseph R. Matthews, Gary S. Lawrence, and Douglas K. Ferguson, eds., *Using Online Catalogs* (New York: Neal-Schuman, 1983); Joseph R. Matthews, ed., *The Impact of Online Catalogs* (New York: Neal-Schuman, 1986); *Subject Access: Report of a Meeting Sponsored by the Council on Library Resources* (Washington, D.C.: CLR, 1982); and Robert N. Broadus, "Online Catalogs and Their Users," *College and Research Libraries* 44 (November 1983): 458-467. Another project was OCLC's Subject Access Project reported by Karen Markey, *The Process of Subject Searching in the Library Catalog: Final Report of the Subject Access Research Project* (Dublin, Ohio: OCLC, 1983) and Karen Markey, *Subject Searching in Library Catalogs Before and After the Introduction of Online Catalogs* (Dublin, Ohio: OCLC, 1984). The Research Libraries Group recent Patron Access Project is reported in Walt Crawford, *Patron Access: Issues for Online Catalogs* (Boston: G.K. Hall, 1987). Two recent reviews are Christine L. Borgman, "Why are Online Catalogs Hard to Use? Lessons Learned from Information-Retrieval Studies," *JASIS* 37 (November 1986): 387-400; Charles R. Hildreth, "Online Public Access Catalogs," *ARIST* 20 (1985): 233-285. For one description of needed capabilities in OPAC's see Charles R. Hildreth, "Beyond Boolean: Designing the Next Generation of Online Catalogs," *Library Trends* 36 (Spring 1987): 647-667.
17. For a discussion of the labor required to create an AMC record see Patricia Cloud, "RLIN, AMC and Retrospective Conversion: A Case Study," *Midwestern Archivist* 11 (1986): 125-134.
18. Richard De Gennaro, "Library Automation and Networking: Perspectives on Three Decades," *Library Journal* (1 April 1983): 631; Charles R. Hildreth, "Library Networking in North America in the 1980's. Part 1: The Dreams; the Realities," *Electronic Library* 5 (August 1987): 226.
19. Avra Michelson, "Description and Reference in the Age of Automation," *American Archivist* 50 (Spring 1987): 192-208.
20. See Donald B. Cleveland and Ana D. Cleveland, *Introduction to Indexing and Abstracting* (Littleton, Colo.: Libraries Unlimited, 1983): 67; John Rothman, "Index, Indexer, Indexing," in *Encyclopedia of Library and Information Science* (New York: Marcel Dekker, 1974); F.W. Lancaster, *Vocabulary Control for Information Retrieval,* 2nd ed. (Arlington, Va.: Information Resources Press, 1986): 24, 26; Lois Mai Chan, *Library of Congress Subject Headings: Principles and Application,* 2nd ed. (Littleton, Colo.: Libraries Unlimited, 1986): 17-20.
21. David Bearman, *Towards National Information Systems for Archives and Manuscript Repositories: The National Information Systems Task Force Papers 1981-1984* (Chicago: society of American Archivists, 1987), 5.

22. The Working Group on Standards for Archival Description was established in 1988 with funding from the National Historical Publications and Records Commission (NHPRC) to explore the need for standardized descriptive practices and procedures for developing or evaluating such standards. A final report was in preparation in the summer of 1989.
23. The Government Records Project, also with funding from NHPRC, is exploring, among other things, the development of form-of-material/series authority files to be used through the Research Libraries Information Network in intergovernmental identification and appraisal of archives. The Government Records Project is an expansion of the Seven States Project which sought to share, also through RLIN, records about holdings, functional descriptions, etc.
24. Steven L. Hensen *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries* (Washington, D.C. Library of Congress, 1983). A second edition is in preparation.
25. David Bearman and Richard Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (Winter 1985-86): 14-27.
26. Richard H. Lytle, "Report to SAA Council, March 1978," quoted in Bearman, *Towards National Information Systems,* 3.