Maximizing Metadata
Exploring the EAD-MARC Relationship

Katherine M. Wisser and Jennifer O'Brien Roper

Encoded Archival Description (EAD) has provided a new way to approach manuscript and archival collection representation. A review of previous representational practices and problems highlights the benefits of using EAD. This new approach should be considered a partner rather than an adversary in the access-providing process. Technological capabilities now allow for multiple metadata schemas to be employed in the creation of the finding aid. Crosswalks allow for MARC records to be generated from the detailed encoding of an EAD finding aid. In the process of creating these crosswalks and detailed encoding, EAD has generated more changes in traditional processes and procedures than originally imagined. The North Carolina State University (NCSU) Libraries sought to test the process of crosswalking EAD to MARC, investigating how this process used technology as well as changed physical procedures. By creating a complex and in-depth EAD template for finding aids, with accompanying related encoding analogs embedded within the element structure, MARC records were generated that required minor editing and revision for inclusion in the NCSU Libraries OPAC. The creation of this bridge between EAD and MARC has stimulated theoretical discussions about the role of collaboration, technology, and expertise in the ongoing struggle to maximize access to our collections. While this study is a only a first attempt at harnessing this potential, a presentation of the tensions, struggles, and successes provides illumination to some of the larger issues facing special collections today.

The Association of Research Libraries' spotlight on special collections in "Building on Strength: Developing an ARL Agenda for Special Collections" has brought to the forefront many of the challenges and strengths that special collections have to offer to the research library setting (Association of Research Libraries 2001). The concern over access issues expressed in the Action Agenda highlights the need for a renewed focus with a realistic understanding of the process and the expenditure of resources. The agenda has confirmed a long-standing consensus among curators and archivists that access is a primary goal in the archival endeavor.

In the past, archivists have largely been left to their own creative devices in communicating the contents of their collections to the public. To facilitate access, they created card catalogs, inventories, registers, indexes of various types, calendars, file plans, and the ubiquitous, nebulous "finding aid." Print catalogs made it possible to advertise beyond the confines of the repository, but this was spotty exposure at best. Repositories needed to work steadily to expose potential researchers to the location, content, and contextual information of collections.

Online catalogs provided one ray of hope for repositories. Repositories could create collection-level representations of their holdings to entice patrons to their doors. This served two purposes. First, patrons using an online catalog to

Katherine M. Wisser is an NCSU Libraries Fellow, North Carolina State University Libraries, Raleigh. **Jennifer O'Brien Roper** is Special Collections Cataloger, Wake Forest University, Winston-Salem, North Carolina.

Manuscript received October 16, 2002; approved February 28, 2003.

research a topic would be directed to the special collections material because of its relevance, when they previously may not have considered archival or manuscript materials as germane to their research. Second, with the advent of the Web and the suspension of geographical and temporal limitations, distant patrons could find the manuscript or archival material that they sought. Electronic union databases aided in this effort, but patrons needed to know that those databases existed and how to search them.

Catalog records, though, make poor substitutes for the traditionally rich descriptive documents that archivists create to represent their collections. The limitations of MARC, including field- and record-character limitations in some integrated library systems and a lack of hierarchical structure, mask the true intellectual work of arrangement and description done by archivists. MARC was created to make representations of items, while repositories wanted to describe collections of items. Soon archivists got the 'Web-bug' and started re-creating their finding aids in hypertext mark-up language (HTML) to present them to a wider audience. These representations were flat, though, and yielded only nonprecise searching capabilities.

It was not until 1995 that the seeds for Encoded Archival Description (EAD) really took root with an online finding aid project at Berkeley. At this point the creation of a metadata mark-up language became a marriage between archival theory and information technology. EAD was intended to supplement—not replace—existing representational structures. Throughout the evolution of information technology applications, including Gopher sites and HTML, the relationship between the online catalog record and the electronic finding aid was consistent (Encoded Archival Description Working Group 1999, 6). Technological advances allowed practitioners to refine that relationship while developing better representations at the same time.

North Carolina State University (NCSU) watched these developments with interest and decided to commit fully to EAD in 2001. In its implementation, the NCSU Libraries sought to define commonalities existing between metadata representations, to shape a workflow that would take full advantage of expertise, and to enhance productivity. With collection level MARC records affirmed as the local policy, staff felt it would be desirable to create a consistent crosswalk from EAD to MARC (Woodley 2000; St. Pierre and LaPlant 1998). In other words, an EAD template needed to be constructed in such a way that the information entered would be consistent with the requirements of a MARC record format. That process should include both the mapping of descriptive elements and the appropriate use of standards for content.

This paper focuses on two aspects of this project. It includes both the process we undertook and a discussion of lessons learned from that process. A detailed description of the methods used includes the collaboration of standards, the formation of a template, both the output to and input from the cataloger, and the examination of that data for effectiveness. The discussion section offers insight into the commonalities between descriptive metadata schemata and the methods of creating them, and the ways in which people from varying perspectives on a project can offer basic knowledge that will enhance each other's understanding of their own objectives. Coordinating two gateways of access to collections provides an arena for catalogers and archivists to learn from each other while streamlining eventual processes to the benefit of each metadata schema.

Process

To achieve both aims, members of the NCSU Libraries' Cataloging and Special Collections departments reviewed the collection-level information included in an EAD document in conjunction with standards for archival cataloging. These included *Archives, Personal Papers, and Manuscripts* (Hensen 1989) to understand the descriptive structure of manuscript cataloging standards and the MARC21 structure for manuscript and archives to find the commonalities between an EAD document and a cataloging record.

Based on these reviews, the departments constructed the template using the *relatedencoding* attribute for the <archdesc> element and strategically placing *encodinganalog* attributes throughout the <archdesc> section of the EAD instance. Figure 1 provides a snapshot of the <archdesc> section of an EAD finding aid using these attributes.

The values for the *encodinganalog* represent the corresponding MARC field for the data within that element. Table 1 presents a summary of the mapping between MARC fields (*encodinganalog* values) and elements from EAD. A complete table representation of USMARC to EAD crosswalking is available in the *Encoded Archival Description Application Guidelines*, Version 1.0 (Encoded Archival Description Working Group 1999, 240–42).

The formulation of the template was by far the most labor-intensive step of the crosswalk process. Once a standard template had been established, staff used NoteTab Pro's clip library capabilities to cloak the encoding of attributes. For instance, the technical support team member created scripts to apply the appropriate *source* and *encodinganalog* attributes and values automatically for the <controlaccess> section of the finding aid. The encoder selects which kind of index term is being entered (i.e., personal name, corporate name, subject, geographical place, or genre form), and the script places the appropriate attribute values in the element. For stable MARC fields, such as the 5xx fields, the clip pro-

gramming scripts are straightforward. For dynamic ones, *such as the 1xx and 6xx fields, the programming is more com*plex, and the clips pose questions that provide the parameter for the element. The detailed programming throughout this process was achieved through the work of the information technology specialist within the department, and he remains available for necessary adjustments to the clips and template as the EAD implementation coordinator and special collections cataloger work on fine-tuning the process.

Generally, during the processing of collections and encoding of finding aids, the crosswalk is invisible to the archivist. Processors follow the template and are not required to have a full knowledge of the correlation between EAD and MARC. This allows the special collections processing staff to concentrate on the task of arranging and describing rather than trying to negotiate two metadata languages.

Once an EAD instance has been completed and parsed against the EAD document type definition, the instance is run through an Extensible Stylesheet Language for Tranformations (XSLT) that produces a .txt file that serves as the basis for the creation of the catalog record. That .txt file includes the MARC fields and the corresponding textual information, but does not include any EAD tags. The .txt file is an excerpt from the EAD instance for the information that both EAD and cataloging metadata structures share.

The mapping of the template at NCSU focused solely on the collection-level information in the finding aid. Because NCSU has few if any catalog records for its special collections materials, a policy decision to represent those materials at collection-level (at the outset) was made. The ease of mapping collection-level information also came into play. For more sophisticated crosswalking, series-level or item-level cataloging can be done. This would require a separate XSLT program that targeted specific areas of the EAD instance. As well, item-level descriptions are not common for collections, particularly at NCSU. In a crosswalk from item-level description from a finding aid, the cataloger would be provided with less information for the item (i.e., physical description information) and would be required to revisit the item itself. Archival description at NCSU is necessarily "top heavy," and therefore the collection-level description was considered most suitable for our access needs.

The cataloging process begins once the .txt file has been created. The .txt document contains the MARC field tags followed by the relevant information for that field. Thus, the majority of the work has been done and is provided for the cataloger to copy and paste into the appropriate fields in the MARC record. It is not necessary for the cataloger to spend time searching for the appropriate MARC fields in which to put information. Furthermore, the information provided is complete in terms of descriptive accuracy of the collection, including subject analysis.

```
<archdesc level="collection" langmaterial="eng" relatedencoding="MARC">
<did id="d_summary">
<head>Descriptive Summary</head>
<repository>
<corpname encodinganalog="710">North Carolina State University. Libraries. <subarea>Special
Collections Dept.</subarea></corpname>
</repository>
<origination label="Creator">
<persname encodinganalog="100">Brooke, A. Wayne</persname>
</origination>
<unittitle label="Title" encodinganalog="245">The A. Wayne Brooke Collection, <unitdate
type="inclusive">1948 - 1986</unitdate></unittitle>
<unitid label="Call Number" encodinganalog="090">MC 268</unitid>
<physdesc label="Extent">
<extent encodinganalog="300">3 archival boxes, </extent><extent>3.75 linear feet</extent>
</physdesc>
```

Figure 1. Excerpt from an EAD Document

Table 1. MARC Fields and EAD Elements

MARC Field	EAD element
1xx	<origination>[<persname>, <corpname>, <famname>]
245	<archdesc><did><unittitle>
710	<archdesc><did><repository><corpname>North Carolina State University. Library. Special Collections Dept.
090	<archdec><did><unitid>
300	<physdesc><extent>
520$a	<abstract>
506	<admininfo><accessrestrict>
530	<admininfo><altformavail>
541	<admininfo><aquinfo>
561	<admininfo><custodhist>
524	<admininfo><prefercite>
500	<admininfo><userestrict>
545$a	<bioghist>
351	<scopecontent><organization>
6xx	<controlaccess><list><item>[<subject>, <persname>, <corpname>, <famname> or <geogname>]
655	<physdesc><genreform>

This is not to say that the cataloger can simply copy and paste from the .txt file to the MARC record, create holdings, and be done. Although the bulk of the work may be completed, the devil is in the details, and it is the details to which the cataloger must pay close attention. First, while the .txt file does contain MARC field tags as well as most of the subfield tags, it does not provide fixed fields or indicator tags and lacks some subfields. Since the fixed fields and indicators are critical for user searching and accurate search limitation, the cataloger must consider the content of the record and accurately supply the missing data. This illustrates the need for an experienced cataloger to convert the .txt file to the MARC record, as these important details might be easily neglected or omitted by someone less familiar with MARC and its functionality. Figure 2 and figure 3 demonstrate the transition from the .txt file to the MARC record.

The second area of detail that the cataloger needs to verify is the accuracy of the access point forms. This applies equally to subject access and to personal and corporate

040 $a NRC $c appm $c NRC

090 $a MC 268

100 $a Brooke, A. Wayne

245 $a The A. Wayne Brooke Collection, $f 1948 - 1986 $h [manuscript]

300 $a 3 archival boxes, $a (3.75 linear feet)

351 $a The papers are divided into eight series with a bibliography appended. The Correspondence series consists of letters from two periods of A. Wayne Brooke's involvement with the Selective Sequence Electronic Calculator (SSEC). The first is during the period in which he worked on the SSEC, and the second was when he was preparing a manuscript detailing the importance of the SSEC to the history of computing. The Writings series contains a draft of the unpublished manuscript detailing the importance of the SSEC to the history of computing. The Research Notes series contains notes and charts that refer to documents in the Writings series. A>

Figure 2. Excerpt from the .txt Document

Type: p Bib l: c Enc l: I Desc: a Ctry: ncu Lang: eng Mod: Srce: d
Form: Dat tp: i Dates: 1948 1986 Control:

040; ;a NRC $e appm $c NRC $
090; ;a MC 268 $
049; ;a NRCC $x JOR $
100; 1;a Brooke, A. Wayne $
245; 14;a The A. Wayne Brooke Collection, $f 1948-1986 $h [manuscript] $
300; ;a 3 $f archival boxes $a (3.75 $f linear ft.) $
351; ;a Organized into seven series: Correspondence, 1950-1986. Writings, 1952-1984. Research notes. Selective Sequence Electronic Calculator. History of computing. Organizations, 1950-1987. Miscellaneous. Walk East on Beacon ; $b Chiefly chronological arrangement. $

Figure 3. Excerpt from the MARC Record

name access. Depending upon the authority acumen of the EAD creator, names may appear in invalid or outdated forms, and the cataloger must catch these anomalies to ensure proper collocation in the catalog. The cataloger is also responsible for notifying the EAD coordinator when authority errors are encountered. Most of the authority errors identified at the NCSU Libraries during this process were the result of a one-time inaccuracy in the inputting of information into the style sheet, which was then used by various creators. Thus, timely recognition of authority errors by the cataloger can dramatically improve the quality of future .txt files.

Aside from occasionally editing biographical or historical notes when the system is unable to accept the information due to its length, the cataloger does not create and/or alter the information provided in the .txt file. Rather, it is more important that the cataloger be attuned to what information is *not* provided and appropriately fill in those gaps. This process allows the person most familiar with the source materials, the processor/EAD encoder, to provide the best descriptions and analysis possible, and the person most familiar with the MARC standard, the cataloger, to create the optimal access tool for the library's catalog.

The connection between the MARC record and the EAD instance is complete when the holdings record is created, including an 856 field pointing to the EAD-encoded finding aid. Just as EAD is not considered a sufficient replacement for MARC, the MARC record should serve as a gateway for the patron to retrieve further information about a collection. Direct access to the information-rich finding aid is one way to help patrons assess the suitability of a resource for an information need.

Discussion

Encoded Archival Description was created with the structure of the materials in mind. The goal of EAD is to describe the archival collection accurately and completely in accordance with archival theory and practice while taking full advantage of technological innovations. Archives and manuscript collections provide different challenges from secondary sources in terms of complexity and volume. Archivists add intellectual value to their collections by representing them in coherent levels of description to communicate the relationship between the materials. Records do not exist in isolation from one another. They are created by individuals or institutions; they are sources of communication—conversations between agents; they are the human record and are necessarily as complex as humans. In order to make the records of an individual or an institution accessible for researchers, archival theory addresses the inherent nature of the material and provides parameters for arrangement and description that illustrate that nature.

The MARC record, on the other hand, was created to automate cataloging done according to existing and accepted standards of bibliographic description. The information provided is intentionally straightforward and uniform in appearance. The value of the MARC record is not intrinsically in any one record, but rather in the compilation of many MARC records into one database. It is through the power of the catalog that users are able to discover and identify materials on their topic from among thousands of other library holdings.

A primary difference between the EAD and MARC standards is the conceptual level at which each metadata language exists. Archival description encompasses several different conceptual levels, whereas a bibliographic description (represented by a MARC record) exists on one level. The parallel between the two, which allows for the crosswalk application described above, is the collection-level description. The archival work done at the subcomponent level informs that collection-level description, which in turn informs the bibliographic description.

Because EAD and MARC standards are addressing different goals, they are not interchangeable. While both are

metadata schemata designed to create a surrogate for a variety of material, they do that task quite differently. The EAD finding aid creates a surrogate that is the equivalent of a model replica of the materials. The user can see the material as a whole, as well as get an in-depth glimpse into the structure and complexity of the material. Alternately, the MARC record provides the equivalent of a photograph of the material. The user can see the material as a whole, as well as a glimpse of the description, but the MARC record is flatter and less complex than the EAD surrogate. With archival material, the power and value of the MARC record lies in its ability to provide a serendipitous connection between users and materials via the catalog. The advantages of MARC cataloging include a more mature, robust standard and practitioners who understand the importance of the details necessary to fully exploit the MARC standard. Both metadata standards have strengths and weaknesses, but acting in concert, the combination of structures provides fuller access to special collections materials and a more comprehensive and intelligent depiction of the collections.

In order to benefit from the intimate knowledge a processor gains when arranging and describing a collection, a work pattern was established that allows information to flow from one level to another. To achieve this flow, archival processors work with the EAD implementation coordinator to assure that collection-level information input into the EAD document conforms with the MARC requirements output to the .txt file. The EAD coordinator and the cataloger balance the collection description, collaborating to serve the goals of each metadata structure.

In conjunction with this process of documents informing each other, the crosswalk process has given the practitioners a chance to learn from each other. In areas such as authority and controlled vocabularies, the cataloging processes have informed the EAD encoding and the implementation of consistency across departmental finding aids. In creating MARC records for collections, the encoding has helped catalogers better understand the structure of archival description and the nature of manuscript and archival materials.

The issue of authority is an excellent example of how this process requires interaction between the standards and the practitioners, not just simple translation of encoding standards. In NCSU's implementation, the departments encountered problems with the mapping of the <repository> tag to the 710 field for the NCSU Libraries' Special Collections Department name. While the form used in EAD was part of the template, the catalogers consistently had to change it to the authoritative form of name to ensure consistent collocation in the catalog. After several instances of discovering that the .txt file had yielded a nonauthority form, the cataloger approached the EAD coordinator to discuss the issue. Through this consultation process, the cataloger learned that the EAD process is not driven by uniformity. The EAD coordinator was able to discern how important the authoritative form was to the cataloging record and worked to integrate some of the MARC-driven uniformity into the EAD template.

On a conceptual scale, the issue of authority control provided even more opportunity to exchange expertise. Archival description does not require authoritative forms, but it can benefit from their use. Users should not have to use multiple variations in terminology to search for the same concept (person, place, subject, etc.), and authoritative forms can strengthen links between disparate collections. In particular, the use of the Library of Congress Name Authority File has taught both the EAD coordinator and the cataloger something about each other's task. Manuscript materials are collected from a variety of institutions and persons, not all of whom have recognized authoritative forms. Recognizing the need for an authoritative form of name and providing information to create the authority record allow each person to "trade places" with the other and appreciate the expertise that each brings.

In addition, the source of authority forms became an important issue. From an administrative standpoint, the EAD coordinator needs to supply the source of the controlled vocabulary and to communicate that source to the cataloger in order to assure that vocabularies were being used consistently. The addition of a *source* attribute for <controlaccess>, where multiple vocabularies were used, provided this framework.

The cataloging procedure just described is in an early phase of development. With full knowledge that it is possible for the process to be more fully automated, this initial procedure served as a pilot to determine if the implementation would succeed. In an upgraded iteration of this process, the MARC record will be directly imported into the integrated library system. This should further expedite the mechanical process and allow the cataloger to focus more intensively on the cataloging details.

Conclusion

When the MARC standard was introduced, it was in itself an innovative idea. Two decades before the Internet explosion, the concept of digitizing information previously only available in print was revolutionary. By the time Encoded Archival Description was created, digitization of information was the norm, not the exception. The implementation of EAD is not as simple as encode-and-go, though. Metadata standards do not exist in isolation, and practitioners benefit from an investigation of these standards' commonalities.

In a world of cut-and-paste and application toggling, both physical and mental processes have changed. The

collaborative process has become increasingly important in terms of maximizing expertise and workflow and creating a congruence of standards and objectives. When the NCSU Libraries sought to implement EAD, it was seen as an opportunity for cooperation and partnership between two types of metadata (EAD and MARC) and two types of library professionals (archivists and catalogers).

Some may view this workflow as eliminating the cataloger from the process of defining the content of the record. This is true to a certain extent, but this process aims to prevent the duplication of effort. Since the collection has already been described for one type of discovery tool, the cataloger is able to use this information and concentrate on refining the resultant MARC record. At the same time, the information generated during the processing and encoding of a collection provides detailed collection representation and ample information from which a catalog record can be created. Other information is standardized and does not need the attention of either the EAD coordinator or the cataloger beyond its established format. A collaborative approach allows both EAD and MARC implementations to learn from each other, to develop along the same descriptive lines, and to create coherent representations of the department's holdings. Each effort allows the representation process to flow seamlessly between two standards and to enhance service to the user by building a more sophisticated gateway to the collections.

Future research on this project should include examining the effectiveness in real terms of multiple access points to collections. Does patron interest in our collections increase? Is that interest a direct result of the creation of multiple metadata representations of special collections materials within NCSU Libraries? Throughout the creation of this process, both the cataloger and metadata coordinator worked under that assumption that increased access would increase usage. Does that assumption bear true? Reflection on the effectiveness of increased metadata representation is one of the next great frontiers in information science research.

Works Cited

Association of Research Libraries. 2001. Building on strength: Developing an ARL agenda for special collections. In *A proposed action agenda for special collections*. Accessed Aug. 6, 2002, www.arl.org/special/action.html.

Encoded Archival Description Working Group of the Society of American Archivists. 1999. *Encoded archival description application guidelines*, version 1.0. Chicago: Society of American Archivists. 6.

Encoded Archival Description Working Group of the Society of American Archivists. 1998. *Encoded archival description tag library*, version 1.0. Chicago: Society of American Archivists.

Hensen, Steven L. 1989. Archives, personal papers, and manuscripts: A cataloging manual for archival repositories, historical societies, and manuscript libraries. Chicago: Society of American Archivists.

St. Pierre, Margaret, and William P. LaPlant, Jr. 1998. Issues in crosswalking content metadata standards. Bethesda, Md.: National Information Standards Organization. Accessed Feb. 20, 2003, www.niso.org/press/whitepapers/crsswalk.html.

Woodley, Mary. 2000. Crosswalks: The path to universal access?" In *Introduction to metadata: Pathways to digital information*, Murtha Baca ed. Los Angeles: Getty Institute. Accessed Feb. 20, 2003, www.getty.edu/research/institute/standards/intrometadata/2_articles/woodley/index.html.

Index to Advertisers