

Collecting Collections: Search Performance and Interface Usability in Six Online Archival
Reference Resources

Melissa A. Lafranchise
San Jose State University
Submitted June 30, 2013

Abstract

This paper reviews six online archival reference resources. Following a discussion of background and scope for each resource, search performance is evaluated based on a known-item search and a general topic search. The review concludes with a comparison of user interface and search functionality based on heuristics gleaned from both web usability and library/archive literature. Improvements in all areas will move these resources and the archives they represent further along the path to being truly user-centered.

Keywords: online archival reference resources, web usability, interface heuristics, archives, information retrieval systems, search performance, precision, United Kingdom Register of Archives, National Union Catalog of Manuscript Collections, OCLC WorldCat, Repository of Primary Sources, Online Archive of California, ArchiveGrid

Collecting Collections: Search Performance Interface Usability in Six Online Archival Reference
Resources

The following review examines six archival reference resources, including the United Kingdom National Register of Archives, the National Union Catalog of Manuscript Collections, the OCLC WorldCat, Repository of Primary Sources, the Online Archive of California, and ArchivesGrid. Following a discussion of background and scope for each resource, search performance is evaluated based on a known-item search and a general topic search. The review concludes with a comparison of user interface and search functionality based on heuristics gleaned from both web usability and library/archive literature. The goal in harvesting various archival repositories' collections into a single resource is to enable efficient searching of those disparate catalogs, and to a greater or lesser degree the six resources accomplish this. There is, of course, room for improvement in interface usability as well as search functionality. Improvement that will move these resources and the archives they represent further along the path to being truly user-centered. As Dearstyne attests "if use of records is the ultimate goal of all archival work, then researcher services are the essential means to that end" (1993, pp. 174-175), and harnessing the technology of the Internet and web to provide those services is paramount in maintaining viability in today's difficult funding environment.

Background and Scope

While not directly comparable in background and scope all six of the reference resources are efforts to bring together multiple smaller and geographically widespread archival collections under the auspices of a single web site. The goal is to enable efficient searching or browsing of these disparate resources.

Established in 1945, the National Register of Archives (NRA, <http://www.nationalarchives.gov.uk/nra/>) is a resource that maintains catalogs and lists of collections relevant to British history. Indexes of these archives are available to search online through the NRA site and link (when possible) to the electronic catalogs of the various repository institutions across the UK and in other countries where the manuscripts and records are held. Where electronic catalogs are not available, record reference numbers are provided. The NRA indexes grow through contributions from institutions and organizations, due to discoveries made and shared by researchers, and through the collection of annual accession reports. There are over 44,000 unpublished lists and catalogs in NRA indexes (The National Archives, n.d., About).

The National Union Catalog of Manuscript Collections (NUCMC, <http://www.loc.gov/coll/nucmc/index.html>) is a cooperative cataloging effort managed by the Library of Congress. Eligible archival repositories in the US provide descriptions of their holdings, and NUCMC catalogers create MARC21 bibliographic records and make those records searchable through an OCLC WorldCat gateway resulting in international exposure. The gateway only searches NUCMC catalogued materials using OCLC WorldCat technology. Printed from 1959 through 1993, the catalog was made electronic in the early 1990's and became available through the OCLC in 2007 (National Union Catalog of Manuscript Collections [NUCMC], 2011). Since 1959, the NUCMC has created more than 125,000 bibliographic records and worked with 1,800 repositories in the US (Gonzales, 2011, pp. 1-2).

A worldwide library cooperative, the Online Computer Library Center (OCLC) has been member owned and governed since its establishment in 1967 (OCLC, 2013). It launched WorldCat (<http://berkeley.worldcat.org/>) in 1971 (OCLC, 2013, A brief history of WorldCat) as way to provide the public with “intuitive access to a comprehensive collection of the wealth of resources libraries make available” (OCLC, 2013, OCLC WorldCat, para. 1). With the global participation of 72,000 libraries, WorldCat connects users to over 2 billion holdings through more than 300 million bibliographic records (OCLC, 2013, A global library reference). The UC Berkeley Library portal to WorldCat is evaluated in this review.

The Repository of Primary Sources (<http://www.uiweb.uidaho.edu/special-collections/Other.Repositories.html>) provides links to the web sites of archives and other primary source collections around the world. It is compiled through the efforts of Terry Abraham of the University of Idaho with the help of volunteers. The site currently includes links to more than 5,000 archives, museums, and special collections around the world (Abraham, 2013).

Made possible through the contributions of over 200 California archives, museums, libraries, historical societies, and rare book and special collections, the Online Archive of California (OAC, <http://www.oac.cdlib.org/>) allows browsing and searching of detailed collection guides with links to their housing institutions’ sites and direct access to digitized collections when available. Launched in 2002, the site was extensively redesigned in 2008-2009 to incorporate the latest web technologies. It is also a fundamental part of the California Digital Library. The OAC provides access to more than 20,000 collection guides and over 220,000 digital images and documents (Online Archive of California [OAC], 2009, About OAC).

ArchiveGrid (<http://beta.worldcat.org/archivegrid/>) is an experimental search site of OCLC Research that focuses specifically on archival collections as gleaned from OCLC WorldCat and from finding aids on the web. Its goal is to “provide access to detailed archival collection descriptions, making information available about historical documents, personal papers, family histories, and other archival materials” (ArchiveGrid, 2013, About ArchiveGrid, para. 2). ArchiveGrid allows searching across approximately 1.7 million records from around 1,000 archival repositories, mostly in the US and Canada (ArchiveGrid, 2013).

The purpose of each reference resource is to provide one web site from which researchers can search multiple archival collections with ease. The collection focus varies for each resource with the NRA being focused on British history archives, the NUCMC primarily concerned with collections found in US archives, the OCLC WorldCat encompassing global library collections with the ability to filter for archival holdings, the Repository of Primary Sources emphasizing North American archives but with some global coverage as well, the OAC focusing on institutions within the state of California, and the ArchiveGrid concentrating on archival holdings primarily in North America. All sites allow searching of their conglomerate collections except the Repository of Primary Sources, which only provides browsing functionality at the institution level, not the collection or record level.

Search Performance

This section will evaluate search effectiveness for the six reference resources. Test searches were modeled after the presumed behavior of a relatively inexperienced researcher. It is assumed he/she will begin with a simple keyword search, as “novice users [prefer] searching by keyword” (Mi & Weng, 2008, p. 7). Users do not use “complex query syntax” or Boolean operators, and they typically use two search terms per query (Yu & Young, 2004, p. 171). While in an actual search situation the researcher would perform iterative searches and would amend search syntax as he/she learned the particularities of the information retrieval system until sufficient relevant records had been found, in this evaluation, only two simple searches are performed for each resource. While cursory, this small amount of data should prove sufficient for a basic comparison of search performance for the six reference resources. In an attempt to, again, mimic real researcher behavior, two types of searches are tested: a known-item search for a specific topic, and a sample search for a general topic. The known-item search assumes the researcher is specifically interested in finding the papers or manuscripts of Walt Whitman—a narrow search for autobiographical or literary primary sources. The general topic search assumes the researcher is looking for papers and manuscripts of 19th-century poets—a broader query to assist a researcher in narrowing their subject matter based on the content and availability of primary sources. As Harter explains, in general topic searches “context is dynamic . . . the initial context will evolve as the search progresses” (Harter, 1992, p. 606).

The metric for a successful search is defined as its precision score, that is, the percentage of relevant records returned from all records returned (Harter, 1992, p. 602; Morville, 2005, p. 49). To calculate the precision scores in these test searches, total records returned were limited to the top 20 when more than 20 results were presented. Relevant records were considered to be those written or created by Walt Whitman in the case of the known-item search and those written or created by a 19th-century poet in the case of the topic search.

As shown in Figure 1, the NUCMC had the highest precision score for the Walt Whitman search with 55% of results being relevant. At 20%, the OAC had the lowest precision score for the known-item search. Most of the sites did return other results related to Walt Whitman; however, these collections and records were not written or created *by* Whitman. They were written or created *about* him and considered irrelevant for this reason. The ArchiveGrid earned the highest precision score (80%) for the 19th-century poets topic search. The UK National Register of Archives returned zero relevant results and thus had the lowest precision score for the topic search. It is likely that the inability to search by subject was the issue in this search, as it is difficult to believe that the National Archives of the United Kingdom has no papers or manuscripts written by 19th-century poets. Finally, it is not truly possible to compare the Repository of Primary Sources with the other sites, because it does not provide search functionality. Hence, its precision scores are 0% for both known-item and topic searches. Additional detail for individual test searches can be found in Appendix A.

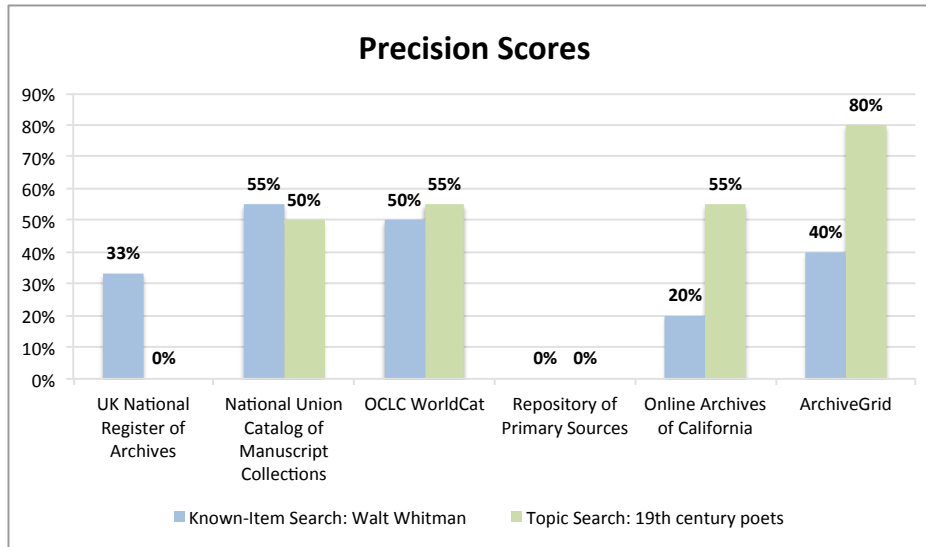


Figure 1: Precision scores allow comparison of the relevance of results among the six repository resources.

Figure 2 presents the total results returned by repository resource. This data provides a glimpse at the total number of collections or records related to the search queries that were not considered relevant by the definition used for this precision analysis. The ArchiveGrid returned the highest number of results for the known-item search with just over 1,400. The lowest number of results was returned by the NRA, but, as Whitman was an American poet, this is perhaps not unexpected. The National Union Catalog of Manuscript Collections provided the highest number of results for the 19th-century poets search with 1,517. Again, as no search is possible, the Repository of Primary Sources can't truly be compared.

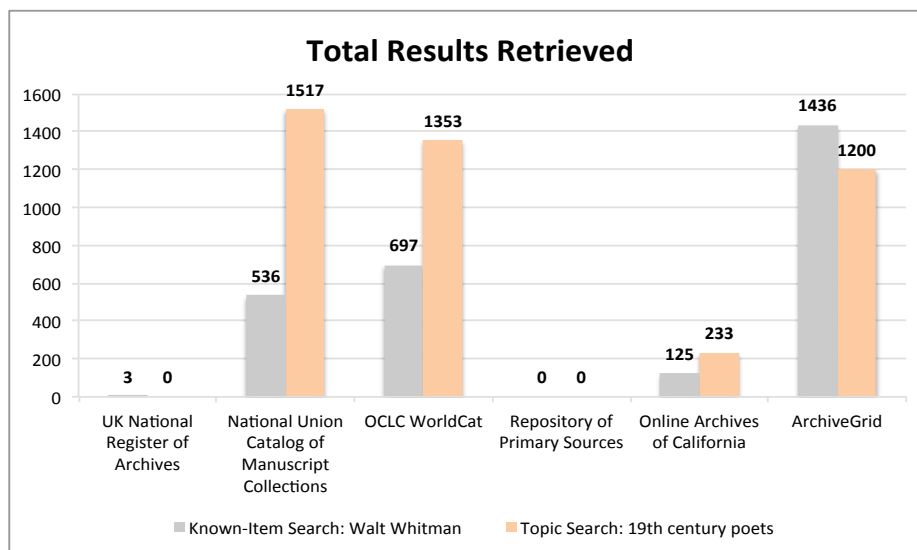


Figure 2: Total results retrieved allows comparison of the number of collections and records related to the search query though not considered relevant by the definition of this analysis.

The NRA's search performed marginally well on the known-item query, though papers by the American poet Walt Whitman are not necessarily expected to be prevalent in the UK. The

search for Walt Whitman returned one result in the personal name index and three collections related to Walt Whitman, though only one of these appeared to contain materials created by him (i.e., most are records of the Bolton Whitman Fellowship and related papers of its founder and members). The system's provenance search restrictions, however, make it all but impossible to search by topic, and, in fact, the FAQs plainly state that the indexes are "**NOT** arranged by theme or subject" (The National Archives, n.d., FAQs). As such, the search for 19th-century poets returned zero results, yet materials could be located by consulting the research aids which *are* subject based. Investigation by subject is plausible; it is just much more time-consuming, requires more manual and piece-meal involvement, and is more likely to result in missed records due to this complexity.

Being a collective catalog of archival holdings in the US, the NUCMC returned many results with just over half being relevant for the known-item search. It was easy to determine that the simple search form for names fields was the most appropriate option for the Whitman search. The topic search for 19th-century poets was also reasonably successful with 50% precision rating. Again, the simple search form for title, notes, and subject fields made the topic search not only possible but also easy. The detail of the catalog records returned varied greatly and was perhaps dependent on the detail of the descriptions provided to the NUCMC by participating repositories.

Comparatively, OCLC WorldCat is quite advanced in terms of matching user expectations for search functionality and performance. Perhaps this is due to the fact that it is not solely an archival catalog but also covers library holdings and, with the larger audiences those holdings bring, comes funding sufficient for frequent web development. The search functionality is specifically intended to mimic that of an Internet search engine. The instructions are elementary: "Use simple keywords as you would on a Web search site" (WorldCat and UC Berkeley Libraries, 2013, How to search, para. 1). Search results for the known-item Whitman search earned a 50% precision score with Berkeley results appearing first; this is by design as the WorldCat catalog is being accessed through the Berkeley portal. The precision score for the 19th-century poets search was a bit higher at 55%. The irrelevant results returned were typically records from the 20th century or critiques and reviews of poets rather than manuscripts and papers created by the poets themselves.

The Repository of Primary Sources is a unique site for browsing archival repositories. It is fundamentally a collection of links to the web sites of individual archives. No search is available on the site, and repository links are organized geographically, predominantly for the United States but also for Europe and other nations worldwide. There is a small section called Additional Links for more topic-based archives (e.g., History of Astronomy). Troublingly, random link sampling does uncover quite a few 404 errors where the linked page is not found or has been moved. This site would be helpful for those who were looking for archives in one particular geographic area, especially if they wanted to search online catalogs and then schedule a trip to the region and visit all archives with materials related to their topic of interest. This site really doesn't compare with the other online catalogs, because it is fundamentally unsearchable. Depending on a researcher's topic, they might be able to browse to an appropriate link and search that archives' holdings, but, for the known-item and topic searches used in this evaluation, the time required to browse through the thousands of links on this site and then search each archival repository individually is quite prohibitive. Nonetheless, given that the site has been in

existence since 1995 and appears to be updated regularly (based on the homepage note regarding holding submissions or corrections until July 5, 2013), it must have a following.

A researcher will need to take care with search syntax to get the most relevant results from the Online Archive of California. The known-item search for Walt Whitman interpreted the keywords with an OR operator. This resulted in many irrelevant results and a low precision score of 20%. Adding quotes around the first and last names to search by phrase rather than word helped to increase relevance. A similar interpretation was made for the topic search, where results were returned for the words 19th century or poets in addition to the entire phrase, but the precision score was solidly high at 55%. Though it varied by institution, most descriptions of collections were detailed and helpful in determining whether or not a record was relevant. Search results are, of course, limited to participating California institutions.

The ArchiveGrid returns results from the OCLC WorldCat, but, frustratingly, it doesn't always repeat the description from the WorldCat listing on the ArchiveGrid results page, requiring the researcher to perform an extra step in order to evaluate the result for relevance. As the ArchiveGrid is searching OCLC WorldCat records, one would assume that the results and precision scores would be quite similar. While the known-item precision scores were comparable (40% for ArchiveGrid, 50% for OCLC WorldCat), the topic search results were not (80% for ArchiveGrid, 55% for OCLC WorldCat). The ArchiveGrid also provides an additional and unique view of results under the Result Overview tab that goes beyond the traditional view of the Result List. The Result Overview provides a filtered/faceted presentation of results by various categories (e.g., People, Groups, Places, and Topics). Within each category, a list of subcategories is provided (e.g., American Poetry – 19th century) with a count of individual results for each. The subcategory name is a link that returns a Result List for only that subcategory. This is a useful and rather more visual method of results filtering.

As Hunter conveys, “much of archival administration has been *material-centered* rather than *client-centered*” (2003, p. 217). Assuming funds can be so prioritized, part of a shift towards more user-centered archives in this digital age would be to adjust online resources to more closely match user needs and expectations. The OCLC WorldCat, ArchivesGrid, and Online Archive of California are leading the way in such an effort. As will be shown in the following section, ideal interfaces make usability the first priority of the design and enable keyword searching across multiple databases with efficiency and flexibility.

Interface Usability and Search Functionality

The evaluation of interface usability, search features and functions, and results presentation will be based on the issues and recommendations highlighted in articles from Nielsen (1995), Yu and Young (2004), and Mi and Weng (2008). A checklist of nine criteria has been derived from these authors' works. Detailed results are included in Appendix B and highlights of strengths and weaknesses by reference resource are discussed below.

1. Consistency with user expectations

With the advent of Internet search engines, both novice and experienced archival

researchers have developed a set of preferences based on commonly experienced interface design and behavior. Archival reference sources should mimic this appearance and functionality to be consistent with user expectations. Doing so increases the usability of the archival source thus facilitating researcher access to and use of the records, thereby fulfilling the archives' mission. Two specific consistency considerations were reviewed in this analysis: the addition of visual and value-added information on the results page and the display of researcher comments and recommendations (i.e., some social/collaborative aspect).

The OCLC WorldCat site is overall most consistent with user expectations. It offers space for visual information like thumbnails for results (though archives without digitized holdings cannot provide such thumbnails). Additionally, it is the only site of the six that also allows researchers to add comments and recommendations to an individual result. The ArchiveGrid and OAC are close seconds in terms of consistency, with the ArchiveGrid providing value-added information and the OAC providing visual information on individual results pages. The NRA and NUCMC make searching rather more difficult in that a single and readily apparent search box—something today's users expect—is not the first thing a researcher encounters. The Repository of Primary Sources is only a link hub; no searching is possible and for that reason it does not meet user expectations. However, when considered as a link hub, the site does perform as users expect in that it provides the most fundamental functionality of the web—hyperlinks.

2. Flexibility and efficiency of use

Enabling flexible and efficient use should be a goal of any site. For these reference resources, efficiency is particularly important, as the point of creating a site that aggregates archival collections is to provide researchers with the most efficient means of searching those collections. Building flexibility into the site adds value to the resource and ensures that researchers will return to use it often. The six reference resources are reviewed for the following characteristics of flexibility and efficiency:

- Metasearch, searching all collections/records from a single search box
- Ability to save searches
- Automatic mapping of keywords to controlled vocabulary
- Ease of switching between individual result and results list
- Search terms and search boxes retained on results page
- Filtering of search results

Overall, the OCLC WorldCat outperforms all other reference resource sites in terms of flexibility and efficiency of use, with the minor exception of switching between an individual result and the results list. When the user clicks into an individual record from any page of results past the first page of results, the back button returns the user to the first page of results instead of the second or third or nth page of results from which he/she started. The ArchiveGrid and OAC perform equally well, adhering to all flexibility and efficiency principles except that neither allows searches to be saved nor maps user-entered keywords to the controlled vocabulary of the database. This last functionality can potentially increase the relevance of search results by suggesting appropriate search terms to researchers as they type into the search box. While the NRA and NUCMC sites do provide functional searches, they are neither flexible nor efficient per

the heuristics used in this analysis. The Repository of Primary Sources is again the outlier in that no search functionality is present. The site organization doesn't make for particularly flexible or efficient browsing, because it typically only allows browsing by geographic area.

3. Spell-checking or correcting

The OCLC WorldCat and ArchiveGrid provide spell-checking, but neither actually suggests the correct spelling on the results page. No other sites offer spell-checking or correcting.

4. Visibility of system status

Only the OCLC WorldCat alerts the user that a task is occurring. For example, pop-ups notify users when results filtering is underway. The ArchiveGrid and OAC do not offer such status messages; however, it could be argued that the results are returned so quickly that no such message is required. The NRA and NUCMC *could* benefit from such a status message as their searches *do* take some time during which the user might not realize the search is underway.

5. Recognition rather than recall for orientation within the site

Users should always know where they are within a site's navigation scheme. Nielsen warns, "if users get lost on a website, they **leave**" (Nielsen, 2010, Why usability is important, para. 1). The fulfillment of this principle varies between all six sites, but all allow researchers to understand their orientation within the site to some degree. Notably, the NUCMC site loses its formatting and menu system within the search dialogue and results, and therefore the search appears to occur outside of the site. Technically, this appearance is accurate as the NUCMC site provides a gateway to a custom OCLC WorldCat search. The OAC provides a unique and helpful type of orientation in that individual collection results pages actually include a breadcrumb trail to the collection guide within the parent institution—a helpful shortcut to either related information or institution information.

6. Help documentation including research aids

Users should be able to quickly locate context-specific help documents. Help documentation is readily available for all sites except the Repository of Primary Sources.

Research aids are present only on the NRA site and are especially important, because the search functionality doesn't allow for subject searching. As such, research aids provide the means for researchers to browse by subject and locate relevant archival collections. The OAC site, uniquely, searches the research aids themselves (which they term collection guides). The NUCMC, OCLC, and ArchiveGrid don't specifically provide research aids, but the search functions on these sites allow for subject searching and, as such, no research aids are required.

7. Aesthetic and minimalist design

Good design should focus users' attention on the important parts of the site; it should never overwhelm content. As Yu and Young explain, "successful screen designs should focus on

presenting the quick-searching options employed by the majority of users first” (2004, p. 169). And Koh’s research revealed “the importance of website design. Eighty-eight percent of respondents valued easy website navigation, making it one of the most important characteristics of services the [archives] offers” (Koh, 2003, p. 190). The OCLC WorldCat, ArchiveGrid, and OAC sites are solid examples of aesthetic and minimalist design. It is assumed that researchers most often come to these sites to search for materials. The search boxes and their placement on these three sites emulate Google’s very successful homepage design by placing the search box front and center—anticipating user needs and providing the means to fulfill them. The NRA’s site design is both aesthetic and minimalist though some additional work on bringing the search boxes to the top of the navigation structure would be beneficial. The NUCMC site is also minimalist and aesthetic, but it doesn’t focus users attention on search. This is perhaps because it is trying to speak to three different audiences on its homepage (i.e., repositories, researchers, and archivists). Targeting the core message of the homepage to one of the three audiences would be beneficial, and such targeting should really focus on researchers as the ultimate customers of the other two audiences. Finally, the Repository of Primary Sources does employ a minimalist design that focuses user attention on the main link categories, even though this design may not be particularly aesthetic.

8. Format

Most sites base their descriptions on some form of MARC or EAD record. The NRA site varies because its results link directly to the record or collection on the individual repository website (but only if an electronic record exists).

9. Open access

The online resources are all open, with the possible exception of the OCLC WorldCat as the site used for this analysis is a portal for UC Berkeley and some records may not be available without a Berkeley login. Physical access to materials will vary based on the rules and restrictions of the individual repositories.

Conclusion

If, as Hunter says “archival records exist to be used” (2003, p. 205), then it behooves the administrators of the six reference resources reviewed in this paper to understand and implement best practices for their web sites. Through a search performance comparison and a heuristic evaluation, this analysis shows that the OCLC WorldCat resource is leading the way. The Online Archive of California and the ArchiveGrid are following closely with usability features and search functionality that meet today’s users’ expectations. The UK National Register of Archives, while performing soundly, can improve by expanding search functionality to cover subject searches, and the NUCMC would do well to focus on researchers as the target audience. Finally, assuming its owner was interested in increasing its usefulness, the Repository of Primary Sources would need to significantly expand its functionality to include search. Fulfilling the vital mission of archives means going beyond preservation to provide expansive access by utilizing today’s technology for the benefit of researchers around the globe.

References

- Abraham, T. (2013). Repository of primary sources. Retrieved from <http://www.uiweb.uidaho.edu/special-collections/Other.Repositories.html>
- ArchiveGrid. (2013). About ArchiveGrid [Web page]. Retrieved from <http://beta.worldcat.org/archivegrid/about/>
- Dearstyne, B. W. (1993). The archival enterprise: Modern archival principles, practices, and management techniques. Available from <http://www.amazon.com/gp/search?index=books&linkCode=qs&keywords=0838906028>
- Gonzales, T. (2011). NUCMC program: annual report, fiscal year 2011 [Adobe Acrobat document]. Retrieved from <http://www.loc.gov/coll/nucmc/NUCMCFY11v3.pdf>
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of American Society for Information Science*, 43 (9), 602-615. Retrieved from <http://www.asis.org/jasist.html>
- Hunter, G. S. (2003). *Developing and maintaining practical archives* (2nd ed.). New York, NY: Neal-Schuman.
- Koh, C. (2003). Reconsidering services for the postmodern student. *Australian Academic and Research Libraries*, 34(30), 184-193. Retrieved from <http://alia.org.au/publishing/aarl/index.html>
- Mi, J., & Weng, C. (2008). Revitalizing the library OPAC: Interface, searching, and display challenges. *Information Technology and Libraries*, 27(1), 5-22. Retrieved from <http://ejournals.bc.edu/ojs/index.php/ital/index>
- National Archives, The. (n.d.). National Register of Archives About [Web page]. Retrieved from <http://www.nationalarchives.gov.uk/nra/aboutapps/nra/about.htm>
- National Archives, The. (n.d.). National Register of Archives FAQs [Web page]. Retrieved from <http://www.nationalarchives.gov.uk/nra/help/nra/nrafaq.htm>
- National Union Catalog of Manuscript Collections. (2011). About the National Union Catalog of Manuscript Collections (NUCMC) [Web page]. Retrieved from <http://www.loc.gov/coll/nucmc/about.html>
- National Union Catalog of Manuscript Collections. (2011). NUCMC timeline: 1947-present [Web page]. Retrieved from <http://www.loc.gov/coll/nucmc/timeline.html>
- Nielsen, J. (1995, January 1). 10 Usability heuristics for user interface design [Web page]. Retrieved from <http://www.nngroup.com/articles/ten-usability-heuristics/>

Nielsen, J. (2012, January 4). Usability 101: Introduction to usability [Web page]. Retrieved from <http://www.nngroup.com/articles/usability-101-introduction-to-usability/>

OCLC. (2013). A brief history of WorldCat [Web page]. Retrieved from <http://www.oclc.org/worldcat/catalog/timeline.en.html>

OCLC. (2013). A global library resource [Web page]. Retrieved from <http://www.oclc.org/worldcat/catalog.en.html>

OCLC. (2013). OCLC WorldCat [Web page]. Retrieved from <http://www.oclc.org/worldcat.en.html>

OCLC. (2013). Public purpose [Web page]. Retrieved from <http://www.oclc.org/about/purpose.en.html>

Online Archive of California. (2013). About OAC [Web page]. Retrieved from <http://www.oac.cdlib.org/about/>

WorldCat and UC Berkeley Libraries. (2013). About WorldCat and UC Berkeley Libraries [Web page]. Retrieved from <http://berkeley.worldcat.org/whatis/local/default.jsp>

Yu, H., & Young, M. (2004). The impact of web search engines on subject searching in OPAC. *Information Technology and Libraries*, 23(4), 168-180. Retrieved from <http://ejournals.bc.edu/ojs/index.php/ital/issue/current>

Appendix A – Detailed Search Performance by Reference Resource

Reference Resource/Search Term	Total records retrieved	Relevant Records	Precision Score (relevant/total)	Comments
UK National Register of Archives				
Topic search: 19th century poets	0	0	0%	Given the search restrictions to provenance categories (i.e. a researcher can only search on corporate name, person name, family name, place name), it is not possible to search by a general topic. Searches result in zero results found. A search of the research guides for "poetry" however brought up a guide to Drama and Poetry and a list of repositories to search individually.
Known-item search: Walt Whitman	3	1	33%	Initial results page shows one result in the personal name index for Whitman. Clicking on this index link brings up three individual listings with links to individual archives websites or directly to the Whitman related collections in the archives' online catalog. While the collections are not specifically papers of Walt Whitman himself, some of his letters and works are included in one and hence that is the only one counted as a positive result.
National Union Catalog of Manuscript Collections				
Topic search: 19th century poets	1517	10	50%	The precision score is calculated based on using the first 20 results as a representative sample. Only poets were counted as positive results; review and commentary written about 19th century poets was not considered relevant. Results returned were global and in more than English which seemed beneficial given that this was intended to be broad topic search.
Known-item search: Walt Whitman	536	11	55%	The precision score is calculated based on using the first 20 results as a representative sample. Records or collections where Whitman was tangentially related are not counted as positive results; only records where he is the main author or active and contemporary second author are counted.
OCLC WorldCat				
Topic search: 19th century poets	1353	11	55%	The precision score is calculated based on using the first 20 results as a representative sample. Only collections where the creator was a 19th century poet were counted as positive results. Most are US poets as the Berkely portal searches the Berkely collections first, then other member associations. There is a definite bias towards American poets.
Known-item search: Walt Whitman	697	10	50%	The precision score is calculated based on using the first 20 results as a representative sample. Only collections that included materials created by Whitman are counted as positive results. Again, Berkely collections are shown first in this Berkely portal to WorldCat.
Repository of Primary Sources				
Topic search: 19th century poets	0	0	0%	No search is available on this site; it is browse only. No precision score will be calculated.
Known-item search: Walt Whitman	0	0	0%	No search is available on this site; it is browse only. No precision score will be calculated.

COLLECTING COLLECTIONS: SEARCH PERFORMANCE AND INTERFACE USABILITY IN SIX ONLINE ARCHIVAL REFERENCE RESOURCES

15

Reference Resource/Search Term	Total records retrieved	Relevant Records	Precision Score (relevant/total)	Comments
Online Archives of California				
Topic search: 19th century poets	233	11	55%	The precision score is calculated based on using the first 20 results as a representative sample. Only collections where the creator was a 19th century poet were counted as positive results. Many results were for 19th century writers or poets from other time periods as the search was based on keywords, not the keyword phrase. Results are restricted to archives found in California repositories. Individual items were not reviewed, but are very helpfully returned on a separate results tab and often include thumbnail images.
Known-item search: Walt Whitman	125	4	20%	The precision score is calculated based on using the first 20 results as a representative sample. Only collections that included materials created by Whitman are counted as positive results. The search defaults to an OR operator instead of an AND, hence many results were for either Walt or Whitman, but not both. Using quotes to search on the phrase/full name increased relevance of results. Results are restricted to archives found in California repositories.
ArchiveGrid				
Topic search: 19th century poets	1200	16	80%	The precision score is calculated based on using the first 20 results as a representative sample. Only collections where the creator/author was a 19th century poet were counted as positive results. This was a very successful search in terms of relevance and would potentially provide much fodder for a researcher. There is an issue with the abbreviated results records in that often collections appear identical in the ArchiveGrid results, but are proven to be different when the OCLC WorldCat descriptions are viewed. This could cause significant frustration and cause researchers to ignore potentially useful material.
Known-item search: Walt Whitman	1436	8	40%	The precision score is calculated based on using the first 20 results as a representative sample. Only collections that included materials created by Whitman are counted as positive results. Many masters theses were returned as well as other contemporary or latter day criticism and correspondence from other creators. Frustratingly, the ArchiveGrid results often simply linked to the OCLC WorldCat for detailed description forcing the researcher to perform an additional step in order to vet the result.

Appendix B – Detailed Evaluation by Reference Resource

Evaluation Criteria		Reference Resources				
	UK National Register of Archives	National Union Catalog of Manuscript Collections	OCLC WorldCat	Repository of Primary Sources	Online Archives of California	ArchiveGrid
Consistency with user expectations	While the search is not exactly reminiscent of an internet search engine, it fundamentally behaves as users expect.	From the home page, finding the search utility is difficult. The search box that is the cue for most people is for the website only, not the manuscript collection.	Overall the site is very consistent with user expectations. It’s up to date with common design standards and search functionality.	If researchers are expecting to visit a site that allows them to search archival collections, they will be disappointed. This site is a hyperlink hub and in that way functions as expected.	The site is very consistent with users expectations in both design and search functionality.	The site is very consistent with users expectations in both design and search functionality, but has some unique features like the Result Overview tab for filtering results.
Visual and value-added information included on results page	No	No	While there are spaces to include visuals on the results page, most archival records do not have thumbnail images associated with them. Individual collection pages do offer value-added features like citation export, sharing, printing, etc.	No	While visual information is not included on the collections results tab, it is included on the individual items results tab in the form of thumbnail images for items when available.	No visual information is included on results pages. Value-added information is included in the Result Overview tab where result filtering is possible. The bottom of each results page also includes helpful links refine the search and search FAQs.
Other researcher comments/recommendations displayed (social aspect)	No	No	Yes, both reviews and tagging are allowed for individual records.	No	No	No

Evaluation Criteria		Reference Resources				
	UK National Register of Archives	National Union Catalog of Manuscript Collections	OCLC WorldCat	Repository of Primary Sources	Online Archives of California	ArchiveGrid
Flexibility and efficiency of use	Search syntax is automatically applied making for a flexible and efficient search. The interface is not exactly flexible or efficient, though; a subject search will likely return no results unless it falls into one of the four provenance search categories.	There are 4 options for searching; the first 2 of mirror keyword searches that users are familiar with. Locating individual collections that aren't at the LoC is quite inefficient requiring 4 or 5 extra steps.	The site and search are designed around common web standards and user expectations for an internet search engine style search. Keywords searches of all member collects are possible. The interface is very flexible and offers a wide variety of methods for filtering search results.	This site simply provides a list of hyperlinks organized geographically and to some extent topically. It is not made to be flexible. If you know the geographic region in which the archives you are interested in is located, then you might find the site efficient. Otherwise, it is a rather inefficient way to browse archival collections.	While it takes a few searchers to understand how search syntax is applied, it's simple enough to learn. The search is flexible and efficient and returns both collections level and item level results at the same time. Filtering options are available on the results pages.	A flexible keyword search and with unique and efficient filtering options available on the Result Overview tab. Individual records frustratingly do not typically include detailed descriptions. The user must click the OCLC WorldCat link for those descriptions, adding a step and decreasing efficiency.
Metasearch	No metasearch. Researcher must select from four databases to search.	No true metasearch, but the first simple search option is almost one in that it covers multiple fields.	Yes. The purpose of WorldCat is to act as a metasearch for all member organizations' collections.	No	Yes. All participating California institutions are searchable from a single search box.	Yes. Searches the OCLC WorldCat records.
Ability to save searches	No	No and results page times out quickly, necessitating re-entry of search terms.	Yes. You can save search results to a List.	No	No	No
Mapping of keywords to controlled vocabulary	No	No	Yes. Alternative queries are automatically suggested.	No	No	No
Ease of switching between individual results and results list	Easy with back buttons in the user interface. Browser back buttons work as well.	Easy with browser back buttons and "More on this record" links from the results list.	While it is easy to switch, there is an issue. Clicking on individual results on page 2 and then clicking back will return you to page 1 instead of page 2, and the filters are removed.	N/A	Easy with the browser back button	Easy with the browser back button

Evaluation Criteria		Reference Resources				
	UK National Register of Archives	National Union Catalog of Manuscript Collections	OCLC WorldCat	Repository of Primary Sources	Online Archives of California	ArchiveGrid
Search terms and search boxes retained on results page	No. Researcher must either go back or click on new search link at top of page.	No. Researcher must click on link to go back to a blank Search form.	Yes. The researcher can perform a new search or amend the existing one.	No	Yes. The researcher can perform a new search or amend the existing one.	Yes. The researcher can perform a new search or amend the existing one.
Filtering of search results	No	No	Yes, many filter options.	No	Yes, several filter options.	Yes, through Result Overview tab.
Spell-checking or correcting	No	No	Spell-check, no correcting	No	No	Spell-check, no correcting
Visibility of system status	No visual cues to the fact that a search is underway.	No visual cues to the fact that a search is underway.	Yes. Pop-ups notify user that a task, such as results filtering, is underway.	No visual cues to the fact that a search is underway.	No visual cues, but results are returned so quickly as to not require any.	No visual cues, but results are returned so quickly as to not require any.
Recognition rather than recall for orientation within the site	The breadcrumb trail does an excellent job of orienting the researcher, but it is otherwise difficult to know where on the site you are located as no navigation options are highlighted in the top menu. It may be difficult to find this page again without a bookmark.	Very limited. The search dialogue appears to be entirely separate from the NUCMC site. There is a link in the footer to return to the NUCMC homepage, but there is no visual and obvious design element to describe your position within the web site.	While it is easy for researchers to see that they are accessing from within Berkeley, and they can tell they are within search results, there is no breadcrumb trail or menu highlights to explain their position within WorldCat.	Headings and subheadings indicate orientation within the site, but no standard menus, nor breadcrumb trails are provided.	Headings clearly indicate which page the researcher is on. Individual collection results actually include a breadcrumb trail to the collection guide within its parent institution, not within the search results—really quite helpful for orientation within the site.	Headings clearly indicate which page the researcher is on. While there is no breadcrumb trail, none is really required as the site is focused almost exclusively on searches. There aren’t many other supplemental web pages.
Help documentation	Search tips are noticeable, clear, helpful, and context dependent. The sites FAQs page covers information about searching, catalog access, copyright, etc.	Search instructions are available through a link on the search form option page. The instructions, while not simple and few, are clear, understandable, and organized in a way useful to researchers.	The search is intuitive enough not to need instructions. Entering keywords like a search engine is all that’s required. There is a librarian chat feature for help as well.	No	Help is easily available from the top navigation menu.	Help is available at the bottom of each search results page. The first option provides quick links to narrow or broaden the search with proper search syntax noted. There are also search FAQs.

Evaluation Criteria		Reference Resources				
	UK National Register of Archives	National Union Catalog of Manuscript Collections	OCLC WorldCat	Repository of Primary Sources	Online Archives of California	ArchiveGrid
Research aids	Aids are present to assist with subject/topical searchers, and direct users to website of specialized archives (e.g., Media Archive for Central England). A sub-search specifically for the research guides is available.	No	There are no research aids apparent, but automated suggestions for related subjects are provided.	No	The search is actually of the collection/finding guides, so the research aids are directly incorporated into the search. Browsing by collection guide and by institution is also possible. Great flexibility.	No
Aesthetic and minimalist design	The visual design is clear, concise, and minimalist. There are no overbearing or confusing images, colors, or patterns. Menus are easy to find and understand.	The NUCMC home page is aesthetic and minimalist if not quite standard interface design which does make it difficult to locate important content, like the search function. The search pages themselves are text-only and minimally designed.	The design is minimalist and aesthetic providing focus for the search box and subsequent results. The interface is well organized and easy to understand.	While it certainly is a minimalist design, it may not be the most aesthetic. The content is, however, organized clearly and understandably.	The design is minimalist and aesthetic providing focus for the search box and subsequent results. The interface is well organized and easy to understand.	The design is minimalist and aesthetic providing focus for the search box and subsequent results. The interface is well organized and easy to understand.
Format	Varies	Based on MARC21	Based on MARC	N/A	EAD	MARC, EAD, PDF, HTML
Open access catalog	Yes	Yes	May require Berkeley login to see some content.	Yes	Yes	Yes