"NISTF II" and EAD: The Evolution of Archival Description

STEVEN L. HENSEN

Abstract: Encoded Archival Description (EAD) is part of the mainstream of archival standards development. The work of the National Information Systems Task Force in creating the MARC AMC format and the concomitant development of *Archives, Personal Papers, and Manuscripts* helped define the basic framework of shared archival description. The architects of EAD have built upon this architecture and provided archivists with technology-based tools capable of producing a fully integrated descriptive apparatus in an Internet environment that are nonetheless firmly rooted in archival principles and traditions.

About the author: *Steven L. Hensen is Director of Planning and Project Development in the Special Collections Library at Duke University. He is the author of* Archives, Personal Papers, and Manuscripts *and of numerous articles and papers in the area of archival description and standards. He is a Fellow of the Society of American Archivists and has been a member of its council. He was a member of the SAA National Information Systems Task Force, the Working Group on Standards for Archival Description, the 1989 Airlie House Multiple Versions Forum, the Bentley development team for EAD, and other groups, and has conducted numerous workshops and consultancies in archival cataloging and description and the use of the USMARC format.*

Introduction

THE DEVELOPMENT OF ENCODED ARCHIVAL DESCRIPTION (EAD) is part of a long process of evolution of archival descriptive standards that had its roots in the early work of the SAA National Information Systems Task Force (NISTF), and which began the integration of archival description into the so-called "bibliographic mainstream." The EAD project is in fact a natural culmination of much that had gone on before; an extension—nay, more of a quantum leap—of standard archival practice that leapfrogs the archival world into the unaccustomed role of being on the "cutting edge" of current advances in networked information access.

The EAD project is maturing at exactly the moment when archivists and manuscript curators both need and will be able to make use of it. When Daniel Pitti first started the Berkeley Finding Aid Project, there was little sense of how such a system might ultimately be implemented. No one had heard of the World Wide Web or HyperText Markup Language, and certainly few could have predicted the current explosion of information resources over the Internet. Today, forward-thinking archivists are developing plans and systems for making information about their holdings—and even the holdings themselves—available over the Internet. These systems are almost universally conceived in traditional hierarchical models of archival description: catalog records with subject indexes, linked to finding aids, linked to digital representations of archival objects. Without the Berkeley Finding Aid Project and its evolution to EAD, such systems would not yet be feasible.

The title of this paper makes a connection between the EAD project and what I refer to as "NISTF II." At the 1994 annual meeting of the Society of American Archivists in Indianapolis, Larry Dowler chaired a session in which I was one of the speakers. The session, somewhat enigmatically titled "Archival Stonehenge," was focused on the tenth anniversary of the MARC AMC format. In commenting on the papers dealing with the past, present, and future of MARC AMC, Dowler remarked that archival descriptive standards and systems had come so far so quickly (and were, in fact, heading even more quickly into unanticipated new directions), that maybe it was time to convene a new National Information Systems Task Force, or "Son of NISTF," as I believe he called it. In February 1995, at the RLG Primary Sources Forum meeting, Dowler was called upon to summarize the presentations and recommendations of the assembled representatives of this group, and he repeated this conclusion. He has since told me that it was the Berkeley Finding Aid Project, coupled with the explosive growth of new modes of networked information retrieval, that were pushing him towards this conclusion. Moreover, Dowler was recognizing subtle parallels and connections between the dynamics of current developments and his own earlier experiences in archival standards development.

It is thus my contention that the Berkeley Finding Aid Project in many of its goals and ultimate aims was, if not the "Son of NISTF" that Dowler was calling for, such a strong lineal descendant that it deserves immediate adoption, if not actual christening, as "NISTF II." As a quondam companion of Dowler's in much of that earlier work, I would like to draw on that experience and briefly examine the principal themes and history of archival standards development over the last ten years by way of more fully appreciating and understanding the importance of the present project and how deftly and systematically it fits into the overall process.

NISTF and the MARC AMC Format

The work of the National Information Systems Task Force[1] was a lengthy and often contentious process. To be sure, the issues with which NISTF was dealing were complex and controversial and required as much attention as could be given them. In addition, however, the world was a more leisurely place twenty years ago; there was not the pace of progress with which we must contend today, in which fundamental social and cultural changes seem to occur at the same pace as (and indeed are often prompted by) changes in computer software and hardware—which is to say an eighteen-month obsolescence cycle. Two recent projects with which I have been involved—the RLG Digital Image Access Project and the EAD project—were frustrated in trying to proceed towards their respective goals as the ground was almost literally moving under their feet. As a consequence, both projects ended up in altogether different environments than those in which they started.

Among the difficult issues with which NISTF grappled was the seeming hostility felt by many in the archival community towards anything that smacked of librarianship, and the firm belief that since archives were unique, they required unique approaches, and standards could thus never be applied. Add to this mix the sentiment that the methodologies and principles of archivists were somehow fundamentally different than those employed by their more library-oriented "manuscript curator" colleagues—perhaps a vestige of the "archives/historical manuscripts" dichotomy that dates to Sir Hilary Jenkinson in the early twentieth century.

Thus, NISTF had to address whether there was any substance in the long-standing dispute between "archivists" and "manuscript curators" over various matters of theory and practice. Towards this end, Elaine Engst conducted a thorough study of descriptive practices in a wide variety of repositories. Her unpublished report, "Standard Elements for the Description of Archives and Manuscript Collections,"[2] clearly demonstrates that there is no significant difference between the descriptive approaches of these two groups and that, in the words of Tom Hickerson, "there are common methods of archival description which could be integrated into a broadly applicable set of standards."[3] More importantly, however, Engst's report helped lay an essential foundation for the subsequent development of a unified data elements dictionary, which was the first step on the road to adapting the MARC format for the purpose of describing (or, more specifically, "cataloging") archives and manuscripts. At the time this work was going on, it was not altogether clear to the members of the task force that it was possible or desirable to describe these materials in the same systems used for describing other library materials, but it was already

[1]NISTF was formed in 1977 by the Society of American Archivists with funds from the National Endowment for the Humanities. Its members consisted of Richard Lytle, chair, and David Bearman, project director, both of the Smithsonian Institution; Maynard Brichford, University of Illinois; John Daly, Illinois State Archives; Charles Dollar, National Archives and Records Administration; Larry Dowler, Yale University; Max Evans, State Historical Society of Wisconsin; Steven Hensen, Manuscript Division, Library of Congress; Tom Hickerson, Cornell University; Charles Palm, Stanford University; and Nancy Sahli, National Historical Publications and Records Commission. For a detailed summary of the work of NISTF, see Richard Lytle, "An Analysis of the Work of the National Information Systems Task Force," *American Archivist* 47 (Fall 1984): 357–65.

[2]Elaine Engst, "Standard Elements for the Description of Archives and Manuscript Collections," unpublished report delivered to the National Information Systems Task Force, 1979.

[3]H. Thomas Hickerson, "Archival Information Exchange: Developing Compatibility," in *Academic Libraries: Myths and Realities, Proceedings of the Third National Conference of the Association of College and Research Libraries*, edited by Suzanne C. Dodson and Gary L. Menges (Chicago: Association of College and Research Libraries, 1984), 64.

obvious that the superstructure used by the library community (the MARC formats) could easily be adapted to archival purposes. The result was the USMARC Format for Archival and Manuscripts Control (MARC AMC).[4]

NISTF also came to the crucial realization that the new superstructure must somehow accommodate multilevel archival hierarchy. There is nothing quite so sacred or central to an understanding of the archival worldview than the principle that the *fonds*, whether consisting of personal papers or government records, are essentially organic in nature, i.e., generated as the natural documentary byproduct of the activities or functions of corporate bodies or persons. From this flows the archival principle of provenance (also known as *respect des fonds*), which holds that the arrangement and description of these materials follows their original function, purpose, and order. Thus, for the archivist, the concept of multilevel description is deeply rooted. In 1964 Oliver Wendell Holmes defined five basic levels of archival arrangement and description—Depository, Record Group and Subgroup, Series, Filing Unit, and Document.[5] And until 1986, when Max Evans effectively destroyed the concept of record groups,[6] this system of hierarchically based levels had, according to Terry Abraham in 1991, achieved the status of "dogma" in the American archival profession.[7] This dogma was based on the essentially hierarchical nature of archives from which, according to Holmes, proceeded distinct descriptive and arrangement requirements inherent in these levels.

Recognizing that any structure that did not accommodate archival hierarchies or levels was both inadequate and doomed to failure, NISTF took a harder look at the MARC formats. There, in some relatively undeveloped fields, they discovered that the structures established to accommodate library analytics (i.e., description of a part of a larger work) were not only perfectly suitable for controlling archival hierarchy but were also, in their "part-to-whole" configuration, philosophically consistent with archival levels of description. This idea, while perfectly obvious now, was an epiphany at the time and paved the way for subsequent full development of the MARC AMC format and the full integration of archival description into heretofore strictly "bibliographic" systems. In the RLIN application of the USMARC AMC format, the Research Libraries Group fully implemented these "linking" fields, and they have become the very essence of effective description of archival material (particularly government records) within RLIN, providing a means to describe materials at any appropriate level while logically associating that description with that of other hierarchically related materials.

No matter how well-suited the MARC AMC format was to archival descriptive needs, it was, however, simply an empty vessel—a "data structure standard," as we now understand these things.[8] To make MARC AMC usable inside the framework within which most MARC records were created, a companion "data content standard" was also re-

[4]This section is based on an earlier article: Steven L. Hensen, "The Use of Standards in the Application of the AMC Format," *American Archivist* 49 (Winter 1986): 33.

[5]Oliver W. Holmes, "Archival Arrangement—Five Different Operations at Five Different Levels," *American Archivist* 27 (January 1964): 21–41.

[6]Max Evans, "Authority Control: An Alternative to the Record Group Concept," *American Archivist* 49 (Summer 1986): 249–61.

[7]Terry Abraham, "Oliver W. Holmes Revisited: Levels of Arrangement and Description in Practice," *American Archivist* 54 (Summer 1991): 371.

[8]Distinctions such as these were first identified in an archival context through the work of the Working Group on Standards for Archival Description (WGSAD). The reports of this group can be found in *American Archivist* 52 (Fall 1989): 431–537 and in *Standards for Archival Description; A Handbook* (Chicago: Society of American Archivists, 1994).

quired. Once again, the forces of serendipity were at work for archivists, when the second edition of the *Anglo-American Cataloguing Rules (AACR2)* was published in 1978.

Archives, Personal Papers, and Manuscripts

Although the publication of *AACR2* cannot be said to have had much direct impact on the archival world, the archival response to it has been of major significance. Most of the archival world took little note of *AACR2*, but this was not the case in the Manuscript Division of the Library of Congress where I was then employed as Senior Manuscript Cataloger. As the Library of Congress was one of the principal partners in the development of *AACR2*, I was more or less obliged to use it. However, a brief review revealed that the rules were written with no obvious input from anyone in the manuscripts or (even more so) archives communities.

The specific problems which rendered *AACR2* essentially unusable for archival cataloging have been described elsewhere.[9] The Manuscript Division's response was to develop an alternate set of rules consistent with sound archival principles while retaining as much as possible the overall spirit and structure of *AACR2*. These alternate rules were subjected to a thorough review within the Library of Congress, as well as by an editorial committee drawn from the American archival community and by a number of other commentators from around the country. The result was the first edition of *Archives, Personal Papers, and Manuscripts* (*APPM*).[10]

This manual, now in its second edition,[11] has been widely accepted by the American archival community as the standard for the cataloging of archives and manuscripts—especially in an automated environment. It is important to understand that this is *not* a manual of general archival description, nor is it a guide for the construction of archival finding aids (though its rules and principles are based upon the existence of such finding aids and upon a general presumption of standardized data elements).

APPM's success is based, first of all, on the fundamental premise that archival cataloging is simply one facet of a larger descriptive apparatus. As noted earlier, the preparation of a variety of internal descriptive finding aids is central to the mission of most archival repositories; no archives or manuscript repository could long survive without such tools, and this manual does not in any way supplant or replace that process. *APPM* clearly states that "in such a system, a catalog record created according to these rules is usually a summary or abstract of information contained in other finding aids."[12] This approach is based upon the assumption that, however effective traditional finding aids might be for describing and controlling our holdings, they are a cumbersome way to share information in a broader information retrieval environment which also includes nonarchival materials. If archival repositories were ever going to share data with the broader research community,

[9]See, for example, Steven L. Hensen, "The Use of Standards in the Application of the AMC Format," 31–40 (from a paper delivered at the annual meeting of the Society of American Archivists, Washington, D.C., 1984) (also reprinted in *A Sourcebook on Standards Information: Education, Access, and Development*, edited by Steven M. Spivak and Keith A. Winsell (Boston: G.K. Hall, 1991); and Steven L. Hensen, "Squaring the Circle: The Reformation of Archival Description in AACR2," *Library Trends* 36 (Winter 1988): 539–52.

[10]Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries* (Washington, D.C., 1983) (hereafter *"APPM"*).

[11]Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, 2d ed. (Chicago: Society of American Archivists 1989) (hereafter *"APPM2"*).

[12]*APPM2*, 4 (Rule 0.7).

summary descriptions, or cataloging records, were, at the time, the most effective way to do this.

Perhaps most important, however, is the fact that *APPM* assumes the legitimacy of archival material as part of the larger universe of cultural artifacts. The introduction to the first edition states:

> A fundamental and compelling rationale for this attempt to reconcile manuscript and archival cataloging and description with the conventions of *AACR2* lies in the burgeoning national systems for automated bibliographic description. If these systems, which are largely based on the descriptive formats for books and other library materials outlined in *AACR2*, are to ever accommodate manuscripts and archives a compatible format must be established. This manual is based on the assumption that, with appropriate modifications, library-based descriptive techniques can be applied in developing this format.[13]

Underpinning this is the conviction that it is both appropriate and desirable to catalog and describe archival materials as a part of those systems which describe more traditional library materials such as books, films, serials, maps, sound recordings, graphics, etc. It is thus now axiomatic from the point of view of access to research information that there are logical, vital, and inextricable relationships among all of these materials, and that it is important to show those relationships in a bibliographic context.

Thus, the acceptance of *APPM* is based upon the ways in which it synthesizes basic archival principles into the broader framework of bibliographic description, fine tuning that framework to transform it into a vehicle for specifically *archival* cataloging. This synthesis is based on four major principles:

First, *APPM* recognizes the primacy of provenance in archival description. This principle holds that the significance of archival materials is heavily dependent on the context of their creation, and that arrangement *and description* should be directly related to the materials' original purpose and function. This results in an emphasis on the use of notes, since the complexities of substance and provenance cannot be captured in the sort of brief formulaic encryption that characterizes most bibliographic description. Moreover, the expanded use of notes is consistent with archival traditions of subjective analysis as an essential part of description.

The second principle embodied in *APPM* is that most archival material exists in collectivities or groupings, and that the appropriate focus of the control of such materials is at this collective level. While the practical effect of this is to relieve the archivist of the overwhelming burden of creating literally millions of item-level catalog records, it also supports the principle of archival unity, in which the significance of individual items or file units is measured principally by their relation to the collective whole of which they are a part. A corollary of this is that the most appropriate place for component-level description and analysis is within the archival finding aid, not the catalog record.

The third principle in *APPM* is that archival materials are generally preserved for reasons different from those for which they were created. They are the unself-conscious byproduct of various human activities and consequently lack "the formally presented identifying data that characterize most published items, such as author and title statements,

[13] *APPM*, 1.

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.''[14] Such identifying data is normally created by the archivist in the course of arranging and describing the material. The principal implication for cataloging is to legitimize traditional archival descriptive systems such as finding aids, guides, and registers as sources of cataloging data, and to move the cataloging process away from the literal transcription of information that characterizes most bibliographic description.

Fourth, *APPM* recognizes that there are ''a number of appropriate levels of description for any given body of archival material. These levels normally correspond to natural divisions based on provenance or physical form.''[15] Thus, the rules provide a framework for multilevel description, making it possible for archival catalogers to prepare consistent records regardless of the level of description. Given the overwhelming importance of hierarchy and provenance, this has been an essential feature of *APPM*, and one which recognizes the significance of NISTF's work to embed in MARC AMC the ability to accommodate multilevel description.

The superstructure provided by MARC AMC and *APPM* for the description and control of archival and manuscript materials would have remained an untested abstraction without some concrete evidence that it actually worked. As noted earlier, many archivists in the United States were still deeply suspicious of the library origins and essentially ''bibliographic'' structure of MARC AMC. Fortunately, however, even before NISTF had completely finished its work, several university libraries that were members of the Research Libraries Group were urging RLG and the National Endowment for the Humanities to support a project that would truly test the viability of this new approach. This early project involving Yale, Cornell, and Stanford quickly proved not only to the archival community, but also to a skeptical RLG and the larger library world, that MARC AMC and *APPM* could be used successfully to integrate archival materials into heretofore strictly bibliographic databases.

Early Standards and EAD

While the history of MARC AMC and *APPM* may seem to have very little to do with SGML and archival finding aids, it seems clear that the foundation provided by these events directly enabled the Berkeley Finding Aid Project by providing the impetus for more archivists to begin exploring the broader world of related standards. Most archivists had survived for years in splendid, idiosyncratic isolation and, but for their homage to a few archival principles, saw no need to standardize the way they went about their business. What the experience of MARC AMC and *APPM* showed is that there was a real benefit in being able to communicate archival information—not only among archivists, but also with the larger world of historical scholarship and research. And perhaps most importantly, that archivists were much more closely allied with other information professionals such as librarians than they had realized.

In retrospect, it seems safe to say that few of us involved in these projects in the late 1970s and early 1980s would have predicted the eventual impact of our work. The task that NISTF had designed for itself was initially very simple; to wit, heading off a

[14]*APPM2*, 5 (Rule 0.11).
[15]*APPM2*, 4-5 (Rule 0.12).

potentially unpleasant jurisdictional dispute between the *National Union Catalog of Manuscript Collections* and the repository guide project of the National Historical Publications and Records Commission. The fact that MARC AMC emerged from NISTF's deliberations was a result more of fundamental pragmatism on the part of the task force than any new vision of the future of archival description: it was easier to adapt the MARC format to our needs than it was to develop an entirely new system to underpin any archival "national information system" that might emerge.

Similarly, my own work in recasting *AACR2* to accommodate modern manuscript and archival cataloging was undertaken with rather more modest goals than those that ultimately resulted. Like NISTF, I was simply looking for a practical solution to what seemed like a relatively small problem; there was little sense that this solution would have wider application or appeal. In addition, though I am somewhat chagrined to confess it, there was also little sense of the vital connection between the work of these two projects. It was by the sheerest coincidence that they were roughly contemporaneous. Thus it was that the combined work of these efforts was presented before the world with a distinct sense of uncertainty and unease. The message of the film *Field of Dreams* notwithstanding, there was little assurance that anybody would come, no matter *what* was built.

These concerns ultimately were groundless. The development of the MARC AMC format and *APPM* has transformed the world of manuscripts and archives—certainly in this country, but also to an increasing extent in Canada, Western Europe, and even into Russia. There are over 475,000 records in RLIN alone from hundreds of repositories in the United States and Europe for previously elusive primary resources and special collections. More significantly, the integration of these materials into heretofore primarily *bibliographic* systems is now understood to have been a logical and necessary evolutionary step.

Moreover, these systems are gradually evolving into integrated research tools in which the entire range of cultural artifacts is both accommodated and encouraged, and where information is accessible without regard to the particular physical form that it might take. The world of research and scholarship has become increasingly interdisciplinary and less concerned with whether information is to be found in traditional printed and published forms or in archives, photographs, motion pictures, videotapes, computer files, or museum registers. It is now recognized that information of all kinds is part of a seamless web, and that service to research and scholarship is optimized when there are no artificial restrictions on the particular form that information takes.

As significant as these advances have been for the world of archives and manuscripts—and I wish in no way to minimize them; they have been spectacular—they have nonetheless been constrained by the limitations of the systems in which they have operated. The MARC format is a thirty-year-old database structure that provides a functional standard through which libraries and cultural repositories can communicate descriptive information. Given the relatively short half-life of more modern database systems, one can only wonder at either the foresight of the early developers of MARC or its stubborn durability in a world not given to easy or sudden changes of direction.

Taking the Next Steps

Is there anything wrong with this picture? Some would argue that this approach has endured because it is effective and serves us well. Several years ago, Richard Pearce-

Moses responded to some statements I had been making on the LCSH-AMC listserv regarding the future of bibliographic description:

> I certainly don't expect to see the baby thrown out with the bath water. But I wonder how much the fundamental paradigms of description and access will really change. The format of description may (finally) evolve away from the card catalog style; yet, that style may have remained fairly constant because it's effective in the way it telegraphs information....Even the notion of hyperlinks to full text would not necessarily dictate change to the bibliographic description. At some point all those e-documents are going to be impossible to find, as would a library of several million volumes be useless without some guide. The bib[liographic] database is an abstraction of the documents, and we will continue to need abstraction to avoid having to search the entire haystack.[16]

Pearce-Moses is correct; we will still need pointers, or "metadata," to get to the information that resides within the collections of our cultural repositories, and cataloging of some sort may still be the way to do this. However, our current cataloging systems are ill-equipped to do this on two counts.

First, as noted above, these systems are based on an approach that focuses almost exclusively on the physical characteristics and manifestation of the thing being described. In a world in which bibliographic "items" or works increasingly exist in many different forms simultaneously, this seems curiously out of step. With our users increasingly demanding and expecting more precise content- and subject-oriented retrieval, an approach that ignores these demands seems suicidal.

Second, these systems are, as also noted earlier, unidimensional in that they are based upon the assumption that there is an object in a library and there is a descriptive surrogate for that object, the cataloging record. That is the "system" in its entirety. The catalog record is used to locate a particular book, and the user, armed with call numbers and library locations, goes off in search of it, hoping (often against hope) that the book will be (a) on the shelf, and (b) contain relevant content.

With the recent explosion of Internet-based information, first via Gophers, then Wide Area Information Servers (WAIS), and now the World Wide Web, the disparity between what we *have* been doing and what we *should* be doing has become all the more acute and increasingly difficult to explain. This is particularly true as libraries become less concerned with managing physical holdings and focus more on connecting users with information—wherever that information might be and in whatever form it may exist. The catalog as a purely physical inventory has little relevance in this environment.

We must therefore reexamine not only the role of cataloging, but also the relationship between cataloging and other forms of metadata. A more archival model for cataloging and description is well suited to solving information access and retrieval problems in the new electronic environment. The reasons for this are rooted, not surprisingly, in the essential principles of archival cataloging touched upon earlier.

First, archival cataloging is almost always part of a larger apparatus of description, which includes a variety of finding aids, guides, registers, calendars, etc. Further, archival cataloging is both derived *from* and points *to* finding aids. These finding aids are not only

[16]Richard Pearce-Moses, "AACR2000," message to LCSH-AMC listserv, 3/10/95 3:39 P.M.

a fundamental and long-standing part of archival practice; they also provide the basis for the understanding that it is neither practical nor desirable for a catalog record to carry the entire burden of description. The archival model, with its hierarchically assembled layers of progressively more detailed information, though postulated in electronic pre-history, is highly suggestive of the architecture of modern information systems. If the catalog record is redefined as a window or gateway to other dynamically linked information resources, then the structure of that record and the access points that lead to it may become something entirely different.

Second, in an archival approach focused on the context of creation, descriptive notes illuminate the complexities of substance and content, particularly as they relate to that context. This approach shifts the burden of description towards content, rather than physical characteristics, which as noted above are increasingly irrelevant in an electronic environment. In addition, by using a system of hierarchically structured metadata that can nevertheless be linked to the catalog record (as with the archival finding aid), it becomes easier to accommodate a richer system of subjective analysis.

Third, in an archival approach more focused on collection-level control, the burden for item-level information shifts to forms of metadata beyond the catalog record, whether finding aids, databases, or even subunit-level cataloging. Such an approach can even be used for cataloging large groups or collections of printed materials.

For example, in 1994 the Special Collections Library at Duke completed a Department of Education Title II-C funded project to catalog the 65,000-item Guido Mazzoni collection of eighteenth- and nineteenth-century Italian pamphlets and monographs which had lain essentially untouched since they were acquired in 1948. While there had been previous sporadic attempts to catalog the collection, the combination of its size, the variety of languages represented, and the fact that it was mostly pamphlets had defeated all attempts to bring it under control. This was particularly awkward, and occasionally embarrassing, since the Mazzoni collection is well-known and contains one of the larger collections of *per nozze*[17] known to exist in the world.

Our solution was to treat the collection archivally. Since Mazzoni had originally organized this material into large, generally subject-based groupings, we would create a series of collection-level cataloging records based on those categories following Library of Congress guidelines on collection-level cataloging of printed materials.[18] Item-level control was then provided in a separate non-MARC SGML database that would be linked to the collection-level MARC cataloging records.

This distinctly archival approach recognizes that, however bibliographically significant individual items within the collection might be, what is *most* important is the collection itself. Mazzoni assembled the material with specific purposes and focuses in mind and, to the best of our ability, we maintained the original structure in our processing and

[17]The term "per nozze" comes from the phrase commonly found in the publication title of these pieces, "per le nozze di...," which means "for the wedding of..." The custom of preparing a gift of verse or prose in honor of a couple's wedding originated with the Greeks, who called these wedding compositions "epithalamia." This tradition continued to develop as a social custom and literary genre in modern times only in Italy, with the exception of a few known examples in France, Germany, and Russia. In Italy, the custom of dedicating verse or prose as a wedding gift began in the late fifteenth century among the nobility, and reached its peak in the nineteenth century, when it was very much in vogue among not only the nobility, but the bourgeoisie as well. See <http://scriptorium.lib.duke.edu/mazzoni/nozze.html>.

[18]"Collection-Level Cataloging at the Library of Congress," *Library of Congress Information Bulletin* (9 September 1991) and *Cataloging Service Bulletin*, No. 53 (Summer 1991).

cataloging in a bibliographic approach to *respect des fonds*. In so doing, adequate access to this collection was provided without the necessity of preparing a full cataloging record for each piece.

Some will argue that certain kinds of research needs will not be met with this approach, that some scholars will be disappointed; this is no doubt true. However, this is what we could afford, and most importantly, at long last the entire collection is accessible.

A more archival approach to cataloging, such as was done with the Mazzoni collection and indeed is done every day with a wide variety of manuscript collections, archival records series, and more traditional archival materials, takes the very practical perspective of preferring limited access to *all* of a repository's holdings rather than detailed control over only some. Or at least this was the case until new network-oriented approaches to information access started to emerge. It is becoming increasingly clear that perhaps we can have our cake and eat it too.

Internet Access to Finding Aids

In the early 1990s archivists and special collections librarians began putting finding aids on networked servers where they could be accessed via the Gopher technology that had come out of the University of Minnesota. It seems likely that MARC AMC cataloging records existed for many of these materials in RLIN and OCLC. As useful as these cataloging records were, however, these archivists and librarians knew that the focus of their descriptive efforts was still—as it always had been—in the finding aids and guides that they prepared and upon which the cataloging records presumably were based. There were, however, two essential problems with these finding aid Gophers.

First, there was no way to logically or dynamically link the finding aids to their corresponding catalog records. A potential user, looking at a repository's on-line catalog (often via a telnet connection) would have to exit the catalog and then log onto the Gopher site to see if a finding aid was there.

Second, the Gophers consisted principally of lengthy text files that were very awkward (and occasionally impossible) to search in any meaningful or structured manner. If the file was accessible via WAIS software, there might be a marginally more robust searching engine, but overall, Gophers were scant improvement over writing to a repository and requesting a photocopy of a finding aid.

Daniel Pitti recognized the essential inadequacy of this Gopher/WAIS-oriented approach to archival metadata when he embarked upon the Berkeley Finding Aid Project, the results of which fit smoothly with and build upon the archival standards and information systems developed in earlier years.

In the early days of RLIN's MARC AMC database there were some interesting attempts to enter entire finding aids into the system, but it didn't take long to realize that not only was MARC ill-suited to the level of detail traditionally found in those finding aids, but also that these huge "pseudo-cataloging" records were totally out of proportion to other records in the system and constituted a somewhat intimidating, if not irritating, presence. The records failed to reflect cataloging's purpose as "summary description."[19] Other institutions entered their entire finding aids into the system on a piecemeal basis, adding a separate record for each item; this approach usually fails to comprehend the

[19]*APPM2*, 4 (rule 0.7).

"significance of the whole."[20] Neither technique reflects the essentials of archival description or cataloging.

To archivists' credit, however, these attempts do reflect our fundamental impulse to make more detailed information on archival holdings more widely accessible. That the MARC format is not particularly effective in accommodating this need helped spawn the impulse on which Pitti's project was based.

It seems useful to note that there are rough parallels between the work of the Berkeley Finding Aid Project in defining an SGML Document Type Definition (DTD) for archival finding aids and the work that Elaine Engst began in the survey of archival descriptive practice that led to the NISTF data elements dictionary. In NISTF's case, that data dictionary became the foundation for constructing the elements of the MARC AMC format. EAD, however, defines the larger universe of finding aid data elements that are at the very heart of archival description.

If this were the only point of Pitti's project, there would have been no need to invoke the power and complexity of SGML. Where NISTF separately developed the data dictionary and MARC AMC and then waited to see whether these instruments could or would be used, the EAD project is combining all these processes together. Document definition, structure (according to an already established standard), and navigational tools are all inherently part of the SGML encoding protocols.

The EAD project has had the benefit of learning from NISTF's experience but has also had the advantage of actually defining itself using the very essence of archival hierarchy: the organic hierarchy of the materials themselves as reflected in the finding aids that describe them. Beyond this, however, EAD has the potential to provide for an unprecedented level of structural hierarchy within the overall descriptive apparatus. By this I mean that it is now possible to fully realize the entirety of that apparatus within our evolving electronic information systems so that the unbroken hierarchy of information is accessible from a single point: from the most general access point in a system, to MARC catalog records, to finding aids, to details within those finding aids, and ultimately—if desired—to linked files of digital images of actual collection materials. The catalog records are already available and, as MARC field 856[21] evolves, so too does the capacity to link those catalog records with related information resources on the Internet. What was most critically and obviously missing in this structure is precisely that which EAD provides: a way to encode those layers of metadata (i.e., the finding aids) that have traditionally existed between description at its most summary and general level and the archival material itself. In addition to providing a mechanism for this linkage, this encoding makes possible a level of navigation that was heretofore unimaginable.

Conclusion

As noted earlier, those involved in the early days of archival standards development had little conception of the eventual impact of that work. A process that started with NISTF

[20] *APPM2*, 5 (rule 0.10).

[21] USMARC field 856, "Electronic Location and Access," is designed to provide an electronic address (e.g., a URL or ftp address) for a digital representation of the material or item described in the catalog record or to additional information about the material or item. "The information identifies the electronic location containing the resource or from which it is available. It also contains information to retrieve the resource by the access method identified in the first indicator position. . . . The information contained in this field is sufficient to allow for the electronic transfer of a file, subscription to an electronic journal, or logon to an electronic resource" <http://lcweb.loc.gov/marc/bibliographic/ecbdhold.html#mcrb856>.

defining a set of descriptive elements and mapping those elements into a bibliographic information communications format has now culminated in a project that has further refined and defined those elements, taking the entire apparatus onto a higher plane. The principal difference between then and now is in our expectations. Because of recent advances in technology and the evident direction of those advances, we now have a much clearer sense of the possible. More than that, however, we now have the confidence and courage to project beyond the possible and to realize that our dreams of a truly ''seamless web'' of information can be realized. It is particularly gratifying to realize that this model appears to have broader applicability to the larger world of cultural repositories and libraries.

Those of us in the archives and manuscripts field have only recently and belatedly come to a fuller understanding of our role in this larger world—especially in the new networked electronic environment. While some of this understanding has come from within our profession, we have also relied on the perspective, goodwill, and assistance of those in the library, museum, and computer systems fields. From RLG's willingness to accommodate the unique needs of archival description in order to develop a more complete cultural information system, to the Library of Congress' assistance in the development of *APPM* and the MARC AMC format, there is a recent history of important furtherance from outside organizations and individuals that have contributed significantly towards our professional evolution. The development of Encoded Archival Description is the most recent example of such assistance and may well be regarded by future generations as one of the most important.