

Purdue University
Purdue e-Pubs

Charleston Library Conference

Destroyer and Preserver, Hear, Oh Hear! Not All Uncirculated Books Must Chariotest to a Dark Wintry Bed: How We Used the OCLC WorldCat Search API to Inform Our Weeding Decisions with Holdings Data

Geoffrey P. Timms
College of Charleston, timmsgp@gmail.com

Author ORCID Identifier: <https://orcid.org/0000-0003-0970-2618>

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

 Part of the [Collection Development and Management Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Geoffrey P. Timms, "Destroyer and Preserver, Hear, Oh Hear! Not All Uncirculated Books Must Chariotest to a Dark Wintry Bed: How We Used the OCLC WorldCat Search API to Inform Our Weeding Decisions with Holdings Data" (2018). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/https://doi.org/10.5703/1288284317016>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Destroyer and Preserver, Hear, Oh Hear! Not All Uncirculated Books Must Chariotest to a Dark Wintry Bed: How We Used the OCLC WorldCat Search API to Inform Our Weeding Decisions with Holdings Data

Geoffrey P. Timms, College of Charleston, timmsgp@cofc.edu

Abstract

Weeding a specialized collection, such as the geography/oceanography subset of the marine science collection at the Marine Resources Library, requires thinking beyond our own walls and users. To ensure potential access to weeded items through other libraries, as well as the preservation of items unique to our own collection, we sought an efficient and free means to incorporate national holdings data into our decision-making process. The OCLC WorldCat Search API enables bibliographic data, as well as holdings from other libraries, to be obtained easily. With a Python script we obtained holdings data for most of our several thousand oceanography items, making more than 10,000 queries of the API over six minutes. We identified holdings of this collection subset within our five peer libraries, NOAA regional libraries, PASCAL (SC state consortium) libraries, LVIS member libraries, and libraries in the United States to inform (not determine) our weeding decisions.

Introduction

The Marine Resources Library (MRL), located on James Island, South Carolina, is a collaborative initiative of multiple marine science research agencies. College of Charleston (CofC), South Carolina Department of Natural Resources (SCDNR) Marine Resources Research Institute (MRRRI), and the National Oceanic and Atmospheric Administration (NOAA) National Centers for Coastal Ocean Science (NCCOS) Charleston Laboratories with participation from the National Institute of Standards and Technology (NIST) comprise the stakeholders in the library. No comprehensive deselection (weeding) project appears to have been undertaken in any subject area at MRL for many years. We decided to initiate a trial weeding exercise with a subset of the collection during the fall 2018 semester.

The MRL is managed by one CofC librarian and one library technical assistant (LTA). The library functions as a branch library, supported by the infrastructure of the CofC Libraries with financial support from the stakeholders. The current online catalog is Innovative Interfaces Millennium. The circulating collection of approximately 32,000 items includes published books and many technical and project reports from a variety of government agencies. The subject areas of focus are primarily oceanography; marine biology, botany, and zoology; microbiology; and aquaculture and fisheries. NOAA material comprises approximately 15% of the print circulating collection, while CofC and SCDNR own the majority of the collection.

Within the MRL circulating collection, 2,936 unique bibliographic records in the Library of Congress

Geography category are represented by 3,487 print items. Of these items 2,743 are classified as Oceanography. We chose the Geography subset of the collection for the trial weeding project. Of all print Geography monographs, 3,252 (93.3%) have not circulated since 1/1/2007. Of these, 2,606 Oceanography monographs (93.9% of oceanography titles) have not circulated in the same time period. If deselection were to be determined solely by use, the majority of these items would likely be eliminated. However, as a specialized library with a focus on the southwestern Atlantic states, and particularly on South Carolina, the preservation of material specific to these regions is also part of MRL's mission.

When weeding a collection, various concerns may be raised. What if items are eliminated, and subsequently requested by library users? Who could provide us with a copy if we needed to request one with interlibrary loan? How scarce are the items we are considering weeding? What if region-specific historical knowledge is lost from the region? Where does responsibility lie to preserve regional information when no regional strategy is apparent? These concerns inspired us to investigate holdings of other libraries to help inform our weeding decisions.

Method

We identified select groups of libraries of significance to MRL, the holdings of which we chose to evaluate. Lists were made of OCLC symbols for libraries represented in each group ("Libraries Very Interested in Sharing," n.d.; "NOAA Library Network," n.d.;

“PASCAL LibGuides: About PASCAL,” n.d.). With a limited interlibrary loan (ILL) budget, we sought to identify libraries that would be willing to loan weeded items without charge, should the need arise. As a member of the Libraries Very Interested in Sharing (LVIS) resource sharing group, we investigated LVIS libraries’ holdings of these titles. Recognizing the wisdom that exists among specialized libraries, we also considered holdings of four peer libraries to identify titles of potential ongoing importance to the subject area. Locally, we sought to ensure that at least one Partnership Among South Carolina Academic Libraries (PASCAL) library currently had a copy of each title selected for weeding. Items owned by NOAA required additional consideration, as the NOAA library network strives to ensure that at least one copy of NOAA publications is available within the network. Thus, we also evaluated holdings of NOAA libraries. Even if weeding decisions were to be made for NOAA publications, this data would inform decisions about the ultimate disposition of the material. Lastly, we established a count of all domestic libraries holding an item.

OCLC WorldCat Search API

Online Computer Library Center (OCLC) is the logical source of holdings data but without a next-generation library system or funds to invest in analytical services, an efficient way of accessing and using this data seemed beyond reach. The extraction and collation of holdings data of LVIS, PASCAL, NOAA, and peer libraries for each item in our collection analysis would be unfeasible if tackled manually in WorldCat. OCLC, however, has made programmatic access to WorldCat holdings data conveniently accessible through one of its many Application Programming Interfaces, or APIs (“API Explorer,” n.d.).

The WorldCat Search API provides access to bibliographic holdings and location data for libraries

represented in WorldCat. Use of the service is free but is limited either to institutions with ongoing subscriptions to WorldCat Discovery and OCLC Cataloging services, or to users who have sought developer access to the API. Application must also be made for an API key, which is used to identify the user’s affiliation when interacting with the API (“WorldCat Search API | OCLC Developer Network,” n.d.). Requests are limited to 50,000 per 24-hour period, although application may be made to increase this limit. For queries generating a response of many records, the maximum number of records returned in a batch is 100 (“FAQs—WorldCat Search API | OCLC Developer Network,” n.d.). Thus, multiple requests must be made to obtain a complete response to a query that yields a total of more than 100 results.

The API can be used to obtain bibliographic record numbers, bibliographic data, catalog URLs for specific bibliographic records in library catalogs, and holding library locations for specific items. Bibliographic records are identified by searching specific WorldCat fields and bibliographic record data are returned by searching for specific OCLC numbers. Catalog URLs and library locations are returned by searching using various identifiers, including OCLC number, ISBN, and ISSN (“WorldCat Search API” n.d.). For the purpose of this investigation, we obtained a count of libraries with holdings of specific items, by OCLC number. A basic query of the API to obtain holdings for OCLC number 2327674 is accomplished with the URL:

```
http://worldcat.org/webservices
/catalog/content/libraries/2327674
?wskey={your-api-key}
```

Other parameters may be added to influence the data returned (see Table 1).

Table 1. Optional parameters used in OCLC WorldCat API.

Parameter	Description
frbrGrouping=on/off	FRBR Grouping includes related OCLC records for the item and is on by default
startLibrary=n	n=1 or more and is used incrementally for returning results in batches where the total result set is larger than the number defined for maximumLibraries. The default value of n is 1.
maximumLibraries=z	Batch size z, if unspecified, is 10 by default; 100 is the maximum value of z
servicelevel=default/full	Full service level searches data for all WorldCat libraries. The default service level searches data for a subset of WorldCat libraries, those that participate in worldcat.org

It is important to understand the consequence of using the FRBR grouping. While the results returned represent multiple bibliographic records for a given title and author, the most widely held record is used to represent the group as a whole. It should also be noted that multiple material formats including microform, print, electronic, and audiovisual are included (“FAQs—WorldCat Search API | OCLC Developer Network,” n.d.). Additionally, the service level should be set to “full” to ensure that the full extent of WorldCat holdings data is queried.

Data is returned in XML format with each holding library represented in the format:

```
<holding>
  <institutionIdentifier>
    <value>VOD</value>
    <typeOrSource>
      <pointer>http://worldcat.org/registry/institutions/</pointer>
    </typeOrSource>
  </institutionIdentifier>
  <physicalLocation>Old Dominion University</physicalLocation>
  <physicalAddress>
    <text>Norfolk, VA 23529 United States</text>
  </physicalAddress>
  <electronicAddress>
    <text>http://www.worldcat.org/wcpa/oclc/2327674?page=frame&url=http%3A%2F%2Fodu-primo.hosted.exlibrisgroup.com%2Fopenurl%2F01ODU%2F01ODU_SERVICES%3Fisbn%3D9780471025405%26checksum%3Db469feela94fd75d900fe5a7e0310c5f&title=Old+Dominion+University&linktype=opac&detail=VOD%3AOld+Dominion+University%3AAcademic+Library&app=wcapi&id=OCPSB-
    </text>
  </electronicAddress>
  <holdingSimple>
    <copiesSummary>
      <copiesCount>1</copiesCount>
    </copiesSummary>
```

```
</holdingSimple>
```

```
</holding>
```

Python Script

The process of reading the many OCLC numbers of the MRL Geography collection, querying the API, synthesizing and processing data, and writing to an Excel sheet was handled with a Python (<https://python.org>) script. To obtain the input data (including OCLC number) for each item, data including circulation, accession date, and OCLC number were generated and exported by Millennium’s Create Lists function into a comma separated value (CSV) file. This was then converted to Microsoft Excel (XLSX) format. In a catalog with SQL access, this data could be obtained with direct interaction with the catalog by a script. The Python script read the XLSX file and queried the WorldCat Search API for every OCLC number found therein. The volume of queries needed to process all items necessitated use of a multithreaded process instead of a linear one. A total of 25 threads simultaneously worked on OCLC numbers and interacted with the API. Linearly, the process would have taken an hour or more, but by using 25 threads, the processing was accomplished within six minutes, making a total of 16,029 queries of the API.

As each holding institution was returned for each OCLC number, the Python script cumulatively enumerated matches of the institution by OCLC symbol with the group lists, while ensuring that MRL’s own two symbols were not counted. This necessitated extracting specific fields from the hierarchy of XML data for each institution to compare to the group lists. The OCLC symbol was used for LVIS, PASCAL, NOAA, and peer library list comparisons, while the physical address was used to identify domestic libraries by matching the phrase “United States” within the text:

```
<holding>
  <institutionIdentifier>
    <value>VOD</value>
  </institutionIdentifier>
  <physicalLocation>Old Dominion University</physicalLocation>
  <physicalAddress>
    <text>Norfolk, VA 23529 United States</text>
  </physicalAddress>
```

OCLCNum	BibNum	CallNum	Barcode	Location						
020	100	245	246	247	257	260	490	600	655	700
#Peer Holdings	#NOAA Holdings	#PASCAL Holdings	#LVIS Holdings	#Domestic Holdings	#Total Holdings					
#CofC Total Checkouts	#CofC Total Renewals	#CofC LVCirc	#CofC OutDate	#CofC DueDate	#CofC OverDueDate	#CofC LastOutDate	#CofC Created Date			

Figure 1. Spreadsheet column headings in spreadsheet created by Python script (identification data, bibliographic data, holdings data, and circulation data).

Once all data were obtained and enumerated, an Excel spreadsheet was written by the script. For each item, bibliographic and circulation data from the original CSV file and holdings data calculated by the script were synthesized in a spreadsheet row (see Figure 1).

We then established criteria to highlight potential candidates for deselection or retention. Patterns of holdings were identified using Excel’s Conditional Formatting feature and spreadsheet rows were colored to flag items according to pattern characteristics. For an item to be flagged as “Scarce,” we determined that no more than one peer library and five or fewer domestic libraries must own the item. This does not indicate that the item has scarcity value. The item may simply be obsolete and has already been eliminated from most library collections. The flag draws our attention to the item with a focus on its potential rarity or obsolescence.

Conversely, for an item to be flagged for potential elimination, we determined that either one or two

peers and at least one PASCAL library, three LVIS libraries, and 10 domestic libraries must hold the item. If three or four peers held the item, we did not flag it for deselection but flagged it as “significant peer holdings,” potentially indicating ongoing importance.

Results

While not reflective of other library holdings, 810 items (23.2% of all items [$n = 3,487$]) meeting any one of three criteria were automatically excluded from consideration for deselection (see Table 2):

- Items added to the collection since 1/1/2007
- Items circulated since 1/1/2007
- Items in which the title (245 field) or publisher/distributor (260 field) matched words or phrases from a list representing:
 - Geographical terms, for example, *Charleston, South Carolina, South Atlantic Bight*, and so on

Table 2. Counts of items in the MRL Geography/Oceanography collection ($n = 3,487$) flagged by criteria in order of precedence.

Criterion (in Order of Precedence)	Recommendation	Number of Items	Percentage of Items
Added since 1/1/2007	Keep	412	11.8
Circulated since 1/1/2007	Keep	184	5.3
Local interest (by word matching in title and publication information)	Keep	214	6.1
Subtotal	Keep	810	23.2
Scarcity/obsolescence	Primary Review	215	6.2
Potential deselect	Primary Review	348	10.0
Significant peer holdings	Primary Review	765	21.9
No criteria matched	Secondary Review	1,349	38.7
Totals		3,487	100

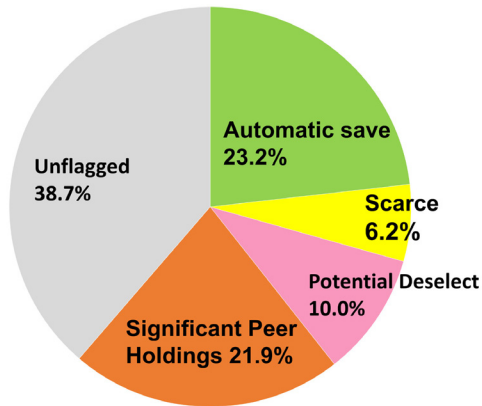


Figure 2. MRL geography collection by flagged criteria.

- Institution names/acronyms, for example, *College of Charleston*, *SCDNR*, *NOAA*, and so forth

Some items met multiple criteria for flagging so an order of precedence was established, as demonstrated in Table 2. The proportions of items flagged are presented in Figure 2. Items not flagged (38.7%) fell between the Scarce and Potential Deselect categories. That is, they are not scarce enough to be flagged as scarce, yet not abundant enough among other libraries to be flagged for potential deselection.

Informed Decisions Based on Holdings Data

After working through each of the three flagged categories, “Scarce,” “Potential Deselect,” and “Significant Peer Holdings” to make retention/deselection decisions, we investigated the categorical differences in the percentages of items identified for retention/deselection (see Figure 3). To test for statistical

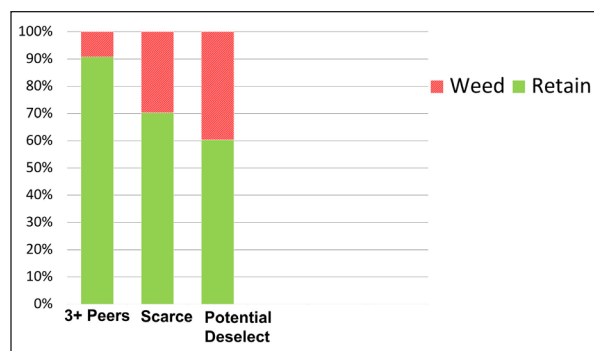


Figure 3. Proportions of MRL geography collection selected for retention/deselection in each of three flagged criteria.

significance, we used a two-tailed Chi square test. The result was a statistically significant dependence between flag category and percentage of items selected for retention/weeding, at the 1% level of significance, χ^2 (2 d.f., $N = 1,328$) = 150.2421, $p < 0.1$.

Discussion

The most notable observation is that approximately 90% of items flagged “Significant Peer Holdings” were retained at MRL. This suggests that peer analysis may be particularly useful in supporting retention/deselection decisions. Items flagged as “Scarce” were weeded at a higher than expected rate of approximately 30%, suggesting that some obsolescence had become a consequence of not evaluating the collection regularly.

The result of 38.7% of the Geography collection remaining unflagged is not ideal. The criteria for flagging are flexible, however. Indeed, the flagging “triggers” for one collection subcategory may not be appropriate for another. The number of unflagged items could be reduced by adjusting the flagging criteria. For example, the definition of scarcity might be broadened to include up to nine domestic copies instead of five. Or the criteria for being flagged as sufficiently abundant for potential weeding could be adjusted by eliminating the single PASCAL holding requirement, thereby including more titles.

Complicating factors include series in which individual items were flagged differently from each other. In some cases, series were not held in their entirety but only select volumes had been added to the collection. In complete series, however, a decision must be made to either fragment the series, or retain or eliminate it in its entirety. This, naturally, depends upon the nature and significance of the series in the specific research context and any interrelatedness between volumes.

While the unflagged portion of the collection has not yet been evaluated for deselection, we can hypothesize that, due to these items falling between the “Scarce” and “Potential Deselect” categories, which were weeded at approximately 30% and 40% respectively, we might weed an estimated 35% of these items. Considering the 23.2% of the collection granted automatic retention, the net weeding rate for the Geography subset of the print collection can be anticipated to be approximately 22%. This does not seem unreasonable for a print collection that may not have been evaluated for a long time.

It is important to remember that such a strategy is not intended to automate collection decisions. While it could be used to do so, and perhaps might if a substantial pressure called for urgent removal of a large number of volumes to liberate space, it would be at the expense of the subtleties and nuance that define the importance of a specific item in a specific context.

Conclusion

The strategy of flagging monographs according to specific criteria is helpful for informed collection

Note

The Python script used in this project may be found at https://schr.ws/hosted_files/2018charlestonconference/18/TIMMS_OCLC-Search-API-Python-Script.py

References

API Explorer. n.d. Accessed April 23, 2018. <https://platform.worldcat.org/api-explorer/apis>

FAQs—WorldCat Search API | OCLC Developer Network. n.d. OCLC. Accessed April 23, 2018. <https://www.oclc.org/developer/develop/web-services/worldcat-search-api/faqs.en.html>

Libraries Very Interested in Sharing (LVIS) listing of members arranged alphabetically by OCLC symbol. n.d. Accessed April 23, 2018. <http://www.cyberdriveillinois.com/departments/library/libraries/OCLC/pdfs/lvisbysymbol.pdf>

NOAA Library Network | NOAA Central Library. n.d. Accessed April 23, 2018. <https://library.noaa.gov/About/Library-Network>

PASCAL LibGuides: About PASCAL: Member institutions. n.d. Accessed April 23, 2018. <http://pascal.libguides.com/c.php?g=454361&p=5455972>

WorldCat Search API. n.d. Accessed April 23, 2018. <https://platform.worldcat.org/api-explorer/apis/wcapi>

WorldCat Search API | OCLC Developer Network. n.d. OCLC. Accessed April 23, 2018. <https://www.oclc.org/developer/develop/web-services/worldcat-search-api.en.html>

decisions. It is, however, just one part of the overall process. Each monograph must still be considered from various perspectives, including relevance, validity, obsolescence, and in relation to other items held on the same subject, including e-books. The provision of holdings data of other libraries adds a sense of consequence to the decision. OCLC has facilitated this investigation by its provision of a useful API, the application of which can assist with informing collection management decisions with holdings data.