Research Articles

Archives and Manuscript Control with Bibliographic Utilities: Challenges and Opportunities

DAVID BEARMAN

Abstract: The MARC Format for Archives and Manuscript Control (MARC-AMC) was designed with the potential for recording a broad range of information and contains structures for a variety of implementations. These hidden potentials are being explored in its implementation within national bibliographic networks, especially within the Research Libraries Information Network (RLIN) by the Seven States Project. The AMC format has the potential both to support successful automation of archival control and to transform the bibliographic utilities as we know them today. Specifically, AMC overthrows the bibliocentricism, the political hegemony, and the existing financial/functional underpinnings of the bibliographic networks. It opens the ways to multimedia cultural information systems and new local/national systems architectures. As a consequence, archivists are likely to play a critical role in redesigning library information networks in the decades to come.

About the author: *David Bearman is the editor of* Archives and Museum Informatics *and* Archives and Museum Informatics Technical Reports *and is the founder of Archives & Museum Informatics, a consulting firm in Pittsburgh, Pennsylvania. He was previously Deputy Director of the Office of Information Resource Management at the Smithsonian Institution.*

A version of this article was first presented at the annual conference of the National Association of Government Archives and Records Administrators (NAGARA) in Atlanta, Georgia, on 24 July 1987. The author wishes to thank Kathleen Roe for suggestions on an earlier draft and the anonymous reviewers for this journal for their helpful comments on the paper as submitted.

Introduction

ARCHIVAL INFORMATION EXCHANGE DID not find its ideal implementation through the kind of rational planning and testing suggested by the National Information Systems Task Force of the Society of American Archivists in 1981.[1] Instead, existing national bibliographic networks were selected as the vehicles for archival information exchange by default, with the assumption that they could subsequently be shaped by their users. Without doubt, the embracing of the MARC Archives and Manuscript Control format (MARC-AMC) and bibliographic networks that has occurred since 1984 has defined a fundamental strategy for the profession. The question now is, how can archivists manage its long-term ramifications? Since no independent archival information sharing networks were constructed, archivists must now identify ways for bibliographic utilities to carry the data they need without doing violence to the data, the information needs, or the utilities.

Archivists came late to bibliographic networking for which an established base of financial support and political mechanisms supporting that base were already in place. Library economics defined the cataloging of new materials as a community benefit for which no charge was applied. Copy cataloging—or duplicating records made by others—saved local institutional resources and so was billed as a network service. This economic equation was clearly not transferable to archives. Although no one suggested that archival materials would be copy cataloged, the bibliographic utilities were willing to consider archival information exchange in the hopes of realizing a secondary, if marginal, benefit. By including archival and manuscript holdings of their member institutions within their databases, they could support an integrated catalog for member libraries. The costs to the utilities of providing this added benefit would be modest so long as archives and manuscript control could be implemented with few changes to the rest of the system. After all, the number of records anticipated was quite small by comparison with the more than fifteen million bibliographic records maintained by the larger networks.

Partially as a consequence of this last assumption, the alliance between archives and the library technical services community has been very fragile. The senior partner has little interest in or understanding of archivists' needs, and archivists have not yet been able to identify their strategic objectives. Both archivists and the library community fear that the needs of archivists, if acted upon, would undermine the current financial, political, and conceptual basis of the bibliographic networks.[2] Indeed, if all the implications of supporting archival information exchange were perceived the strains would be greater, because the MARC AMC format is a vehicle for revolutionary change in library practice.

This article explores that revolutionary potential, in part by examining the experimental implementation of archives and manuscript control on the Research Libraries Information Network (RLIN) by the Seven States Project (SSP). In 1985, the Research Libraries Group (RLG), and the state archives of Alabama, California, Minnesota, New York, Pennsylvania, Utah and Wisconsin received funding from the National Historical Publications and Rec-

[1]David Bearman, *Towards National Information Systems for Archives and Manuscript Repositories: The National Information Systems Task Force (NISTF) Papers, 1981-1984,* (Chicago, SAA, 1987), 119p.

[2]Since this article was accepted, RLG has established a new basis for charging for RLIN which reflects the dissolution of the copy cataloging foundation of bibliographic utilities and recognizes the new basis for the utility as an information service. Charges from January 1989 will be based on searches of the database rather than processing of records. *Research Libraries Group News* 16 (May 1988): 2–6.

ords Commission (NHPRC) to conduct a study of the benefits of using RLIN to share information about their holdings, and the appraisal activities. An examination of the SSP will illustrate how archivists, by generalizing what they believe must take place for archives to use library networks well, can both serve their own short-term interests in getting the networks to adopt mechanisms suited to archival and manuscript control and provide a platform from which the bibliographic networks of the 1970s and 1980s can become the knowledgebases of the 1990s. This argument could, therefore, be addressed as much to librarians as to archivists, because it is librarians who will need to see a long-term benefit in the kinds of (costly) modifications that archivists will propose as a result of the SSP and similar projects. But first it will be necessary for archivists to recognize and understand the possibilities.

I. Present Systems: What Are Library Bibliographic Networks and What Are Archivists Doing in/to Them?

Much to the surprise of those who tried to plan national bibliographic networks in the 1960s and early 1970s, library networks grew like Topsy and became national without the benefit of national planning.[3] OCLC rode to success on the wings of shared cataloging—a method of distributing the unit costs of using an authority file of bibliographic entries that is particularly cost effective for organizations able to locate in this fashion the cataloging data for almost all the materials they acquire. OCLC was willing to cede to regional networks other cooperative services, including support for bibliographic processing and value-added reprocessing of cataloging and processing data, thus forging an even stronger dependence on OCLC through the enhanced services provided by regional network to their members. Because large research institutions with foreign language and specialized acquisitions and large specialized, unique collections were less able to benefit from shared cataloging, they eventually formed RLG, an organization whose cooperative benefits were not to be limited to shared cataloging. Trends in local systems during the 1970s and early 1980s supported the growth of national networks. In these local systems, catalogers entered MARC formatted data, acquisitions modules depended on ISSN and ISBN identifiers, and public access catalogs were loaded from network-provided tapes.

In the early 1980s, when archivists first broached the possibility of contributing data on archival and manuscript holdings to the RLIN database, RLG was also just beginning to move beyond shared cataloging and the pains of its formative period to explore the benefits of sharing research databases and making cooperative use of processing/cataloging data. It was easy to avoid asking whether the archives database was going to be another bibliographic file, a research database, or a processing tool, and it was to everyone's advantage to beg a question that might have required a hard, financially based, decision.

It was generally accepted that if archival records could be treated as another bibliographic file, such a file would be sufficiently small that cost-recovery issues could be absorbed within the cost recovery for the much larger bibliographic files. If, contrary to projections based on the size of the National Union Catalog of Manuscript Collections (NUCMC) and the NHPRC SPINDEX databases, it grew a great deal, the file would become valuable as a research database (and, it was hoped that cost-recovery issues would then be irrelevant within the context of RLG which was organized precisely to encourage such research data-

[3]David Bearman, "Archival and Bibliographic Information Networks," *Journal of Library Administration* 7 no. 2/3, (Summer/Fall 1986): 99–110.

base development). Even if archivists eventually used such a system heavily for processing control (under the then existing structures for payment), the implementation would be cost-justified because archivists would pay the costs of maintaining the file. Each of these scenarios, however, made forecasts for what became the AMC file as if it would forever be insulated from the rest of RLIN. Few considered what must now be taken seriously, that AMC might be a Trojan horse, or a virus.

The AMC format looked innocent enough to most librarians because in almost all surface respects, it appeared to be in conformity with other MARC formats. But dangers to the status quo were built into several, barely submerged, reefs. The first, and in the long run most important, threat was a conceptual innovation. Archival methods did not respect the boundaries established by bibliographers between physical information formats, each represented by its own MARC format. The MARC-AMC format was created to accommodate the requirements of a community that described cultural materials without considering their physical format as a primary feature.

The second danger arose from the politically novel terms under which the format was adopted. The AMC format, as adopted by the library community, came with a tether to the archival profession.[4] It was to be jointly owned and jointly administered. None of the other formats were restricted in this way, and the full implications of the constraint are still being explored. On the one hand the idea of involving other communities beyond libraries in developing MARC formats has proved infectious, but the nature of the constraint it imposes on librarians, if any, is not clear.[5]

The final danger was that the AMC format had been launched with the full knowledge that it was financially adrift. Its designers built a vessel that could hold bibliographic item descriptions, bibliographic collection descriptions, inventories, organization and person authority records, and novel processing data. Its actual data cargo would be chosen by its owner and would be up for auction for several years. Along with what data would ultimately be exchanged, what functions would be required were unknown. True, these records, as legitimate MARC records, seemed to fit neatly into a category already defined by librarians. But, on closer examination, and as archivists have begun to work out how best to exploit their containerized information transport vehicle, it is evident that the lack of prior description standards provides an opening for archivists to employ the format in ways that are extremely novel, including the creation of pseudo-bibliographic records for forms of material or for record series schedules. In addition archivists are intensively exploiting several areas of the format, such as the processing control data in field 583 and the notes fields, 5xx. It now appears that these practices could become the drivers for economical archival information exchange. If so, AMC could transform library network economics.

Since neither librarians nor archivists are fully aware of these three submerged reefs, or what they might mean, each is analyzed further below and then discussed in the context of the SSP experiment on RLIN.

[4]The meaning of the agreement between the Library of Congress and SAA has been further obfuscated by "format integration," a principle adopted by the advisory committee on the MARC formats, MARBI, at its meeting in July of 1988. For archivists, however, the integration largely represents a triumph of their views since its major implication is to extend collective description and the other features of AMC described in this article to all the other formats.

[5]Library of Congress, MARC Standards and Network Development Office, "Format Integration Considerations," Discussion paper #16 (Washington DC, 15 May 1987) recognized this political constraint. It is worth noting, in the summer of 1988, that the political agreement struck in 1983 may have been undone by format integration, but it remains too soon to say.

A. Exposing Conventions. To name is to differentiate. To differentiate is to bring something into existence from an inchoate mass. In this fashion, the archives community named the library conventions embedded within MARC, differentiating them from its own, and thereby giving them birth as independent realities. Implicit within library cataloging practice, they became bound up in the MARC formats, and their implications there had previously gone unrecognized.

The AMC format explicitly named and differentiated its practices from those of libraries in a number of ways. By stating that archival conventions could govern the description in an archival record, it identified the Anglo-American Cataloging Rules, 2nd. ed. (AACR2) as a description convention, existing outside the MARC formats, to which library records adhered. In so doing, it called into question all aspects of the format which presumed AACR2-based descriptive practice rules, rather than serving as neutral communication mechanisms. By providing for explicit identification in a record of the conventions being followed, it opened the way to yet undefined conventions of other communities.

While there remains no code of archival description practice akin to AACR2 that serves as a universally accepted description standard, it is already apparent that AMC has differentiated itself in practice by rarely employing fields considered nearly essential in bibliographic item cataloging. The fact that an archival collection would frequently, if not usually, lack a "title proper" or that archives do not, by their inherent character, respect the of and about distinction reflected in the use of fields 6xx and 7xx are signs that something very different is being practiced by archivists. The presence of these elements in the AMC format, however, for those rare occasions that require them, disguises what practice is now revealing.

What is emerging, and will doubtless be codified and accepted by archivists in the coming years[6], is an archives-and-manuscript-control cataloging convention distinguished from the two fundamental bibliographic conventions of item cataloging and serials control. The guiding principle of bibliographic descriptive, or item, cataloging is that the cataloger *transcribes* cataloging information from the *item in hand* and may assign to the item topical subject headings describing what the item is *about*. The guiding principle of serials control is that bibliographic information from the publisher, usually but not always printed somewhere within the journal, is used to *establish an entry* for items in hand *and* for items that will be produced in the future as part of the series. It is possible for individual items within the series to be analyzed separately, and have topical subjects assigned, but it is by no means common.

The guiding principles of archival control are provenance and life-cycle management of records. The items in hand are described in terms of the *activity out of which they arose* and the *actions that have been taken* on them. They are not described, in themselves, except tangentially and as a matter of convenience. Therefore, archivists do not distinguish between what the materials are of and what they are about since they are by inference about whatever they are of and the volume of materials

[6]Steve Henson's *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies and Manuscript Libraries* (Washington, DC: Library of Congress, 1983) achieved the status of a de facto standard for most MARC AMC users. Its SAA sponsored second edition, scheduled for issuance in 1989, will probably be adopted formally. The desire of archivists for such standards was demonstrated by a vote at the 1987 annual meeting calling upon Council to seek support for definition of description standards that led to the NHPRC funded project, directed by Larry Dowler of Harvard University. This project is also expected to report in 1989 on standards needed by the profession.

typically precludes description of what they are about but not of.

Each of these traditions, and the cataloging rules they invoke, can describe the same physical things—for example, a government report on land use in a particular county may be a government document (book), a part of a series of land use reports, and a part of the archives of an agency. The potential difference between these descriptions reveals that cataloging is directed by implicit functional requirements and that different purposes are intended to be served by the different methods of documentation of holdings. As a consequence, articulating what conventions are being followed becomes essential if the descriptions are to be available for the intended purposes and audiences. The construction of descriptions for distinct implicit ends will become more evident as a result of technological changes that challenge our concepts of *published* and *serial*; these developments are erasing the lines between archival and bibliographic and forcing managers of repositories holding these materials to confront the purposes of cataloging in selecting the conventions they will follow. For instance, serials control was confined to printed materials when *journal* meant periodically issued printed text publication, but now is extended to include serial music, photographic or graphic serial works, and serial maps, and the concept will be extended increasingly in the future as electronic serial publications become commonplace. As the act of publication itself becomes publication on demand or remote publication, the difference between delivering a single copy and publishing, usually the difference between manuscripts and publications, is obscured.

B. Franchising the Periphery. Librarians often ask archivists, "why can't everyone just describe this item the same way?" without recognizing the bibliographic description bias inherent in the question. Obviously, if there were no functional basis for the different approaches, then everyone could adopt the same conventions. In practice, the differences between these conventions are only examined when more than one might be applicable to the same materials. But the question hides the real issue, which is the politics of *validating* an approach. Librarians have controlled the means of production and dissemination of bibliographic data; why, we should ask, have they permitted themselves to be put on the defensive and to accept approaches to this data that reject their standards?

The answer lies in the chimera that has been the psychological motivation for library automation—the unified catalog, integrated finding aid, or national bibliographic database. Long an elusive goal of library automation, the completely automated integrated union catalog became a realizable objective in the early 1980s when most major academic libraries had automated cataloging of traditional bibliographic materials and were finishing or planning retrospective conversion efforts to capture large bodies of materials cataloged previously. They were planning, installing, or dreaming about on-line public access catalogs. And as they looked around for materials that were left out of the scope of their unified catalog, they were keenly aware that archives and manuscript collections, often considered the jewels of their institutional holdings, were missing.

Thus, when archivists and, later, visual materials curators, approached the library community to consider including their holdings within the scope of the bibliographic utilities, the library community was open to them, in spite of the revisions to the MARC formats that each group felt were necessary. On the other hand, the political risk of negotiating with the library community required that those representing the SAA insist on some guarantee of mutual responsibility for the formats as a condition

for undertaking their revision. Because archivists agreed to use the normal channel for revisions to MARC (the MARBI committee process), the novelty of joint ownership was not unacceptable to librarians.

As a consequence of the proliferation of formats, of which the AMC was one, a movement developed within the library community to consolidate, or merge the formats, which was accepted in principle by the MARBI committee in July 1988. The implications of the differences between AMC and the other formats, however, will not be fully appreciated until (and unless) there is an integrated implementation. Because the unification process will result in a format that incorporates fields from all the formats, no single prescription of descriptive practice will be reflected in the unified whole, and the externality of descriptive standards and rules will be reinforced. Also, because it creates a single large and complex format, unification is increasing the pressure for simplification, a concept that, while desirable in the abstract, can serve as a cover for imposing on specialized users the value judgments of the majority, which sees certain elements of information as extraneous.

The position of archivists in the unification discussions, which they fully supported, was that format integration should not require communities to sacrifice the satisfaction of their disparate needs. We have certainly not seen all the implications of this political realignment.

C. Exploding Bibliocentricism. The last threat that AMC poses for the other bibliographic description formats is extremely subtle and has thus far escaped the attention of most archivists, as well as librarians. It concerns the presumptive architecture of a database in which AMC records reside.

Traditional library descriptive catalogs consist of a master file with indexes. Several methods of access may be provided to enable a researcher to query the file by authors, titles, subjects, and other bibliographic elements, but the subject of the query is bibliographic items and bibliographic items are what will be retrieved. If the bibliographic system is implemented with a degree of authority control, full authority records for names and subjects may also be in separate databases used by catalogers, but the public catalog, which will have only headings used in bibliographic records, is just an index. If the environment is one in which acquisitions and circulation are integrated with the catalog, transaction records of processing history may be implemented as a separate file. In any case, within a library system, each of these records will point to the bibliographic data that is clearly the master file.

Materials described using the AMC approach are not the center of their database. Already it is evident that provenance records (history of agencies and persons) and form of material or series records (which exist across jurisdictions) are more primary to the archival universe than their invocation in a particular document or record. Just as Copernican heliocentricism overthrew the privileged position of the earth, so AMC overthrows the privileged position of the bibliographic record. And, in so doing, it fundamentally alters the universe, not just for archivists, but for librarians as well. The implications of this revolution are explored further in the final section of this article.

II. What Does the Seven States Project Do?

A. The Project Itself. The work of the SSP illustrates many of the potential challenges to traditional library practices and bibliographic networking that are inherent in AMC. The project is not simply constructing a database of descriptions of state and local government records from seven states, but is also testing its utility in an archival context.

First, the participating state archives are

examining the power of a functions vocabulary to augment subject access. The vocabulary, consisting of terms for *sphere of activity* and *process*, describes not the contents of the records, but the characteristics of the activity that generated them. This activity is a property of the agency that created the records, not of the records themselves, but is being assigned to record series descriptions in the absence of office/activity of origin authority records. Employing sphere of activity and process terminology enables a researcher to cut across jurisdictional boundaries and structural accidents to identify commonalities of human action in state governments over time. Such a descriptive language for documented activities makes it possible for the reference archivist to fulfill the obligation to identify evidence of activity, to locate documentation of accountability, directly rather than through circuitous inferences.

The creation of agency history records with such functions identified gives the archivist an information resource needed as much by those in managing the organization on a daily basis as by archival researchers. With it, the archivist is potentially invested with the power of an information manager, responsible for and able to provide information on the current missions, objectives and activities of the institution.[7]

Second, the SSP is exploring the potential of sharing data on records scheduling between jurisdictions. In phase one of the project, they implemented this without a common record series or form of material authority file, but such a file is planned for the phase of the project proposed to begin in late 1988. The project, providing a way for archives to link their own records, whether in-house or scheduled, to their own schedules, is exploring the viability of generic records descriptions. All that is required to make the generic schedules into authorities for form of material is to externalize the retention directives (which are particular to each state) and provide a scope note within the records description for divergences from the authority. In this way, it is hoped that each state could exploit a set of common county, court, state, and federal records series descriptions as the basis for retention determinations, giving specific access by series titles or form of material for researchers and providing the basis for joint appraisal and cooperative acquisition decisions. Without such authority terminology, searches for similar materials will lack precision, and cooperative appraisal projects will involve unacceptable overheads or require great faith in the finesse of search strategies.[8]

Third, SSP has been exploring the agency history record. Because the experiment being undertaken by the SSP is constrained to operate within the current bibliocentric system, it is creating these records as pseudo-bibliographic records within RLIN, so it must be imagined what a provenance-centered government records control system might look like. What is more interesting for the longer term is whether RLIN will be able to implement agency history records as full partners in a bibliographic records database, so that the center depends upon the users' view.

Last, the SSP is exploring the full use of action field 583, which permits archives to treat the lifecycle of the records as integral to their intellectual control, not merely as a matter of housekeeping, which is how circulation or acquisitions data is viewed by librarians. The RLIN implementation of these fields is the fullest expression of their

[7]David Bearman and Richard Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (Winter 1985/6): 14–27.

[8]David Bearman, "Authority Control: Issues & Prospects," unpublished paper presented at the Society of American Archivists annual meeting 2 October 1988.

potential yet.[9] For a variety of political as well as technical reasons, it is important that the state archives employ these fields as fully as they can. In the long run, the benefits of archival networks may lie as much in cooperation in storage, conservation and distribution of documentation as in exchange of information for reference.

Each of these uses the SSP is making of the RLIN network was specifically suggested five years ago in working papers of the National Information System Task Force:

> The lesson which archivists have drawn from library practice, that archives cannot benefit from inter-institutional information exchange because their holdings are unique, is unwarranted. The proper lesson is that archivists will benefit from the exchange of archival authority data just as librarians have from the exchange of library authority data.[10]

Archivists are urged in that paper to pursue the potential of corporate history and form of material authority files. The suggestions for corporate history files with function vocabularies as access points have been followed by the SSP project and the results are gratifying especially now that it appears the National Archives will join in the effort using the same vocabulary. Everyone will benefit from the effort when it is fully analyzed. The attempt to exploit the commonality of forms of material across modern bureaucracies has departed somewhat from the approach anticipated in 1982:

> The obvious potential implied by this similarity for modeling retention schedules from similar environments has not been lost on records managers who regularly consult record schedules from other jurisdictions in compiling their own retention regulations, but its potential for archival appraisal has been less appreciated. . . . The similarity of such records from one environment to the next makes selective retention a tool for sampling such records of a particular sort. . . . One state could choose to keep physicians licenses for the first year of every decade, knowing that neighboring states are retaining such records for the third, fifth, seventh and ninth years of each decade without any real loss of informational value if definitions of, and information about retention policies for, forms of material were shared between similar institutions.[11]

In 1983 it was recommended that the states and the National Archives examine the potential of form of material control, especially to test the applicability of the combined use of form-of-material and function to identify duplicative documentation of federally funded, state-administered, post-1960 social programs, which have contributed vast quantities of documentation to the national records storehouses and which will be the leading candidates for statistical sampling and cooperative appraisal projects for the next decade.[12] The second phase of the SSP, scheduled to begin in December 1988 with participtiion of seventeen states and the federal government, will test scheduling driven by form-of-material authority control. This has the potential of facilitating shared appraisal in addition to providing one additional, and important, form of intellectual access.

B. Implications of a Multiple Peer File Database. The most radical aspect of what

[9]A radical implementation of the concept, adding procedural control to the actions, has been implemented on a local system, Collection, sold by Vernon Systems, a New Zealand based museum automation firm.

[10]Bearman, I. "Alternative Models," NISTF Working Paper #1, 13.

[11]Bearman, II. "Opportunities and Requirements," NISTF Working Paper #2, 26.

[12]Bearman, "Problems, Policies & Prospects," NISTF Working Paper #1, 89–106.

SSP participants are doing is what they are not doing—they are not practicing how to use a convention established in the abstract, but rather are trying through practice to evolve conventions that will be presented to the archives community. If these conventions can exploit the potential of bibliographic networks, they will be presented by archivists to the library bibliographic community.

If archivists were simply going off in their own directions after a couple of years as junior partners in library bibliographic networks, it would hardly be worthy of note. But they aren't. Nor are they following libraries blindly. The SSP illustrates that archivists are moving toward multiple peer files in the bibliographic networks, a direction in which many librarians are rapidly realizing it is essential for them to move as well. As such, the developments in the archives community are now of interest not only to archivists, and to RLG, which is assisting its realization, but also to librarians.[13] Libraries, especially the university libraries, are increasingly the hub of an information distribution network in which electronic databases, including the library catalog as a database, play a growing role. This is forcing them to consider the relationship between information from other sources, information from other sources, information that the library can readily acquire, and information about the holdings of the library.

If the library catalog is a database among others, new research opportunities are created that in turn threaten its privileged position. For instance, in a world of biographical databases, it is possible to conduct research about groups of individuals and use the results of such studies to generate a query to the bibliographic database for background citations, while in the universe of bibliographic databases and name authority files, it was not possible to conduct any significant research about persons (except perhaps acquiring a subset by dates), and so the bibliographic file remained primary.

As librarians have come to recognize, maintenance of authority files is time consuming and very expensive. But one man's authority file is another man's database proper, so if architectures can be designed that do not require any given file to occupy a privileged position, the value of each file can be exploited with respect to every other file. This will make possible the construction of cultural databases that will support education in the future, not by the cataloging of every item, but by importing data from disciplinary authority sources and organizing it in databases that serve broad cultural information ends. The issue is not whether the networks can afford to mount such reference databases, but whether librarians can continue to afford networks that require them to make all authority files de novo. Put this way, the recognition by archivists that such databases will have to be brought into the environment of archival descriptions in order to enrich them and provide access to the descriptions, should attract library attention.

OCLC, which has just formed a new Electronic Publishing and Information Delivery Division, and RLG, are already moving to make themselves into general information utilities and gateways into yet broader information networks.[14] Databases, such as the Eighteenth-Century Short Title Catalogue or the Avery Art Index, which will not be mounted on commercial on-line services, are finding homes within the bibliographic networks. Efforts to link

[13]Carol A. Mandel, *Multiple Thesauri in Online Library Bibliographic Systems. A Report Prepared for the Library of Congress Processing Services* (Washington, DC: Library of Congress, 1987).

[14]OCLC, *Campus of the Future* (Dublin OH, OCLC, March 1987). Also, see the report of an analogous RLG meeting in *RLG News*, April 1987.

these files are being actively pursued; a research project underway at RLG is examining geo-referents as a method of linking files, which has long been one of the more problematic and enticing potential connections between databases and is one that is critical in scientific, social scientific, and humanistic research. Databases, such as the Art and Architecture Thesaurus, that can serve as bridges between other files are being put into place on the networks, and mechanisms for pointing to them as the authority sources within bibliographic records have been adopted.

If the library databases of the future will contain numerous files built by different disciplines and inter-related along links made by one serving as an "authority" to the next to form a semi-coherent whole, they might support the needs of archives. Because of the costs of authority files, and the effort that goes into them, I have urged state archives to begin now to devote efforts to acquiring databases on the functions of state agencies, the finances of state government, the rulings of state courts, and the events of state history. Acquiring does not necessarily mean creating. If archivists are clever, they may be able to avoid making their own authority files altogether.

For example, state archivists are well aware that official publications and newspapers are critical sources of information about state history that librarians are willing to catalog using traditional serials control approaches. Why not let them? Because state archivists are aware that the acts of state legislatures and courts are being indexed in great detail and are available through electronic service bureaus, why aren't they planning to integrate such information into archival finding systems? State archivists are aware that guide to the functions of state agencies, their bureaus, and other state controlled or funded entities are needed by the legislature and executive of state government. If others are making them already, then archivists need to link to these databases.

An opportunity is being squandered by the National Archives, which, as the agency responsible for issuing the Federal Register every day, is creating a database of the missions, functions, and activities of its parent jurisdiction, but not using it as an authority file.[15] How many other archives create, or could easily obtain, data about their parent institutions that could be similarly employed? It is critical for archivists, if they are to survive as providers of information services within their own institutions and as a profession, to address how they can structure such databases for the purposes of locating activity by sphere and process.

In seeking to populate the information environment in which descriptions of records will thrive and take on meaning, archivists need to adopt a new, proactive, strategy. They need to identify the kinds of information and files that are needed and to canvass their own organizations to identify others who might already be making such a product (and need only make modest changes to it to meet the needs of the archives). If others are not making these files, then government archivists need to "sell" the value of the information product to state or federal legislatures or literally to sell the idea to commercial distributors—if possible, it is always preferable to get someone else to make what is needed!

To interest commercial information services in making some of these databases isn't impossible, but it will require some coordinated effort on the part of the states. Commercial information services are begging to be given greater access to state records in order to make (and sell) just the sorts of information products the state ar-

[15]David Bearman, "The National Archives & Records Service: Policy Choices for the Next Five Years," NARS Assembly Presentation, 15 December 1981.

chives need.[16] But it is in the nature of their markets that they cannot find a way to sell a product (for example regulations regarding records retention or a functional index to state governments) that is limited in scope to one or a handful of states. Commercial information vendors are finding that many states erect barriers to their access to the information that would make up such products because state governments have not yet realized they would benefit from private information distributors creating value-added products from state information sources. The states need to realize that purchasing them from commercial services (which has the added value of generating jobs and taxes for the state) would probably not be as expensive as making them in-house. After all, the commercial publishers are spreading their costs over a broader market. Archivists could play a role within their respective states in making state information accessible to value-added resellers, and in the process do themselves a favor.

On the other side of this equation, cooperation between the state archives could also have benefits. Existing information products are not being exploited as well as they could because of the complexity of negotiating arrangements for their use by any one archives. How many state archives or historical societies have made arrangements with Marquis Who's Who to provide biographical information (including locations of personal papers) to the dozens of Who's Who publications produced from Marquis databases in return for electronic biographical files on state personages? How many state archives are engaged in discussions with the U.S. Census Bureau and commercial publishers planning the release of 1990 census data on CD-ROM to assure that this information will be structured in such a way as to facilitate integrated access to archival holdings through this vast electronic information resource?

Here archives are again touching an area of concern that has the potential of allying them closely with the libraries. Not only are these same issues confronting librarians, but as the very concepts of published and unpublished are dissolving in the electronic age, database access, which doesn't respect that boundary, becomes the central question. As access to information replaces collections as a professional concern, archives and libraries will come to see the databases they construct concerning their holdings as tools for tying together other resources, and as knowledge bases, made powerful by the conceptual genius brought to their design and by the power of the languages forged to connect them with relevant information from other sources.

III. Foundations and Futures

At this moment in the history of archival automation, two agendas compete for attention. The first is to make it work. Projects like the SSP have advanced it far along this path already. If such projects report their findings, archivists will be well positioned to develop solutions. The second agenda is to realize a dream, a vision of an information future in which archives have a role far beyond that which they play today and in which archival knowledge bases are, indeed, the cultural memory and its most valued heritage.

The incorporation of archival control within the family of MARC formats was explicitly undertaken as a step toward the incorporation of museum artifacts.[17] Li-

[16]Peter Marx, "State Public Records: Database Goldmine or Landmine?" *Information Times* (April 1987): 21, 28.

[17]Discussions the author had with Henriette Avram at the Library of Congress early in 1980 prior to embarking on the MARC AMC format explicitly introduced museums as did much subsequent work on the format itself.

brary of Congress staff were keenly aware of the inadequacy of the format for photographic material in 1982-83 and had already launched the review that subsequently gave birth to a new format for visual resources, since modified (inadequately, but it's a start) to accommodate three dimensional materials. This expansion of the scope of the formats is continuing and will eventually lead to the acceptance of MARC as a communications vehicle for object collections, because libraries (and archives) hold such materials and the pressures for an integrated catalog will once again come into play. The demand in higher education for cultural databases that cross disciplines and media is insatiable.

But the cultural information systems that will temporarily slake the thirst of universities cannot consist simply of document and object surrogates, representations of information-laden holdings; they must consist of the things themselves. Thus, images will need to be carried along with their descriptions, texts along with their cataloging, and data files along with the descriptions. Such multimedia information bases will make it possible to listen to a speech and read the notes for it or watch a politician approach a crowd and see the next day's news analysis juxtaposed to the politician's diary entry.[18]

The number of archival and museum objects of interest to scholars is vast. No institution can possess them all, because the items in museums and archives are unique. Therefore, scholars need to travel widely in order to conduct research or find materials with which to teach. Digital data storage and telecommunications, the same technological developments which have fueled the information explosion of the past two decades, provide a potential relief from this burden: new ways to distribute information widely. This distribution is not only applicable to the document surrogate, or catalog record, but to the information contents (including images and sounds) of the evidence itself. Archives and museums could benefit from this distribution even more than libraries, precisely because of their reluctance to circulate unique holdings or even to make them available for browsing.

The tremendous volume of the materials provides the incentive for highly discriminating retrieval capabilities and mechanisms to rapidly distinguish materials needed from those that are provided in error. This is one of the reasons for using multiple interdependent reference files as authorities, but it is also an impetus for libraries to move beyond traditional bibliographic references to full text, and for museums and archives to capture their images, and even texts, of their holdings.

The potential of such sound, image, and text bases to transform higher education has been recognized for some time, and now that a technology in which one can develop hypermedia learning materials is available, many university educators are working on programs to allow students to travel independently through our cultural heritage. One substantial barrier, at the moment, is that the necessary image, sound, and text bases are not available. The means for providing them is, of course, resident in the libraries, archives, and museums. Those institutions that find a way to store and transmit the information contents of their holdings to these potential users will reach an audience and exert an influence much larger than any they have previously had.

While these kinds of databases may seem distant the technical barriers are fast becoming trivial. Storage costs, which have declined by a factor of three in the past decade, tumbled another order of magnitude in the last two years. Transmission bandwidths too expensive to imagine five

[18]The user driven association of such multimedia objects from diverse sources still poses some intractable problems, but image and data bases with concrete audiences are being developed at numerous universities with impressive results.

years ago are now routinely available in over supply due to optic fiber installation. Images available with an art or museum catalog no longer are novelties;[19] in a few years they will simply be assumed. Sound recordings are being digitized for preservation purposes anyway and soon will be integrated.[20] More than a quarter of all text information is now created in machine-readable form and intelligent scanning devices are making it possible to capture other text with ever decreasing effort. Significant hurdles must still be overcome to use these multimedia bases intelligently if they are created, but the desire for them is stong and the pressure to incorporate archival and museum information into library/information services is overwhelming.

There is no reason to stop with higher education; secondary schools and the general public have as great a use, and need for, access to the cultural knowledgebase. Delivery vehicles, in the form of cable television systems and fiber optic telecommunications networks, are already installed. The challenge of the next century will be to transform archives from repositories to intermediaries.[21] Archivists must ask how they can best position their institution to deliver information in all its richness to the citizens of tomorrow who may then learn from, and work with, archives as a primary cultural resource. Information delivery begins with the reexamination of finding tools and access points, but it doesn't end until the information itself is provided directly to patrons, in their own intellectual framework, on their own terms, and wherever they may be.

[19]David Bearman, "Museum Automation at the AAM," *Archival Informatics Newsletter* 1 no. 2 (Summer 1987): 2–4.

[20]Jeremy Silver and Lloyd Stickells, "Preserving Sound Recordings at the British Library National Sound Archive," *Library Conservation News* 13 (October 1986): 1–3.

[21]Stewart Brand, *The Media Lab: Inventing the Future at MIT* (Cambridge MA: MIT Press, 1985).