Michelle Rigual
University of New Mexico Law Library

OCLC: WorldCat Collection Analysis Service and the Desert States Law Library Consortium

In November, 2005, the eight libraries of the Desert States Law Library Consortium (University of New Mexico, University of Arizona, Arizona State University, University of Colorado, University of Denver, University of Utah, Brigham Young, and University of Nevada Las Vegas) licensed OCLC's WorldCat Collection Analysis Service (WCA). Each library was curious to see what the WCA would tell them about their collection's strengths and weaknesses compared to other libraries. As a consortium, we sought to identify our collection gaps by benchmarking the combined consortial holdings against large individual libraries and other consortia. Further, we hoped to maximize each library's unique holdings within the consortium, limiting the amount of money and effort spent on purchasing, processing and preserving duplicate materials.

WCA is designed to provide objective information concerning the makeup of a library's collection, including subject coverage, age of materials, languages of publication and material types and audience level. Multi-institution comparisons provide detail and summary views of how two or more collections overlap or differently specialize, subdivided by subject areas, languages of publication, material types and dates of publication. Desert States librarians who worked with the service considered it a wonderful concept and promising tool but discovered that it had significant limitations.

Topical analysis within WCA is based on the OCLC Conspectus which includes 32 broad *divisions* such as 'Law.' The Conspectus subdivides further by approximately 500 *categories* such as 'United States, Federal' and 'Public International Law.' At its narrowest are approximately 7,000 *subjects* such as 'Intellectual Property' and 'Public Land Law.' The information is displayed in a chart, clicking on the numbers in the chart allows you to drill down to the title level and see a list of short records. This list can be exported to Excel for graphing or other manipulation.

Using WCA, we were able to see both the number and percent of unique holdings of each consortial library broken down by subject and time. Similarly, we were able to view how much our holdings overlapped with each other. It is easy to use WCA to obtain some interesting, but mostly superficial, data; it is quite a bit more complicated to get meaningful information. Most librarians did not learn to use the service very well because it takes a lot of time and initiative to turn your searches into information that you can act upon. Even after substantial time working with the training materials and the service itself, our most ambitious users still did not feel that they were able to make good use of the data export features. Those investing in this product should budget additional funds and time for training in order to get the most out of it. At the time we purchased there were few active training options, but OCLC and regional providers now offer web seminars and workshops.

Data Categorization Problems

To me, the most frustrating aspect of the WCA is that the Conspectus *divisions*, *categories*, and *subjects* are based on LC call number and divided up in ways that are not useful. Some *subjects* are far too narrow and should be combined with others. For example, no library has more than a handful of titles in the *subjects* of 'Indian Lands,' 'Computers–Law & Legislation,' 'Engineering Law,' 'Manufacturing,' 'Public Property– Public Restraints on Private Property,' and 'Roads and Bridges.' On the other hand, some *subjects* are so broad that meaningful comparisons cannot be made. The most notorious *subject* is 'Law–United States, Federal' (not to be confused with the *category* of the same name). Some topics are not broken out, making them very hard to find. Examples include: 'Supreme Court,' which is hidden in the *subject* of 'Courts, Procedure.' Similarly, 'Human Rights' is safely nestled away from view in the *subject* of 'Comparative Law, International Uniform Law.'

A fellow librarian in the Desert States Consortium has suggested that a good deal of the existing subject organization difficulty could be alleviated by adding a feature that allows the searcher to customize subjects according to call number range. His example is that he would like to be able to search just for "military law" contained in call numbers KF 7201-7749. Currently, this is not possible.

My biggest wish for WCA is different, but related. I would like the service to provide the ability to compare topics that are meaningful in the context of my law school's areas of expertise. Unfortunately, these are topics that do not match up with the Conspectus. For instance, I would like to assess our collection's relative strength in the area of Indian Law. University of New Mexico's Indian collection contains holdings that fall throughout the LC classification, from BL to Z but WCA includes only KF 5660-5662 in Indian Lands and KF 8201-8228 in Indian Law. The Indian Law materials outside this narrow definition cannot be included in comparisons. However, a search on subject headings (FirstSearch's descriptors) would neatly allow comparative analysis of these materials.

Data Integrity Problems

A more serious shortcoming of WCA is the debatable integrity of the data it produces. OCLC provides a WCA users' list that was recently the site of a thoughtful, well-documented report from a librarian who calls into question the trustworthiness of the numbers provided by the service. In five tests over a period of months, the exported records included numerous blank entries. When compared against their corresponding display records on FirstSearch, some of these matched up to entries labeled "No Title Available" while others matched up to perfectly visible records. The first few times these blank records were reported to OCLC, the library received a reply email stating simply that the problem was resolved, though it was not. A similar report was made to OCLC in mid-September, OCLC indicated that it would respond within a few days but still has not done so. The blank entries also create problems down the road. Because Excel cannot process the exported blanks, a manual search and removal of blanks is necessary before any data manipulation can occur.

I also experienced blank entries but was not nearly so conscientious about tracking down the source of the blanks or reporting them to OCLC so cannot comment on the record from which the blanks originate or on OCLC's response to problem reports.

An obvious limitation of the data is that only items entered into OCLC can be compared, and it is up to each library individually to ensure that the records of its holdings are as complete, clean, and up to date as possible. Of course, there are all sorts of holdings that are not reported to OCLC. One law library encountered problems because, at the behest of their University, their policy was to add a holdings symbol to an OCLC record only if it was unique to their collection. If the title was already in the University's collection, they only added an item record to the OPAC. More commonly, serial holdings reporting is only possible at the title level and exempts large numbers of volumes from possible consideration even though subscriptions represent the majority of the library expenditures.

Data Manipulation Problems

The final source of problems with WCA is the display and export of information. In its current format, only numerical reports are available. This is fine for purely numerical comparisons but more detail is available if you link from the number in the chart to the list of short titles. Unfortunately, access to this detail is lost when running comparisons. The best improvement OCLC could make in this area would be to provide a means to display title-by-title comparisons without complex and time-consuming manipulation of data exported to large and unwieldy spreadsheets.

For example, when running comparisons such as overlap and uniqueness you cannot easily determine specifically which titles are owned by whom. Your own library's records are marked but to find out who holds the records that are not marked, or other libraries holding the same title, you have to open a longer record then enter that information, title by title, into the title list that you will have already exported to Excel. An improvement would be to have the system produce a title list that indicates who holds what. A less acceptable solution, but still an improvement, would be to indicate in the short title list which libraries

hold each title.

Similarly, there is no mechanism for comparing conspectuses. It can be done by merging two or more Excel spreadsheets, making sure to create some indicator of which library has which title, such as bolding, underlining, or using a different font. In the merge, you will be able to determine how many titles are owned in common by having multiple entries of the same title (each in a different font which indicates the source library). Even more manipulation is required to produce a desiderata list - you need to delete all titles you own from the comparison spreadsheet, which you produced as above.

Practical Applications

Despite the problems listed above, WCA has been a useful tool to numerous libraries. It is very easy to use the system to get a general idea of where your library falls in comparison to others. In a recent ALL-SIS list posting one library reported that it produces data the deans and law faculty can understand and pointed to another library that used it to compare holdings to the five Google libraries and discovered that something like 35% of the collection was unique (and would not be on Google). This caught the attention of the university administrators and helped counter perceptions that everything will be digitized.

OCLC makes frequent enhancements to the system in response to the needs identified by subscribers. Many of the features that the Desert States Consortium wished for when we first subscribed over a year ago have since been added. A significant recent enhancement is creation of a list of libraries who have given blanket permission for their collections to be used in one-to-one comparisons and no longer have to be contacted directly. OCLC also recently created two predefined peer groups for benchmarking purposes that are particularly useful for law libraries: (1) Top 10 Law Schools and (2) Google 9 (the nine institutions currently participating in the Google Book Project). An interesting recent enhancement is the addition of more subject breakdowns for the Conspectus *category* 'Law, General' and narrowing down the *subject* 'Comparative Law, International Uniform Law.' I applaud the recognition that the law topics are insufficient but also suspect that this means that they are more interested in tinkering with the topics rather than upgrading the search capabilities.

Since collections are unlikely to change in a significant manner relative to peer institutions over short periods of time, this tool will likely be useful only infrequently and there is no need to purchase it on a continuing basis. The subscribing institution would be best served to run this type of analysis every few years or in preparation for reaccreditation.