

Developing a Cooperative Intra-institutional Approach to EAD Implementation: The Harvard/Radcliffe Digital Finding Aids Project

LESLIE A. MORRIS

Abstract: The Harvard/Radcliffe Digital Finding Aids Project (DFAP) was established in 1995 to explore options for making finding aids available remotely over the Internet. This paper reports on why SGML/EAD became the delivery medium of choice; the process of selection of SGML authoring software; the ongoing development of cross-Harvard standards and guidelines for application of EAD; plans for cross-repository indexing and development of a search engine; and implementation challenges faced by the eight Harvard/Radcliffe repositories participating in the project. Also discussed are the development of an EAD template for future Houghton Library finding aids and problems with retrospective conversion of existing finding aids.

About the author: Leslie A. Morris is Chair of the Harvard/Radcliffe Digital Finding Aids Project (DFAP) and Curator of Manuscripts in the Harvard College Library, Houghton Library, Harvard University. This is a slightly expanded version of a paper given on 29 August 1997 at the annual meeting of the Society of American Archivists held in Chicago.

Archives at Harvard University

TO THE WORLD OUTSIDE, the words “Harvard University” summon an image of a monolithic institution, one of the country’s oldest and wealthiest universities, an organization that speaks with one voice. From the inside, the picture is somewhat different. One of the first pieces of university lore a new employee learns is the rather unlovely saying “Every tub on its own bottom.” The “tubs” are the individual faculties, and they have virtually complete autonomy. The librarians for each faculty report to their respective deans, not to the university librarian. Their budgets come from the faculties as well, which means that the business and law libraries have access to more funding than the education and divinity libraries. There is, however, a “Harvard University Library.” It is a department of the Central Administration, and it has a purely coordinating role. But, if one of the faculty libraries decides it does not want to be coordinated—as when the Business School decided not to participate in HOLLIS, the Harvard on-line catalog—the university library has no coercive powers.

When one refers to Harvard libraries and archives, therefore, it is important to remember this highly decentralized nature. There are not only ninety-eight separate libraries in the Harvard system, there are forty-nine separate archival repositories, including archives that are part of libraries, museums, hospitals, and a forest. Each has its own traditions, and each has its own budget. *Users* of Harvard collections, however, are almost completely unaware of this situation and how it affects their research. Since the mid-1980s, the manuscript and archives community at Harvard has worked very hard to ameliorate the effects of such decentralization of collections on users. The first step was the Harvard/Radcliffe Manuscript Survey and Guide Project (1984–1986), a project to make many archival and manuscript collections accessible via collection-level MARC records in HOLLIS, the then-new Harvard on-line catalog.

This was a major step forward in providing access to information about Harvard collections, but Harvard archivists still hoped for more. Easy access to the more detailed collection information available in finding aids is as important to us as is access to summary descriptions. In 1994, when Gopher sites for finding aids were all the rage, Harvard archivists received permission to set up an archives site on the central library server and received a few hours of training from the Office for Information Systems (the Harvard University Library systems office) on how to maintain it. While everyone felt the Gopher site was useful for delivering copies of individual finding aids, it was soon apparent that such unstructured text offered only clumsy and frustrating searching, particularly with large files.

While the Gopher site was under development, the university library announced that work would soon begin on the design of the next generation of the on-line catalog, known as HOLLIS II. In response to intense interest from the archives and manuscripts community (which felt that its needs largely had been ignored in HOLLIS I), the Harvard University Library Automation Planning Committee appointed a Special Collections Task Force. This group was charged to examine the automation needs of Harvard repositories that acquire and make accessible *collections* of material (as opposed to monographs). The report of that task force in November 1994¹ contained many recommendations, but the one that is key to the Harvard EAD story was the statement that the information in finding aids must

¹<http://hul.harvard.edu/hollis2/task_groups/specialprpt.html>.

be as easily available to scholars in electronic form as is the collection-level MARC record, followed by the recommendation that options for doing this should be pursued.

During the preparation of its report, the task force investigated work being done at the University of California, Berkeley, using Standard Generalized Markup Language (SGML) to encode electronic finding aids. Berkeley's work demonstrated that SGML was a powerful yet flexible tool that could address the need of the large Harvard community of archivists to produce electronic finding aids and make them available remotely in order to meet the growing expectations of our research clientele. The Automation Planning Committee agreed with the task force's recommendation and established the Digital Finding Aids Project in February 1995.

The Harvard/Radcliffe Digital Finding Aids Project (known as "DFAP") is a group with members from eight Harvard repositories: Historical Collections, Baker Library, Business School; Special Collections, Loeb Library, Design School; Manuscripts and Archives, Andover-Harvard Theological Library, Divinity School; Library of the Gray Herbarium, a special library of the Faculty of Arts and Sciences; Manuscript Department, Houghton Library, Harvard College Library; Manuscript Division, Law School Library; Schlesinger Library on the History of Women in America, Radcliffe College; and the Harvard University Archives. In addition to these collection representatives, the group includes a member from the Office for Information Systems (OIS), and the electronic texts librarian for Harvard College, whose position is in Research and Bibliographic Services.²

The group's charge is to "plan and oversee the design and deployment of a new computer application system to store, search, and retrieve digital finding aids in SGML format at Harvard/Radcliffe." Membership during this initial project phase is designed to provide broad representation of the Harvard archival and electronic systems community. Once DFAP has established standards and procedures, contributions to the finding aids database will be sought from all corners of the Harvard archival and library community.

Selection of SGML Authoring Software

DFAP's first task was to review and recommend SGML authoring software. The group wanted software that could be used by all repositories, regardless of hardware platform, thus enabling Harvard archivists to develop and share a common experience and knowledge. As "minority" members of the library community at Harvard, archivists are accustomed to depending largely on themselves and each other, rather than on understaffed computer support offices (for those who have such support available), for our specialized software needs. The authoring software had to be inexpensive, since several members of the group run single-person operations with minimal financial support. The ability of the software to import existing ASCII and word-processed files also was a factor, since the need for conversion of existing finding aids was part of the project's charge.

After investigating many SGML authoring packages, project members selected WordPerfect 6.1 SGML Edition (running under Windows 3.x) for PC platforms. Perhaps the most important reason for this was cost. Copies could be purchased for thirty dollars each under Harvard's site license with Novell (who then owned WordPerfect; it has since been sold to Corel). Also, about half of the group were already using WordPerfect for their finding aids and so felt very comfortable with the software. Two members of the

²A list of current members, as well as much information about Harvard's implementation of EAD, is available on DFAP's website, <<http://hul.harvard.edu/dfap/>>.

group were Mac-based, and this posed a difficulty. The only package it was possible to recommend was Author/Editor—at \$750, a much more expensive alternative, but still cheaper than many of the other authoring packages tested.

Since 1995, of course, the field has changed. WordPerfect now bundles its SGML module with Versions 7 and 8 (Windows95 or NT required), and many archivists obtain it “free” as standard software on local area networks. Several members of the project have recently purchased Author/Editor and are beginning to use it. The Research Libraries Group makes available to members a package including SoftQuad’s Author/Editor (SGML authoring), Panorama Pro (SGML browser and style sheet editor), and HotMetal (HTML authoring) at a cost of \$375 for the first copy and \$600 for subsequent copies. As the DFAP project needed only one copy of Panorama Pro to write the SGML display style sheet and did not need HotMetal, the Harvard systems office negotiated a price with SoftQuad of \$275/package for Author/Editor, if at least five copies are purchased at one time.

Finding Aids on the World Wide Web

The second task of the DFAP was to decide on a method for “publishing” the SGML-encoded finding aids on the Internet. After evaluating access issues, we decided that the finding aids will be available through two “gateways.”

The first “gateway” is a link from the collection-level MARC record in HOLLIS, Harvard’s public catalog. HOLLIS is already the database our researchers search first for information about manuscript and archival collections held at Harvard, and the project has mandated that every SGML finding aid must have a corresponding collection-level MARC record in HOLLIS. One might assume that in a place like Harvard, particularly with the Manuscript Survey and Guide Project, such records would already be in place. Not so! And Houghton (my own repository) is perhaps the worst offender. The survey project covered many, but not all, Houghton collections, and between the end of the project in 1986 and my arrival as curator in 1992, there was no one in the Manuscript Department who knew how to create a MARC record. A lack of support for MARC cataloging is still a problem for many of the smaller Harvard repositories, and this is something that Harvard will need to address programmatically. Keep this in mind if you begin your own EAD project—it will (or should) force you to think about your entire system of collection control, and there may be additional work other than EAD itself that you will need to anticipate.

The second “gateway” to the SGML finding aids is the search interface for a separately searchable finding aids database. It is extremely important to us to be able to do cross-finding-aid, cross-repository searching. This database will reside in HOLLIS Plus, a large group of databases outside the on-line catalog, and will be available at no cost to those outside Harvard. To implement this, an SGML search engine is necessary.

Initially, the DFAP finding aids were to serve as a test database for OCLC’s SiteSearch, which the university library was considering purchasing for use with a number of Harvard databases, including the finding aids. But we found that the finding aids were simply too complicated, and too long, for SiteSearch to handle well without an enormous investment in programming time. We have now selected OpenText’s LiveLink—the same package Yale is using—but do not yet (as of October 1997) have a search engine in place. The reason for the delay is that OpenText changed its pricing. Harvard needs only LiveLink Search, for which we were originally quoted a price of \$7,500. Subsequently, however, OpenText said that it would not sell the search module separately; we must buy

the entire suite of products for \$50,000 (this includes an academic discount). While Harvard continues to negotiate with OpenText, we also are exploring other search engine options.

In the meantime, the finding aids can be viewed individually in SGML using Panorama Free, as well as in HTML. A program converts the SGML into HTML documents “on the fly.” Until SGML browsers are readily available—and Extensible Markup Language (XML) offers that possibility in the not-too-distant future—individual finding aids will be viewable in HTML. (The indexes used by the search engine to locate the appropriate individual finding aids will, however, be drawn from the SGML tagging.) The major drawback to viewing a finding aid in HTML is that one loses the powerful navigational features that SGML browsers such as Panorama offer. Browsing a two-hundred-page finding aid in HTML can be tedious; browser utilization of SGML markup makes going back and forth in an electronic finding aid much easier.

Doing SGML Markup with EAD

The Harvard project’s third task, undertaken concurrently, was to mark up a variety of Harvard finding aids using SGML. It quickly became clear that, given the divergence of traditional practice at Harvard repositories, a Harvard standard (based, of course, on the Berkeley-developed standard, by this point in time called EAD) was needed. This standard does not so much prescribe the content of the finding aid (if you look at the Harvard guidelines on the DFAP website, almost every tag is optional), as it does the structure of the finding aid. We discovered, once we trained ourselves to think about the intellectual content of the finding aid divorced from how it “looked” (and this is, truthfully, the single most difficult thing to do in applying the EAD), there were not that many differences in content. Rather, what differed was the order in which information was presented and the level of descriptive detail provided.

One of the initial difficulties was how to make sense of an incredibly detailed and complex SGML Document Type Definition (DTD). I remember very clearly our first few meetings, when it was difficult to see the forest for the trees—the trees being hundreds of pointy angle brackets—and I felt very depressed at the prospect of trying to make sense of it all. This feeling passed once I realized that it was not necessary to apply *all* the EAD tags, and as I began to understand the structure of the DTD. We began the job of understanding EAD by passing around copies of what we each considered to be “typical” finding aids, and then we analyzed their various parts, trying to map them to the many EAD tags. This process will, I hope, be much simpler once formal application guidelines for EAD are available.³

But even with guidelines, I would urge every archivist to undertake this exercise of evaluating finding aid structure; I found it an illuminating and stimulating experience. Archivists are accustomed to doing things a certain way, not examining why or whether it is the best way, or if what we are doing is comprehensible to our user community. Learning how to apply EAD forces one to think hard about all these issues. I believe that over the next few years, finding aids will change—and not only the finding aids themselves, but also collection-level MARC records, and what information goes in which place, and how much overlap there is between them.

³A draft, prepared by Anne Gilliland-Swetland and edited by Thomas LaPorte, is available at <<http://scriptorium.lib.duke.edu/findaids/ead/guidelines/index.html>>.

By analyzing and mapping the elements in each of the finding aids, we soon decided that less was more. We chose to mark the structure of the front matter of the finding aid at a fairly general, not at the most specific, level. We use optional "level" and "attribute" tags only if we feel a need to do so for indexing purposes. There are an enormous number of options available in EAD; the challenge is to use the minimum level of markup that will produce a useful electronic document. This is not to say that a Harvard repository is prohibited from doing more detailed markup; it can be as detailed as the repository deems appropriate, as long as the basic structure conforms to the Harvard guidelines. But in terms of broad access to the information in the finding aid, we reached consensus that in many instances, implementing EAD to its finest, most granular level really was not necessary.

There is greater detail, and more variety in practice, in the markup of the container listing portion of the finding aid. Here, each repository judges the level of access needed by its user community. With literary collections, for example, every correspondent's name may be an important access point. In a corporate archives, name access may not be as important, and the markup of names may be much more selective. Thus far, the participating repositories have marked up all names and all dates; many form and genre terms are also marked up. Most problematic is use of the <subject> tag, since this leads into a discussion of the usefulness of indexing uncontrolled vocabulary terms (this also is an issue with indexing names, as not all repositories do name authority work). We have not yet determined the ideal level of markup, but we continue our discussions and our experimentation. Resolution of these issues will be facilitated when a search engine is in place and we can see the results produced by different markup decisions.

As a general principle, we mark up finding aids to a level that allows the building of useful indexes (see Appendix 1—preliminary indexing decisions). There is a school of thought that says that full-text keyword searching of finding aids, not field-specific indexing, provides adequate access. Harvard's experience with HOLLIS, with some 8 million bibliographic records, led us to favor field-specific keyword indexes, which we believe will provide more rapid and more precise retrieval of pertinent information.

In January 1997, the DFAP did a survey to attempt to quantify the scope of a possible Harvard retrospective conversion project for finding aids. Twelve of the forty-nine repositories responded. (This included all the largest collections, with the exception of the Medical School, which is undergoing renovation and reorganization.) There are more than fourteen thousand finding aids in these twelve Harvard repositories. Only 21 percent are already in electronic form, 50 percent are typed pages or cards, and the remaining 29 percent are handwritten pages and cards. We do not know exactly how many pages of data this represents, but we estimate at least five hundred thousand. In terms of numbers of alphanumeric characters, a Harvard finding aids database will be roughly equal in size to HOLLIS. Field-specific indexing will allow more precise retrieval within a very large database. Full keyword searching also will be an option, and it will be interesting to compare the uses of each approach.

In planning EAD projects, it is important to keep this question of scale in mind. For Harvard, given the potential size of a finding aids database, tagging at a level that will produce field-specific indexes is important. In the long term, if we think on a national or even international scale about access to electronic finding aids, it seems equally important. But for smaller repositories interested primarily in making finding aids available locally, detailed markup may not be seen as necessary; certainly developing a separate search engine would not be a wise use of resources. A few consortia, including RLG and SOLINET, have announced plans to provide a host site for finding aids marked up by their

members; this may be the best solution for small- and medium-sized repositories who want to make their collections available to the wider research community.

The Harvard encoding guidelines have been in place (and available on the DFAP website) for more than a year, although they are frequently modified and clarified as project members gain more markup experience (see Appendix 2). Individual repositories also are working on their own implementation standards, translating the Harvard guidelines into local practice. A style sheet, which displays all Harvard finding aids in a uniform style, was developed using Panorama Pro; this work was done largely by project member Susan von Salis of the Schlesinger Library with the assistance of MacKenzie Smith of the Office for Information Systems. More than sixty finding aids had been encoded as of August 1997.

The diversity of archives at Harvard and the considerable experience of the repositories' staffs have made project meetings lively forums for informed discussion and creative problem solving. Having participated in such a collaborative approach to understanding and implementing this technology, I have intense admiration for those who have undertaken EAD projects largely on their own initiative. At Harvard, we have found it necessary to meet regularly: for the first year we met for two hours every two weeks, and more recently, while we wait impatiently for our search engine to become operational, about every two months. We need to give each other moral support, because we all work on this project in our "spare" time. If it were not a congenial and extremely hard-working group, none of us would have the energy to keep going, given everything else we are expected to do.

Having many people involved in the project also makes its likelihood of survival much higher. When a project is the brainchild of one person, it can atrophy or die if that person leaves the organization. At Harvard, there are now enough people involved and committed that one of us can leave (and some have), and the project will continue. I would recommend that any "lone arranger" who is considering EAD implementation try to find at least one other local archives with whom to collaborate. It is helpful to have someone with whom you can talk face to face (e-mail is good, but not really the same). As mentioned earlier, implementing EAD forces you to rethink how your finding aids are structured and what they accomplish; this takes considerable thought and energy, and it helps enormously to discuss the issues with other archivists.

Producing Houghton Finding Aids Using EAD

Over the past ten to fifteen years, Houghton finding aids were produced using a fairly standard format. In planning for the inclusion of EAD tagging in our workflow, I took the Harvard EAD guidelines, went through a typical Houghton finding aid, and created a template with SGML tags (see Appendix 3). Instead of the manuscript cataloger having to know a lot about EAD markup in order to insert the appropriate tags, the tags appear already on the screen as part of the template, and the cataloger simply "fills in the form." The template in Appendix 3 is for a collection of personal papers, which is the bulk of what we acquire, and I anticipate that a separate template will be useful for other kinds of archival records. The template can receive text as well as tags, so any repeating text also can be inserted. This reduces the amount of keying and also helps enforce which tags should and should not be used. While inevitably there is some "tweaking" of the markup that needs to be done, the bulk of it is done quickly and easily. The template works equally well in WordPerfect and Author/Editor (we are using both at the moment).

One of the advantages of SGML is that data (and thus templates) can migrate between authoring packages with a minimum of fuss.

While the template approach works well for new finding aids, it is not particularly useful for Houghton's "legacy" finding aids. A finding aid created five years ago is formatted differently than one created thirty years ago. While the two contain much of the same information, the difference in the "look" makes it difficult to formulate routines for automatic conversion. This is even more the case when one looks at all the Harvard repositories over time. It will be necessary to formulate conversion guidelines not only for each repository, but for different time periods and different kinds of collection finding aids, within each repository. Conversion of existing finding aids cannot take place within each archivist's day-to-day work; outside funding and staff clearly will be needed.

The Larger Context

I will conclude with some observations about the Digital Finding Aids Project and EAD within the larger Harvard community.

My first point has to do with the "inevitability" of EAD implementation within the Harvard context. Harvard is wary of innovation, as are many other institutions. It has been critical for us to stress that EAD is an outgrowth of all the work that has gone before. Without the Manuscript Survey and Guide Project to create collection-level MARC records, without experimentation with Gopher sites, and without the recommendations of the Special Collections Task Force within the context of Harvard's preparations for a new on-line catalog, DFAP would not (and probably should not) have happened at Harvard.

Secondly, the long-term "survivability" of DFAP depends on its relevance within the overall Harvard electronic information environment. We are not solely a group of archivists. With the electronic texts librarian for Harvard College as a DFAP member, we keep in close touch with what is being planned in electronic information resources and can anticipate how our project can utilize and enhance these efforts. We obtained approval to switch the proposed search engine to LiveLink because that software was to be used for a resource much in demand by the faculty, the *Oxford English Dictionary*, and so we were able to argue that we should not stay with SiteSearch. Archives and manuscripts will continue to be a minority group within the larger electronic information system, and it is up to us to keep current with the larger context and to understand how it can serve our needs.

It also has been essential to have a good relationship with OIS, the university's library systems office, given that individual archives at Harvard will never have the staff and expertise to maintain a large and complex database over time. The university library has been very supportive of what the manuscript and archives community is attempting to accomplish. But equally, members of DFAP have put much time and effort into the project; it certainly has not been a question of archivists saying, "We want this, please give it to us." The archivists in DFAP have been largely responsible for writing the style sheet (Susan von Salis, Schlesinger Library), and the SGML to HTML conversion program (MacKenzie Smith from OIS, with assistance from Kim Brookes of Radcliffe College), are maintaining the project's website (Kim Brookes) and doing all of the actual markup without external programming support, and have invested much time in developing standards. OIS is providing the programming expertise for the design and implementation of the search engine, something most archivists cannot reasonably be expected to do themselves. But our contribution has been considerable.

The archival community is in a good position to get needed support from library systems offices at the moment, because what we want to do fits in well with other projects many libraries are undertaking, or are planning to undertake. At Harvard, one member of DFAP is now the university library's digital projects librarian, four members are part of various task groups responsible for selecting the next-generation Harvard on-line catalog, and two members are part of the college library's new Digital Library Projects development team. The nationwide effort to define and develop a National Digital Library will highlight the contribution that archives and EAD can make; it also will draw needed resources into the development of fully functional electronic finding aids for all kinds of collections.

This brings me to my third point: the necessity of explaining the project to non-archivists within an institution. By making presentations to various library managers groups, we have garnered interest and support among the upper management tier, as well as potentially bringing into the project many additional finding aids from visual, microform, and audio-recording collections across the university. We regularly issue "press releases" on Harvard listservs and have had enthusiastic feedback, particularly from reference librarians, who are now much more aware of manuscript and archives collections when giving research advice to students and faculty. Particularly in a large and decentralized institution such as Harvard, such "specialized" collections can become isolated from the larger research context. Electronic finding aids, integrated into the overall information structure of the university, enable us to demonstrate that "special collections" are essential to the research mission of the university, not luxury items that are peripheral to the "real" business of research. This, to me, is the most compelling argument for implementing EAD.

Appendix 1

Harvard/Radcliffe Finding Aids Database

Preliminary Indexing Decisions

(February 1997)

Non-public indexes

1. <eadid> used for updating by OIS
2. <admininfo><acqinfo><corpname> famname, persname role=donor OR role=source>: to provide donor/source index
3. <processinfo><name role=processor>: to provide access to collections cataloged by a particular cataloger.
The term “processor” is to be added to the Harvard-supplemented USMARC Code List for Relators as a standard term to be used as an attribute.

Public indexes

1. Names (keyword), including:

<name> anywhere
 <persname> anywhere
 <corpname> anywhere
 <famname> anywhere
 <archdesc><did><origination>

Dates and relation information found in name elements will be indexed as keywords together with the names. Names should be marked up to allow for optimal retrieval (i.e. normalized as much as possible, especially punctuation). This issue needs to be discussed further as we work on the database.

2. Repository name, to be used to limit searches

<frontmatter><titlepage><author>

3. Call number/accession number

<findaid><did><unitid>
 <admininfo><acqinfo><p><num>

4. Subjects (keyword) (names are not included in the subject index)

<findaid><scopecontent>
 <findaid><bioghist>
 <occupation> anywhere

Appendix 1 — *Preliminary Indexing Decisions—Continued*

<c>...<unittitle> but not subelements within <unittitle>
 <c>...<note>
 <genreform> anywhere
 <subject> anywhere
 <geogname> anywhere

5. Geographic name index

<geogname> anywhere

6. Titles of works (as opposed to cataloger-created descriptions)

<title> anywhere

7. Dates

<c>...<unitdate>
 <date> anywhere except in <frontmatter> and <admininfo>
 <findaid><did><unitdate> gives inclusive dates of collection materials
 <admininfo><acqinfo><date>

Dates will need to be four-digit numbers, and only those numbers will be indexed. That is, if a date field contains 1920-1930, a search for 1925 will not point to that field.

Notes

Ignore punctuation in indexing () . , [] —

The user should probably be presented with a list of finding aids that met her/his search criteria, with each finding aid only appearing once on the list (unless we can implement weighted search results, which is unlikely). The user would then enter each finding aid at its top and be directed to do a key word search using the browser's "find" feature within that finding aid.

At this point, all of the indexes are envisioned as being key word indices (as opposed to exact string indexes) because of the lack of standardization in the forms of these elements and the lack of subelements (such as surname and forename) to help. If we would like phrase searching we may need to rethink how we have defined the subject index.

Appendix 2

Harvard/Radcliffe EAD Application Guidelines (for EAD version 1; March 1998)

Notes on the guidelines

Punctuation: keep punctuation together with the word(s) it modifies, inside the tags... i.e. treat it as character information that belongs with the text, not as separate white space. Punctuation, meaning the symbols such as commas, periods, quotation marks, brackets, etc. and NOT meaning whitespace (tabs and blanks and carriage returns) will be normalized out of the indexes so that users won't have to know the punctuation to retrieve documents.

If there is extra text that doesn't logically fit in the heading, or in a `<c>` element, like a label or a running title that was included to make the print version more readable but that has no function in the finding aid, use an `<odd>` element to enclose it. For example, if a box number only refers to a physical location (not the intellectual order of the material), use `<odd>`. If, on the other hand, the box number uniquely identifies the material being described, tag it as a `<unitid>` within a `<c>` that begins before the box number and ends at the end of the box.

In general, do not use rendition attributes (`<emph>`, `<lb>`, etc.). Rely on the stylesheet instead.

Use a `<head>` element for information that helps the reader with navigation, or provides useful information, but should not be indexed, such as labels or folder summaries. For information that is there for retrieval purposes (i.e. should be indexed) use a `<title>` element such as the `<unittitle>` defined in the `<c>` element.

For the attribute "role," use the USMARC Code List for Relators. The term "processor" may also be used.

If the finding aid is in a language other than English you should record that fact with the `LANGUAGE="en"` attribute on the element. If the finding aid is in mixed languages you can't really reflect that now, so choose the predominant language and use the attribute if it's not English.

Use internal links (`<ref>` and `<ptr>`) only. Do not use external references, as these will undoubtedly change; wait until the situation is more stable.

Additional notes about our preliminary indexing decisions are also available.

Guidelines

`<ead>`

`<eadheader>`

Appendix 2 — *Harvard/Radcliffe EAD Application Guidelines—Continued*

<eadid>xxxxxxxx</eadid>

[Unique 8 character id number for the digital version of the finding aid, where xxx is the HOLLIS loc code (ex. hou for Houghton, sch for Schlesinger, or des for GSD), and nnnnn is a five digit number for the finding aid. These numbers will be used to update the union database of finding aids, so it's essential that they be unique within each repository.]

<filedesc><titlestmt><titleproper> use the same title as in the <titleproper> element from the <frontmatter> section below </titleproper></titlestmt></filedesc>

These elements are needed because the DTD requires them, but we won't be using this information. At least a call number must be encoded in the <titleproper> element.

<profiledesc><creation><date>mm/dd/yyyy</date> encoder's name [optional] </creation> </profiledesc>

[The <date> subelement should contain the creation date of the digital version of the finding aid. The encoder is the person who did the SGML markup and is optional.]

<revisiondesc><change><date>mm/dd/yyyy</date><item><name> encoder's name [optional] </name></item> </change></revisiondesc> [optional]

[This element set holds date and (optionally) name information for revisions to the SGML-encoded finding aid. Its function is similar to that of a publishing history for a book, and it allows you to retain the information about when and by whom the original version was created, as well as verify the most recent version.]

</eadheader>

<frontmatter>

<titlepage>

<num> collection/accession number </num> [optional]

<titleproper> official finding aid title </titleproper>

[drawn from the MARC 100 and/or 245 field, if they exist, followed by a colon and words distinguishing this as descriptive material, not the material itself (e.g. register, contents list, guide, inventory, or finding aid). For example, Jane Doe Papers: A Guide.]

<author> repository name </author>

<publisher> Harvard University </publisher>

<date> date finding aid was created </date> [optional]

<p>© YYYY The President and Fellows of Harvard College</p> </titlepage>

</frontmatter>

Appendix 2 — Harvard/Radcliffe EAD Application Guidelines—Continued

<archdesc level="collection">
<did>

[NOTE: the following optional elements can be used in any order that makes sense for the finding aid in hand. Those marked "optional, recommended" are elements that will be required in the RLG union database, used to generate brief information about the collection.]

<repository> repository name </repository> [optional, recommended]
<unitid> call number, accession or collection number (can also be HOLLIS#)
</unitid> [optional]
<physloc> shelving designation</physloc> [optional]
<unittitle> collection title [optionally can use a subelement for <date> dates of coverage </date>] </unittitle> [optional, recommended]
<unitdate> date(s) of collection materials </unitdate> [optional]
<origination> creator(s) of the collection, with or without dates. Names can be repeated if more than one creator was involved (use <persname>s; or <corpname>s, etc.
</origination> [optional, similar to the MARC 100 field]

NOTE: the collection creator can be input as a <persname role="originator"> name
</persname> if this information is wanted for local indexing.

<physdesc><extent> information about the size of the collection (can optionally include a <num> element with a call number for a microfilm version, etc.) </extent>
</physdesc> [optional, recommended]
<abstract> A very brief summary of the collection </abstract> [optional, recommended]

</did>

<admininfo> [optional]

<processinfo> [optional]

<head> "Processed by:" </head> [optional]
<p> collection processor (person who created the original finding aid) </p> [optional]
<p> date of creation </p> [optional, and additional <date> tags can be used in the <p> if desirable]

</processinfo>

<acqinfo> [optional]

<p><num> accession number, or details of acquisition </num></p> [optional]
<p><persname role=donor> donor </persname></p> [optional]
<p><date> accession date </date></p> [optional]

</acqinfo>

Appendix 2 — Harvard/Radcliffe EAD Application Guidelines—Continued

<accessrestrict> restrictions on access, if any </accessrestrict> [optional]
 <userrestrict><p> restrictions of use of materials, if any </p></userrestrict> [optional]
 </admininfo>

<bioghist> [optional]

<head> heading such as Biography or History of Organization </head> [optional]

<geogname, persname, etc.> Name [optional <date> dates </date>] </geogname, persname, etc.> [optional]

<chronlist> [optional]

<chronitem><date></date><event>...</event><</chronitem> [repeatable]

</chronlist>

NOTE: Use this approach, instead of the simpler <persname> approach, when the biographical note consists of a chronological list.

</bioghist>

<scopecontent> [optional, repeatable]

<head> heading such as Notes or Scope and Content Note </head> [optional]

<geogname, persname, etc.> Name [optional <date> dates </date>] </geogname, persname, etc.> [optional]

<organization> description of the structure of the finding aid, e.g. "Organized into the following series: I. Correspondence; II. Biographical files." </organization> [optional]

<arrangement> description of the filing sequence, e.g. alphabetical, chronological (if in doubt, use the broader <organization>) </arrangement> [optional]

</scopecontent>

<dsc type="insert appropriate value here"> [repeatable, type attribute value will usually be "in-depth"]

<head> heading for collection-level info [recommended]

[Note: preferred to using <dsc><c><head>]

<c level=series, subseries, etc.>

<did>

<head> [optional]</did>

<scopecontent> </scopecontent> [optional]

Appendix 2 — *Harvard/Radcliffe EAD Application Guidelines—Continued*

<c level=file/item/etc.><did>

[optional and repeatable. The default assumption is that level=item, and that “items” are whatever is the lowest level of description that you have chosen to do (so, for example, if you only describe to the folder level, then you don’t need to encode level=“file” for each folder since the “item” display treatment should work equally well)]

<container> or <physloc> Box, etc. designation, or location; can go with the unitid information. This can go before or after the <unitid> element. </container> or </physloc> [optional]

<unitid> id number for box, folder or other physical unit. This can be the item number, if one is assigned, or an item range like 1-9, or any other information that points to the physical thing being described. </unitid> <unitdate> dates at the <c> level </unitdate> [optional]

[NOTE: <unitdate> can be placed inside the <unittitle> if appropriate, and will display inline with the title. For other kinds of dates, use the <date> element inside the <unittitle> or other <c>-level subelement.]

<unittitle> title information for box, folder, series name, title or description
</unittitle>

NOTE: if your description has only a <unitdate>, put the <unitdate> within a <unittitle>. Use of the following tags is optional, but recommended, within <unittitle> for indexing purposes:

<persname></persname>

<corpname></corpname>

<famname></famname>

<genreform> Use terms from the DFAP list of recommended form and genre terms. </genreform>

<physdesc><extent> size of unit being described (eg. 1 folder, 12 folders, 2 boxes, etc.) </extent></physdesc> [optional]

<note> <p>... </note> </p> [optional]

</did></c> [for level=file, item, etc.]

</c> [for series components, if any]

<controlaccess> [optional]

[use for terms that should be indexed, but that do not appear explicitly in the finding aid (e.g. LCSH, NAF forms of names). All <controlaccess> terms are under authority control. <Controlaccess> can appear at many points in a finding aid.]

Appendix 2 — Harvard/Radcliffe EAD Application Guidelines—Continued

<name> ... </name> [optional, repeatable]
 <famname> ... </famname> [optional, repeatable]
 <persname> ... </persname> [optional, repeatable]
 <corpname> ... </corpname> [optional, repeatable]
 <geogname> ... </geogname> [optional, repeatable]
 <occupation> ... </occupation> [optional, repeatable]
 <subject> ... </subject> [optional, repeatable]
 <genreform> ... </genreform> [optional, repeatable]

</controlaccess>

</dsc>

<add> [optional]

<index>

[<index> is a **list** of terms, usually at the end of a finding aid. Terms are not necessarily under authority control]

<indexentry><namegrp><name> ... </name><name> ... </name>
 </namegrp></indexentry> [optional, repeatable]
 <indexentry><name> ... </name></indexentry> [optional, repeatable]
 <indexentry><famname> ... </famname></indexentry> <ptr></ptr> [optional, repeatable]
 <indexentry><persname> ... </persname></indexentry> <ptrgrp><ptr>
 </ptr><ptr></ptr> </ptrgrp> [optional, repeatable]
 <indexentry><corpname> ... </corpname><ref target="source ID">
 </ref></indexentry> [optional, repeatable]
 <indexentry><geogname> ... </geogname></indexentry> <ptrgrp><ref>
 </ref> <ref></ref> </ptrgrp> [optional, repeatable]
 <indexentry><occupation> ... </occupation></indexentry> <ptr></ptr>
 [optional, repeatable]
 <indexentry><subject> ... </subject></indexentry> <ptr> [optional, repeatable]
 <indexentry><genreform> ... </genreform></indexentry> <ptr></ptr> [optional, repeatable]

[Each variation of <ptr>, <ptrgrp>, and <ref> may be used in each element of an <indexentry>. Use these tags for internal links.]

</index>

</add>

</archdesc>

</ead>

Appendix 3

EAD Template for Houghton Library Finding Aids

The following template does not include all possible options, but only those elements for coding information usually found in Houghton finding aids. Text in [] is to be supplied by the cataloger; other text and punctuation marks outside the [] are required.

Macros are defined to repeat <c level="item"> ... </c> tags. Level="item" is the default value.

=====

```
<ead>
<eadheader>
<eadid>hou[nnnnn]</eadid>
```

```
<filedesc><titlestmt><titleproper>[Main entry, title, and inclusive dates of collection,
as in MARC record; followed by the size designation (b,f,pf) & call number in paren-
theses].
```

```
Guide.</titleproper></titlestmt> </filedesc>
```

```
<profiledesc><creation><date>[mm/dd/yyyy] </date>[encoder's name]</creation>
</profiledesc> </eadheader>
```

```
<frontmatter>
<titlepage>
<num>[size designation and call number]</num>
<titleproper>[same as "titleproper" in "filedesc" above, without the size designation
and call number in parentheses]. </titleproper>
<author>The Houghton Library</author>
<publisher>Harvard University, Cambridge, MA 02138</publisher>
<date>[year finding aid was created in form yyyy]</date>
<p>&copy; [YYYY] The President and Fellows of Harvard College</p>
</titlepage>
</frontmatter>
```

```
<findaid>
<archdesc level="collection">
<did>
<unitloc>[size/shelving designation, e.g. b,f,pf]</unitloc>
<unitid>[call number]</unitid>
<unititle>[use same title as in "titleproper" in "titlestmt"]. </unititle>
<unitdate type="inclusive">[date(s) of collection materials]. </unitdate>
<physdesc><extent>[xxx] boxes ([xxx] linear ft.) </extent></physdesc>
</did>
```

Appendix 3 — EAD Template for Houghton Library—Continued

```

<admininfo>
<processinfo>
<head> Processed by: [your name here]. </head>
</processinfo>

<acqinfo>
<p><num>[accession number]</num></p>
<p><persname role="donor" or role="source">[donor name and/or source of
collection (select corresponding attribute)]. </p>
<p><date>[accession date, in form YYYY]</date>[month day]</p>
</acqinfo>

<accessrestrict><p>[restrictions on access, if any]. </p></accessrestrict>
<userrestrict><p>[restrictions of use of materials, if any]. </p> </userrestrict>
</admininfo>

<dsc type="in-depth">
<c level="collection">
<c level="series">
<head>[I. Correspondence or Letters]</head>

<c>
<did><unitid>([item #])</unitid>
<unittitle><persname>[name of correspondent as in LCNAF: last name, first name],
</persname></unittitle><unitdate type="inclusive">[YYYY-YYYY; or put in single
date and change attribute to "single"]. </unitdate><physdesc><extent> [#] folders.
</extent></physdesc>
</did>
<note><p>[Note on contents, if needed]. </p></note>
</c>
</c>

<c level="series">
<head>[II. Compositions by xxx]</head>
<c>
<did><unitid>([item #])</unitid>
<unittitle><title>[title of work] : </title><genreform>[AMs, Ts, etc.], </genre-
form></unittitle><unitdate type="single">[YYYY]. </unitdate>
<physdesc><extent>[# of folders]. </extent></physdesc>
</did>
<note><p>[note on contents if needed. Markup names and genre/form as appropriate].
</p></note>
</c>
</c>

```

Appendix 3 — *EAD Template for Houghton Library—Continued*

</c>
</dsc>
</archdesc>

<add>
<index>
<indexentry><persname> ... </persname></indexentry>

</index>
</add>

</findaid>
</ead>