Standards for the Exchange of Descriptive Information on Archival and Manuscript Material in the United States

by **STEVEN L. HENSEN***

It is now just over ten years since the advent of automated archival description in the United States. None of us who were involved in its inception would have dared believe that it would have come as far as it has today, or that its influence would become so pervasive. Evidence of this influence exists everywhere: sessions relating to automation-related topics or to standards have come to dominate professional meetings during the past few years. Between the two major bibliographic utilities (RLIN and OCLC) there are nearly 500,000 bibliographic records describing previously unknown and inaccessible archival and manuscript materials in hundreds of widely separated and diverse repositories. In the case of RLIN, at least, the information contained in these records is accessible to scholars literally anywhere in the world via a computer and modem. There are now several electronic computer "bulletin boards" subscribed to by hundreds of archivists, manuscript librarians and interested scholars, which are devoted to little else but detailed (and occasionally tedious) discussions of issues relating to various aspects of automation and its impact upon a world that I, for one, would have once thought impervious to the dubious allure of automation. Finally, moreover, as difficult as it may be to believe, the Society of American Archivists is now a full member of the National Information Standards Organization (NISO).

How can we account for these changes? The fact is that ten years ago, when all of this started, there were so many factors working against the automation of archival description in North America that the fact that it has come as far as it has it can only be seen to be just short of miraculous. Moreover, unlike Dr. Johnson's notorious comparison between women preachers and walking dogs, however surprising it may be to many of us that automated description is done at all, it is perhaps even more so that it is actually done pretty well. The reason why this is so is related to recent substantial changes in archival thinking in the United States.

From the beginning, the principal motivating factor behind almost all library automation in the United States was in the economies of distributed or shared cataloguing. In this model, a library (often the Library of Congress) would catalogue a book and create a machine-readable MARC record for it. That catalogue record would then be entered into one of the national bibliographic utilities, whece other libraries which had the same

 Archivaria 34 (Summer 1992)

book would simply derive the existing catalogue record, put their own location symbol on it, and then move it into their own local catalogue (which could be either manual or on-line). This obviated the need for each library to do all of its own original cataloguing, while, not coincidentally, also guaranteeing a fair amount of uniformity in the cataloguing. That such a system, through the sheer accumulation of bibliographic information, might have some usefulness as a research tool was only a secondary consideration if it was thought of at all.

Given that archival materials (under which rubric I include everything from literary manuscripts to personal papers to corporate and government records) are, virtually by definition, unique; furthermore, that very uniqueness makes it extremely unlikely that one repository would need to "derive" the cataloguing of another repository, the shared cataloguing model was never very relevant or useful in an archival context. As Alan Tucker has pointed out, "the same historical factors which explain the emergence of . . . library-oriented standards . . . also explain the absence of a similar level of standardization among archivists. The repetitive cataloguing of thousands of copies of the same item in thousands of institutions generated needs and solutions which have none of the same impact in an environment in which virtually all of the materials being described are unique."[1]

Moreover, since some of these systems existed for little more than to serve as a mechanism for derivative cataloguing, little interest was shown in exploiting their research utility until fairly recently. For archivists and manuscript librarians for whom this primary design had little relevance, the fact that these systems actively discouraged their use in the only way that *did* make sense seemed somewhat backward, if not perverse. As a consequence, archival and manuscript repositories were nearly as uninterested in the automated cataloguing of archival material as the library world was in accommodating it.

Another factor contributing to this onetime archival disaffection with library automation was the fact that the latter was based almost exclusively on a library cataloguing model. Beyond the historical (and not entirely unjustified) antipathy between the archival and library professions, such an approach was seen by most archivists and manuscript librarians in the United States as not only irrelevant to their own descriptive needs, but positively antithetical to the very foundations of archival theory and practice. Archival description in the United States had evolved into elaborate systems of internal finding aids which were built upon the principles of hierarchy, provenance and informed subjective analysis. Archival "cataloguing," if done at all, tended to be locally developed and highly idiosyncratic. The library model, as enunciated in the so-called "Paris Principles" of 1961 (upon which the modern *Anglo-American Cataloguing Rules* were built), focused instead on the objective analysis of the artefactual and physical information characteristics of the bibliographic item, and the controlled transcription of that information into the library cataloguing record.

Furthermore, past protestations to the contrary notwithstanding, the MARC formats were developed as a direct extension of the cataloguing rules. (Until recently, the Library of Congress took the absurd official position that USMARC and *AACR2* were completely separate and unrelated systems. The *Cataloging Service Bulletin*, however, now acknowledges the existence of the MARC formats as it explains the impact of cataloguing rule changes upon affected MARC fields.) In fact, it is now perfectly plain to the outside world that these formats are little more than an automated articulation of those

descriptive cataloguing principles. Whatever flaws might be found in these formats are, in fact, mostly related to the inherent flaws of the nineteenth-century card catalogue upon which it was modelled.

Thus it was that a combination of all these factors—the unrelenting library orientation of everything from descriptive standards (otherwise known as cataloguing rules) to the very purpose and orientation of the systems—seemed almost consciously designed to produce maximum archival frustration and irritation. How is it, then, that archivists are now working gladly (for the most part) within those very systems, and are eagerly embracing and developing standards in areas in which the very idea of standards was so recently anathema? Have they simply thrown themselves prostrate under the juggernaut of library automation? Or is what is happening part of a reconceptualization not only of archival description, but also of the larger information universe?

The first step in this transformation in the United States was the realization and admission that archivists, manuscript curators, records managers and everyone else who looked after the nation's documentary heritage had more in common with each other than generations of minor quibbles might have suggested (bad enough that the archivists fought with the librarians; they also quarrelled among themselves). As part of its work in building a data dictionary, the Society of American Archivists' National Information Systems Task Force (NISTF) commissioned a study of descriptive practices across the entire range of repositories having custody of what was very loosely defined as "archival" or "original" material. This study, which was conducted in 1980 by Elaine Engst of Cornell University, clearly demonstrated, to the mild astonishment of the members of the Task Force, that there were indeed broad areas of common descriptive practice among all the institutions, and that state and government archives followed pretty much the same practices in controlling their holdings as did small historical societies or university manuscript collections. More important, however, was the discovery that, horrifying as it might seem to more traditional archivists, these descriptive practices had obvious direct parallels with library descriptive cataloguing. Based on this finding, the Task Force was able to move confidently in adopting (and adapting) the MARC Formats for Bibliographic Description for the purposes of archival description. Thus, the actual process of developing the MARC AMC format became a relatively straightforward one-to-one correspondence of existing MARC bibliographic tags with the NISTF data dictionary.

There were some significant differences, however. In these differences can be found the seeds of some profound changes, not only in the theory of archival description, but also in the overall foundations of the national bibliographic systems. I shall focus briefly on two of these differences. The first of these was the idea of process control. Unlike other bibliographic materials that are more or less confined between their covers, archival materials are mutable over a time/space continuum. In other words, things happen to them after they enter institutional custody (they are weeded, sorted, re-sorted, added to, conserved, preserved, etc.), and these changes have a direct impact on the description of the materials. The bibliographic description of such material must, therefore, be able to accommodate and record these changes. That a bibliographic record could be anything other than carved in stone was a profoundly revolutionary idea to the library community at the time. Since then, however, the library preservation group has understood the utility of this concept as the ravages of time on wood-pulp paper has shown that the bibliographic artifacts in institutional custody are, indeed, sadly mutable. The

current discussions over multiple versions is rooted in the idea of bibliographic objects existing in different variants over space and time—a hitherto distinctly archival idea.

The second difference was in the archival application of what were then undeveloped fields in the MARC format for accommodating analytics. Archival description, as I noted earlier, had developed along distinctly hierarchical lines (reflecting, no doubt, the fundamentally hierarchical nature of most archival records and of the bureaucratic organizations that generate those records). Thus we have the American hierarchy of record group, series, subseries, subgroup, folder item, and the Canadian hierarchy of repository, fonds, series, filing unit and item. The concept of *respect des fonds* on which these descriptive hierarchies were built is so fundamental to archival practice that any descriptive system that could not accommodate it was automatically doomed to failure from an archival point of view. In examining the MARC formats, NISTF discovered that the structures established to accommodate library analytics were perfectly suitable for controlling archival hierarchy. This idea, while perfectly obvious now, was an epiphany at the time and really paved the way for subsequent full development of the MARC AMC format. The Research Libraries Group fully implemented these "linking" fields in the RLIN execution of the USMARC AMC format. These fields have since become the very axis of the description, particularly of government archival records within RLIN, providing a means to describe materials at any appropriate level, while logically associating that description with other descriptions of hierarchically related materials. All that remains now is to sort out with the library community some fundamental misunderstandings over archival use of the term "series."

Simultaneous with this process of defining descriptive elements and developing a format, another project was underway at the Library of Congress that was attempting to "reconcile manuscript and archival cataloging and description with the conventions of *AACR2*." The rationale for this work was based on the idea that if the "burgeoning national systems for automated bibliographic description . . . are to ever accommodate manuscripts and archives a compatible format must be established . . . [and that] with appropriate modifications, library-based descriptive techniques can be applied in developing this format."[2] I should say at this point that 'format' here was defined in a strictly cataloguing sense; I am embarrassed to confess that the work of NISTF was largely seen as a separate and unrelated activity at the time. It is only now, given our contemporary understanding of "data structure" and "data content," that their utter interdependence is fully understood. In any event, the result of that enterprise was the first edition of *Archives, Personal Papers, and Manuscripts* (otherwise known as *APPM*).

Without sounding immodest, the success of this cataloguing manual has been little short of stunning—particularly in a profession for whom the words "cataloguing" and "standards" have traditionally been, if not anathema, then certainly redolent altogether too much of librarianship. Now in its second edition, *APPM* was the first standard formally adopted by the Council of the Society of American Archivists, while going on to become the Society's all-time best-selling publication. Moreover, I am currently giving a series of workshops for the Graduate School of Library and Information Science at the University of Texas in the practical application of these cataloguing rules. The demand for these workshops has been almost overwhelming, and reflects the reality that archival institutions of all sorts and sizes have not only accepted the idea of standardized cataloguing of archival material, but are also actively participating in greater and greater numbers in the bibliographic systems which require the use of such standards.

The success of *APPM* has nothing to do with those qualities normally associated with other "best-sellers." It is a cataloguing manual, after all, much of the prose of which has been borrowed directly from *AACR2*. What has made it so popular and useful to archivists is the way in which it integrates basic archival principles into the broader framework of bibliographic description, modifying that framework in order to transform it into a vehicle for specifically *archival* description. Such a synthesis is based on three major principles:

First, it recognizes the primacy of provenance in archival description. This principle means that the significance of archival materials is heavily dependent on the context of their creation, and that the arrangement and description of these materials should be directly related to their original purpose and function. This translates into a basic rule for choice of main entry in which archival materials are entered "on the basis of provenance, under the name of the person, family, or corporate body chiefly responsible for [their] creation."[3] This rule is fully consistent with the *AACR2* principle that bibliographic materials are entered under the entity "chiefly responsible for the intellectual or artistic content of a work."[4] It also translated into a heavier emphasis on the use of notes in archival cataloguing, since it is difficult to capture the complexities of substance and provenance in the sort of brief formulaic encryption that characterizes most bibliographic description. Moreover, the use of notes is more consistent with the archival tradition of subjective analysis.

Secondly, it acknowledges that most archival material exists in collectivities or groups and that the appropriate focus of the bibliographic control of such materials is at the fonds or collection level. This approach is practical not only in that it relieves the archivist of the overwhelming burden of providing item level catalogue records for records series or manuscript fonds more frequently measured in linear metres, but also in that it supports the principles of archival unity, in which the significance of individual items or file units is measured principally by its relation to the collective whole of which they may form a part. In a library setting, where the publications explosion is contributing to ever-increasing cataloguing backlogs, the idea of aggregate-level control of certain classes of material is increasingly being looked on with some favour. Even beyond these practical considerations, however, libraries are starting to realize that traditional item level bibliographic control may not always be the most logical way to provide optimal access to its holdings. In fact, the Library of Congress has just issued guidelines for aggregate-level cataloguing in its *Cataloging Service Bulletin*, which borrow heavily and directly from the principles laid down in *APPM*. Duke University, moreover, has a proposal before the federal government to provide funds for a rare book cataloguing project which would employ essentially aggregate-level archival techniques for describing a large collection of eithteenth- and nineteenth-century Italian pamphlets. Item level access would be provided in this project, not through the catalogue, but through an independent and related internal computer database.

Thirdly, and finally, these rules recognize that archival materials are preserved for reasons different from those for which they were created. They are the unself-conscious by-product of various kinds of human activity, and consequently lack "the formally presented identifying data that characterize most published items, such as author and title statements, imprints, production and distribution information, collations, etc. Personal or corporate responsibility for the creation of archival materials (another way of saying provenance) is generally inferred from, rather than explicitly stated in the

materials."[5] The principal implication of this for the cataloguing of archival materials was to legitimize traditional archival finding aids, guides, registers, etc. as sources of cataloguing data and to move the cataloguing process away from the literal transcription of information that characterizes other bibliographic description.

In consequence of the groundwork laid by NISTF in developing the MARC AMC format, and by *APPM* in providing an approach to archival "cataloguing" that is consistent with *AACR2*, archivists in the United States have begun to re-examine some of their most cherished beliefs and prejudices. The world of libraries and librarianship is no longer viewed as the enemy, but instead is being embraced as the natural ally in the world of information management that it has in fact been all along. Archivists are now taking an active role in such things as the development and maintenance of national information standards, participation in national authority files and thesaurus construction. Finally, the education and training of most archivists today, in spite of lingering hostility and suspicion among more atavistic colleagues, is largely being done within the curricula of our graduate schools of library and information science.

Most important, however, has been the very real evolutionary changes that this process has caused in our national information landscape. Initially, the entry of bibliographic information on archival and manuscript material into the national bibliographic databases was viewed with suspicion and hostility. It was seen, in fact, as a kind of Trojan Horse or virus, which would somehow infect the purity of these library catalogues at the very least, these cataloguing records just "looked funny" and somehow distasteful. Such suspicions were, of course, well-founded, but the consequences have been anything but negative. The world of research and scholarship (which most of us serve) has become increasingly interdisciplinary and less concerned with whether the information it seeks is to be found in traditional printed and published form or in archives, photographs, motion pictures, videotapes, computer files or museum registers. Information of all sorts is now regarded as a kind of seamless web, and it is becoming increasingly clear that service to scholarship and research is optimized when there are no artificial restrictions on the particular form that information takes. The Research Libraries Group is formally committed to the concept of RLIN evolving into a database of what they are calling "cultural artifacts," a term that includes not only the more traditional textually-based holdings of research libraries and archives, but also various media and artifacts, ranging from photographs and fine-art prints to archaeological specimens.

The same process that the American archival community began over ten years ago is currently underway in the American museum community. The Art Information Task Force (sponsored by the Getty Art History Information Project and funded by the National Endowment for the Humanities) is currently working to develop a MARC-compatible data dictionary and descriptive standards for "art objects and their visual surrogates"; the Common Agenda of the American Association for State and Local History is attempting to do the same thing for history museums.

The recent communication that the British Library has joined RLG is very good news indeed. It bodes well for the further extension of the developments I have just been describing into the very important research collections in Great Britain and on the Continent. In the most significant archival development of the last half-century, American archival repositories and manuscript libraries have become full participants in automated information exchange, and have taken their rightful place in the information community. The

experience of the American archival community over the past ten years has demonstrated not only the vital importance of archival participation in these growing national and international information networks, but also the relative ease with which the necessary accommodation and arrangements can be found.

It is perhaps a sad commentary on modern society that libraries and, most especially, archives have traditionally been among the most marginalized of institutions. As we move into an era in which society will almost certainly be placing greater value and sharper focus on the control and exchange of information, it is only logical to assume that such institutions, as repositories of vast amounts of information, should become ever more powerful. This will not happen, however, without active participation. Those libraries and archives which, through either inertia or active opposition, or simply by clinging to impractical and outmoded ideas and prejudices, decide to withhold their information from the global information stream, risk only worse marginalization and isolation.

Notes

* Paper originally presented at the Liber Conference ("Information Exchange in the Age of Automation") held in London, England, 7-8 April 1992.

1 Alan M. Tucker, "The RLIN Implementation of the Marc Archives and Manuscript Control Format" (in "Academic Libraries: Myths and Realities": Proceedings of the Third National Conference of the Association of College and Research Libraries, Seattle, Washington, 4-7 April, 1984).

2 Steven L. Hensen, *Archives Personal Papers and Manuscripts: A Cataloguing Manual for Archival Repositories, Historical Societies, and Manuscript Libraries* (Washington, 1983), p. 1.

3 Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloguing Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, 2nd ed. (Chicago, 1989) [hereafter *APPM*], Rule 2.1, p. 38.

4 *Anglo-American Cataloguing Rules*, 2nd ed., 1988 rev. (Chicago, 1988), Rule 21.1A, p. 312.

5 *APPM*, Rule 0.11, p. 5.