EAD Testing and Implementation at the Library of Congress

MARY A. LACY and ANNE MITCHELL

Abstract: This case study chronicles the Library of Congress's participation as an early implementer of Encoded Archival Description (EAD). The activities described include testing the alpha version of the EAD DTD, tagging approaches used in two test finding aids in the Manuscript and the Prints and Photographs divisions, a pilot project in the Manuscript Division which encoded ten finding aids, exploration of how EAD accommodates finding aids in various formats, and the work of the library's EAD Task Force in resolving organizational and technical issues. The authors' conclusions are based on their two-year involvement with EAD at the Library of Congress.

About the authors: Mary A. Lacy is an archivist in the Preparation Section of the Manuscript Division of the Library of Congress and serves as the division's liaison to the National Digital Library Program. She has been active in the digital conversion and distribution of Manuscript Division finding aids since 1995 and has participated in the library's testing of EAD and encoding efforts during the last two years.

Anne Mitchell is a cataloger of pictorial materials with a specialization in architecture, design, and engineering collections in the Prints and Photographs Division of the Library of Congress. Prior to her participation in testing the use of EAD for pictorial collection finding aids, she completed an eight-year project to arrange, preserve, house, and catalog architectural drawings relating to the Washington, D.C. metropolitan area. This article is a revision of a paper given on 30 August 1996 at the annual meeting of the Society of American Archivists held in San Diego.

Introduction

THE LIBRARY OF CONGRESS (LC) is one of several institutions that tested the alpha and beta versions of the Encoded Archival Description (EAD) Document Type Definition (DTD) in 1996 and 1997. The library's goals were to explore the use of Standard Generalized Markup Language (SGML) as a tool for encoding finding aids, to evaluate the EAD DTD's structure, noting any problems or inconsistencies, and to see how well EAD could accommodate a variety of finding aid formats. Staff also sought to determine the ease or difficulty of converting a finding aid to EAD.

LC has an immediate interest in using EAD. As it begins to digitize its extensive special collections, the digital object-linking features of EAD offer a promising approach for its National Digital Library Program (NDLP) projects, given that many of LC's special collections are far too large for each component to receive individual cataloging.

In addition, LC is interested in making its finding aids broadly available whether or not a particular collection's contents are digitized. LC has formally published selected registers over the last forty years, and since 1995, its special collections divisions have made additional finding aids available on the Internet through LC's Gopher and World Wide Web sites.[1] LC is using SGML encoding, however, as a better and more lasting investment than it would be to encode finding aids using HyperText Markup Language (HTML). As an SGML DTD, EAD is platform- and program-independent and provides a stable storage environment for data. By supporting more sophisticated display and navigation than is possible using either HTML or plain ASCII text, EAD helps the hierarchical structure of a finding aid convey the depth and complexity of an archival collection in an intelligible way. In addition, EAD-encoded finding aids offer researchers an opportunity to simultaneously search multiple finding aids within and among institutions.

LC has found EAD to be a worthwhile method of automating finding aids in several divisions and has benefited from participation in the alpha and beta testing of the DTD. Staff learned to use the EAD DTD by experimenting with various approaches to tagging and learned that encoding can progress rapidly once finding aid creators are familiar with the tags and structure of EAD. The learning curve has been steep, however, and substantial technical support has been needed to understand how to configure the SGML files to interact properly. LC has tested varying approaches and strategies for converting existing finding aids and now has thirteen encoded finding aids. LC's EAD Task Force is disseminating information internally about EAD and finding aids; fostering discussion of common standards for content, coding, and display; and investigating how best to distribute EAD finding aids via the World Wide Web.

This case study focuses on LC's participation as one of several institutions that tested EAD by encoding finding aids.[2] The first section addresses LC's test of the alpha version of the DTD, describing common activities and decisions; the differing approaches to tagging taken by the Manuscript and the Prints and Photographs divisions in encoding sample

---

[1]As of September 1997, the Manuscript Division has made ASCII versions of over one hundred finding aids available through Gopher and Web gateways. The American Folklife Center, the Prints and Photographs Division, the Music Division, and the Motion Picture, Broadcasting, and Recorded Sound Division also have made numerous finding aids available via LC's Gopher and websites.

[2]This case study does not describe LC's participation in the Berkeley Finding Aid Project or on the team of researchers who worked at the Bentley Historical Library on the transition from the FindAid DTD to EAD, or in pre-alpha testing of EAD. Neither does it address the role of LC's Network Development and MARC Standards Office as the maintenance agency for EAD in conjunction with the Society of American Archivists.

finding aids is also analyzed. The second section describes the encoding activities of the Manuscript, Prints and Photographs, and Music divisions following the alpha test. The final section discusses organizational and technical issues in the cooperative exploration of EAD at LC and summarizes the authors' conclusions, based on their two-year involvement with EAD at LC.

Alpha Testing at the Library of Congress

Group Activities and Decisions

Three divisions at LC participated as early implementers of EAD beginning in January 1996. By May 1996 (no one worked full time on the project), LC had two prototype finding aids on-line. Finding aids selected for conversion included the moderately sized traditional register for the Shirley Jackson Papers (Manuscript Division) and the larger finding aid for visual materials from the NAACP Records (Prints and Photographs Division). The very large finding aid for the Federal Theatre Project (Music Division), begun during alpha testing, was completed during beta testing and will be discussed in a later section.

The participants took a group approach, blending the skills of staff in NDLP and Information Technology Services, who were familiar with SGML from tagging book texts, with the skills of staff in the Manuscript and the Prints and Photographs divisions, who were familiar with finding aids. The number of EAD tags was a shock! It gradually became apparent that finding aids often merit more markup than the historical narrative texts that were tagged in the NDLP projects because finding aids, which are documents about other documents, have more structure and more varied content. It also became clear that staff members who were unfamiliar with finding aids would have a hard time doing markup. The parts of finding aids cannot be as easily identified as a chapter, heading, or paragraph in a book, for example, and the headings for each finding aid section do not always match the generalized EAD tag names (e.g., "Description of Series" vs. "Description of Subordinate Components").

The participants gained some familiarity with SGML and EAD by looking at standard SGML reference sources,[3] the pre-alpha EAD DTD, and eventually the alpha DTD and tag library. Marking up finding aids on paper at this stage helped the testers learn the DTD structure and tag library before going on-line. A four-day SGML overview course offered by NDLP later in 1996 helped some staff. LC found, however, that once software is configured and sample EAD finding aids exist, the training focus for most staff can be on EAD itself.

At first, all tagging was done at one digital library development workstation on which two software tools, SoftQuad's Author/Editor and Panorama PRO (SGML editor and viewer software) had been installed. The alpha version of the EAD DTD was loaded, and rules files were generated to enforce the syntax of the DTD during tagging. As additional copies of software were installed, alpha testers were able to work with the SGML editing software in their own divisions, while continuing to work with the SGML viewer on the digital library workstation.

---

[3]Charles F. Goldfarb, *The SGML Handbook* (New York: Oxford University Press, 1990); Eric Van Herwijnen, *Practical SGML* (Boston: Kluwer Academic Publishers, 1994).

The earliest encoding was done in weekly "play sessions" attended by staff from the three divisions that were encoding finding aids; staff with technical expertise also participated.[4] Working without adequate documentation was difficult (the alpha tag library became available in late February 1996), but testers were aided by the expertise of two LC participants in the Bentley Fellowship Program who were available to help interpret the DTD.[5] Additional insights were gleaned from postings on the EAD listserv.[6]

Staff in each of the divisions invariably focused on different encoding problems, and these informal, experimental sessions were excellent opportunities to see things from another angle, to pick up practical tips, and to resolve many issues. Topics included tagging dilemmas, use of WordPerfect macros to insert tagging, SGML conventions, how to configure the different components (i.e., the DTD, entity files, finding aid files, and files required by the viewing software) to interact on-line, and how to create and edit style sheets and navigators.

An important part of the test was making provisional decisions about how to encode; specifically, how to interpret the tags and how intensively to code the finding aids. Data retrieval and navigation needs were considered as much as were data display features. Staff purposefully explored different encoding options, with the hope of demonstrating by experimentation that EAD could be applied at various levels of detail. The register for the Shirley Jackson Papers was marked up by the Manuscript Division at a high level of specificity up until the container list. The Prints and Photographs Division explored a briefer tagging style, "EAD Lite," for its longer NAACP finding aid, which included an index, an image sampler, and additional material. The Music Division's Federal Theatre Project finding aid, discussed below, minimized its container list, tagging even less than the others to reduce the bulk of this very long finding aid.

The alpha group agreed to certain common practices. All appropriate data would be supplied in the EAD Header <eadheader>, the tag which holds information identifying the finding aid. The Manuscript and the Prints and Photographs divisions agreed on a title page display generated from the header that would provide a uniform look and feel to on-line LC finding aids.

All LC EAD finding aids begin with summary elements in the Archival Description-level Descriptive Identification <archdesc><did> element to give researchers a quick, general sense of what each collection includes. These elements, not previously found in LC's paper finding aids, ensure that basic data can be readily located on-line; EAD encourages inclusion of this data in all finding aids. The extent statement was moved from

---

[4]Many specialists from throughout LC advised staff during these sessions: National Digital Library Program staff lent their expertise in SGML and SGML software, gained from implementing the American Memory DTD; Automation Planning and Liaison Office staff installed and configured the required hardware and software; and Information Technology Services staff assisted with configuring software to read the EAD DTD, creating the navigator, and testing ways to point from the finding aid to the Prints and Photographs Division home page.

[5]Janice Ruth (Manuscript Division) and Helena Zinkham (Prints and Photographs Division) were members of a team of experts in archival descriptive standards sponsored by the Bentley Historical Library Research Fellowship Program, which first met July 1995 in Ann Arbor, Michigan. The team collaborated in the production of finding aid encoding standard design principles, a revised finding aid data model, a revised finding aid Document Type Definition (which became EAD), and finding aid encoding guidelines and examples. This group continued to shape the development of EAD; most members now belong to the Society of American Archivists' Committee on Archival Information Exchange EAD Working Group, which maintains intellectual ownership of EAD.

[6]Information on the EAD listserv and instructions for subscribing are available at <http://www.loc.gov/ead/eadlist.html>.

the administrative information page of the paper finding aid to the descriptive identification element. In other words, the use of EAD is changing and improving former paper finding aid practices.

Both divisions chose to enrich their finding aids with a list of Controlled Access Headings <controlaccess>, which are authorized forms of names, topical subjects, and other terms taken from the USMARC catalog record for a collection. Finding aid searches limited to these specific fields will retrieve collections which are particularly strong sources of material for these subjects. The working group decided that detailed tagging of names, dates, and places in the container list would be unlikely to yield enough retrieval benefits to justify the time required by available tagging methods.

Finally, the two divisions chose to enhance navigability among the sections of the finding aid by using reference tags to provide hypertext links between the scope and content note and description of series, and between the description of series and container list. Cross-references in the container list also were linked.

Manuscript Division: Shirley Jackson Papers

For purposes of the alpha test in early 1996, the Manuscript Division focused on finding aids previously converted from WordPerfect to plain ASCII for mounting on the LC Gopher site. The register for the Shirley Jackson Papers is typical of a finding aid for a literary author; at nineteen pages, it is of moderate length and complexity.

Tagging began by importing an ASCII version of the register into Author/Editor. The SGML editor enforced correct tagging and prompted participants with the tags available at any point; this interactive method allowed the group to learn the DTD thoroughly while making tagging choices. Once the tagged finding aid was validated, it served as a model from which to devise more efficient tagging procedures.

In order to use as many different elements as possible to evaluate their usage and to suggest changes for the beta version, the Manuscript Division coded personal <persname>, corporate <corpname>, family <famname>, and other names, geographic places <geogname>, and dates <date> in the Administrative Information <admininfo>, Biographical Note/Agency History <bioghist>, and Scope and Content Note <scopecontent> elements. Bibliographic references <bibref> were coded throughout the finding aid, as were titles of works <title>; for the latter, attributes were used to distinguish between titles which should be rendered in italics (e.g., book titles) or with quotation marks (e.g., article titles). Detailed tagging of names, dates, and places in the container list was deferred.

Paragraphs in the Administrative Information <admininfo> section were coded as Acquisition Information <acqinfo>, Processing Information <processinfo>, and Restrictions on Use <userestrict>, as appropriate, although coding them as Paragraphs <p> also would have been acceptable.

Both the Description of Subordinate Components <dsc type="analyticover"> and the Container List <dsc type="in-depth"> were coded with numbered rather than unnumbered components (e.g., <c02> rather than <c>) for ease of tag tracking and proofreading. The register for the Shirley Jackson Papers contains five nested levels of components, and many other finding aids feature even deeper hierarchical nesting. Unit dates <unitdate> were tagged only at the series and subseries level.

The Manuscript Division staff chose to code the container list using Display Rows <drow> and Display Entries <dentry> in hope of enabling a tabular display of data,

with container numbers in a left column and contents in the right column. This model was an attempt to impose tabular display while tagging the components of the container list according to their hierarchical value.

Prints and Photographs Division: NAACP Finding Aid

The sixty-seven-page print version of the finding aid for the visual materials in the National Association for the Advancement of Colored People (NAACP) Records, completed in 1994, was created in a traditional finding aid format. The Prints and Photographs Division selected this finding aid because the NAACP collection is of high interest to the division's users, and staff hoped that the navigation and searching capabilities of SGML browsers would allow researchers easier access to specific information in this long document. The finding aid's traditional format facilitated mapping to EAD.

Division staff explored four main issues while encoding the NAACP finding aid: testing "EAD Lite,"[7] developing techniques for converting a traditional-style paper finding aid into an on-line version, showing how EAD accommodates some special qualities of pictorial collection finding aids, and examining the linking of digital images to finding aids on the Web.

"EAD Lite" is an attempt to save time (and thereby reduce costs) by avoiding tagging that would result in relatively few display and retrieval benefits; staff can create more EAD finding aids by tagging each at a basic level. They can focus on understanding EAD core structures, and nonarchivists can successfully be employed to provide basic tagging. The full EAD DTD allows for very detailed content tagging when more intricate retrieval is needed (for example, every instance of "NAACP" as a corporate name, and all personal names could be tagged as <persname>). "EAD Lite," in contrast, permits paragraph-level information to be tagged as minimally as possible; attributes are seldom used. The focus is on links, collection summary elements, and large structures rather than on tagging every personal name and date.

"EAD Lite" stresses the importance of "heads" and "links" to ensure easy navigation within the parts of the finding aid. Prints and Photographs Division staff created hotlinks on call numbers in the series description to lead users directly to more detailed information in the contents list, as well as on call numbers in the Photographer Index to lead users to the series-level description. These links were created by tagging the call numbers as references.

The "Contents List" in the NAACP finding aid was encoded as a minimalist container list to test the use of nested, unnumbered components. Staff found it difficult to keep these unnumbered components properly nested. They also discovered the hazard of components that appear to be at two different levels because they are displayed on two lines in the paper finding aid, but which are, in fact, only a folder title and subtitle of the same component. The Prints and Photographs Division's finding aids will use numbered components in the future.

Tagging was done using Author/Editor and WordPerfect software. It helped to manually key tags for "large" elements (e.g., <ead>, <archdesc>, <findaid>) before im-

---

[7]Available as an appendix in Anne J. Gilliland-Swetland, *Encoded Archival Description Document Type Definition (DTD) Application Guidelines* (Washington, D.C.: Library of Congress and the Society of American Archivists, 1996). Unpublished draft available from <http://scriptorium.lib.duke.edu/findaids/ead/guidelines/index.html>.

porting the text to Author/Editor in order to side step the time-consuming "wrapping" of text. WordPerfect's search-and-replace function was useful for inserting tags in long lists when the tags could be applied consistently.

The division developed an "Image Sampler" (or digital archival object group <daogrp>) of eight images to give researchers an idea of the typical images in this collection (Figure 1). (Paper finding aids similarly often include photocopies of images because visual data for visual images is so important.) The division will continue to link digitized images of items in collections to its EAD finding aids, highlighting the images of most interest to researchers.

Within the <daogrp>, the digital archival object <dao> points from the document to an individual "gif" file for each image. The sampler featured images of approximately 50 KB each rather than thumbnails. Entity references were declared for each image at the beginning of the SGML document. (An SGML declaration document file would be used to handle the entity references for names of digital image files if more than ten images are included.) Naming conventions for digital image files are still evolving.

Unfortunately, these images cannot be included in the version of the finding aid that is currently mounted on LC's public website because they were produced within the last fifty years, in some cases by unknown photographers, and the rights to most of the images are not clear. Therefore, the version with the images is available only on LC's internal "intranet" website.

Results of alpha testing

The most concrete result of alpha testing at the Library of Congress was a preliminary version of the register for the Shirley Jackson Papers, which was loaded on the American Memory website on April 26, 1996; the NAACP register followed on May 22. These registers have been revised to conform to the beta EAD DTD released in fall 1996. The third finding aid, for the Federal Theatre Project Records, was made available in February 1997. The Shirley Jackson Papers register was submitted as an example for use in the EAD application guidelines.[8]

The basic EAD structure seems to fit LC finding aids well. LC reported only a few problems and proposed a number of minor changes to the EAD DTD, most of which were implemented in the beta version. Most of these DTD changes resulted in new names for elements and attributes, and the alpha EAD finding aids were updated with simple search and replace procedures.

Divisional Case Studies

Manuscript Division: EAD Pilot Project

In September 1996 the Manuscript Division approved an EAD pilot project. In support of EAD's long-term goal to improve onsite and remote access to information contained in archival finding aids, the short-term objectives of the project were:

1. To increase the division's production of encoded finding aids, giving LC a larger pool of documents from which to test searching methods, linking mechanisms, and other display and access questions pertinent to LC's implementation of EAD;

---

[8]Gilliland-Swetland, *Encoded Archival Description Document Type Definition (DTD) Application Guidelines.*
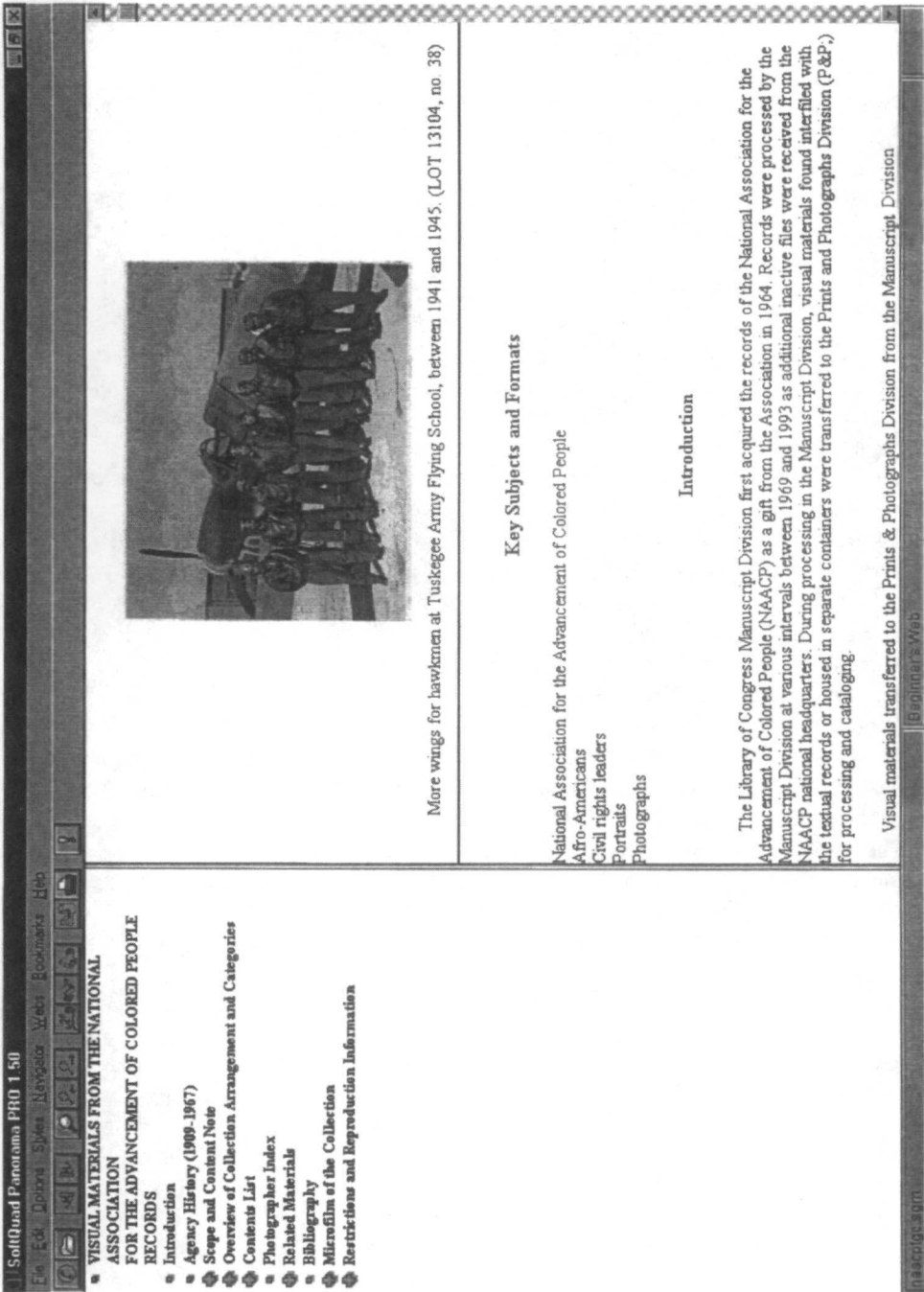
Figure 1. Image Sampler, NAACP Finding Aid, displayed in Panorama.

2. To identify strategies, staffing levels, and workflow patterns that the division may wish to implement for full-scale production of EAD finding aids; and

3. To gather additional comments, criticisms, and recommendations about the beta version of EAD. (In fact, few comments resulted from the beta testing, due to the success of alpha testing.)

The Manuscript Division pilot project was conducted in the fall of 1996 and was largely completed by spring 1997. Ten finding aids already available in plain ASCII and WordPerfect versions were identified by reference and subject specialist archivists as being of high research value and general interest. An advisory group of division reference and processing managers, curators, catalogers, and processing archivists met at the beginning of the project to reevaluate the coding decisions made during alpha testing. The group continued to advise on levels of specificity and extent of tagging in an attempt to balance the improved access provided through more sophisticated searching against the decreased productivity that resulted from more detailed content tagging. Group members also provided input in editing the style sheet for EAD finding aid display. Manuscript Division catalogers reviewed the ten collection-level cataloging records and completed authority work for all headings in the records.

The project coordinator worked with the division's automated operations archivist, who developed a suite of powerful and creative macros to perform the bulk of the tagging within WordPerfect. These macros were run on WordPerfect documents which had already been prepared for export as plain ASCII versions by removing page headings and substituting "fake underlines" for emphasis codes such as bold and italic. Conversion macros inserted tags based on tabs, indents, and other "hooks;" the macros relied upon regular tab settings and hard returns to convert the context of the container list from white space to tagged components. Each finding aid, tagged thusly within WordPerfect, was ready to be imported into Author/Editor for validation. Conversion of documents to SGML was facilitated by the prior existence of detailed divisional guidelines for consistent preparation of registers and by extensive use of word processing macros and templates in the production of the original WordPerfect finding aids.

The project coordinator trained and directed a manuscript technician, who worked half-time on the project for its duration and ably encoded the ten selected finding aids. The technician kept a careful record of the time required for each step in the project, which helped the project team streamline the workflow and improve the evolving macros and documentation.

After the tagged WordPerfect finding aid was printed out and proofread against the original document using an editing checklist, it was exported as generic ASCII text and imported into Author/Editor; this process identified additional errors. Staff then added appropriate content tags to names, places, bibliographic citations, and types of information in the <admininfo> section using Author/Editor (this step requires intellectual input and is not easily done by conversion software). Where links were needed for navigation or container list cross-references, the technician wrote on the paper finding aid a short name for the attribute called "id" for each element to which a reference pointed; those values were added to the EAD finding aid in Author/Editor. After work in Author/Editor was complete, the coordinator reviewed the Panorama display of the finding aid against the original paper finding aid and selectively checked tagging in Author/Editor. The Panorama style sheet, a file that controls how the data is displayed, was edited as needed.

Substantial experimentation was done with the Panorama style sheet and the navigator file, a file which determines which EAD elements are pulled into the on-line table of contents. The division wished to convey the finding aid information as clearly as possible, and staff attempted to replicate, on-line, the tabular appearance of the printed container list. This tabular formatting could be approximated on-line using either <drow> or <did> (Figure 2), but true tabular formatting required the assignment of empty <dentry> tags contrary to the DTD. The division thus decided to abandon this <drow><dentry> approach and changed all eleven Manuscript Division finding aids to a <did> presentation.[9] This simpler approach makes coding and proofreading easier, as well as decreasing total file size. Once staff identified the necessary file directory structures and had server space assigned, the EAD finding aids were loaded and made available via the Manuscript Division finding aid page on the LC website.

The pilot program's preliminary strategy, therefore, tested the use of existing software, equipment, and staff for the retrospective conversion of current finding aids already available via LC's Gopher and websites in ASCII text. The division identified minor changes to paper finding aids which ease the conversion process and more accurately reflect the content of sections of the finding aid. Converting a finding aid to EAD should be viewed as the final step in processing collections and creating, inputting, and editing the registers. The time required to complete this step will continue to decrease.

In the summer of 1997 the Manuscript Division identified additional high-priority finding aids for SGML conversion; staff also scanned or rekeyed registers not yet in machine-readable form in preparation for EAD encoding. The division will encode these finding aids by the methods already developed. Staff expect to refine the process and streamline workflow; identify hardware and software needs; design styles, macros, and templates; and investigate autotagging software to speed document conversion. Word processing operations are being transformed to make EAD conversion faster and more reliable. Concurrently, LC is upgrading its WordPerfect software to versions which include SGML modules.

Still to be determined is the best method of creating new finding aids from which both the paper and electronic EAD products can be efficiently derived. LC also will explore methods for storing, indexing, and retrieving EAD finding aids as part of its NDLP repository efforts. How EAD best meets reference needs also will be evaluated.

Prints and Photographs Division: Experiments with Alternate Finding Aid Formats

Encoding activities in the Prints and Photographs Division following the completion of alpha EAD testing have centered on experimentation with a variety of divisional finding aids for photographic and other pictorial collections.

Test encoding of the finding aid for the Chadbourne Collection of Japanese Prints explored how well EAD could accommodate detailed item-level descriptions for each of the 188 prints in this collection (this level of description also could be handled by item-level USMARC records, but the descriptions in this finding aid are preliminary and more informal than those in the division's USMARC catalog records). Staff included USMARC encoding analog attributes in element tags, but individual names and titles are not con-

---

[9]For instance, a line of the finding aid that had been encoded:
<c04><drow><dentry spanname="c8-20"><unittitle>Letters sent</unittitle></dentry></drow>
</c04> was encoded more simply: <c04><did><unittitle>Letters sent</unittitle></did></c04>.
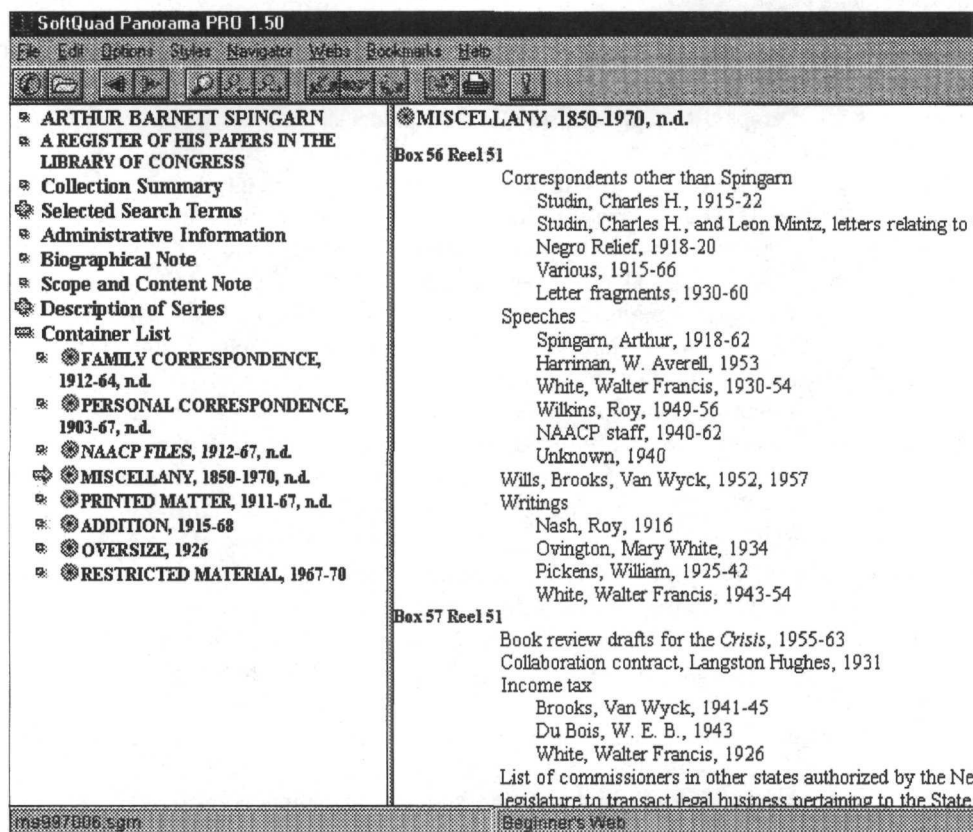
Figure 2. Style sheet and navigator displays of finding aid in Panorama. The navigator panel (left side) is controlled by the navigator file. The Container List entry has been expanded to show the series title; the arrow points to the user's place in the document. The display of the right panel is controlled by the style sheet. Labels have been supplied for box and reel numbers, which do not appear on the same line as the corresponding titles.

structed in the same way as the terms found in library catalogs. The division believes that EAD finding aids provide the flexibility to accommodate those cases in which it is necessary, perhaps only selectively, to provide fairly detailed item-level descriptions; examples could include caption lists or other types of pictorial inventories. The Chadbourne finding aid contains consistent categories of information, and so most tags could be inserted fairly quickly using WordPerfect macros and the search/replace function. Staff also tested use of EAD's label attribute to improve users' comprehension of the meaning of specific information categories (for example, to explain that a particular name is an "artist").

A finding aid currently is being compiled for the division's baseball card collection containing 2,100 early baseball cards. This collection also is being digitized; images of both the front and back of each card will be available on the American Memory website. The finding aid contains an item-level caption list for each card that includes the player's name, position, date, city, team name, and the card's call number. Because this information has been organized in a Paradox database, nearly all of the tagging can be generated from a database report. Staff constructed a hierarchical report form that reflects the four levels of nesting in the finding aid. Tags were added to the form after determining valid tag

structure for the finding aid. The finished report is imported into SGML authoring software as an ASCII file for final tagging.

Other finding aids are slated for EAD conversion in the next few years. The division's finding aids come in a wide range of formats: some are straightforward lists containing detailed item-level descriptions (such as that for the Chadbourne Collection of Japanese Prints); others contain long lists of folder-level headings (such as the finding aid for the photo morgue of the *New York World Telegram and Sun* newspaper). The division also creates finding aids in a more traditional finding aid format. Most describe collections at the folder- or item-level, since visual materials often lack captions, and the finding aid serves as a caption list. Reference staff have found that providing more detailed description reduces the need to browse original images.

These projects will present new challenges. Encoding the finding aid for the *New York World Telegram and Sun* photo morgue will require the creation of multiple SGML files, given that it consists of a ten-thousand-page folder list for the one-million-item collection. Creating multiple files will help address concerns about how well Web browsers can handle such an enormous finding aid.

Test encoding of the NAACP finding aid has demonstrated that EAD is flexible enough to accommodate finding aids for photographic and other pictorial collections, even though these finding aids do not always fit the traditional finding aid structure. The EAD structure is standardizing the use of core finding aid elements across LC. The Prints and Photographs Division has begun to include some important pieces of information not currently found in some of the division's older finding aids, such as the collection summary information mentioned earlier in the Manuscript Division context. Researchers using EAD finding aids on the LC website will not have access to reference staff who assist researchers onsite, so remote access adds the new requirement to provide a fuller context for understanding and using these collections.

Music Division: Federal Theatre Project and Digital Images

The Music Division, assisted by the National Digital Library Program, has encoded LC's largest EAD finding aid to date, the three-hundred-page published register for the Federal Theatre Project (FTP) Records. This finding aid was selected because significant parts of the collection are being digitized—by 1998 more than thirty thousand digital objects should be on-line. The finding aid does not currently link to these digital objects, but links will be added in the next few years.

The first part of the markup project, begun during EAD alpha testing in 1996, involved the overall view of the document structure—how it was organized and how much "nesting" of information was involved. The FTP finding aid has an introduction and an extensive container list with many layers of description. This structure suggested using <did> rather than <drow><dentry> to decrease the number of tags required, as well as numbered rather than unnumbered components to keep better track of hierarchical levels in the extensive container list. Thus the basic pattern was set for the markup: determine the outermost boundaries first, then fill in the details.

The question, "What are the research needs?" guided the tagging philosophy for this finding aid, which evolved over several months. After initial uncertainty regarding how to use the Physical Description <physdesc> tag for folder counts, staff decided to use it only for titles that overflowed into multiple containers to avoid using an excessively

large number (and possible overload) of tags. Names and dates were tagged only in <admininfo> and <chronlist>.

The markup of this document required several phases. It was crucial to do a preliminary markup on paper, highlighting the higher-level elements (the <c01>s and <c02>s) before tagging in Author/Editor, as it was much easier to see the breakdown of the sections on paper than on-line. The finding aid was divided into five files during encoding: one file for the introductory material and four separate files for the container list. Smaller files were easier to manage in both word processing software and Author/Editor; and also allowed more than one person to do markup simultaneously. A ''fake'' SGML heading, including the SGML declaration and header for each of the smaller files, was added to permit the sections to be validated independently. Once the tagging was finished, these files were reconnected by eliminating the temporary headers to make a single document, which was nearly one megabyte in size.

Because of the large amount of data, it was easier to manipulate the finding aid first in a word processing program. Preliminary tagging was accomplished through the use of WordPerfect macros. The combinations of indents and returns already present in the electronic document were searched and replaced with strings of SGML tags, thus enabling much of the tagging to be automated. The text editing software Codewright, which includes features to facilitate autotagging, was used concurrently with word processing software. Codewright was faster to use than word processing software, as it is capable of more elaborate search and replace functions, but it required that staff learn to use yet another program.

The SGML editor was used mainly for validation of the files. The WordPerfect files invariably contained formatting errors which translated into errors in the initial tagging. Cleanup of these errors, as well as refining the tagging, required a time-consuming line-by-line check of the entire document.

The finding aid for the Federal Theatre Project Records was made available on the American Memory EAD website in February 1997. On June 28, 1997, the NDLP unveiled the initial version of ''The New Deal Stage: Selections from the Federal Theatre Project, 1935–1939.'' Access to ten thousand digital objects is currently being provided through indirect HTML links, pending resolution of LC digital repository issues to facilitate direct links between the finding aid and the images.

LC EAD Task Force

In June 1996 the acting director of Public Service Collections (PSC), the directorate to which LC's special collections belong, charged a joint task force of National Digital Library Program and PSC representatives to further expand LC's EAD implementation. The group's charge included the following: identify types and quantities of finding aids that should be encoded; help each division that compiles finding aids to establish appropriate work flows to convert or author EAD finding aids; establish software and hardware needs; determine training needs; work with the NDLP, LC's Automation Planning and Liaison Office, and Information Technology Services to implement finding aid server capability that can be accessed by each LC reading room and by offsite users; and assist each division in establishing a work plan, as well as contributing at least one finding aid to the LC EAD website.

The task force learned that each of LC's seven special collections divisions has created archival finding aids to control some portion of its collections. The Manuscript

Division has by far the most finding aids, and creates the most new finding aids each year. LC's earliest candidates for conversion to EAD include the almost two hundred on-line finding aids already available in plain ASCII format. The universe of paper finding aids is much larger: there are more than 2,000 finding aids in the Manuscript Division alone, another 100 in Prints and Photographs, and about 125 total among the American Folklife Center; Geography and Map; Motion Picture, Broadcasting, and Recorded Sound; Music; and Rare Book and Special Collections divisions. The size of these finding aids ranges from one page to ten thousand pages and more.

One of the EAD Task Force's charges is education, and monthly meetings have served to educate members from the participating divisions. One such educational activity was the "live" tagging of a finding aid from the American Folklife Center using Author/ Editor. A subgroup identified training opportunities using LC and external resources such as the Research Library Group's FAST (Finding Aids Archival SGML Training) course. A half-day EAD overview workshop developed by LC staff was taught first at the Mid-Atlantic Regional Archives Conference (MARAC) in Charlottesville, Virginia in May 1997, and is now being offered to staff throughout LC, as well as at subsequent MARAC meetings.

The conversion of finding aids to EAD has slowed while a number of underlying technical issues are resolved. LC's EAD Task Force appointed a technical issues subgroup in March 1997 to address questions relating to the library's digital repository, including digital object naming; indexing schemes; structural and administrative metadata; levels and types of description; and relationships among USMARC records, EAD finding aids, related description files, and the digital objects to which finding aids can be linked. Links between LC's catalog records and finding aids are central to the successful integration of these bibliographic and descriptive tools. LC is implementing a digital repository which assigns permanent names called "handles"[10] to EAD finding aids and digital objects; this is in lieu of Uniform Resource Locators (URLs), because the latter change too often to be practical addresses in a digital repository.

Accomplishments of the technical subgroup thus far include formulation of a method of creating unique identifying names for the mandatory <eadid> element to make it compatible with LC's naming scheme for objects in its digital repository, and the development of an SGML file system for central storage of the DTD, entity files, style sheets, navigators, and image files which are shared by finding aids stored within each division's file structure. The Manuscript Division created a model file structure for its finding aids and edited "catalog" and "entityrc" files to point to the SGML file system. Files are made available for internal use by divisions wishing to adapt these for their own needs, as well as for use from outside LC via anonymous file transfer protocol. A new LC EAD webpage,[11] released in August 1997, provides access to all EAD finding aids at LC and also leads researchers to other LC on-line finding aids and special collections information. This webpage supplements the EAD home page maintained by the Network Development and MARC Standards Office, which makes available the EAD DTD, guidelines, and other background information.[12]

---

[10]"Handles" are Uniform Resource Names (URNs) developed by the Corporation for National Research Initiatives (CNRI), Reston, Va. CNRI is working with the LC to build a prototype digital repository system.
[11]<http://lcweb.loc.gov/rr/ead/eadhome.html>.
[12]<http://lcweb.loc.gov/ead/>.

There are still many issues to be resolved at the Library of Congress. LC's reading rooms only recently introduced public access Internet stations, and efforts to obtain user feedback about on-line finding aids has been limited thus far. Also, a freely accessible method of delivering finding aids on the Web is not yet in place; in general, industry development of SGML viewers has lagged behind expectations. For now, users must download a free version of Panorama, an SGML viewer that currently runs only under Windows, and not all potential users are willing or able to download this software. One way around this problem is to convert SGML documents to HTML versions for display, but to facilitate retrieval by allowing users to search the original SGML versions and to formulate search queries which take advantage of the rich structural SGML tagging. It seems that EAD finding aids may be convertible to Extensible Markup Language (XML), an emerging standard that promises to serve SGML on the Web as readily as HTML, while taking advantage of SGML tagging. Finding aids would be delivered by XML viewers currently under development by commercial vendors.

Searching SGML documents at LC presents additional problems at present. LC uses INQUERY from Sovereign Hill Software to index and search NDLP collections. INQUERY cannot read a DTD and lacks the capability to exploit some basic SGML features, such as nested searches or querying on attributes. Not until indexing issues are resolved will it be possible to create search forms that allow researchers to exploit the rich SGML tagging without expert knowledge of the tags.

Other SGML users at LC confront similar issues of indexing, conversion, and document management. SGML activities at LC include developing the American Memory DTD that the NDLP uses to encode historical documents, development of the USMARC DTD by the Network Development and MARC Standards Office, and several Congressional Research Service projects such as encoding the *Bill Digest*. LC's SGML Working Group facilitates the exchange of information among those divisions working with SGML and helps users address SGML needs and issues.

Conclusion

The Library of Congress's "early implementers" of EAD were provided with an opportunity to collaborate with other divisions at LC—the National Digital Library Program, the custodial divisions of LC, and automation and cataloging offices—and to communicate with others from across the profession involved in creating finding aids and testing EAD. Finding aid creators are beginning to discuss common standards for content, coding, and display, and to develop models for encoding finding aids across LC. Knowledge of and interest in archival collections and finding aids has been raised to new prominence in LC. The project also has raised awareness of the benefits of SGML.

What value do these experiences have for newcomers to EAD? What elements of this journey of discovery can the authors recommend for others? First, one should analyze current finding aids in the context of EAD. Do they meet the needs of researchers? Does the EAD model point to ways they can be improved? Consider to what level finding aids should be encoded. Examine printed copies and attempt to mark at least one representative sample manually as a template for future encoding. Use the authoring software as a learning and validation tool.

There are now many resources available to help institutions embark on encoding finding aids with EAD. Most of the work of the early implementers was done with little available documentation and few examples, although the alpha version of the EAD tag

library was invaluable. Institutions now have the benefit of the Research Library Group's training manuals that were created for its FAST workshops, draft EAD application guidelines, and a revised tag library. The authors also recommend that implementers study the retrospective conversion guidelines and toolkits prepared by the American Heritage Virtual Archive Project[13] and the University of California EAD projects,[14] and adapt or use them accordingly. Examination of EAD websites and study of the markup of selected finding aids available on-line are useful ways of keeping current with EAD practice. Participation on the EAD listserv continues to be a valuable source of information for EAD implementers as well. As more institutions become involved in using EAD, this pool of resources will continue to expand.

---

[13]<http://sunsite.berkeley.edu/amher/>.
[14]<http://sunsite.berkeley.edu/FindingAids/uc-ead/>.