Finding Archival and Manuscript Collections in the Digital Age (Draft)

Fred Burchsted and Anna Esty

Over the last few years, the discovery of archival collections, original, microfilmed and digitized, has become increasingly complex. This paper suggests tools and strategies for finding archives and manuscripts (AM) focused on the United States. Topical (subject) searches are emphasized here (for example, slave insurrections), as known person searches are less problematic.

In what follows, “catalog record” refers to the kind of record found in library online catalogs, similar to those for books, although sometimes a bit longer. “Finding aid” (sometimes called an inventory) refers to a list of the folder labels for the collection, accompanied by a brief collection overview (scope and contents note) and a biographical (or institutional) note on the creator of the collection. Finding aids may be as long as needed given the size of the collection. They vary considerably according to the practices of individual repositories.

There are four main databases containing AM collection records: Archive Finder, WorldCat, ArchiveGrid, and Social Networks and Archival Context Project (SNAC). We will discuss using each of these tools to locate AM collections, as well as searching the free web.

Archive Finder

Contents
* Archive Finder contains two different sets of records:

   Starting in 1959 archival repositories sent catalog records of their collections to the National Union Catalog of Manuscript Collections at the Library of Congress (http://www.loc.gov/coll/nucmc/). These were published in a series of volumes until 1993. From 1986 on, NUCMC also sent records to the WorldCat database (see below). Archive Finder searches the NUCMC records through 2010.

2. National Inventory of Documentary Sources (NIDS). Since the 1980s many repositories in the US and the UK sent finding aids to the National Inventory of Documentary Sources project of the commercial publisher Chadwyk-Healey. Over 65,000 collections are now represented. The NIDS project created a catalog record with subject terms for each collection. Repositories submitting finding aids for
collections include the National Archives; the Presidential Libraries (which house many sets of personal papers of people serving in Presidential administrations); the Smithsonian Institution Archives; the Manuscript Division of the Library of Congress; state archives, libraries, and historical societies; plus academic libraries and other repositories. These finding aids were put on microfiche. Archive Finder searches the NIDS catalog records which bear the reference number for the corresponding microfiche. The NIDS microfiche are available in many large libraries. Bear in mind that not all repositories have chosen to submit finding aids to NIDS, and those that have may have sent inventories of only a portion of their holdings.

Unlike WorldCat, there are no records for microfilmed or digitized collections.

Searching
Within the Archive Finder interface, choose "Collection Search" from the list on the left-hand side of the screen. On the Collection Search screen, use the Collection Name search if you only want collections consisting solely of your subject's papers. If you want to find papers of your subject housed in other people's papers (letters sent by your subject, etc.) use "All Keywords". Searching for the papers of Robert Frost, it is safest to try both "Robert Frost" and "Frost Robert." "Robert Frost" will search the description field. "Frost Robert" will search the index terms in NUCMC and NIDS. Occurrence of your search terms will be indicated by a red dot. The search Frost and Robert will pick up occurrences of the two words even if they are part of different names, but may be useful if you are unsure of the correct form of the name. Collection searches can be limited by repository, city, or state.

NUCMC and NIDS each have their own subject heading systems. Note the subject headings assigned to your topic and, if needed, repeat your search using the proper subject headings for each database. NUCMC subject headings may be browsed in the Index to subjects and corporate names in the National union catalog of manuscript collections, 1959-1984 (3 vols. Alexandria, VA: Chadwyck-Healey, 1994).

A Repository Search retrieves brief descriptions of repositories, not of collections. Repositories can be searched by city or state. Choose "Repository Holdings Keyword" to search repository holdings statements. There is no standardized subject vocabulary in the holdings statements, so try several terms.

WorldCat

Contents

* WorldCat allows for searching of the NUCMC records (1986- ) and also contains catalog records for non-NUCMC AM material contributed by member libraries.
Searching
There are three interfaces for searching WorldCat AM records: FirstSearch (the full version of WorldCat, which is licensed by many libraries), the public version of WorldCat, and the WorldCat AM records available via the NUCMC site. Whenever possible, use the full version of WorldCat which offers features not available in the other versions (described below). AM collections may appear in three different forms within WorldCat: original collections, microfilmed collections, and digitized collections. The different steps involved in searching for microfilmed and digitized collections are covered after the discussion of original collections.

Full version
To search for AM in WorldCat, go to Advanced Search. As in any catalog search find the proper Subject terms (for example for slave revolts, use Slave insurrections) as well as any keywords you wish to search. Put these in a search box. Then you have two options for limiting to AM material (you must do both searches in order to cover all relevant materials, though the results will overlap to some extent). The reason for limiting is that WorldCat contains millions of records for books and other media which normally overwhelm the AM records.

1. Check Archival Materials at Limit type to: and search.
2. Remove the check next to Archival Materials and then change Subtype limits from Any format to Manuscript. Adjust the menu of another search box to NOT Material Type and type in the box: Thesis/dissertation. Theses and dissertations are tagged as both manuscripts and books and often clutter the results of topical AM searches.

In the first search, setting Limit type to: Archival Materials yields “mixed materials”, that is, typical sets of personal papers and organizational records, together with many scrapbooks. Excluded are individual manuscripts such as manuscripts of literary works, diaries, and other unitary items, usually tagged as manuscripts (which will be included in the second search). Much manuscript material is tagged as both Manuscript and Archival Material, but a good deal is only tagged as Archival Material or as Manuscript. Searching Archival Materials and Manuscript simultaneously will exclude much material, thus the need to perform two searches.

The particular form of an item is specified by the Genre/form designation. You can search a very particular type of material by searching, e.g., Maritime journals as Genre/form (a subcategory of Subject) and limiting (one at a time) to Subtype limits: Manuscript and also (separately) Limit type to: Archival Materials. The Yale University
Library makes available a list of form and genre terms useful in archives and manuscript work [http://www.library.yale.edu/beinecke/manuscript/mscat/appE.htm](http://www.library.yale.edu/beinecke/manuscript/mscat/appE.htm).

A keyword search in WorldCat on your person's name will find both collections of your subject's own papers (which may well exist in several repositories) and letters written by your subject and occurring in their recipients' papers. You can also search for categories of persons by using subject terms: physics, physicists, women authors. Finding aids are often linked from the WorldCat record.

**Public version**

The public (free) version of WorldCat ([http://www.worldcat.org/advancedsearch](http://www.worldcat.org/advancedsearch)) offers the limitation to Archival Materials but not to Manuscript. Thus many collections just tagged as Manuscript will be missed.

**NUCMC version**

The WorldCat AM records can also be searched for free via the NUCMC site ([http://www.loc.gov/coll/nucmc/oclcsearch.html](http://www.loc.gov/coll/nucmc/oclcsearch.html)), but there is no way to exclude dissertations and much other non-AM material (grey literature, etc.).

*Microfilm in WorldCat*

Many archival collections have been microfilmed. Records of these collections are usually not themselves tagged as AM. Search them as follows in Advanced Search:

“slave insurrections” (for example)

AND

Subject: sources or diaries or correspondence or archives or manuscripts or “Personal narratives” or notebooks

And also do a separate search for

“slave insurrections”

AND

Title: papers or records or letters or documents or collection

For both of these searches, change Subtype limits from Any format to Microform.
Online guides (reel contents lists) may be linked from the WorldCat record. If not, search the collection title in Google or check the websites of the microform publisher. Online guides are not always available.

*Digitized collections in WorldCat*

WorldCat offers the most comprehensive search for digitized archival/manuscript material. However, analogous to microfilm, digitized AM collections usually are not tagged as AM. There are several methods, none of which is perfect, of isolating digitized manuscript material.

Use Advanced Search:

"slave insurrections"

AND

Subject: archives or correspondence or diaries or scrapbooks or sources or manuscripts

Check Internet Resources at Limit type to: and search

Most of the results are likely to be ebook versions of books containing primary sources rather than digitized AM. The default results order is by number of holding libraries. Ebooks held by numerous libraries will be toward the top of the list. Jump down in the list until you get to resources held by only one or two libraries. Mixed in with these will be the records of digitized AM collections in the free Web: clumsy but there is no other way to get them consistently.

The difficulties in searching WorldCat for AM material derive from the numerous member libraries each having its own idea of how AM (original, microfilm, digitized) should be tagged. Records for books are much more consistent.

*ArchiveGrid*

**Contents**

* ArchiveGrid contains all the WorldCat AM catalog records together with thousands of finding aids contributed by libraries and archives worldwide but mainly US.

All the WorldCat AM catalog records, with dissertations excluded, are included in ArchiveGrid, together with thousands of full text finding aids. All of this material is searched together. Microfilmed and digitized collections are not generally included.
Searching
ArchiveGrid has the advantage of offering a finer grained search than WorldCat. However, relatively insignificant occurrences of a name or subject term in a finding aid have equal weight to the occurrence of the same term in a catalog record. On the one hand, this makes it easier to find mentions of relatively obscure people or subjects; on the other hand, searches may turn up many results which will need to be examined closely in order to distinguish which ones actually contain the person/subject of interest as a major part of the collection. Thus, searches in WorldCat offer the advantage of yielding a more focused set of collections. It is somewhat analogous to JSTOR in that one may choose between searching a term in full text versus in title/abstract. ArchiveGrid can be searched using Library of Congress subject headings.

Some of the finding aids in ArchiveGrid also occur in NIDS (searched via Archive Finder, above), but NIDS has many that are not available in ArchiveGrid and vice versa.

SNAC (Social Networks and Archival Context)

Contents
* SNAC (Social Networks and Archival Context) http://socialarchive.iath.virginia.edu/

  extracts, structures and displays information on persons, corporate bodies, subjects, etc., from archival descriptions. It does not search the full text of finding aids as ArchiveGrid does, but SNAC displays its information in a more organized form with links to online finding aids. The SNAC Prototype at present holds a relatively small number of records and includes French and British as well as American repositories. SNAC includes all AM records from the public WorldCat up to about two years ago and so offers an alternative method for searching most WorldCat records.

Searching
Searching a subject (e.g., “civil rights”) will not lead to general collections around this subject. Rather, it will lead to a results list of collections focused on persons, corporate bodies, and organizations related to that term. This may be useful in discovering connections between entities that were not otherwise apparent.

Limitations of the Four Major Databases

Even these extensive databases do not cover all the bases. There are some limitations of Archive Finder, ArchiveGrid, WorldCat, and SNAC. First, the records in these four databases are brief and so cannot list everything in the collections. Named correspondents and other associates often include only well-known people or persons represented by large amounts of material. Second, the analysis necessary to allow
preparation of a record in one of these databases is time-consuming, and many unanalyzed and unreported collections await attention in the storage rooms of most repositories. Third, numerous repositories do not report their collections. Many of these are small local repositories but may have significant collections.

There is no sure way around these limitations. Subject guides, compiled by experts in special areas, often list unreported collections. Printed subject guides can be found by determining the appropriate subject terms. In library catalogs and WorldCat, the Subject terms applied to published subject guides are variable, but the following Subject searches are often effective:

- Anthropology and "archival resources"
- “Electronic data processing” AND (Archives OR Manuscripts) AND (Directories OR Catalogs)

A list of selected subject guides arranged by subject is available [http://guides.library.harvard.edu/content.php?pid=175047&sid=1473447](http://guides.library.harvard.edu/content.php?pid=175047&sid=1473447). Otherwise you will need to identify possible repositories, often from clues in biographical sources, and enquire directly.

**Other Sources for Catalog Records and Finding Aids**

The following sources include AM not in the major four databases. Some of them may offer subject browsing or other features useful even where they overlap with Archive Finder, WorldCat, ArchiveGrid or SNAC.

Many repositories are publishing finding aids in state and regional archival consortia databases without submitting them to Archive Grid. These must be searched individually. Many of these are linked from EAD Aggregators (Encoded Archival Description is a format for digitized finding aids) [http://library.wcsu.edu/staff/wiki/EADAggregators](http://library.wcsu.edu/staff/wiki/EADAggregators). A few sources not listed there include:

- PACSCL Finding Aids (Hidden Collections Initiative for Pennsylvania Small Archival Repositories) [http://sla.library.upenn.edu/dla/pacscl/index.html](http://sla.library.upenn.edu/dla/pacscl/index.html)

- Uncovering New Chicago Archives Project (UNCAP) delivers hundreds of finding aids from the archival collections of the University of Chicago and Northwestern University libraries. [http://uncap.lib.uchicago.edu/](http://uncap.lib.uchicago.edu/)
ArcCat http://arccat.uvm.edu/ is a catalog of Vermont archival and manuscript collections.

Hidden Collections Registry
http://www.clir.org/hiddencollections/registry#c12=all&amp;b_start=0 reports newly processed archival and special collections.

The Digital Public Library of America searches online material emanating from numerous US libraries. One cannot effectively limit to archival/manuscript sources, so broad topical searches are difficult. Specific searches work much better. Searches for proper names often yield individual items (photographs, letters) within collections.

There are many state and regionally focused link collections of primary source digital collections. These four sites overlap, but each lists sites missed by the others.

State Digital Resources: Memory Projects, Online Encyclopedias, Historical and Cultural Materials Collections
http://www.loc.gov/rr/program/bib/statememory/

Digital Libraries by State or Region
http://www.princetonlibrary.org/explore/online-resources/genealogy#state

250+ Killer Digital Libraries and Archives

State Archives Collections Online: Directory of Digital Collections, Virtual Exhibits, and State Memory Projects
http://www.statearchivists.org/arc/education/online_coll.htm

OAISTER is the largest search engine for online academic material. It includes digitized books and journal articles, open access publications, manuscript/archival material, photographic images, audio and visual files, data sets, and theses. It includes such a vast range of resources that digitized archival and other primary sources are lost in the abundant results if a broad topical term is used. So it is best to use a narrow term or proper name. Thus searching "Act-Up" yields archival letters. OAISTER searches digitized AM often down to item level. It is possible to limit a search to Archival Material, but we have not found this to be useful.

**Finding Archives and Manuscript collections online**
(Advanced Google Search)
Like dandelions in the springtime, AM collections seem to spring up everywhere, on every possible subject, yet not all of these collections have been formally reported to a larger organization for inclusion in an online catalog or database. General Google searches may yield many results, but it may take much sifting through the results in order to find relevant items. Using Google Advanced Search with specific search terms can help yield more focused results.

Within Google Advanced Search, use the “all of these words” box and the “this exact word or phrase” box to enter keywords for your subject as appropriate, then in the “any of these words” box, enter some combination of the following terms: “archives” or “manuscripts” or “correspondence” or “diaries” or “scrapbooks” or “sources.” Also consider searching “digital library” or “digital collection” or “digital archive” or “digital exhibit” or “online collection” or “online library” or “online archive” or “virtual library” in order to get at whole collections of items (note that the first set of words may be more likely to lead to any sort of AM collection, while the second set would lead to a digitized AM collection). Also, to find AM collections within a specific geographic region, consider searching “archival collections” OR “manuscript collections” along with the region or state name in the “any of these words” box. For any of these searches, use the “site or domain” box to search within just .edu, .org, or .gov sites.

Searching in this manner often gets at smaller, more specialized collections or portions of larger collections related to your topic (for instance, the civil rights materials within a presidential library or university archives). While the collections themselves may be smaller in scope than the ones turned up by searches in ArchiveGrid and Archive Finder, if your topic is specialized enough, or if you are not finding material elsewhere, this may be the route you need to take.