

# EAD and the Small Repository

ELIZABETH H. DOW

**Abstract:** Impressed by the potential EAD offers for intellectual access to manuscript collections via publication of their finding aids on the World Wide Web, the University of Vermont Libraries supported Special Collections in accumulating the tools and knowledge needed to convert printed inventories to Web-publishable documents. Staffed by one faculty member and two undergraduate students, all on a part-time basis, and with support from the library's systems staff, the EAD Project published its first inventory in March 1997. The project raised many questions, such as what constitutes accurate and appropriate use of EAD markup, what are the differences between publishing on paper and publishing the same information electronically, how to handle vocabulary control, and what is the most appropriate technology for electronic access to a variety of finding aid formats.

*About the author:* Elizabeth H. Dow is Public Services Librarian in Special Collections at the University of Vermont. Some of the opinions and text in this article have appeared in the July 1996, December 1996, and August 1997 issues of SAA Manuscript Repositories Newsletter. The paper is based on a presentation made on 29 August 1997 at the annual meeting of the Society of American Archivists held in Chicago.

EXCEPT FOR A CERTAIN genetically driven pig-headedness,<sup>1</sup> I think I'm fairly typical of other small repository archivists and librarians.<sup>2</sup> I know my way around a computer, the world of databases, and the Internet in a more than casual but less than expert way. I have no programming background beyond the concepts of command stacking in Dialog, SPSS<sup>3</sup> query structures, and HTML coding—all pretty pedestrian stuff. But as soon as I saw finding aid markup using EAD, I thought it important that Special Collections at the University of Vermont have this technology as quickly as possible, and I knew it would be my responsibility to make it happen. Unquestionably I've done this the hard way, but I didn't see any alternative at the time. To tell this story, I will describe Special Collections at the University of Vermont and the process by which we developed the EAD Project. I will then address some of the issues the conversion project raised.

### Special Collections at the University of Vermont

The University of Vermont (UVM), founded in 1791, is the largest institution of higher education in Vermont, enrolling 7,500 undergraduates, 1,125 graduate students, 375 medical students, and 1,150 nondegree students, led by 850 full-time and 150 part-time or research faculty. The three libraries on campus hold about 1.25 million volumes.

The Wilbur Collection of Vermontiana, located in the Bailey/Howe Library, ranks among the largest collection of Vermontiana in the state. It holds approximately 80,000 books and pamphlets, 200,000 pictures, 7,500 maps, and uncounted thousands of items of ephemera. The collection also includes more than 7,500 linear feet of manuscript materials in some 700 collections, of which 40 percent have been fully inventoried, approximately 40 percent have been arranged, and about 20 percent await processing. Boxed collections range in size from single five-inch boxes to 400 one-foot containers.

As with most collections of this sort, patrons and staff alike find it difficult to access the contents of all but the published materials. The primary finding aids to the boxed manuscript collections consist of about ten linear feet of inventories created between 1970 and the present, ranging in size from one page to more than three hundred pages. More than half were created with a typewriter, and no one seems to know the whereabouts of the computer files for all but a few of the rest. The quality varies, as the inventories were prepared by students, interns, and professionals with a broad range of skills.

The Wilbur Collection also has fifteen card catalog drawers referencing smaller manuscript collections. Catalog information for about one hundred of our collections has been submitted to the *National Union Catalog of Manuscript Collections*, and about a dozen have been entered into the university's on-line public access catalog (OPAC). We also have a card catalog for maps. Fewer than half our broadsides are listed in the OPAC, and there is no other finding aid for them. We published guides to our manuscript collections in 1986, to our photographs in 1992, and to our American Civil War manuscripts in 1994.

---

<sup>1</sup>According to my baby book, the first complete sentence out of my mouth was, "I do it *myself*." That sounds like a pretty sophisticated sentence for a toddler, but who am I to doubt the historical record?

<sup>2</sup>Case studies usually reflect the work of teams, and so most reports on case studies use plural pronouns in the course of describing the project. The implementation of EAD at the University of Vermont, however, came about because I decided to do it, knowing that a shortage of staff in the department would make it essentially a one-person project. In writing about it, therefore, I thought plural pronouns sounded royal, if not actually stilted, so I wrote in the first person, generally using the singular pronoun.

<sup>3</sup>*Statistical Package for the Social Sciences*, SPSS, Chicago.

We have nearly four thousand vertical files of clippings and a wide range of short, incomplete runs of defunct serials and other periodical publications such as annual reports, high school newspapers, manufacturers' catalogs, and the like. While none of these has national interest, all could have high appeal to our primary patrons at UVM if they knew about them, but most have no intellectual access at all.

### Why EAD?

A basic pessimism drives my enthusiasm for EAD: I see a dismal future for small repositories that don't establish a presence on the Internet. This pessimism gained support on August 14, 1997, when twenty or more messages on the ARCHIVES Listserv<sup>4</sup> (a large proportion of the traffic that day) consisted of members from all over the country recounting discussion and/or confrontation with researchers who want access to materials discovered via the Web. Undoubtedly, many researchers express diminishing patience with both paper finding aids and the detective work they once accepted as part of the research process. Because of the potential of the Web, they want intellectual access tools, and the actual documents, on-line and searchable. Long-term survival strategies for small repositories must include a Web presence which provides at least a catalog of holdings, preferably inventories for larger collections and probably portions of the collections themselves, because it seems clear that researchers will go to those sites that do have these resources.

My pessimism combines with my desire to achieve a standard of professional service which I cannot provide today but which I see the use of EAD making possible in the future. For decades the archivist's human memory has served as a major tool both for identifying and evaluating collections the researcher should consider. As a result, patrons receive uneven service, depending on who is serving them and how nimbly that archivist's memory functions during the reference interview. From one month to the next, I may or may not remember that some of our Vermont governors' papers have materials on solid waste and landfill issues, but once the computer "knows," it will never forget. While I don't think computers can ever adequately evaluate the "goodness of fit" between a research question and an archival resource, without a doubt they do a much better job of initial identification. Turning our finding aids into searchable databases using both MARC records and SGML documents relieves archivists of the brute memory work and allows us to apply our knowledge of the collections to the scholar's tasks by helping determine what comes closest to his or her actual need.

### EAD at UVM

I had worked in Special Collections at UVM for less than two months when I saw Daniel Pitti's demonstration of the Berkeley Finding Aid Project at the annual meeting of the Society of American Archivists in Washington, D.C. in 1995. When the session ended, I was so excited I literally left the hotel and walked around the block a few times to calm myself and to comprehend the full potential of what I had seen. I had already planned a MARC database to handle item-level and collection-level records, and Pitti's project demonstrated the solution for the manuscript inventories. I knew that I had to make EAD

---

<sup>4</sup>Dean DeBolt, et al. (14–15 August 1997), "Reference Services—Common Complaint [Discussion]," ARCHIVES listserv, available at <<http://listserv.muohio.edu/archives/archives.html>>.

implementation possible for UVM, but that I couldn't expect much assistance from other members of our small department; nobody had the time.

The staff of Special Collections consists of two library faculty, 2.9 FTE staff, and undergraduate work-study student help that totals sixty hours per week during the academic year. The staff maintains a public service desk in two locations for eighty-four hours a week during the academic year; this cuts back to fifty-six hours through the summer. While the library's technical services department handles most processing and cataloging of our print material, all work on archival materials must be done in Special Collections.

A few months after the Pitti demonstration, I learned from a colleague at the university's computing center that Inso Corporation would grant a full suite of electronic publishing software, called DynaText, DynaTag, and DynaWeb, to educational institutions that present an "interesting, innovative, or unique" project with which to use it.<sup>5</sup> My colleague from the computing center and I each developed project ideas, I wrote the proposal, and in mid-summer 1996 we got word that we had received the grant. In September, the two of us were joined by the library's head of systems for a week at Inso headquarters learning to use and administer the software. As soon as I got back from the training, I spent six weeks developing, test-running, and writing an Ameritech/LC American Memory grant proposal<sup>6</sup> full of ideas about how we would use this wonderful new software to post the George Perkins Marsh inventory and selected papers to the Web. In November 1996, I finally got down to work on EAD.

The Inso software will read and display SGML documents but does not have authoring capabilities. As it happened, just as I was learning the Inso software, WordPerfect went public with a beta version of its SGML authoring tool, which we purchased; we later bought a full version of WordPerfect 7 with its SGML authoring module. I started on two documents: 1) a contents list for fifty-six cartons of newly received unprocessed papers from Madeleine Kunin, a former Vermont governor on her way to serve as Ambassador to Switzerland, and 2) the inventory of the eighteen-carton Marsh collection. The Marsh inventory had been created on a typewriter, and we hoped that conventional wisdom would prove wrong and that we could convert it to ASCII text with optical character recognition (OCR) software. While the front matter scanned well enough, the container lists, set up in a tabular format, did not. One software package resolved the problem caused by the container list tables by stacking the columns one after the other. The second preserved the columns by including a lot of proprietary codes which the SGML authoring software couldn't work with and which conversion to ASCII completely destroyed, turning the document into a jumble of alphanumeric characters. In the end, I had a student rekey the inventory in WordPerfect while I worked on encoding the Kunin inventory, which a student had already created in WordPerfect.

I had only WordPerfect 7's authoring software and a hard copy of the EAD Document Type Definition (DTD) to guide me. To learn to interpret EAD, I went to the stacks and pulled the library's three or four books on SGML. I quickly discovered that: 1) I lacked the background to make more than the foggiest conceptual sense of them, and

---

<sup>5</sup>For more information about Inso and its products, see <<http://www.inso.com>>.

<sup>6</sup>Library of Congress/Ameritech, *National Digital Library Competition* (Library of Congress, 1997), available at <<http://lcweb2.loc.gov/ammem/award/>>. We did not receive funding.

2) they were written for people about to write a DTD, not for those of us trying to use one. I longed for Laura LeMay to write a manual on EAD as she had on HTML.<sup>7</sup>

With that thought, I decided to compare the HTML DTD to LeMay's instructions on using it. I expected the comparison would demonstrate how to translate the code into English, and it did. While the subtleties of the EAD code remained obscure, the general structure easily came clear, and I started generating EAD markup in WordPerfect 7, trusting the software's validator to rap my knuckles when I erred.

I stepped my way through the front matter successfully, but when I came to the container lists my heart sank. We needed to apply the following set of tags to 3,500 folder labels: `<c03><did><unitloc>Folder 53</unitloc><unittitle>Campaign Contributors, Caledonia County,</unittitle><unitdate>1988</unitdate></did></c03>`.

I inquired about macros, but nobody I talked to had any experience with WordPerfect macros in a Windows environment, and the documentation wasn't helpful. We slogged on by hand. I had six hours of student help per week, and through the long dark Vermont winter the students and I took turns at markup duty. Talk about mind deadening! But, as spring 1997 rolled around, we finished the markup and I moved on to style sheet creation.

A readable SGML document requires three modules: 1) a marked-up document, which defines how each element relates to other elements; 2) a style sheet, which tells the browser how to display the marked-up document and; 3) a browser, which provides the actual display. Web browsers in 1997 can read only HTML markup, not other SGML DTDs. DynaWeb does an "on the fly" conversion from EAD to HTML, so after I created the EAD style sheet, I mapped the EAD tags to HTML tags. Such mapping can be a fairly simple process of equating a tag in one tag set to a tag in the other (e.g. `unitloc = h4`; `unittitle = h4`; `unitdate = h4`), but it can also be a lot more complicated. When I finished tag mapping, our systems department set up the DynaWeb server, we loaded the files . . . and it worked. I can't imagine that any corner of the Earth failed to hear the shouting!

After a rush of exuberance upon seeing our first finding aid on the Web, I found myself overwhelmed by the enormity of the task of converting all the old to the new. Although our collection inventories generally resemble one another, when you get down to the details, they contain enormous variations and are occasionally highly idiosyncratic. Though the EAD developers consciously included enough flexibility to accommodate many such variations, each of us who use EAD must decide how to deal with variation on a case-by-case basis.

By late spring I keenly felt a need for contact with others working on similar projects, so I registered for Pitti's class in EAD at the University of Virginia's Rare Books School. His pre-course reading led me to the Berkeley SunSITE, which I had vaguely known about but had ignored. There I found the American Heritage Virtual Archive Project, and the EAD "Retrospective Conversion Guidelines."<sup>8</sup> The American Heritage Project joins Stanford University, the University of California at Berkeley, Duke University, and the University of Virginia to create "a prototype 'virtual archive,' integrating into a single source,

<sup>7</sup>Laura LeMay, *Teach Yourself Web Publishing with HTML in a Week* (Indianapolis: Sams Publishing, 1995), with an update, *Teach Yourself More Web Publishing with HTML in a Week* (Indianapolis: Sams.net Publishing, 1995), and a second edition (*Teach Yourself Web Publishing with HTML 3.0* (Indianapolis: Sams.net) in 1996.

<sup>8</sup>University of California at Berkeley, EAD "Retrospective Conversion Guidelines" (1997), available at <http://sunsite.berkeley.edu/amher/upguide.html/>.

hundreds of archival finding aids . . . from collections . . . held by four major academic research libraries.”<sup>9</sup>

When I compared what I had done to the American Heritage Project guidelines, I discovered that I had created valid, but thin, EAD markup. WordPerfect wouldn’t let me do anything illegal, but it didn’t prompt me for optional but important elements that I had overlooked or ignored. I had created “Stranger in Paradise” when *Scheherazade* was called for.

Determined never again to subject myself, or the students, to marking up container lists by hand, just before I went to Virginia I contacted a programmer I had worked with in another department of the university; he had retired because of medical disability, needed some diversion, and found this new area of work a challenge. When I returned from Virginia, he had completed the alpha version of the first of what will become a set of utilities for container list markup.

Thanks to those utilities, UVM now has six inventories on-line.<sup>10</sup> At this point I’ve stopped to reassess everything I’ve done, to revisit all decisions and assumptions I’ve made, to work with Alvin Pollock at UC Berkeley to create a template for our “front matter,” and to document the final decisions which will establish the UVM standards for further work. That’s where the project stands at this writing.

## Lessons Learned

How do I think we did? To answer that question, I’ll make a variety of observations from an evaluation stance.

### *What Effect Did the EAD Project Have on Department Workflow?*

At UVM, the major workflow change came about when the university hired a librarian (me) to succeed a historian/bibliographer. As someone with a history of building databases in archival repositories in Vermont, I was asked specifically to bring Special Collections into the twenty-first century. The EAD project consumes about 30 percent of my time, as well as six hours weekly of student time, enhanced by the efficiency supplied by a volunteer programmer.

### *What Stand out as Successes?*

1. That I can publish finding aids on the Internet at all feels like success.
2. That I involved our systems people; they know what the project involves, they support it enthusiastically, and they respond promptly to my needs.

### *What Would I Do Differently?*

1. I would begin using the Inso software immediately after completing the training in its use. As I designed the Marsh project for the LC/Ameritech reviewers, I learned a lot that I’ll eventually need to know when we reach the point of attaching full documents

<sup>9</sup><http://sunsite.berkeley.edu/amher/proj.html>.

<sup>10</sup>Special Collections, University of Vermont, *UVM Special Collections* (University of Vermont, 1997), available at <http://sageunix.uvm.edu/~sc/>.



and illustrations to the infrastructure provided by the inventory. In the course of sorting all that out, however, I forgot a lot of detail about the Inso software. Consequently, I found I had to relearn much that we had covered at Inso.

2. I would work less independently; I succumbed too much to my “do it *myself*” mentality. If I weren’t so independent, for example, I might have pushed the macro issue harder. I now know that I should have gone from the computer-types I asked about macros to some top-notch secretaries who had worked intensively with WordPerfect. If I didn’t now have a programmer, I’d be furiously writing keyboard macros in WordPerfect. Further, if I hadn’t been so independent, I might have found the Berkeley SunSITE earlier and monitored it more closely.

### ***What Problems Continue?***

1. Working alone obviously has its drawbacks. I monitor the EAD and the Inso Listservs, but the phenomenon of the “teachable moment” comes very much into play. I frequently failed to comprehend the importance of discussions and information posted there because I lacked a context for it. Invariably some months later I would reach the “need to know” point and then would have to go back and dig it out of the listserv archives.

2. I have decided to adopt the American Heritage Project protocols. That consortium has experts to debate the details and establish a range of acceptable practice; I see little reason to second-guess its decisions. Even so, the American Heritage guidelines provide a range of practices, not a set of rules. I face many decisions that require knowledge of the software, the DTD, and the document I’m working on. I make those decisions alone, because I have no choice, but I wish I had someone locally to discuss them with.

3. SGML depends heavily on indirect logic and references to external files. SGML browsers expect to find those external files precisely placed on the system. Loading and managing files takes more time than I ever imagined.

### ***What Changes or Refinements has EAD brought to Inventory Creation?***

1. As Pitti points out in his introduction to the American Heritage Project’s EAD “Retrospective Conversion Guidelines,” using EAD does not in and of itself ensure that machine-readable finding aids will function well in a union database. Inventories and other finding aids in union databases will need to share a degree of uniformity to make them easily intelligible for users moving from one institution’s documents to those of another. More fundamentally, they will need to share a degree of uniformity so that computers can manage them; this uniformity applies to both intellectual content and markup protocols. Therefore, as I look forward to entering our inventories into a union database in the future, I must stay current with new developments.

2. I first directly translated the look and feel of the paper inventory into two electronic documents we marked up, which means that I included the box number only once—at the top of its container list—followed by folder numbers only. This approach works poorly on-line, because SGML display technology in 1997 does not retain the box number at the top of the screen the way typists do when they carry over information from page to page. Unless the display includes the box number on the screen with every folder number, the patron quickly loses track of which box an item resides in.

3. Most of our finding aids blur the distinction between intellectual and physical arrangement. EAD was designed to give primacy to intellectual organization, and so as we convert our data from paper to electrons we have an opportunity to reorder the inven-

tory to take advantage of this. I think that is the right thing to do from a theoretical point of view, but I question whether taking the time to rethink and reorder every inventory for conversion will yield enough benefits to justify the cost of the process.

4. I would like all my data to work using one style sheet. While conceptually easy to understand, the minute detail necessary to develop a good style sheet takes a lot of time and a considerable amount of expertise. I would prefer to spend that time just once. Given that inventories do have unavoidable variations, however, I may have to develop up to three: one for collections processed in classic series/subseries order; one for those processed in chronological and/or physical location order, and one for totally unprocessed collections.

### *What New Considerations Has EAD Brought to the Inventory Creation Process?*

1. Collection inventories contain lists of names associated with political or ideological action/causes, and the folder labels make the location of those lists quite clear, as in my earlier example for "Campaign Contributors, Caledonia County, 1988." In theory, despite publication on the Web, nothing about this finding aid has changed. The inventories direct researchers to the contributor lists and are public information, as they always were. But in reality, before the Web, the public could not so readily access that information. It was confined to a notebook in our reading room which was open to the public only about a third of the daytime hours each week, and so the information simply didn't fall into the hands of anyone who didn't consciously, and at some inconvenience, seek it out. On the Web in a searchable inventory available every minute of the year, the information is now easily and globally available. After Web publication, the probability that someone will find and use membership or contributor lists, benignly or otherwise, increases substantially. I have begun to wonder what responsibility, if any, we have for people's privacy when we make such information globally available. This can be particularly troubling in collections acquired before global availability even seemed possible, and our deeds of gift say nothing about this sort of information dissemination.

2. With Web publication, a paper inventory becomes a searchable database, necessitating good database maintenance practices. EAD includes protocols for controlled access points at all levels, but is it worth the time it will take to include them? On the other hand, can we, in good conscience, *not* include them? How much should we gear our markup toward the search engine? Should we begin standardizing our folder label language? What about authority control for proper nouns? If we're looking to produce one or more union catalogs of finding aids, we cannot simply ignore vocabulary issues. Nor can we each arrive at our own conclusions independently.

3. At about the same time the Berkeley Finding Aid Project started to develop EAD, a group of digital library researchers, Internet networking specialists, content specialists, and librarians—all concerned about absence of librarianship on the Web—started to develop what has become known as the Dublin Core, "an ongoing effort to form an international consensus on the semantics of a simple description record for networked resources. It is expected that a simple and widely-understood set of elements will promote interoperability amongst heterogeneous metadata systems and improve resource discovery on the Internet."<sup>12</sup> Like anyone who has used any of the Web search engines, I applaud

<sup>12</sup>"The 4<sup>th</sup> Dublin Core Metadata Workshop Report," *D-Lib Magazine* (June 1997), available at <<http://www.dlib.org/dlib/june97/metadata/06weibel.html>>.



the effort to develop such a protocol, but it's not yet clear to me how to integrate the EAD and Dublin Core elements.

4. Repositories process collections for three reasons: First, going through the materials and storing them in a tidy manner in archival enclosures helps preserve them, identifies items of particular value, and brings to light items that need conservation. Second, a well-ordered collection brings like materials together, making them convenient to present to patrons. Third, an inventory of the collection provides intellectual accessibility. Given that UVM fully processes manuscript collections only when we have special grant money or other beneficial circumstances, there is no real hope we'll ever truly process many of our collections. Therefore, I see creating and publishing the existing container lists in "raw" order as a way of providing at least some intellectual access where there would be absolutely none otherwise. I can hear the real archivists among you bemoaning the devaluation of the neighborhood, and I sympathize. Granted, all of us would be better served if collections were fully processed. However, in a world in which that may not happen, a searchable finding aid makes the collection *usable* without full processing. Further, tracking which unprocessed collections get repeated use can give us data for choosing those collections on which we will spend our scarce processing resources.

5. As Richard Szary has suggested, the introduction of EAD has served as a catalyst for discussion of fundamental issues regarding what a finding aid is.<sup>13</sup> When does a MARC record adequately serve as a finding aid, and when do we need something more? When do we need individual MARC records, and when can we represent our holdings more efficiently but just as effectively through SGML-encoded lists? I've created a half dozen or so MARC databases for materials such as our vertical files, uncataloged town reports, photo files, etc. What is the relationship between our on-line public access catalog, the website on which we publish SGML finding aids and these databases? The UVM library system will soon move its public catalog from a mainframe computer to a client-server system. When that's in place, of course we will link the MARC records and the published inventories, but it is unclear how we can blend all aspects of our intellectual access tools into one integrated user-friendly system for the patron.

## Conclusion

EAD implementation at UVM began when it did because one faculty member took complete responsibility for it and received total support from the administration at both the departmental and library levels. Within the next few years we will convert all our inventories, as well as container lists for many of our unprocessed or semi-processed collections. We also have lists of holdings, such as nineteenth-century newspapers, diaries, ledgers, and other materials, for which the list constitutes the finding aid, and I'm not sure how to handle these. Should we load them into a database, mark them up in a perverted EAD, or take on TEI,<sup>14</sup> a DTD more suited to general documents such as lists? After we've published the inventories, to what degree do we want to link scanned images of the documents? We don't know, but when we reach the point of addressing those questions seriously, we'll find answers—probably by watching what others have done.

<sup>13</sup>Richard Szary, "Improving Access to Finding Aids: The Encoded Archival Description Project," *NEA Newsletter* 23 (October 1996): 4-8.

<sup>14</sup>*Text Encoding Initiative*, a project to develop guidelines and an SGML DTD for encoding electronic texts for scholarly use, available at <<http://www.uic.edu/orgs/tei>>.

If the University of Vermont is a small repository, the state of Vermont is dotted with micro-repositories. Virtually every one of the state's approximately three hundred municipalities has a historical society or library or municipal office with a collection of locally important historical documents. Some have materials of much broader importance, but the holdings of most remain totally inaccessible. Having begun my archival career in one of them, I care about these repositories a great deal. Although UVM Special Collections would face marginalization without a Web presence, it would not face extinction because of its position within the university. Vermont's micro-repositories, however, have no such institutional insulation. Many struggle to survive now and will surely follow the town band into oblivion as the users and financial supporters on which they depend completely come to rely on repositories like UVM that can serve their needs by remote access through the Web.

EAD is quickly becoming fundamental to the Web presence of a historical repository. I'm very grateful for the work large repositories have put into EAD, and their willingness to share their tools and insights with those of us in small repositories. Despite my illusions of having done it "myself," I know that nothing this large, or even my own little success, can happen with only one person or one institution's efforts, and that we as a profession still have huge questions yet to address. The workable answers to the questions raised by the authors in this issue of the *American Archivist* will require all of us to relinquish some of our more idiosyncratic practices for uniform standards that will serve us all. Comprehending and adhering to those standards will come hardest to the small- and micro-repositories around us, because we are largely understaffed, underfunded, and poorly trained. As I have benefited from the direct and indirect help of UC Berkeley, the University of Virginia, Harvard, and Yale, I must now reach out to Vermont's micro-repositories.