Using WorldCat API to Identify Unique and Distinctive Materials within Collections

Philip White¹ & Juleah Swanson²
University of Colorado Boulder Libraries

The Problem

While libraries may have a sense of what is rare, unique, and distinctive within their special collections, identifying what sets a library apart from its peers within the circulating collection may be undervalued or unknown.

Efforts to identify such pockets of distinction have previously involved costly tools (Genoni & Wright, 2011), or labor-intensive work (Barnes, Kelly, and Kerwin, 2010).

Our research sought to determine if the WorldCat API can be used as an efficient and cost-effective solution.

What is an API?

• Application Programming Interface.
• Method of data exchange, usually in XML or JSON format.
• The World Wide Web relies on APIs for a variety of tasks, like embedding other sites’ data in webpages (e.g., Google Maps, National Weather Service data).

How to Work with WorldCat APIs

OCLC provides access to many APIs with a wide array of uses including metadata creation, searching bibliographic information, finding nearest copies of library materials, and more.

To work with WorldCat APIs you will need to read the documentation carefully to understand how to construct queries and what parameters are available to search. You will also need an API key. API documentation and access information is available at oclc.org/developer.

Option 1: OpenRefine

• OpenRefine is free and has the ability to send requests to URLs, making it easy to send batch requests to an API from a spreadsheet-like interface.
• OpenRefine is easy to use and ideal if your data set is less than a thousand.
• Parsing JSON response data is relatively straightforward, but it is more difficult to work with XML response data.

Option 2: Programming Language

• Using a scripting language (e.g., Python, PHP, JavaScript, etc.) provides maximum flexibility when querying an API.
• Creative throttling and breaks can be programmed—useful for larger amounts of queries.
• There is a learning curve to writing code, and it takes practice to learn how to parse XML or JSON data returned by an API.

Case Study: Gifts in Kind

• In 2018, we conducted a retrospective analysis of gift-in-kind items donated to CU Boulder over the past 10 years to develop a data-driven decision process for accepting future gifts.
• We queried 16,000+ gifted items to determine how many copies were available in WorldCat and nearest copies to CU Boulder.
• Using the WorldCat Search API LibraryLocations search method, we collected the 50 nearest locations of gift items by querying each item’s OCLC number.

• We discovered that gift items contribute unique strengths in the collection in Asian language materials, with many of the nearest copies of these items available nowhere else in North America—confirming usefulness of gifts-in-kind for building non-English collections.
• Most gift items, however, were also available within 200 miles of CU Boulder (approximately the area of our regional borrowing network), indicating usefulness of method as prescreen of future gifts-in-kind.
• This work resulted in an automated process for using WorldCat Search API to prescreen potential gifts-in-kind based on items’ OCLC number or ISBN to determine uniqueness of those items and proximity of copies to our library.

Implications

• OCLC’s WorldCat APIs put a wealth of data at our fingertips.
• Making use of the LibraryLocations method can lead to new discoveries of strengths, uniqueness, and value in our library collections.
• A data science approach to collection assessment using APIs and other methods brings greater efficiencies and the ability to conduct analyses at larger scales than ever before.
• Leveraging WorldCat data may soon bring deeper insights to collection assessment and new directions in research and analysis.
• Potential future research should seek to combine other data points like interlibrary loan statistics, citation analyses, and consortial collection data with WorldCat data to perform large, wholistic assessment of library monograph or serial use.
