



# **Leveraging WorldCat Web Services for Collection Analysis: An Introduction**

Presentation to the ALCTS Collection Evaluation and  
Assessment Interest Group

ALA Annual Conference, San Francisco, CA

June 28, 2015

Richard Entlich (rge1@cornell.edu)

Collection Analyst Librarian

Cornell University

# Data sources and tools for collection analysis

- Some tasks are best suited to local catalog data
  - Characterizing ones own collection
    - Statistical reporting for internal use or for ARL, ACRL, NCES, etc.
      - Absolute holdings (whole collection or subsets)
      - Total circulation
      - Changes in any of the above over time
    - Internal duplication
    - Profile or internal comparison (by language, subject, etc.)

# Data sources and tools for collection analysis

- Some require a look at the collective collection (usually as represented by WorldCat)
  - Overlap/duplication between individual libraries or within consortia
  - Identification of rare or unique items (regionally, by country, or worldwide).
  - Determination of relative collection strength (by subject, language, etc.)

# Some options for querying, downloading, and/or analyzing WC data

- WorldShare Collection Evaluation tool
- FirstSearch in 'Expert' mode
- Connexion Client or Browser
- z39.50
- WorldCat APIs

# WorldCat APIs

- API = Application Programming Interface
- WorldCat APIs are web service APIs (or just web services) developed by OCLC
  - Submit a specially formatted URL via HTTP, get back a response via HTTP
- Primarily designed to allow discovery systems and other software tools to access WorldCat bibliographic and holdings information
- No fees for non-commercial use by cataloging members of OCLC

# Exploring the WC APIs

- Best place to start

<https://www.oclc.org/developer/develop/web-services.en.html>

- Lists all available APIs and for each
  - A link to the documentation
  - A link to the WorldCat API Explorer

## WorldCat Search API

Status: Production    Sandbox access: Yes

Developer-level Access to WorldCat - for bibliographic holdings and location data

Try the API »

[Read the documentation »](#)

API access to: [WorldCat](#)

# Exploring a specific WorldCat API

## WorldCat Search API

**Status:** Production

**Sandbox access:** Yes

Developer-level Access to WorldCat - for bibliographic holdings and location data

[Try the API »](#)

[Read the documentation »](#)

# Other important WorldCat API links

- How to request a WSkey
  - <https://www.oclc.org/developer/develop/authentication/how-to-request-a-wskey.en.html>
- API FAQs
  - <https://www.oclc.org/developer/develop/web-services/worldcat-search-api/faqs.en.html>
- Service levels
  - [https://www.oclc.org/content/developer/worldwide/en\\_us/develop/web-services/worldcat-search-api/service-levels.html](https://www.oclc.org/content/developer/worldwide/en_us/develop/web-services/worldcat-search-api/service-levels.html)

# Using the APIs

- No real user interface
  - Experiment with using API Explorer, or any web browser (once you have a WSkey)
- Develop a small program or script to submit requests and parse or store results
- Test on small batches first

# Sample script

Query submission is handled by a Unix/Linux shell script, an outline of which is shown below. The two lines highlighted in red handle sending the individual queries to the API and parsing the response to extract the holdings count.

```
#!/bin/bash
```

```
querysrc=$1
```

```
echo -e "QueryNum\tHoldingsCount" >> ${querysrc}_out.txt
```

```
while read line
```

```
do
```

```
  querynum=`echo $line|cut -d"~" -f1`
```

```
  query=`echo $line|cut -d"~" -f2`
```

```
  response=`wget -O - -q $query`
```

```
  response_count=`echo "${response}"|grep numberOfRecords|cut -c 18-|cut -d"<" -f1`
```

```
  echo -e "${querynum}\t${response_count}" >> ${querysrc}_out.txt
```

```
done < ${querysrc}
```

# Parts of a WC Search API query (library catalog location mode)

- **Base URL**
- **Parameters**
  - maximum libraries
  - FRBR grouping
  - service level
- **Authentication key**

# Sample library location query

- [http://www.worldcat.org/webservices/catalog/content/libraries/11784417?maximumLibraries=100&frbrGrouping=off&servicelevel=full&wskey=\[redacted\]](http://www.worldcat.org/webservices/catalog/content/libraries/11784417?maximumLibraries=100&frbrGrouping=off&servicelevel=full&wskey=[redacted])

# Real life example #1: Find unique and rare items in a collection

- Create set of queries using OCLC numbers for desired items
- Feed queries to search API in batches of 50,000 per day (about two weeks to complete)
- Extract relevant holdings counts from results

# Sample results (after processing)

BIB_ID	TitleBrief	LOC	Hldgs_nonFRBR	Hldgs_FRBR
7550721	Pengantar sejarah kebudayaan	ech	1	30
7550724	Hommage to Prof. Dr. Sumitro [	ech	1	6
7551081	Khrōngkān suksā phūa kānpraki	ech	1	1
7551217	Lǎeng bōrānnakhadī lǎe bōrānsa	ech	1	1
7551246	Pat̄chai thī mī phon tō kānmai p	ech	1	1
7551446	Kartini : sebuah biografi /	ech	1	41
7551454	Manusia dan kebudayaan di Ind	ech	1	41
7551494	Salah asuhan /	ech	1	55
7551504	Pergolakan : roman /	ech	1	37
7551558	Katak hendak jadi lembu /	ech	1	8
7551565	Sengsara membawa nikmat /	ech	1	32
7551606	Bentrokan dalam asrama /	ech	1	7
7551620	How to grow rattan /	ech	1	1

# Example #2: Identify publishers by ISBN using xISBN API

ISBN prefix	Publisher	Region	City
087590	American Geophysical Union	English speaking area	Washington, D.C.
087591	Ann Arbor-Humphrey Science Publishers	English speaking area	Ann Arbor
087593	Trail-R-Club of America	English speaking area	Beverly Hills, Calif.
087595	Oregon Historical Society Press	English speaking area	Portland, Or.
087596	Rodale Press	English speaking area	Emmaus, Pa.
087597	Crescendo Pub. Co.	English speaking area	Boston
087598	Pierpont Morgan Library	English speaking area	New York
087599	S.G. Phillips	English speaking area	New York
087601	Carnation Press	English speaking area	State College, Pa.

# Example #3: Decode OCLC symbols using Registry API

OCLCsym	Name	City	State	Country	type	reg_index
DE-604	BibliotheksVerbund Bayern (BVB)	München		DE	5	ISIL
DEBBG	BibliotheksVerbund Bayern (BVB)	München		DE	5	oclcSymbol
DIA	Defense Intelligence Agency	Washington	US-DC	US	8	oclcSymbol
DIBAM	Dibam Biblioteca Nacional de Chile	Santiago		CL	6	oclcSymbol
DJB	Joint Bank Fund Library	Washington	US-DC	US	5	oclcSymbol
DLC-R	Library of Congress, RCCD CJK	Washington	US-DC	US	6	marcOrgCode
DLM	University of Delaware Library	Newark	US-DE	US	1	oclcSymbol
DMM	Dayton Metro Library	Dayton	US-OH	US	3	oclcSymbol

- 
- 1 ARL Library
  - 2 Academic Library
  - 3 Public Library
  - 4 School Library
  - 5 Other
  - 6 State or National Library
  - 7 Library Networks or Processing Center
  - 8 Government Library
  - 11 Vendor
  - 13 Museum
  - 14 Archives
  - 15 Regional Service Provider
  - 16 Correctional Facility Library
  - 17 Dental Library
  - 18 Hospital Library
  - 19 Law Library
  - 20 Medical Library
  - 21 Music Library
  - 22 Nursing Library
  - 23 Veterinary Library
  - 24 Corporate Library
  - 25 Special Library
  - 26 Health Library
  - 29 Group
  - 30 Consortium
  - 31 Collection
  - 32 Program
  - 33 Project
  - 999 Unknown

# The bottom line

- WorldCat APIs
  - Are flexible
  - Are free to use for most purposes by OCLC cataloging members
  - Are an excellent way to analyze single or multiple collections within WorldCat and to enhance the analysis of local collections
  - Require modest technical skills to utilize
  - Require close attention to detail. Expect some trial and error before starting large-scale projects