## AN INTRODUCTION TO ENCODED ARCHIVAL DESCRIPTION

Susan J. von Salis Schlesinger Library Radcliffe Institute for Advanced Study 10 Garden Street Cambridge, MA 02138

ABSTRACT: Encoded Archival Description (EAD) is a newly-developing standard that increasing numbers of archivists are using to "mark up" finding aids in Standard Generalized Markup Language (SGML). Unlike finding aids, existing on-line public access catalogs provide only summary descriptions of our collections. The finding aids that archives maintain in-house contain detailed contents lists, and provide a context and a hierarchy in which manuscript materials can be more easily identified and comprehended. By using a standard Document Type Definition (DTD), archivists are able to make finding aids available on the World Wide Web in a way that allows for complex searches both within and across repositories; the finding aids retrieved by any given search can be presented so that they are both intelligible and easily navigable.

# **Encoded Archival Description**

Archival and manuscript finding aids are valuable, information-rich documents that provide in-depth information about the contents of manuscript collections. Yet until recently, these documents were available to researchers only when they had already determined what specific collections they were interested in, and then only on-site at a repository. For the past five years, the archival community has been developing Encoded Archival Description (EAD), a Standard Generalized Markup Language (SGML) Document Type Definition (DTD), for use in marking up finding aids. SGML is a "markup language" that enables us to identify the parts or "elements" of an electronic document by embedding codes. We can then use various computer applications to recognize the codes to provide the document with some "intelligence," allowing us to provide complex searching and display capabilities.

## Harvard/Radcliffe Digital Finding Aids Project

There are over 49 archival repositories at Harvard/Radcliffe and, like the rest of "decentralized" Harvard, they have been run fairly autonomously, developing their own collecting policies, administrative systems, and cataloging records independently of one another. But there are of course times when we recognize the benefits of working on projects cooperatively! In 1994 some of us wanted to make the contents of our finding aids available and searchable across repositories in the same way that collection-level records were searchable across the university via HOLLIS, Harvard's OPAC. Our first

attempt was a gopher site. We eventually put up 150 finding aids, but became frustrated with gopher's limitations:

- navigation within a document was cumbersome
- one couldn't see a folder or item description within the context of the whole finding aid
- searching was primitive

We decided to look for a new way to make our finding aids available. Our discovery of the emerging DTD for archival finding aids spurred the creation of the Digital Finding Aids project (DFAP) in the spring of 1995. Today there are 10 repositories represented on DFAP, in addition to staff from Harvard's Office for Information Systems. Together we have mounted almost 200 finding aids.

# **Finding Aids**

I assume that everyone here is familiar with archival finding aids. These documents contain descriptive information about the contents of manuscript collections, such as

- where did the collection come from?
- how is it arranged?
- what does it contain?

A finding aid differs from a MARC record in that it includes a detailed list of a collection's contents, usually by folder or file, while MARC provides more of an overview of a collection as a whole. The structure of a finding aid is usually hierarchical; it might list the entire collection (Pauli Murray papers), a series (work), a subseries (law), and then individual files (case files). The point, of course, is to provide a tool so that researchers can find specific information about the contents of a collection so they don't have to come to the library and wade through individual files in many boxes to locate the specific items they are interested in. Since SGML is also very hierarchical, it is an almost tailormade tool for making finding aids available on the web.

Let's take a moment to look at a finding aid. While the "look and feel" of finding aids can, and do, differ significantly between repositories, we at DFAP discovered that the "parts" of our finding aids are more similar than it appeared at first glance. For example, here's a finding aid from the Schlesinger Library [see slides, Appendix A]: we can identify the sections that provide information about the size of the collection, its provenance, any restrictions on access or use, a biography of the person and a scope and contents note, followed by a folder list. The limitations of a paper finding aid are similar to those of gopher: it's hard to navigate through the document, to know "where you are" in context, and it's impossible to search.

## Using EAD

As I mentioned, most finding aids include common components such as provenance, scope and contents notes, and access restrictions. It is these "parts" that comprise the elements of the SGML DTD. Markup itself is simply a matter of wrapping the correct tags around the proper text. Presumably you're familiar with HTML markup; since HTML is just another DTD (like EAD), the markup process is similar, although EAD is a significantly more complex DTD than HTML. So let's look again at the Buttenwieser finding aid [see slides, Appendix B]. To mark it up, we identify a body of text as an "element," let's say acquisitions information. The accession number and the paragraph about the donor are acquisitions information, so we wrap the tag around that chunk of text. By identifying and marking up the text chunk by chunk, we eventually mark up the whole document.

You should know that markup itself is only a small portion of an EAD project. One still needs to provide access to these encoded finding aids, by considering the following:

- storage
- searching
- display
- navigation

## Storage

By storage I mean that you need to have a physical machine where the EAD files, which often get quite large, can be saved. You also need to be able to provide users access to these files as they search for finding aids. Of course, you also need to provide a search engine that is capable of the kinds of searching needed to search for complex criteria across large numbers of files, and a comprehensible front end.

# Searching

One can, of course, search EAD finding aids with a web search tool. For example, if you perform a simple search of the Harvard/Radcliffe finding aids for the word "church," you'll get hits from a number of repositories, but you can't differentiate between when "church" is part of a corporate name, a personal name, or a subject. This is hardly a useful search result. We want to be able to take advantage of the markup we've been investing our time in, so we need to be able to perform intelligent searches, based on the intellectual components of a finding aid (using the EAD tags). We want users to be able to specify when they are searching for names, subjects, form and genre terms, and so on. It is also important to provide ways to search for these terms in certain **contexts**. Does the researcher want to find every finding aid that includes the genre term "letter," or only those that have the term "letter" in a folder description? Are they interested in the name "Mary Smith" when that person is the creator of a collection, or a donor, or a processor?

Then there's "proximity searching." Finding aids are not regulated by any kind of controlled vocabulary. A finding aid may refer to "Henry James" in the biography

section, and "James, Henry" in a list of correspondents. Proximity searching allows us to let researchers search for the name Henry whenever it appears near the name James, in any order, whether or not there is an initial (Henry J. James) or date (James, Henry, 1811-1882) connected to the name. This is how a proximity search might look in our search interface, called OASIS (Online Archival Search Information System) [see slides, Appendix C]. Proximity searching can also allow one to search for terms other than names. For example, it is very useful to be able to perform a search for the date 1914 near the subject "suffrage parade" and the form/genre term "photograph."

# Display and navigation

Once a researcher finds a finding aid, it is important that she views it in a comprehensible way. This is where display and navigation come in. Currently one needs to download and configure a "helper application" that works with the web browser to interpret a stylesheet, rendering the finding aid in a desired way. The best-known application is Panorama. Once one has identified the components of a finding aid in SGML, one can, as we have seen, use this "intelligence" to search across finding aids. One can use the same tagging to render the finding aids in a comprehensible way. Here is an example of how a finding aid looks in Panorama [see slides, Appendix D]. Panorama displays a split screen, with the full text of the finding aid on the right, and a navigational tool on the left. As one scrolls through the finding aid, the yellow indicator on the left moves, showing one exactly where one is in the finding aid at any given point. One of the limitations of a paper finding aid is that it is difficult to keep track of "where one is" in the often complex hierarchy of the arrangement. Is this particular folder description, for example "correspondence between EK and Gordon Browning" part of the "Personal and Biographical" series or part of the "Democratic Party" series? With a stylesheet, one can simply look at the navigator to identify the specific series and subseries one is reading about in the folder list. The navigation display can also "expand" and "collapse" groups of folders, making it easier to view an overall outline of the collection's arrangement at a glance.

In addition, one can use the stylesheet to take advantage of SGML's linking capabilities. One can, for example, link an index entry to the folder it refers to. One can also link a "see also" reference to the folder or series it pertains to, enabling the researcher to jump directly to a related group of materials. These links can also be used for external references as well as those within a finding aid itself. The external reference can be an EAD finding aid for a related collection, or it can link to an image or sound file associated with an item described in the finding aid.

The drawback at this time of using such a helper application is that it is sometimes difficult to download and configure the software. Archivists are waiting expectantly for the full implementation of eXtensible Markup Language (XML) in the major web browser software; this will enable us to create stylesheets that do not require additional software for the user. For those who do not wish to install Panorama, we are also making the finding aids available in HTML. While doing so removes all of the EAD tagging we

worked so hard to encode and "dumbs down" the document so that we can no longer display it with any level of accompanying navigation, it does allow access to the text of finding aid to a larger audience.

That is EAD in a very small nutshell. Embarking on such a project involves a significant commitment of staff time and IT resources, and includes much more than simply markup. At Harvard/Radcliffe we were able to implement DFAP only because we were fortunate enough to have the resources and technical expertise of Harvard's Office for Information Systems available to us. Many other EAD projects have been undertaken by repositories that also have significant systems assistance, who have full-time it staff available inhouse, or who are involved in cooperative projects with other repositories so that they can share the workload.

I should point out that nobody sees EAD as a replacement for MARC. There is still a need for enabling researchers to perform collection-level searching using broad strokes with the use of a controlled vocabulary. In fact, at Harvard/Radcliffe all collections to be included in OASIS must have a corresponding record in HOLLIS. The HOLLIS record must include a 856 field with a link directly to the finding aid in OASIS. Here are the results of a HOLLIS search for "women lawyers--New York (State)" [see slides, Appendix E]. By clicking on the listed URL, we can go directly to the text of the finding aid.

I hope that you have gotten a sense from this brief talk of why archivists are so excited about EAD. Please visit the OASIS website, or any of the others on the handout [see Appendix F].

# HELEN (LEHMAN) BUTTENWIESER, 1905-1989 Papers, 1909, 1940, 1968-90 2 file boxes, 1 folio folder, 1 folio folder

Processed: February 1994

By: Bert Hartry

Accession number: 90-M123

The papers of Helen L. Buttenwieser were given to the Schlesinger Library in July 1990 by her heirs via her son and executor, Lawrence B. Buttenwieser.

# TERMS OF USE

Access. Unrestricted.

Copyright. Radcliffe College holds copyright for the Schlesinger Library in papers written by HLB; copyright in other papers in the collection may be held by their authors, or the authors' heirs or assigns. Researchers must obtain the written permission of the holder(s) of copyright and the director of the Schlesinger Library before publishing quotations from materials in the collection.

Copying. Unrestricted.

## **BIOGRAPHY**

Helen (Lehman) Buttenwieser, lawyer, civil libertarian, children's rights advocate, social worker, and philanthropist, was born on October 8, 1905, in New York City. She was the second of three daughters of Adele (Lewisohn) and Arthur Lehman (AL), members of two of the leading German-Jewish families in the city. ALL's father, Adolph Lewisohn, made a fortune in copper and the shipping of mining products, and was a leader in prison reform. AL's father, Mayer Lehman, was a founder of the New York Cotton Exchange, a partner in Lehman Brothers bank, and the father of three prominent sons: AL, a banker; Herbert, elected governor of New York and later to the U.S. Senate; and Irving, chief judge of the New York State Court of Appeals....

### SCOPE AND CONTENT

These papers partially document HLB's life between 1968 and 1990, with both professional and, especially, personal papers. They yield no information about her parents, and very little about her siblings, education, friendship with Alger Hiss, or legal cases. There is, however, some family correspondence as well as material about HLB's philanthropic interests.

The most interesting correspondence is between HLB and Helen Suzman, South African liberal, Member of Parliament, and co-founder of the Progressive Party. Two letters from publisher Alfred A. Knopf deal with abortions for poor women and the state of the union (U.S.A.).

The HLB papers have been divided into two series:

Series I, <u>Personal and biographical</u> (#1-15f), includes a curriculum vitae, entries in law directories, a lengthy interview with HLB, some writings and speeches, photographs, clippings, information about two HLB oral histories, and some documentation of her involvement with the Legal Aid Society, as well as related correspondence.

Series II, <u>Correspondence</u> (#16-56), is divided into two sections: correspondence grouped by individual or subject, and arranged chronologically within each category; and general correspondence arranged chronologically.....

# Additional catalog entries

American Civil Liberties Union American Jewish Committee Ames, Oakes Association of the Bar of the City of New York Auchincloss, Louis Beggin, Gussie...

Grandparent and child Hiss, Alger Jewish families--New York (N.Y.) Jewish women--New York (N.Y.)

Women lawyers--New York (State)

#### INVENTORY

## Series I. PERSONAL AND BIOGRAPHICAL

Box 1: 1-29

- 1. Biographical: curriculum vitae, affidavit, Who's Who in American Law and Martindale-Hubbell Law Directory entries, etc., 1981-89, n.d. See also #35, 50, 53, 54.
- 2f\*. Certificate admitting HLB to practice before U.S. Supreme Court, 1968

- 3. Re: HLB oral histories: American Jewish Committee (1977), Federation of Jewish Philanthropies (1982); includes correspondence, 1978-83.
- 4-6. 1989 interview; includes correspondence, 1989-90.
- 7. Honors: citation, invitation, correspondence, 1975-89 (scattered). See also #10-12, 14, 26-28.
- 8. Photographs of HLB: alone, with family, with others, n.d. FILED IN PHOTOGRAPH DRAWER.
- 9. Writings, 1971-76; includes article re: Alger Hiss by Allen Weinstein.
- 10. Speeches, 1983, n.d.

15f. 1987-88

# Series II. CORRESPONDENCE.....

## INDEX OF SELECTED CORRESPONDENTS

This index includes the names of selected writers and recipients. Information about persons and subjects is not indexed.

Key No symbol = Writer

\* = Writer and recipient

+= Recipient

The numbers refer to folders.

American Civil Liberties Union 32, 50\* See also Glasser, Ira; Carliner, D.

American Jewish Committee 3

Ames, Oakes 22, 25+, 33, 37, 40. See also Connecticut College.

Association of the Bar of the City of New York 45, 55

Auchincloss, Louis S. 1+, 22, 25+, 46+

Bank Street College 48

Barnard College 7\*, 49

Bass, Audrey De Jong 43

Bates, Bethlyn 33

Beggin, Gussie 22, 25+, 33, 34+, 40\*, 52\*, 53+, 55\*

Bernhard family 23, 25+, 33+, 46\*

Bernstein, Marver H. 39

Bingham, June and Jonathan 22, 25+, 48+

Brandeis University 7, 38, 39

......

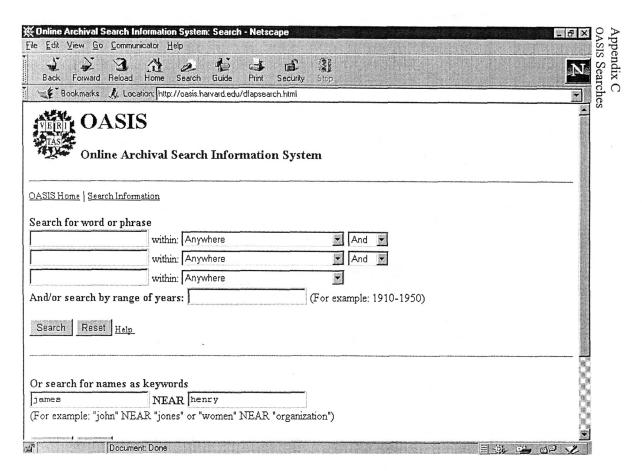
Appendix B

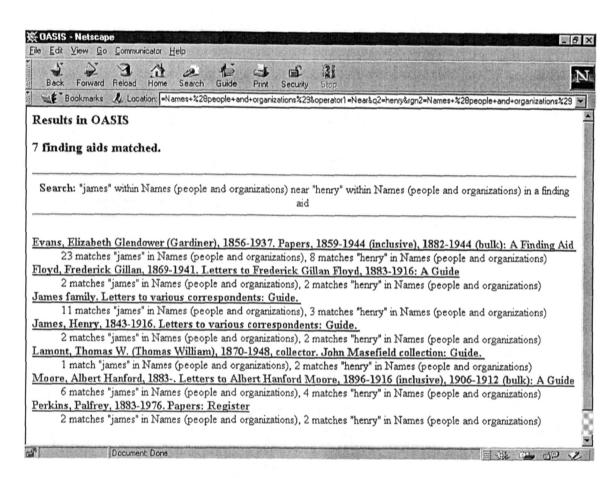
Schlesinger Library Finding Aid with markup

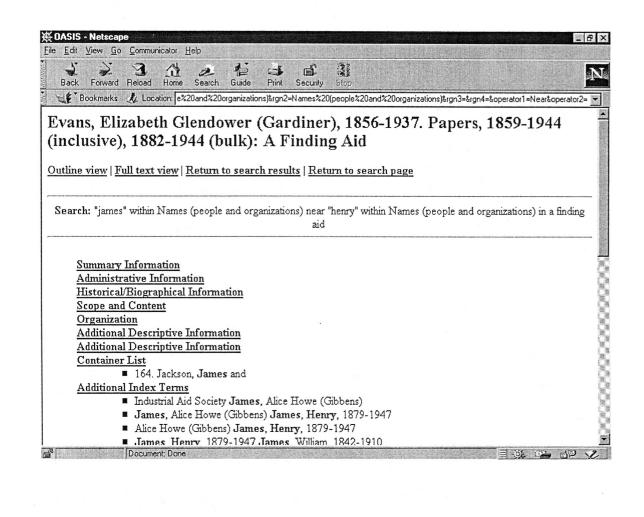
```
<ead>....
<archdesc level="COLLECTION">
<hib>
<repository>Schlesinger Library, Radcliffe College</repository>
<unitid>MC 425</unitid>
<origination><persname role="originator">HELEN (LEHMAN) BUTTENWIESER,
1905-1989</persname></origination>
<unittitle>Papers, <unitdate>1909, 1940, 1968-90</unitdate></unittitle>
<physdesc><extent>2 file boxes, 1 folio folder, 1 folio+ folder/physdesc>
</did>
<admininfo><processinfo>
Processed: <date> February 1994</date></processinfo>
Sy: <personame role="processor">Bert Hartry</personame>/processinfo>
<acqinfo>
Acession number: <num> 90-M123</num>
The papers of Helen L. Buttenwieser were given to the Schlesinger Library in July
1990 by her heirs via her son and executor, <persname role="Donor">Lawrence B.
Buttenwieser</persname>.</acqinfo>
<accessrestrict><head>TERMS OF USE</head>
Access. Unrestricted.</accessrestrict>
<userestrict>
Copyright. Radcliffe College holds copyright for the Schlesinger Library in papers
written by HLB; copyright in other papers in the collection may be held by their authors,
or the authors' heirs or assigns. Researchers must obtain the written permission of the
holder(s) of copyright and the director of the Schlesinger Library before publishing
quotations from materials in the collection.
Copying. Unrestricted.
</userestrict></admininfo>
<br/>bioghist>
<head>BIOGRAPHY</head>
Helen (Lehman) Buttenwieser, lawyer, civil libertarian, children's rights advocate,
social worker, and philanthropist, was born on October 8, 1905, in New York City. ...
</bioghist>
<scopecontent>
<head>SCOPE AND CONTENT</head>
These papers partially document HLB's life between 1968 and 1990, with both
professional and, especially, personal papers. ....
```

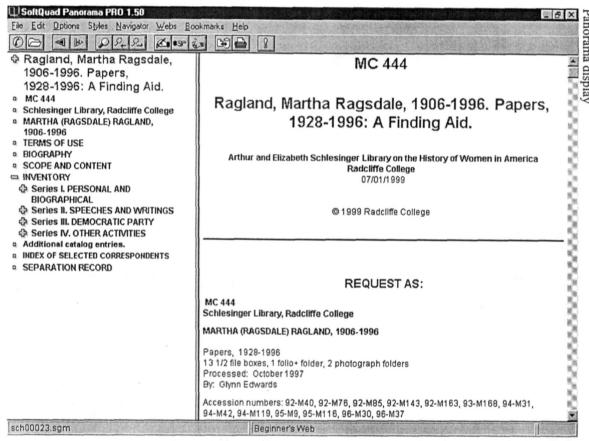
```
The most interesting correspondence is between HLB and Helen Suzman, South
African liberal, Member of.... 
</scopecontent>
<controlaccess>
<head> Additional catalog entries.</head>
The following catalog entries represent persons, organizations, and topics
documented in this collection. An entry for each appears in the Harvard On Line Library
Information System (HOLLIS) and other automated bibliographic databases.
<corpname>American Civil Liberties Union </corpname>
<corpname>American Jewish Committee </corpname>
<persname>Ames, Oakes</persname>
<subject>Women--Charitable contributions/subject>
<subject>Women lawyers--New York(State)</subject>
</controlaccess>
<dsc type="IN-DEPTH">
<head>INVENTORY</head>
<c level="SERIES">
< did>
<head>Series I, PERSONAL AND BIOGRAPHICAL</head>
</did><c level="otherlevel"><did>
<container>Box 1: 1-29</container>
</did></c>
<c id="sch00005f1"><did><unitid>1.</unitid> <unittitle>Biographical: curriculum
vitae, affidavit, Who's Who in American Law and Martindale-Hubbell Law Directory
entries, etc., <unitdate>1981-89, n.d.</unitdate> See also #<ref
target="sch00005f35">35, </ref><ref target="sch00005f50">50, </ref><ref
target="sch00005f53">53. </ref><ref target="sch00005f54">54</ref>.
</unittitle></did></c>
                                         2f+.</unitid><unittitle>Certificate
<c id="sch00005f2f"><did><unitid>
admitting HLB to practice before U.S. Supreme Court,
<unitdate>1968</unitdate></unittitle></did></c>
<c id="sch00005f4-6"><did><unitid>4-6. </unitid>
<unittitle>1989 interview; includes correspondence, <unitdate>1989-90.
</unitdate></unittitle></did></c>
<c id="sch00005f7"><did><unitid>7.</unitid>
<unittitle>Honors: citation, invitation, correspondence, 1975-89 (scattered). See also
#10-12, 14, 26-28.</unittitle>
```

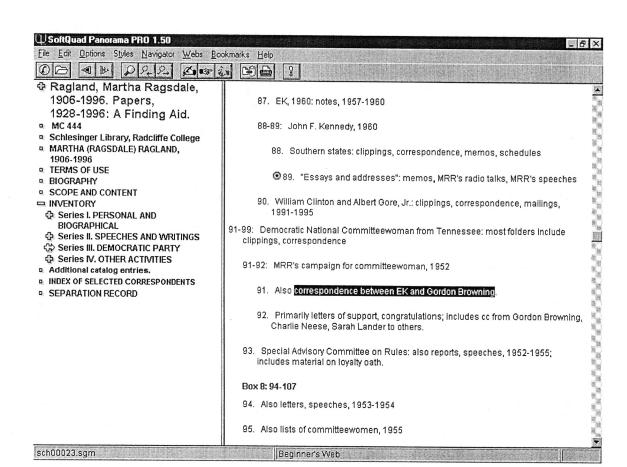
```
</did></c></c>
</dsc>
<add>
<index>
<head>INDEX OF SELECTED CORRESPONDENTS</head>
This index includes the names of selected writers and recipients. Information about
persons and subjects is not indexed.
Key
               No symbol = Writer
* = Writer and recipient
 = Recipient
The numbers refer to folders. 
<indexentry><corpname>American Civil Liberties Union</corpname><ref>see also
Glasser, Ira; Carliner, D. </ref></indexentry>
<indexentry><corpname>American Civil Liberties Union</corpname>
cptrgrp><ref target="sch00005f32">32,</ref><ref</pre>
target="sch00005f50">50*</ref></ptrgrp></indexentry>
<indexentry><corpname>American Jewish Committee</corpname>
<ref target="sch00005f3"> 3</ref></indexentry>
<indexentry><persname>Ames, Oakes</persname><ptrgrp><ref target="sch00005f22-
23">22,</ref><ref target="sch00005f25">25,</ref><ref
target="sch00005f33">33,</ref><ref target="sch00005f37">37,</ref><ref
target="sch00005f40"> 40</ref></ptrgrp></indexentry>
</index>
</add>
</archdesc>
</ead>
```

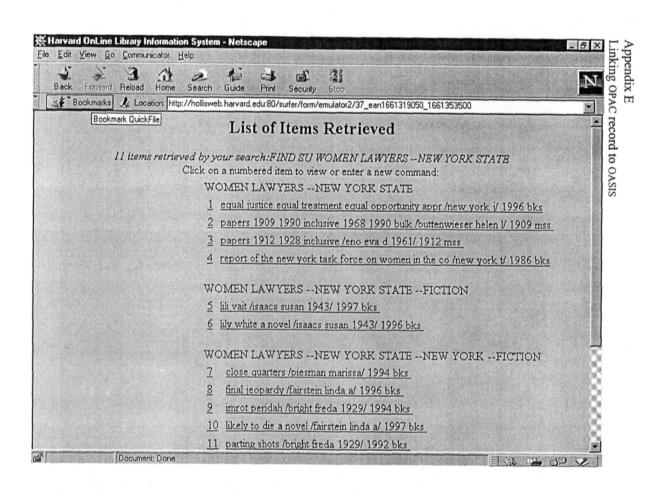


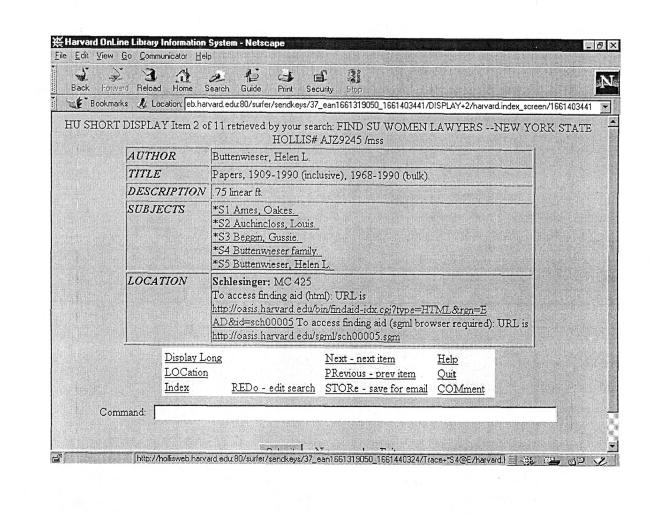


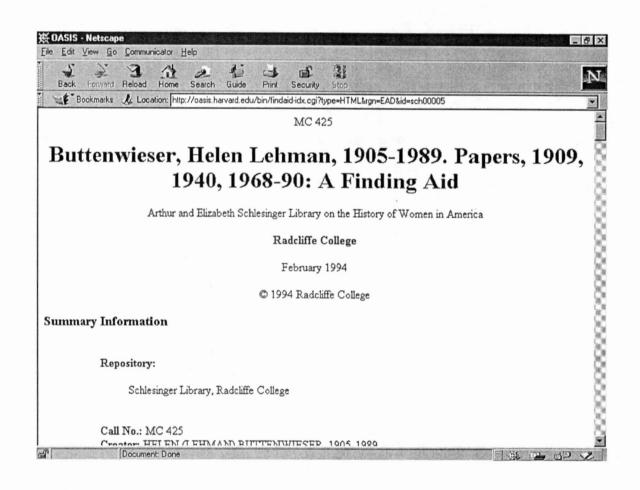












## Some useful URLs

http://www.lib.umb.edu/newengarch/InternetResources/technology.html

This page on the New England Archivists website has links to many EAD-related sites, including the site to download the DTD, the Library of Congress "official" EAD page, and many home pages of repositories involved in EAD projects.

http://hul.harvard.edu/hul/dfap/dfapinfo.html

This is the URL for the Harvard Digital Finding Aids Project; Here you can find a list of participating Harvard/Radcliffe repositories, and information about the history of the project and markup and display of finding aids.

http://oasis.harvard.edu/

This is the site where you can search all of the finding aids that have been marked up so far by Harvard/Radcliffe repositories.

http://lcweb2.loc.gov/faid/

Search some of the Library of Congress's EAD finding aids

http://iefferson.village.virginia.edu/ead/

The Society of American Archivists' EAD help pages

http://www.sil.org/sgml/

To learn more about SGML go to "The SGML Web Page"