

# The Cost of Converting to MARC AMC: Some Early Observations

---

PATRICIA D. CLOUD

THE MARC ARCHIVAL AND Manuscripts Control (AMC) format is rapidly becoming the centerpiece for the new orthodoxy in archival automation. Archivists who wish now or in the future to adopt automated intellectual control procedures or to exchange descriptive information with other archivists will be subject to practical and professional pressures to accommodate the MARC AMC format. New possibilities for local implementations are emerging; the size and significance of online archival databases are growing; national standards, documentation, and guidelines are in place. Everything, it seems, is ready to go.

Nevertheless, critical issues surrounding automation with MARC AMC format remain unexamined. One of the chief issues is the cost in time and resources of using the new format. At a time when so many repositories are planning first-time automation projects using MARC AMC, it may be wise to review some of the experiences of archives who were pioneer users.

The cost of creating MARC AMC records will vary greatly among institutions, depending in part on whether or not a bibliographic utility, such as OCLC or RLIN,<sup>1</sup> or a local system, such as NOTIS,<sup>2</sup> is used; if a local system is used, costs will be affected by the type of system and by the anticipated uses for the MARC records. However, in addition

---

Patricia D. Cloud is Assistant University Archivist, Northwestern University Archives, University Library, Evanston, Illinois.

to the major costs of participating in a bibliographic utility or installing a local system, or the purchase of hardware, software, or subscription fees—issues that are beyond the scope of this essay—staff time will be the most significant cost factor in the creation of MARC AMC records. One recently completed project provides the opportunity for making some tentative generalizations about this important factor.

From 1984 through 1986, twelve major research libraries participated in the retrospective conversion of approximately 21,000 existing data records describing archival and manuscripts materials using RLIN, the bibliographic utility of the Research Libraries Group (RLG). This retrospective conversion project, referred to as the AMC RECON Project, was funded by the National Endowment for the Humanities (NEH) Research Resources Program and the Pew Memorial Trust. It targeted materials currently available for scholarly use and particularly those of significant research value to scholars in the humanities. Participants were asked to refrain from creating AMC records principally related to official university records of administrative activities.<sup>3</sup>

All records created for the project were to conform to a minimum cataloging standard recommended by RLG.<sup>4</sup> The *RLG Standard* established required fields and mandated the use of Steven L. Hensen's manual, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries (APPM)*<sup>5</sup> and the *Anglo-American Cataloguing Rules*, 2d ed. (AACR2),<sup>6</sup> for bibliographic descriptive practices. According to the *RLG Standard*, records created in the RLIN AMC file must contain a main entry, title, date, and physical description, as well as an identification of the institution creating the record.

While this standard allows AMC records to be quite abbreviated, it was assumed that records created for the AMC RECON Project—most of which represented collection level descriptions of significant research materials—would contain considerable detail, and that the average record would require between ten and fifteen name and topical subject entries. It was also assumed that most records would contain collection name; type of materials; creation or span dates; size, occupation and flourishing dates of the collection's creator or creating agency; summary or scope note; list of prominent persons represented in the collection; and notes on any restrictions, provenance, and finding aids.<sup>7</sup>

The *RLG Standard* requires, in addition, that name headings be established according to AACR2. Participants were required to search the Library of Congress Name Authority File (LCNA), which is avail-

### *Cost of Converting to MARC AMC*

able online, for name form and to follow AACR2 for establishing names not found there. Topical subject headings, which comprise the traditional bibliographic approach to creating subject access, were not required. However, if any topical subject headings were present in a record, which was expected to be the case for most records created for the AMC RECON Project, then at least one heading was required to conform either to *Library of Congress Subject Headings*, 9th ed. (LCSH)<sup>8</sup> or to *Medical Subject Headings* (MeSH).<sup>9</sup> This policy would ensure the integration of AMC records with RLIN's files for books, serials, recordings, scores, visual image materials, and machine-readable records. Given this profile of the records to be created following these standards, project participants expected to spend an average of approximately two hours to produce each AMC record.<sup>10</sup>

This estimate was based in part on the experiences of an earlier pilot project: in 1983 and 1984, four RLG members participated in a Title II-C funded project designed to establish the use of the new format in RLIN. These members, Yale, Cornell, Stanford University, and the Hoover Institution, created the first AMC records in RLIN's AMC file during 1984.<sup>11</sup>

The twelve participants in the AMC RECON Project were the libraries of Brigham Young University, Brown University, Columbia University in New York, Dartmouth College, Johns Hopkins University, New York University, Northwestern University, Rutgers University, the State University of New York (SUNY) at Buffalo, the New York Historical Society, and the Bentley Historical Library of the University of Michigan. For the purposes of this paper, each of the participating institutions was contacted and queried as to the length of time they spent creating records and the authority routines they followed. While not all of the information provided by the participants could be analyzed quantitatively, some general observations can be made on the strength of their responses.

Retrospective conversion is an activity firmly established in library practices. It is the process whereby card catalog records of library holdings are converted to machine-readable records, and involves the transformation to MARC format of information stored in the traditional card catalog. Whereas archivists are familiar with the techniques of reappraisal of existing collections, they have had as little experience with retrospective conversion as they have had with MARC format and, therefore, some fundamental observations are in order.

Retrospective conversion of archival and manuscripts material is similar to retrospective conversion of published material in that for

both the basic idea is to create machine-readable records to replace preexisting manual records; however, the procedures followed and, not coincidentally, the costs incurred are very different. While published matter is produced in quantity and, in most cases, with standardized, preexisting bibliographic data, manuscripts are unique.

Librarians considering a retrospective conversion project evaluate the degree to which they are likely to find preexisting machine-readable records for the items in their collection. The Library of Congress, RLIN, OCLC, or other regional bibliographic utilities can be counted on to have records for some or most of the items to be converted. Once a preexisting MARC record is found, all that remains is to verify its information and duplicate it. Most libraries follow this procedure for well over 50 percent of the records they need to convert;<sup>12</sup> only a small percentage of any library collection is "unique," or so rare as to not be cataloged in MARC. In comparison, virtually all archives and manuscripts collections are, by definition, unique. This means that for archives and manuscripts retrospective conversion projects, most records to be converted will be "tagged" or coded in MARC by the repository during the conversion.

For both libraries and manuscripts repositories, the "chief source of information" for cataloging purposes is not the material itself, but rather a preexisting description of it—be that a card, inventory, or some other form of finding aid. However, since Hensen has established the finding aid as the chief source of information for cataloging of archival and manuscripts collections,<sup>13</sup> the distinction between original cataloging and retrospective conversion of these materials can appear somewhat artificial. The assumption underlying the AMC RECON Project was that virtually all records would be created from preexisting finding aids.

Six participants reported that their work on the project took longer than they had anticipated. Crucial factors affecting the amount of time it took to create records seem to have been identified similarly by all repositories, whether or not they found the project unexpectedly time-consuming. The factor that seemed to have the most effect on creation time was the integrity of the existing finding aid (see table 1).

Finding aids used in archives can be catalog cards, descriptive inventories with extensive narratives, or a number of other kinds of lists. Needless to say, the "integrity" of the existing finding aid is independent of its form. Several participants—including those who did not take longer than anticipated to create records—reported that the existing finding aids proved inadequate in a significant number of cases. Seven

## Cost of Converting to MARC AMC

TABLE 1  
FACTORS AFFECTING CREATION TIME OF AMC RECORDS

<i>Factors Affecting Creation Time</i>	<i>Number of Institutions Reporting</i>
Integrity of Finding Aid	8
Authority Work	5
Cataloging Problems	3
Staff Turnover	1

repositories mentioned the need to go beyond the available finding aid and consult the collection or manuscript itself in order to determine critical information such as date, language, and extent. One participant reported the finding aid provided insufficient information for nearly 20 percent of the records; two other participants cited 30 and 40 percent.

The second factor that slowed record creation time was authority work, that is, the process of verifying name and subject headings in local authority files and in LCNA. Depending on the standards established at a particular institution, authority searching for the project could involve as many as nine separate steps (see table 2). All project participants were required to search the LCNA (available on RLIN as the Library of Congress Resource File [LCR]) and LCSH, in the event that topical headings needed to be created. Most of the archives reporting their procedures also checked local authority and bibliographic files in addition to LCNA. Procedures beyond this varied considerably. Repositories that checked further consulted at least one of the following sources: OCLC or RLIN Books files, the RLIN AMC file, or manual files such as the *National Union Catalog of Manuscript Collections* (NUCMC).<sup>14</sup>

There have been some difficulties associated with using NUCMC as a source for establishing name forms. Traditionally, archival and manuscript collections use the fullest possible form, with qualifying information where available, and NUCMC name entries have been so established. This procedure, however, is at variance with AACR2, which calls for establishing the most common form of name—which may or may not be the fullest form. Only recently, in conjunction with preparation to put NUCMC online through RLIN, have NUCMC entries been contributed to LCNA. The entries for names occurring in manuscript collections are being established using the fullest form; nevertheless, NUCMC remains an inadequate and possibly misleading

TABLE 2  
SOURCES FOR AUTHORITY WORK IN AMC RECORD CREATION

<i>Sources Used for Authority Work</i>	<i>Number of Institutions Using</i>
LCSH*	12*
LCNA	12
Local Authority File	6
Local Bibliographic File	6
RLIN AMC	3
RLIN Books File	2
OCLC Bibliographic File	1
NUC	1
NUCMC	1

\*Three repositories did not report details concerning their work. It is assumed that they followed the RLG Standard and searched LCNA and LCSH; all other figures reflect actual responses.

source for names which have previously been established strictly according to AACR2.<sup>15</sup>

The extent of authority searching is dependent on the likelihood of finding a given heading in an established authority file. Since authority files are generally designed to serve the bibliographic community and, therefore, contain primarily authors' names, many archival repositories will discover relatively few headings established for their collections. Extra searching will extend the time required for authority work, as will the need to use manual, rather than automated, authority files. Manual authority files can exist on cards or microfiche (the Library of Congress makes its name authority files available on microfiche as well as online), or indeed in lists such as RLG's "Form Terms for Archival and Manuscripts Control."<sup>16</sup> All types of authority files, however, should permit access to an established name form through variant forms and should clarify any additional difficulties the name might present. Files available online in MARC format are often easier and faster to use than manual files. Two repositories reported that authority work took roughly a third of the total record creation time.

Cataloging problems—including the difficulty of creating corporate headings according to AACR2—were mentioned as significant factors by three repositories. The final element, staff turnover, was reported by only one repository. However, this factor may emerge as a far more significant setback as more projects are undertaken and it becomes more difficult to find staff trained to code AMC materials.

### *Cost of Converting to MARC AMC*

While knowledge of the various cataloging tools and techniques is crucial and can be learned at some cost, the coding of AMC format requires a sophisticated approach to the material being described. There is a strong likelihood that candidates for jobs on retrospective conversion projects—often part-time temporary positions—will have some but not all of the proper background and training; staff turnover in these positions could easily have a profound effect on project schedules.

There are several ways to compile statistics on record creation time; project participants, who were not required by the grant to keep statistics, reported their statistics in two ways. Two participants kept track of each increment or stage of the record creation process and can report minutes spent coding, authority searching, and inputting. A comparison of these statistics shows the two institutions had very similar experiences with the project (see table 3). The Bentley Historical Library and Northwestern University Archives had virtually identical authority procedures. The shorter creation time reported by Northwestern University may be due in part to the higher proportion of single-item or small collection records created by Northwestern (80 percent of 1168 records, as compared to 20 percent of 3000). These records, which represented small collections or series, posed minimal difficulties in terms of description and subject access, and could be created far more quickly than those representing larger collections. Another possible explanation is that Bentley's statistics were based on a sample taken one month after the project began, whereas Northwestern's statistics were compiled using figures from the entire project. The average time needed for record creation dropped over the course of the project at Northwestern as catalogers gained familiarity with AMC and with the various work routines imposed by the project.<sup>17</sup> Time spent on authority work by Bentley may reflect their reliance on manual, in-house authority files, whereas Northwestern accomplished virtually all authority searching online.

TABLE 3  
CREATION TIME STATISTICS (REPORTED IN MINUTES PER TASK)

<i>Institution</i>	<i>Creation Time per Task</i>		
	<i>Coding</i>	<i>Authority Work</i>	<i>Input</i>
Bentley	14	13	18
Northwestern	15	4	9

While it is possible to record minutes spent on various aspects of record creation, it is cumbersome, and becomes nearly impossible when problems arise with a record. Cataloging questions, complicated authority work, and the need to consult collections directly, all can kick records out of the workstream. Problem solving can take many hours and is not an activity that readily lends itself to the careful recording of statistics. Therefore, some repositories calculated their record creation statistics after the fact by dividing the number of records created by the number of staff hours spent on the project.<sup>18</sup> Three repositories reported their statistics in this manner, and of these three, the total time per record varied considerably, from a low of 14 minutes per record at SUNY-Buffalo to a high of 2.9 hours at Dartmouth (see table 4).

The low creation time at SUNY-Buffalo is explained in part by the extraordinary brevity of their records: most of the Suny-Buffalo records recorded single items with from three to five added entries. SUNY-Buffalo profited in addition from earlier cataloging experience, having begun a project with monographic cataloging on OCLC in 1982. At Northwestern, project staff had minimal previous cataloging experience and were required to become acquainted with AACR2 and RLIN after the starting date of the project.

Authority searching was not considered a problem at SUNY, despite their seven-stage searching process. They used automated authority files, which are helpful in keeping search time under control. In addition, most of their records were for literary manuscripts; hence, most of the names were readily discovered in the standard authority files.

### Some Considerations

The implications of these statistics are not particularly hard to discern. For those institutions evaluating the total projected cost of

TABLE 4  
CREATION TIME STATISTICS (REPORTED PER RECORD)

<i>Institution</i>	<i>Time per Record</i>	
	<i>In Minutes</i>	<i>In Hours</i>
Dartmouth	175	2.9
Northwestern	165	2.7
SUNY-Buffalo	14	.24

## *Cost of Converting to MARC AMC*

creating MARC AMC records, it would be wise to consider the following issues.

### *Integrity of the Existing Finding Aids*

Any repository considering a retrospective conversion project should evaluate their finding aids: will dates of creation, extent, nature, and language of material be readily available from the existing finding aid? If not, will a less rigorous standard for record creation be acceptable? What logistical difficulties are posed by the need to consult the items directly? At what point will something like "original cataloging," or description directly from the manuscript or archival collection, be either desirable or feasible? If the existing finding aids are inadequate, problems and time spent will multiply.

While for retrospective conversion it is common practice to rely whenever possible on a preexisting description of the materials, often it may be necessary to consult the items themselves. Depending on the type of manuscript, single items may be converted very rapidly by direct consultation: SUNY-Buffalo coded directly and efficiently from individual manuscripts for most of the duration of the project. Direct consultation of archival and manuscript materials is most troublesome in large collections requiring considerable examination.

### *Authority Work*

The procedures associated with authority work are unfamiliar to most archivists. Although some repositories may have already established authority files, many will need to begin establishing such files when they embark on a retrospective conversion project. It will be important to establish what standards will be adhered to, which files will be consulted, and in what form these files will be.

One crucial question is, how strictly should a repository adhere to national authority standards? It has been pointed out that automated systems have a "low tolerance for idiosyncrasy and individualism."<sup>19</sup> As more and more research libraries adopt local automated systems, and as more efforts are undertaken to exchange information through machine-readable records, archivists face the option of integrating their research materials into a growing body of readily accessed bibliographic records. In order to do so, they must adopt some level of standardization, local or national. In certain cases, locally established headings may be preferred, even where a repository is using one of the national bibliographic utilities. The relative rigor that standardization imposes will have a significant impact on record creation time. Repositories with access to

LCNA online will spend far less time searching than those who use LCNA on microfiche. Another significant issue will be the number of names occurring in a given collection that are likely to be established headings in LCNA or in other local authority files.

### *Experience*

To say that a retrospective conversion project has taken longer than expected is not to say that it was not worth the time taken; a number of benefits can result from the experience.

Previous experience with MARC format, and training of some sort in the use of AACR2 and MARC AMC are obvious advantages in any retrospective conversion project. However, the lengthy process of learning MARC AMC and AACR2, which can so dramatically affect the time required by such projects, may pay dividends long after retrospective conversion is complete—particularly if a repository decides to adopt the AMC format for its ongoing procedures. In addition, a retrospective conversion project can help identify difficulties with preexisting guides and finding aids—a systematic review and elaboration of existing finding aids to accommodate an externally devised descriptive standard can be of immense and enduring value. The same may be said of the imposition of authority files. How AMC will affect users of archives and manuscripts material, in terms of improved access, or repositories, in terms of a potential increase in user demand, remains dubious. The databases of the bibliographic utilities are not, as yet, routinely consulted by researchers, nor is it clear when or how these databases will be made directly available to patrons. Of great interest for the next several years, however, will be the measurable impact, in terms of queries and direct consultation, that may be traced to the presence of AMC records in both local and national systems.

Measuring the benefits of using MARC AMC will be as complex as measuring the costs. It is unlikely that the whole story can be forecast by any one factor—be it record creation time or increased patron inquiries—although evidence of these will be eagerly awaited.

## References

1. The Online Computer Library Center (OCLC) and the Research Libraries Information Network (RLIN) are two of the largest and most influential utilities. They provide resources, primarily to libraries, for cooperative cataloging, interlibrary loan and collection management. Participants subscribe to the utilities and must meet system-wide standards for the creation of records.

## Cost of Converting to MARC AMC

2. Northwestern Online Total Integrated System (NOTIS) is a software package used, generally by libraries, for control and access of MARC records. Local systems are often marketed with hardware, as well as software, and are referred to as "turnkey" systems. Local systems do not allow direct access to other institutions' databases, whereas bibliographic utilities do allow such access. For more information, see Sahli, Nancy. *MARC for Archives and Manuscripts: The AMC Format*. Chicago: Society of American Archivists, 1985; and \_\_\_\_\_. "Interpretation and Application of the AMC Format." *American Archivist* 49(Winter 1986):13.
3. This information is drawn from the grant proposal which governed the AMC RECON Project, which is not available for public consultation.
4. Research Libraries Group. *RLG Standard for Use of the AMC Format* (RLG Document 85-67). Stanford, Calif.: RLG, 1985.
5. Hensen, Steven L. *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies and Manuscript Libraries*. Washington, D.C.: LC, 1983.
6. Gorman, Michael, and Winkler, Paul. *Anglo-American Cataloguing Rules*, 2d ed. Chicago: ALA, 1978.
7. See reference 3.
8. Library of Congress. *Library of Congress Subject Headings*, 9th ed. Washington, D.C.: LC, 1980.
9. National Library of Medicine. *Medical Subject Headings*. Bethesda, Md.: National Library of Medicine, 1960- .
10. See reference 3.
11. For a more detailed discussion of this project, see Engst, Elaine. "The Development of the RLIN AMC System." *Documentation Newsletter* 9(Fall 1983):1-6; and "AMC Offers New Access to Nation's Archival Resources." *The Research Libraries Group News* 5(Sept. 1984):3-5.
12. Peters, Stephen H., and Butler, Douglas J. "A Cost Model for Retrospective Conversion Alternatives." *Library Resources & Technical Services* 28(April/June 1984):149-62.
13. The chief source of information for single manuscripts is, as with published material, the item itself. Hensen, *Archives, Personal Papers, and Manuscripts*, p. 8.
14. Library of Congress. *National Union Catalog of Manuscript Collections*. Washington, D.C.: LC, 1959- .
15. On recent developments concerning NUCMC and LCNA, see Hensen, Steven L. "The Use of Standards in the Application of the AMC Format." *American Archivist* 49(Winter 1986):31-40.
16. Research Libraries Group. *Form Terms for Archival and Manuscript Control*. Stanford, Calif.: RLG, 1985.
17. The opposite effect has been noted in certain bibliographic retrospective conversion projects wherein record creation times actually increased over the course of the project due to the tedium of the work. See Peters, and Butler, "A Cost Model for Retrospective Conversion Alternatives," p. 155.
18. Northwestern University Archives calculated record creation time both incrementally and after the fact.
19. Hensen, "The Use of Standards," p. 32.

This Page Intentionally Left Blank