Professor: David de Lorenzo

Student: Jason Sue

Class: INFO 256-10 Summer 2020

Evaluation of Reference Resources for Archival Retrieval

When it comes to advancing research through the study of primary resources, there is not a single resource or platform that stands out as peerless in their ability to locate relevant archival records or personal papers. This paper evaluates six reference resources Discovery from The National Archives (UK), Finding Aids (Library of Congress), WorldCat (OCLC), ArchiveGrid (OCLC), the Mountain West Digital Library (MWDL), and the Online Archive of California (OAC) in regards to their efficacy in aiding researchers in locating primary resources in the custody of manuscript and archive repositories.

This first section of this paper discusses the scope, as well as, how thorough and useful those records are to a researcher. The second section of this essay evaluates the process of each platform and focuses on their usability (focusing specifically on what features—or lack thereof—are present to enhance either the ease of the search process or the quality of retrieved records). The final section assesses the strengths and weaknesses of the six reference resources in relation to their retrieval of relevant archival material.

# Section 1: Evaluate the Scope and Content of the Resources.

## Discovery (The National Archives of the UK)

One of the references sources under examination as a source of archival material is aptly named “Discovery” and provides online access to the finding aids of The National Archives (UK). In addition, this archival resource contains records from 2,500 other archives. In total, this “Discovery” tool has accumulated over 9 million records that are available for download (National Archives, 2020). The thoroughness of any specific finding aids retrieved from the tool Discovery is largely dependent on the practices of the specific contributing institution in creating descriptive finding aids for the collections under their care. From this author’s cursory exploration of the tool Discovery, finding aids originating from The National Archives of the UK are generally thorough.

## Finding Aids (Library of Congress)

The second reference resource for archival collections that this essay analyzes is called Finding Aids and is provided by the Library of Congress (LoC). Finding Aids allows researchers to search for the finding aids that describe the Library of Congress’s archival material (Library of Congress, 2020). This search is limited to the collections under the direct custody of the institution.

Finding aids retrieved from this reference resource are thorough. They not only provide full descriptions of the content but also warnings of when material is stored offsite and when access restrictions apply. In addition, the copyright of the institution’s collections is clearly stated in the finding aids for the consideration of the researchers.

## WorldCat (OCLC)

A third reference resource for archival material is WorldCat—the Online Computer Library Center’s global union catalog, which contains the records of 10,000 institutions. The OCLC boasts that this resource is the “world's largest network of library content and services” and contains the records of several archive and manuscript repositories (Online Computer Resource Center, 2020d).

The WorldCat database retrieves bibliographic records of which current records are created in accordance with Machine-Readable Cataloging (MARC). Older bibliographic records are included that are formatted in accordance with older standards. In general, these records lack the depth of a finding aid but serve as a good starting point and offer links to the website for the institution(s) that currently own the manifestation of the resource being described.

## ArchiveGrid (OCLC)

Another reference resource for reference material is ArchiveGrid. ArchiveGrid was previously offered as a subscription-based discovery by the OCLC in 2006, but this practice was discontinued in 2012 (OCLC Research, 2020). Despite this, the database for ArchiveGrid is still being updated with new records on intervals of six weeks. These updates consist of MARC records extracted from the WorldCat union catalog. ArchiveGrid is now freely available to the public through the OCLC website.

ArchiveGrid provides references for over five million archival items (Online Computer Resource Center, 2020a). WorldCat supplies the majority of archival material descriptions that are accessible through ArchiveGrid. These bibliographic records come primarily from the collections of institutions that reside in the United States. This reflects the volume of participation of archive and manuscript repositories of the United States in comparison to other countries (OCLC Research, 2020).

The ArchiveGrid reference resource retrieves the title of relevant collections and items from the participating institutions. Each retrieval is accompanied by a primary contributor, an institution and a basic description. If a researcher is interested in the record they may select the blue “Contact Information” button to go to the institution’s website or select the green “View the Catalog Record” button to view the bibliographic record (e.g. a MARC record).

## Mountain West Digital Library (MWDL)

Another reference resource under examination by this paper comes from the Mountain West Digital Library (MWDL). The MWDL hosts a union catalog, which draws from the bibliographic records of public libraries, universities, museums, historical societies, and government agencies in Hawaii, Idaho, Nevada, Oregon, Utah, Wyoming, and other locations consistent with the western United States (Mountain West Digital Library, 2020). Collectively, the Mountain West Digital Library provides bibliographic access to over one million resources. These collections, which have been contributed to MWDL’s database, are primarily academic or archival in nature.

Retrievals consist of the bibliographic record of the collection or item listed as well as a link to the finding aid of the collection which provides a more thorough record than a basic bibliographic record. Copyright authorization does not seem to be consistently listed among the records.

## Online Archive of California (OAC)

The last archival reference resource that this paper will analyze and describe is the Online Archive of California (OAC). The OAC provides free access to the descriptions of collections of primary resources maintained by more than 200 contributing institutions. These institutions include libraries, archives, historical societies, and museums throughout the state of California. Of some note is the inclusion of the archival collections maintained by the 10 University of California (UC) campuses. The OAC features more than 20,000 collection guides. Cumulatively, the OAC contains over 250,000 documents and digital images (Online Archive of California, 2020).

Retrievals describe far more than what is expected from a basic bibliographic record. This includes statements on how access is obtained to the materials as well as who notes on the copyright ownership. The records viewed by the author of this paper though did not describe the relevant collections arrangement of the collections at the box level and therefore fall short of the definition of a finding aid.

# Section 2: Evaluate the Resources using a search query term

## Discovery (The National Archives of the UK)

The National Archives (UK) tool Discovery provides a simple search easily locatable at the top of the webpage. To enhance usability, the webpages of the tool Discovery easily adapt to changing window sizes and are able to accommodate both mobile and desktop screen sizes. The simple search provides a field for keywords and two fields to specify a time period in which researchers may limit retrievals to. Lastly, the simple search has a dropdown menu that allows the patron to search the records of all participating archives, The National Archives of the UK, or all participating archives except for the holdings of The National Archives (UK). This layout allows the application of multiple filters for an initial simple search

If the simple search is not adequate, the researcher may select the links “Advanced search” or “browse” for more options. The tool Discovery’s advanced search has two different forms to outline a search. The first form focuses on the attributes of the records and the second form focuses on the records creators.

The unrestricted keyword search results for “Holocaust AND internment” retrieved 18 finding aids of which one result was from The National Archives (UK). The thoroughness of the retrieved finding aids was largely dependent on the practices of the contributing institution with the finding aid from The National Archives (UK) being the most thorough. All retrievals were relevant indicating a high level of precision. The Discovery tool supported Boolean searching. As proof, the author swapped the Boolean operator “AND” for the Boolean operator “OR.” The resulting keyword search for “Holocaust OR internment” generated 178,066 retrievals, which is consistent with Boolean search theory.

The Discovery tool organizes the results initially by relevance, but researchers have the option of reordering them in accordance with title (ascending and descending), and date (ascending and descending), and reference. Searches may be refined through filtering. The options for filtering provided by the Discovery tool of The National Archives of Great Britain are limited, but useful. Filtering by date is restricted to predetermined time periods and filtering by source is limited to the three options discussed earlier in this tool’s simple search. The Discovery tool does follow the best practice of accompanying each filter with a figure to assist users with narrowing their search results to a manageable number.

The author of this paper has made attempt to evaluate all the reference resources using the same criteria. The author assumed that there would not be adequate time to conduct an evaluation on the reliability and consistency of these reference resources, but an honest assessment of this reference resource must make note of the shortcomings that were encountered during the evaluation of the Discovery tool of The National Archives. These shortcomings are detailed in the next paragraph.

When searching for the scope of the database and basic information on the institution the author encountered several dead links. In addition, when the author was attempting to replicate the search results to verify a search feature of the archive’s interface the author was unable to connect to the search tool, because the website was timing out. The author verified that the institution’s website was the source of the technical difficulties by surfing through several news sites without incident. After several hours, the author was able to access the tool, but the keyword search query for “Holocaust AND internment” retrieved only a single finding aid—namely the finding aid for the collection under the custody of The National Archives. The author carefully started a new search and got the same singular result. An hour later, the author made another attempt at a new search and replicated the initial 18 retrieved results. Clearly, the consistency of the Discovery tool and the reliability of The (UK) National Archives website which hosts the Discovery tool could both be improved.

## Finding Aids (Library of Congress)

The Library of Congress’s Finding Aids interface allows researchers to conduct simple searches by keyword and “Finding Aid ID.” It should be noted that these webpages are not mobile-device friendly and have difficulty with narrow windows. The reference tool Finding Aids’ simple search is able to filter results by medium and by certain ethnic or geographic divisions as designated as part of an existing LC research center. The Finding Aids tool lacks a comprehensive advanced search option but does allow patrons to browse the Library of Congress’s archival collection by date, location, name, and subject.

Using the Library of Congress’s Finding Aids tool to search for “Holocaust AND internment” retrieved five results of which all were relevant indicating a high level of precision. The Library of Congress’s Finding Aids is not capable of Boolean searching. If it were, the search result for “Holocaust OR internment” would have retrieved a significant quantity of finding aids (including finding aids from collections concerning the Japanese internment in World War II) instead of the same five results. A significant handicap to the ability to refine such results of the tool Finding Aids is the inability to filter said results. Another shortcoming is that the reference instrument Finding Aids does not indicate by what standard it orders a list of retrieved finding aids and that there is no way for a patron to specify by which attribute they prefer to use when ordering retrieved finding aids.

## WorldCat (OCLC)

The Online Computer Library Center (OCLC) has assembled an impressive compilation of bibliographic records in their WorldCat catalog; but despite the technological sophistication that such a name implies, the WorldCat website does not adapt to differing window sizes and thus is not friendly to users of mobile devices. WorldCat’s simple search allows a person to filter results by medium (e.g. books, DVDs, CDs, articles) and only allows the application of one filter. It also lacks the ability to immediately narrow the results of a search query to archival material. WorldCat does offer an advanced search option that allows the application of multiple filters including limiting retrievals to only archival material.

The author’s initial exploration of WorldCat began with a simple keyword search “Holocaust AND internment” with a full expectation that the overwhelming majority of the results would not refer to primary sources (from archival institutions). This initial search retrieved 25,022 results. This large number of retrievals supports the premise that the WorldCat union catalog has the highest recall of the world’s relevant material—abet—mainly of secondary sources. To verify the ability of WorldCat to handle Boolean operators, the author modified substituted the Boolean operator “AND” with the Boolean operator “OR.” The number of retrievals for the basic keyword search for “Holocaust OR internment” exploded exponentially to 46,153,384 retrievals. The difference in the volume of results when comparing the two searches confirms that WorldCat does indeed have Boolean functionality.

Obviously, the search results for the initial query of “Holocaust AND internment” had a low precision with many irrelevant resources. Such searches can easily be refined to retrieve archival material through the application of the “Format” filter set to the trait “Archival material.” Other filters include author, year, language, audience, and topic. The WorldCat website allows researchers to utilize multiple filters to refine their results. The “Format” filter has an additional degree of flexibility as it allows the patron to select as many (or as few) formats as they need to narrow their search results. In addition, there is an “Open Access” filter that identifies items ready for consumption without the barrier of payment. In addition, each filter is accompanied by a number, which represents the quantity of records that match the trait for a specified attribute. This serves to assist the research process by allowing individuals to narrow their results to a manageable quantity given the time that they have available to dedicate to their research.

Returning back to the 25,022 bibliographic records retrieved for the “Holocaust AND internment” query, the application of the “Archival material” format to the list of retrievals narrows down the record to 1,725 results. This category can be narrowed even further the use of the subcategory “Downloadable archival material” to 725 results. No archival material is cataloged as open access. Application of these results were organized initially by relevance, but could be ordered by title, author, or date (ascending or descending). Of some interest is the ability to keep track of one’s research. Individuals may create a (free) WorldCat.org account and saving their search results to a list for later reference.

## ArchiveGrid (OCLC)

ArchiveGrid is optimized for mobile screens as demonstrated by its behavior of dropping elements from the webpage layout and resizing elements so that mobile users will only have to scroll (or swipe) up or down to view all of a webpage’s content. ArchiveGrid’s simple search is simple indeed as it consists of a single field for keyword search and does not accommodate any filtering without directly specifying it within the field through specific syntax (e.g. “"mark twain" location:virginia”). ArchiveGrid’s keyword search can use Boolean operators as is indicated by their “How to Search” webpage (Online Computer Library Center, 2020b). Should this simple search be inadequate to the patron’s needs, said patrons cannot utilize an advanced search as ArchiveGrid does not offer an advanced search as an option.

What ArchiveGrid does offer is on option to geographically search for archival institutions and then browse their collections. The homepage offers three different methods to select an archival institution. The first method is to simply pick one from a global map. The second option is to use a field that allows for the retrieval of participating institutions by place or zip code. The third and last method is to select from a comprehensive list of institutions grouped by either US states or global countries.

Keyword search results are presented in two formats: a “List View” and a “Summary View.” The List View resembles a traditional list of retrieved records from a catalog. When conducting a search for the query “Holocaust AND internment,” ArchiveGrid retrieved 256 results. The retrievals of the first page were all relevant to the requested topic and indicated that this reference tool is capable of a high level of precision.

An interesting design choice by the OCLC was to strip the List View of filtering capability to further refine results. It is interesting, because ArchiveGrid has not disregarded the usefulness of filters but has instead grouped the functionality of filtering results into a separate Summary View. The Summary View consists of tabulated lists where each list is analogous to a filter. These lists may be organized either alphabetically by name or by quantity (where higher attributes with a higher number of corresponding retrievals are listed first). Each filter in the Summary View is accompanied by numerical figures that let users know how effective an application of a filter be for narrowing results.

It should be noted that the homepage of ArchiveGrid provides instruction on how to limit a keyword search query retrieval of images, sound recordings, or other online materials increasing the possibility that a researcher may be immediately able access primary resources to advance his or her research. ArchiveGrid does not appear to have a way to save search results.

## Mountain West Digital Library (MWDL)

The union catalog of the Mountain West Digital Library (MWDL) features a simple search that is accessible through its homepage and consists of a single field with no discernable way to add filters to an initial search query. This simple search does support Boolean searching. As an example, searching for “Holocaust AND internment” yielded three finding aids, but swapping the Boolean “AND” for the Boolean “OR” exponentially increased the number of results to 758 retrievals.

Effort was made to make this website mobile-friendly. When narrowing the window of the MDWL homepage, this webpage never required horizontal scrolling. Restricting navigation to upward and downward scrolling to view the content of a webpage made navigation easier for users of mobile devices, but the effort to optimize the website for mobile devices was incomplete. If the window was too narrow, elements of the homepage layout would still extend off the webpage.

One example of this was the website’s horizontal menu bar. This horizontal menu bar also provided access to the website’s advanced search. This advanced search is well structured and allows the user to create multiple fields to define the search query and connected each new field to the existing query using Boolean operators. Following the hallmark of advanced searches, MWDL’s advances search allows the application of multiple filters during the initial search query.

As mentioned earlier the MWDL retrieved three findings aids when searching for “Holocaust AND internment.” All three of these finding aids were from the correct time period, two of them focused mainly on the Japanese internment and were irrelevant indicating low precision. While so few results do not require refinement or assistance in ordering, the MWDL catalog allows researchers to order results in accordance with relevance, title, date, and author. The platform also allows multiple filters to be applied to the results. These filters include resource type, subject, creation date, author/creator, location, language, and digital collection (i.e. institution). Numerical figures accompany the filters to help researchers select appropriate filters to refine their search results. There is no way to save historical search results, but it was possible to view certain digital object, most notably photos.

## Online Archive of California (OAC)

The catalog of the Online Archive of California (OAC) has a simple search that consists of a single field. OAC’s catalog does not have an advanced search, but it does allow browsing of its content by collection and by contributing institution. The website of the OAC has a mobile optimized layout where all content is assessable by scrolling (or swiping) up and down.

The OAC does not appear to respond to Boolean logic. Using either the query “Holocaust AND internment” or the query “Holocaust OR internment” for OAC’s catalog generated the same list of 44 results. The first page of results referenced 20 collections of which approximately three quarters of the result were relevant and achieves an acceptable level of precision. These results can be ordered by the traits of relevance and title. Should retrievals be to many or imprecise, the OAC interface only allows the application of filters based on institution, date (or more specifically decade), and whether the resource is available online. The OAC does not allow a researcher to save search results for later reference.

# Section 3: Evaluate the strengths and weaknesses of the resources

## Discovery (The National Archives of the UK)

One of the advantages of the Discovery tool is the breadth of archival materials that this research instrument grants access to from the 2,500 contributing institutions with the entirety of these accessible records being archival in nature. Another advantage of the Discovery tool of The National Archives is that it is optimized for mobile devices, but this stands in contrast to the poor performance of the website for issues more related to the operation of a website than the maintenance of a reference tool. One of the downfalls is the existence of “dead links.” This includes the “Discovery help” page to assist patrons with familiarizing themselves and the institution’s “About” page. These are not minor webpages and should never lead to a dead link. In addition, this authors personal experience with the tool seems to indicate that the Discovery reference instrument is plagued with service outages and generates inconsistent results.

Another downside of the Discovery tool is the limited ability to refine results through filtering. Despite this, navigation through the Discovery tool of The National Archives is excellent and includes alternative search strategies including browsing by either government department or participatory archival institution. While less of a concern, one of the other downsides to having 2,500 contributing institutions is the variance in the quality (and credibility) of the records of the catalog.

## Finding Aids (Library of Congress)

One weakness of the Finding Aids archival reference instrument is its limited scope. While the scope of materials searchable by this Finding Aids interface is limited to the collections of the Library of Congress, this shortcoming is outweighed by the thoroughness of the finding aids retrieved. These finding aids not only provide full descriptions of the content but also warnings when material is stored offsite and when access restrictions apply. In addition, the copyright of collections is clearly stated in the finding aids for the consideration of the researchers. Another drawback of the archival collections of the Library of Congress is that some of these collections contain classified material and have had restrictions applied to their access.

A minor weakness of the Finding Aids search interface is its reflection of the Caucasoid ethnocentric perspective. As someone of Asian descent, the author appreciates the convenience of being able to filter results from the initial search to retrieve results corresponding to the author’s Asian heritage, but it is noticeable that the simple search lacks for Arab, Native American, or African cultures. Clearly, the Library of Congress’s search for finding aids is influenced by a Caucasoid ethnocentric perspective and depending on the subject being researched could significantly bias search results.

The Finding Aids interface is also in dire need of an update as it lacks Boolean functionality, a comprehensive advanced search function, as well as the ability to refine results through filtering. Combined with the fact that the Finding Aids reference tool provides no indication on which attribute results are ordered by or allow researchers to organize a list of results in accordance with their preference, Finding Aids provides practically no way for patrons to manage their search results. Additionally, the Library of Congress website is also notably unfriendly with narrow “mobile-sized” screens.

## WorldCat (OCLC)

The primary advantage of WorldCat is the diversity of artifacts and resources catalogued in its bibliographic records. If a patron’s research strategy is to cast as wide a net as possible for relevant material on the subject matter, it is imperative that they consult WorldCat catalog to acquire relevant primary sources in the custody of archival institutions.

From an information visualization perspective, the WorldCat interface needs to be reworked to enhance usability. An example of this is the “Advanced search” link found on WorldCat’s homepage.

The black text for the “Advanced search” link blends in with the surrounding boldly darkened, blue background of the initial search interface. There needs to be greater contrast between this link and its immediate background to highlight the location of this important search instrument.

Overall, the vast majority of bibliographic records within the WorldCat union catalog do not correspond to primary sources. This contribution of institutions such as public libraries coupled with the diversity of the types of materials the bibliographic records (many of which are secondary sources) increases the risk of having search results with low precision when specifically seeking out primary sources (i.e. archival materials) relevant to one’s research topic. Applying the filter format filter set to the trait “Archival material” greatly increases the precision of retrieving primary sources.

Filtering is one of the strengths of the interface of the WorldCat catalog. Multiple filters can be applied of which the filter with the greatest significance for the purpose of retrieving primary sources is format filter. Another useful filter is the interface’s “Open Access” filter, which identifies items ready for utilization without the barrier of payment. In general, the bibliographic records accessed through WorldCat lack the depth of a finding aid, but these records offer links to the institutions where more thorough records are likely to be found.

## ArchiveGrid (OCLC)

There are two primary advantages that ArchiveGrid has over other archival references resources. The first advantage is that most of the records that can be accessed through ArchiveGrid are archival in nature. The second is that the resource is highly optimized for mobile devices. Particularly, the use of Bootstrap's front-end framework to consolidate filtering capability in a “Summary View” is an excellent example of how reference tools should be formatted to work best with the touchscreens of mobile devices (OCLC Research, 2020). On the flipside, simplifying functions to increase usability for mobile devices may negatively impact the experience of users of laptops and desktops. Consider the simple search that has been stripped down to a single field. Users of laptops and desktops could benefit from a more robust simple search.

Because ArchiveGrid draws the majority of its records from the WorldCat catalog, it is susceptible to many of the vulnerabilities and weaknesses of WorldCat. This includes the accumulation of standards (and cataloging errors) that may prevent ArchiveGrid from scraping records of archive material from WorldCat and the European/American/Western emphasis of participating archival institutions. In general, the bibliographic records accessed through ArchiveGrid lack the depth of a finding aid, but this is mitigated by emphasizing links to the institutions the maintain custody of the manifestation.

## Mountain West Digital Library (MWDL)

The MWDL does not stand out an any particular way as a source of archival material. The catalog does provide acceptable access to the records of contributing institutions and an attempt has been made to make this reference resource accessible to mobile devices. The contributed records of the institutions are either academic or archival in nature but limited. Depending on the topic, the MWDL may generate useful results. Results can be refined through the application of multiple filters.

## Online Archive of California (OAC)

The Online Archive of California (OAC) provides free public access to detailed descriptions of primary resource collections and is optimized for mobile use. As a reference tool, the OAC catalog falls short. The search engine does not respond to Boolean logic and only allows the limited use of filters to refine results. The one way that OAC stands out is its emphasis of materials that are immediately assessable such as images. The subject coverage of the records of the OAC does appear to be adequate, but that may depend on the topic.

# Conclusion

Nothing truly in the world can replace the efficacy of visiting the archives that physically hold the collection(s) of interest. These reference resources are best used to identify the best physical collections to advance one’s research and form a strategy on how to access said collections. The archival community has a long way to go before a majority of all of their records will be found online; but as time goes on, use of these reference resources will allow researchers the ability to advance their research without having to travel to the controlling archival institution; but when this does come to pass it still likely that the records of collections of archival institutions will be fragmented across multiple catalogs and that a thorough search on a topic will require consultation of multiple archival reverence resources.

# References

gov.uk. (2020, July 4). The National Archives. <https://www.gov.uk/government/organisations/the-national-archives>

Mountain West Digital Library. (2020, July 2). *Mountain West Digital Library*. <https://mwdl.org/>

The National Archives. (2020, July 2). *Discovery*. <http://discovery.nationalarchives.gov.uk/>

Library of Congress. (2020, July 2). *Search finding aids*. <https://findingaids.loc.gov/>

OCLC Research. (2020, July 3). *ArchiveGrid*. <https://www.oclc.org/research/areas/research-collections/archivegrid.html>

Online Archive of California. (2020, July 3) *About OAC*. <http://www.oac.cdlib.org/about/>

Online Archive of California. (2020, July 3). *Welcome to the Online Archive of California*. <http://www.oac.cdlib.org/>

Online Computer Library Center. (2020, July 3). *ArchiveGrid*. <https://researchworks.oclc.org/archivegrid/>

Online Computer Library Center. (2020, July 9). *How to Search*. <https://researchworks.oclc.org/archivegrid/help/>

Online Computer Library Center. (2020, July 3). *What is WorldCat?* <https://www.worldcat.org/whatis/default.jsp>

Online Computer Library Center. (2020, July 2). *WorldCat*. <https://www.worldcat.org/>

Online Computer Library Center. (2020, July 3). *ArchiveGrid*. <https://researchworks.oclc.org/archivegrid/>

University of Florida. (2020, July 4). *Evaluating search engines for and with K-12 students*. <http://plaza.ufl.edu/ginger71/criteria.html>