

How Do I Find Thee, Let Me Count the Ways:

A Manuscript Collections User Survey at Six Research Institutions

Susan Hamburger

August 9, 2000

INTRODUCTION

The rapid expansion of the World Wide Web allows research libraries and archives to encode the finding aids to their archival and manuscript collections in SGML and mount them on the Web for online access. Some archivists have been reticent to accept or embrace *Library of Congress Subject Headings* and the *Library of Congress Name Authority File* for cataloging their collections; manuscript repositories, closer in mindset to libraries, have had more success with retrospective conversion projects and standardized access points. Finding aids and guides to collections have lacked a standard format and content until as recent as the last ten years as MARC:amc fields influenced the data needed from a finding aid to create the catalog record. Researchers have had to cope with forty or more years of inconsistent description of collections with nonstandard vocabulary. The move to full text of finding aids online presents the researcher with varied

approaches to access: keyword searching throughout the finding aid, non-boolean search engines, search strings, and the opportunity to utilize standardized vocabulary. How are repositories that are using SGML Encoded Archival Description (EAD) to mount their finding aids on the Web dealing with standardized subject headings and name authority?

BRIEF DESCRIPTION OF RESEARCH PROJECT

I conducted two simultaneous surveys, one with the archivists/librarians at selected repositories who are actively mounting a substantial portion of their finding aids on the World Wide Web using EAD- and/or HTML-encoding, and the other with their on-site user population. I wanted to determine what search strategies library/archives patrons use for finding manuscript collections online and what the current practice is among libraries/archives for standardizing access through controlled vocabulary or if they are relying on keyword access.

A literature search of library and archival journals since 1995 reveals a dearth of information about the use of controlled vocabulary in manuscript finding aids. Encoded Archival Description moved out of the beta-testing phase into Version 1.0 in 1999 and, aside from the working papers online

distributed among the EAD-user population addressing the mechanics and philosophy of field tags that culminated in the *EAD Tag Library* (1998), *EAD Application Guidelines* (1999), and two special issues of *American Archivist* devoted to introducing EAD, there are no other articles focusing on the contents of finding aids online.

PLAN OF WORK

I checked the websites for repositories that encode in EAD and determined that New York Public Library, Library of Congress, University of Virginia, and Yale, Cornell, Harvard, and Duke universities would be the best representatives on the east coast. I arranged to interview catalogers in all of these repositories (except Cornell who never responded to email and phone requests) about their manuscripts cataloging practices for online catalog records in MARC (using LCSH, NAF, AAT), finding aids online (in EAD, HTML, XML), and/or databases (such as dBase, Access, or others). How do they provide searchable terms across finding aids (keyword searching, boolean, full-text, controlled vocabulary)? If they embed subject terms, where do they place them in the documents? What limitations do they encounter with software (EAD, HTML), search

engines (Open Text, Dynatext), or viewers (Dynaweb, Panorama, Arbortext)? What feedback, if any, have they received about their finding aids online? How are users finding out about their manuscript collections [OCLC, RLIN, ArchivesUSA (*National Inventory of Documentary Sources in the United States* [NIDS], *National Union Catalog of Manuscript Collections* [NUCMC], and *Directory of Archives and Manuscript Repositories in the United States* [DAMRUS]), OPAC, printed guides]?

Concurrently, I designed a survey to ascertain how users are accessing manuscript collections (finding aids online, catalog record online, paper finding aids, OCLC/RLIN/OPAC, NUCMC paper or online, word of mouth, DAMRUS, phone, e-mail, in-person drop-in). I wanted to know if researchers are using subject headings.

I visited all the repositories except the New York Public Library in April to interview their catalogers and left each one 50 user surveys to distribute and return to me in a postpaid priority mail envelope. I interviewed NYPL representatives on the phone and at the Mid-Atlantic Regional Archives Conference in May. Yale at first preferred to do a generic online survey [**which you see here**] then changed their minds and

declined to participate in that portion of the project. I included Penn State patrons in the user survey after receiving authorization from the Office for Regulatory Compliance to use human subjects in research.

INTERVIEW RESULTS

New York Public Library Manuscripts and Archives Division uses the same LCSH terms in their finding aids that they use in their RLIN record but only at the highest level, not at the series level. They don't display subject terms to the patrons, nor do they make a conscious effort to write their finding aids with keywords that patrons might use instead of Library of Congress subject headings. Because six staff archivists plus grant-funded people create finding aids, and two others do the actual encoding, there is no consistency in terminology use in the MARC record or the EAD finding aid. However, the Performing Arts Division uses controlled vocabulary in the front matter and at the series level in the container list. In the scope and content note they use *Art and Architecture Thesaurus* form/genre terms. Performing Arts has four MARC catalogers, two of whom do markup for HTML or EAD.

Harvard University's Houghton Library provides good name access for their literary collections using LCNAF. Their anonymous or historical manuscripts have some subject headings such as slavery or American Revolution. Rather than put the subject headings in context they provide a name index at the end of the finding aid. Their finding aids don't replicate the MARC record. They perpetuate in EAD what works on paper. The markup of names reflects that the Houghton is committed to personal name access. At the Schlesinger Library they replicate all subject headings at the top of the finding aid and mark up the subjects within the finding aid.

At Yale University, all collections—whether one item or 459 boxes—get MARC records. The finding aids' scope and content notes now include key terms but not necessarily LCSH. They've found that most of the research is name-driven so they deemphasize subject access.

Special Collections at the University of Virginia is using its old guides and finding aids for cataloging. These don't have controlled vocabulary and there is a problem with consistency. Each collection description needs to be completely rewritten and have LCSH applied. They also use a controlled vocabulary list of southern terms that was developed several years ago by a consortium spearheaded in South Carolina.

Duke University has the most consciously thought out process for providing subject access within their finding aids. They write a scope and content note for each series and deliberately include common terminology that patrons may look for that doesn't replicate LCSH that are also inserted at the series level.

The Library of Congress provides a two level access approach. They include LCSH, occupation codes, and subject headings at the collection level. They abstract all terms from their catalog record into their finding aid. Literary collections usually focus on personal names and exclude topics about which the author is writing.

At Penn State I've adopted a combination of the Library of Congress and Duke University models to consciously include common terms but also provide LCSH and to write scope and content notes at the series level to give the researcher the context for the container list of folder headings. [**show NASSH example**]

USER SURVEY

Now that we know how these major repositories are presenting their catalog records and finding aids, do the researchers actually take advantage

of the extra access points? Of 300 surveys, the 131 returned from the six repositories indicate a gap in the sophistication level of researchers across institutions. While 92 percent use the computer daily and 75 percent claim they can navigate through the online environment easily, the majority of them locate manuscripts from footnotes in articles or books. Even though they indicate they use the library's online catalog and the library's website, they are still guessing which library to contact and searching one catalog at a time. They are not availing themselves of new methods of finding collections of materials in repositories whose holdings either don't appear in the older printed sources or who have recently acquired collections. They are woefully ignorant of OCLC and RLIN, and ArchivesUSA is off their radar screen. Twenty-six checked that they search the paper copy of NUCMC that has been superceded by ArchivesUSA online. Faculty responses indicate the traditional research methodology while the computer-savvy undergraduates try to find manuscripts via Internet search engines. As Helen Tibbo pointed out in her presentation at the Society of American Archivists conference last year, the Internet search engines just don't do the job; she took known finding aids and tried to find them using several commercial search engines and got terrible results.

When ranking the order of usefulness of specific tools, 43 chose the online catalog, 23 chose paper finding aids, 11 chose the Special Collections web page, and the next most useful was the manuscript card catalog. Eighty-nine percent of the respondents found what they were looking for. Seventy-eight percent did a keyword search while thirty-one percent searched for a personal name, and twenty-three percent used a subject search. Unfortunately, many of the respondents did not follow directions and write in the terms they used so it is difficult to analyze the accuracy of their searches. In the cases when they did provide the search terms, it is a wonder that they found what they wanted. For example, one researcher at the Library of Congress indicated a topic of National Council of Jewish Women concerning German-Jewish refugees, 1933-1950s. The search strategy moved from checking footnotes to NUCMC, email, and then an onsite visit. But this person searched the National Council of Jewish Women as a title. Duplicating this search, I got zero hits. However, searching for it in "name browse" yielded twenty-five records, one of which is for Records of the National Council of Jewish Women, 1893-1989 (bulk 1940-1981). A Duke researcher looking for information on the naval history of the Civil War searches individual library online catalogs. His subject, Red River campaign, is actually Red River

Expedition, 1864 in LCSH. There seems to be a preponderant dependence on personal names even if the research is topical. For example, a Library of Congress researcher was looking for the U.S. Navy's economic impact on Wilmington, NC during the Civil War and chose to look up three personal names: S.P. Lee, David Dixon, and Louis M. Goldsborough. Only Goldsborough produced a catalog record for his papers that didn't indicate anything involving Wilmington; nor is the finding aid online.

The comments section yielded helpful suggestions. When asked what was missing, one researcher wrote "description of all archival holdings," another wanted "an easier-to-use guide to access/locate manuscripts," and a third wanted "manuscripts available online." Another noted "online find aids/guides are the most helpful to narrowing search." One thought "better connections need to be made between manuscript collection/rare books when the items (pamphlets) are duplicates." Still another pleaded that paper copies of the finding aids continue to be available; "sometimes these are faster and easier for people to use—and a godsend when computers are down."

CONCLUSION

Although I got a 43.6 percent return rate on the surveys, many of the respondents failed to answer the most important questions about their search strategy, how they searched, and what terms they used. From the limited number of responses, it appears that until librarians and archivists better educate researchers on how to search for manuscript materials they will continue to rely on personal names as the entry point for topical queries. While this is perfectly acceptable for literary manuscript collections, historical collections require a more sophisticated approach and search strategy. The majority of researchers answering this survey continues to utilize traditional methods of uncovering primary sources and don't take full advantage of the online resources. At this stage in providing online access to finding aids, our efforts are underutilized. As said in *Field of Dreams*, "If you build it they will come." In our case, we need to provide a roadmap and explicit directions for the stubborn ones who won't ask directions.