Using the WorldCat Collection Analysis Tool: Experiences from
The University of Kansas Libraries

Amalia Monroe-Gulick, almonroe@ku.edu
Lea Currie, lcurrie@ku.edu
Author Final Draft

Introduction
The KU Libraries is a founding member of the Association of Research Libraries (ARL) with 4,235,542 total cataloged items. KU has a sizable special collections library which holds manuscripts, rare books, maps, photographs, and ephemera. KU also holds many unique collections in international area studies, including its East Asian; Spanish, Portuguese, and Latin American; Slavic; and African Studies collections, which combined total more than one-third of the KU Libraries’ collections. KU is a longtime member and contributor to OCLC.
The WorldCat Collection Analysis (WCA) tool is an instrument for evaluating one library’s collection against the holdings in the entire WorldCat database and with selected OCLC members. When the WCA tool was first introduced, OCLC representatives were invited to visit KU and demonstrate its capabilities. After reading negative reviews and hearing that OCLC was promising to make improvements in the future, KU Libraries decided to wait until the problems were ameliorated before subscribing to the tool. KU Libraries’ Associate Dean of Technical Services and the Assistant Dean of Collections and Scholar Services had been paying a attention to the WCA, and when they thought the proper improvements had been made, they announced that KU would subscribe for a year. The KU Libraries began a subscription to the WCA in March of 2009. The subscription included three designated OCLC sites, including KU General Libraries, the Kenneth Spencer Research Library, which houses rare books, manuscripts, and other special collections(6), and the KU Medical Center’s Dykes Library.

KU Libraries has a long history of collecting data to analyze its collections. Circulation and interlibrary loan statistics are collected, as are the numbers of faculty, students, and graduation rates for each KU academic discipline. This data is used to inform allocation decisions each year. The WCA comparison data provides additional information for identifying strengths and weaknesses of the collections for allocation purposes. The results of this analysis are being used to better understand the KU collections and to realign development priorities for the foreseeable future. Librarians at KU will also use the comparison information to identify potential subject areas for collaborative collection development with other libraries in the state of Kansas, most notably Kansas State University.

**Literature Review**

The available literature on the WCA addresses the functionality of the product and case studies of individual libraries. There is very little literature discussing a large institution implementing the tool across collections. The leaders of the of the implementation committee reviewed articles that addressed WCA utilization, strengths, and weaknesses. All subject librarians participating in the project were encouraged to read the articles to develop a better understanding of the process and what types of analyses could be conducted. The literature also informed the KU Libraries’ implementation committee’s documentation and training sessions.

The data generated from WCA can be used in several different types of analyses to help achieve the goals of individual institutions. Sneary (2006), an OCLC Creative Services Analyst, suggests that WorldCat Collection Analysis (WCA) can help ensure that current collection development is in alignment with the “strategic goals of the university.” With the data collected from the WCA, libraries can determine if certain subject areas are heavily collected.

Comparing local holdings with other institution’s collections is the most common approach to utilizing the WCA but there are other goals as well. Intner (2003) suggests that the knowledge gained from evaluations, specifically comparison projects, allows for informed justifications when discussing collections at the university level or requesting additional library funding. Discovering that a library collection is smaller or older than its peers’ collections is a simple and understandable way to communicate to university administration that additional library resources are needed. At Saint Leo’s
University, Henry, et. al. (2008) used the WCA as a collection evaluation in response to the university’s Institutional Effectiveness Plan. One component of their project was the comparison of their holdings to holdings at similar institutions. As a secondary result from gathering WCA data, lists of titles currently owned by SLU were generated for weeding purposes. St. Leo’s librarians plan to conduct a second analysis in a few years to provide the library with additional data and a longitudinal study of the collection. Spires (2006) used the WCA to run comparisons with a defined peer group of libraries in the Consortium of Academic and Research Libraries (CARL) in Illinois and Bradley University. The librarians at Bradley University found the WCA helpful in comparing collection size, age of collection, collection overlap, and collection uniqueness with libraries in CARL. At Colorado State University, Culbertson and Wilde (2009) used the WCA and other metrics to assess the library’s support of doctoral programs in twelve disciplines. The purpose of the study was to support a request for additional funding to the university administration. Librarians worked with teaching faculty to identify comparable institutions. Using the WCA, the monograph collections were evaluated, but evaluating journal collections was challenging because of the lack of accurate serials records in WorldCat. For example, WorldCat does not include records for electronic serials in aggregator databases. Librarians incorporated faculty input, accreditation criteria, Journal Citation Reports, statistics from CSU’s open URL server, Local Journal Utilization Reports, and interlibrary loan statistics to develop a list of essential core journals.

One of the most important aspects of the literature review is the identification of the problems and drawbacks of the WCA, and ways to circumvent them. Negrucci (2008) identified some of the weaknesses of the WCA. The two biggest problems identified were the over- or under-reporting of unique titles because of multiple editions and formats, and the rigidity of the OCLC subject conspectus. The author was not able to apply the WCA Functional Requirements for Bibliographic Records (FRBR) filter to specific comparison groups, only to the general WorldCat Analysis, because FRBR was not available for individual subjects at the time. When FRBR is not working, it is easy to misinterpret the data, because it appears there are quite a few overlapping titles, when they are actually different editions or formats. At North Carolina State University, Orcutt and Powell (2006) used the WCA to run comparisons against groups of research libraries in their consortium. They found that, implicit to its design, the WCA works better when running comparisons to a single institution and not as well with multiple institutional comparisons. NCSU librarians also realized that WCA data is only updated quarterly and the tool could not accommodate sampling methods (8). Obtaining workable data required gathering information within restrictions inherent to the tool. They were forced to exclude all formats other than monographs because of inconsistent reporting to OCLC. They also had to exclude titles with imprint dates within the most recent two years in order to account for differences in cataloging and acquisition rates across their consortium. In many cases, the WCA subject categories were inflexible and often less than helpful.

Some academic libraries chose to evaluate specific collections in their libraries. Beals (2007) used the brief test assessment model and the WCA to evaluate zoology collections at three universities. The brief test assessment was developed based on the Research Libraries Group (RLG) Conspectus model as an attempt to quantify collection strengths. Like Orcutt and Powell (2006), Beals had similar difficulties when using the WCA. The author encountered problems running analyses
for subjects with multiple call number ranges. The author argued that combining both the brief test assessment and the WCA provides a more complete picture of a collection since they each fulfill a different role in collection assessment and provide a more complete picture of a collection. Cox and Gushrowski (2008) used the WCA to determine the publication date span and median publication date for a weeding project in a dental library. Library staff were able to determine the quantity and age of titles in specific subject areas and compare them to the collection as a whole.

A less documented but useful way to utilize the product is the use of interlibrary loan statistics. Way (2009) used the WCA in a unique manner. Grand Valley State University Libraries used the interlibrary loan (ILL) analysis in the WCA to generate a list of titles that had been borrowed. A review of the list seemed to indicate a large number of titles would likely be appropriate for their collection. As a result, the library decided to pursue the development of a patron-initiated purchase program via ILL to enhance the library’s collection.

Implementation
Following the decision to purchase a year-long subscription to OCLC’s WorldCat Collection Analysis (WCA) tool, the Associate Dean of Technical Services and the Assistant Dean of Collections and Scholar Services appointed an implementation committee that was led by the Head of Collection Development and a Social Sciences Librarian. The implementation committee reviewed and identified institutional goals and priorities related to the project, and taught subject librarians in a classroom setting and one-on-one sessions to use the WCA. The committee wanted to establish and facilitate an efficient process because all subject librarians with collection development responsibilities (12) were required to complete reports using WCA.

The committee compiled a document outlining the overall goals of the project, a timeline, and a list of limitations within the WCA product. The priorities, goals, and timeline for KU Libraries were as follows:

1. Between May 1, 2009 and March 31, 2010, compare our collections with ARL peers and other groups or individual libraries as identified by subject liaisons (see Appendix A) (10).
   a. Identify strengths and weaknesses that characterize our collections generally.
   b. Identify unique collections or unique material.
   c. Identify gaps in our collections, overlaps, and duplication.
   d. Identify resources needed to support new and expanding programs and to support formal accreditations.
   e. Identify possibilities for collaborative collection development within a region or consortium.

2. By June 1, 2010, create a formal collection development plan resulting from the analyses conducted
   a. Recommend specific areas for increased budget allocations.
   b. Recommend specific areas where collections budget will be cut.
   c. Report on significant weaknesses and subject areas where collections may be stronger than necessary (due to the goals and mission of the institution).
3. During Fiscal Year 2011 and beyond, coordinate discussion with other regional libraries regarding cooperative collection development opportunities.

Before the WCA project was implemented, fallacies and inconsistencies of the WCA product and records in OCLC and problems unique to KU were identified to better inform the collection analyses. Those included:

- WCA uses the WorldCat accession number as the unique identifier of a bibliographic record for matching purposes, and not the title, author, or edition statements. Negrucci (2008) describes how (36) the same edition of a work may have multiple bibliographic records in WorldCat, resulting in a comparative analysis that over-reports uniqueness and underreports overlap. OCLC has attempted to mitigate the over-reporting of uniqueness by providing the option of applying a Functional Requirements for Bibliographic Records (FRBR)algorithm so that the same titles with different formats or editions are compared .(34)
- Orcutt and Powell (2006) (36) reported difficulty in obtaining reliable samples from different subject areas. Libraries have been forced to use the limit function to exclude non-book formats and recent publication dates to obtain workable data sets for their core collection analysis.
- Orcutt and Powell (2006) complained about (36) the rigid conspectus structure of WCA. The lack of functionality to conduct a user-defined search of LC subjects and classifications limits the detailed view needed for an in-depth collection analysis.
- The currency of WCA data is also an issue. The tool relies on an extract from WorldCat taken once per quarter. For more current imprints, the infrequency of updates precludes tenable comparisons. WCA is not the tool for comparing recent acquisitions.
- The WCA does not permit sorting by language.

Problems identified that are unique to KU collections included:

- Many of KU’s electronic resources, including e-record sets like Early English Books Online, , 18th Century Collections Online, ACLS (American Council of Learned Societies) (15) humanities e-book, etc. are not in WorldCat.
- A few of KU’s e-journal titles will be found in WorldCat; however, the KU Catalog’s record for the e-journals does not contain an OCLC number. (15) (34-omitted Voyager)
- Most U.S. government documents and international documents are not in WorldCat.
- Many of the East Asian collection titles are not in WorldCat. (15,16)
- Approximately 50,000 maps in the Map Library are not in WorldCat.
- Many microforms are not in WorldCat.
- Most of the sound recordings from the KU Archive of Recorded Sound are not in WorldCat.
- Over 3 million photographs from the KU Libraries are not found in WorldCat.
- Some manuscript collections from Kenneth Spencer Research Library are not in WorldCat, particularly those in Special Collections.
- OCLC records without a Dewey or LC call number will be designated as “Unknown Classification” in WorldCat Analysis. This includes most Kenneth Spencer Research Library
special collections materials, theses and dissertations, some microforms, some sound recordings, as well as catalog records that were contributed to OCLC but which lacked a Dewey or LC call number in the master record.

- Some records from Kenneth Spencer Research Library appear in the main KU General Libraries WorldCat Collection Analysis account. They may, or may not, also appear in the Spencer Research Library WorldCat Collection Analysis account because an item record may be connected to both OCLC records

In addition, the committee broadly outlined the process that subject librarians would take to achieve the priorities and goals stated above. Subject librarians were directed to:

1. Conduct a basic WorldCat comparison.
2. Run reports comparing collections to our ARL Peers. (see Appendix B)(10)
3. Choose other appropriate pre-identified groups (Big 12, Regents Libraries (four year public colleges and universities in Kansas), etc.)(see Appendix B)(17,10)
4. Consult with teaching faculty to compile lists of peer institutions appropriate for the unique disciplines to conduct additional analyses.
5. Choose from authoritative lists, such as Best Books for College Libraries.
6. Consider different types of tools for unique collections – all collections are not created equal!
7. Share ideas for analysis with other subject librarians while working through the process.

As the participants began working on the project, it became clear that more guidance and training were necessary. Therefore, additional documentation was created throughout the project in response to questions from subject librarians. At their request, a report template was created to help analyze the data collected (see Appendix A). The committee agreed that a template would provide a certain amount of consistency in the data reported.

**Training and User Support**

In addition to the documentation, three different kinds of support were offered to further assist subject librarians with using WCA: workshops, user groups, and desk-side coaching. All three provided basic information on the product, how to use the product, and useful Excel features. In addition to the initial documentation previously mentioned, a wiki was created that included login information, lists of the libraries in each comparison group, a bibliography of articles discussing WCA projects and product reviews, and later included completed reports.

Introductory workshops designed specifically for KU Libraries’ needs were conducted by three members of the WCA implementation committee and a library technology instructor, who trained the subject librarians to use advanced options in Excel. During the workshops a practical demonstration was presented on using the WCA. Subject librarians were shown how to choose comparison groups, limits, and display options.

The monthly WCA Users Group meetings were held for any interested WCA project participants. These meetings were designed for users to come and work on reports, ask questions, and share ideas.
The meetings also provided users with a venue to express their concerns and the problems they encountered when running reports. Later in the year, in this venue, subject librarians shared their completed reports and the WCA features they used to analyze their own collections. These demonstrations created a context and provided support for subject librarians who had been struggling with the project. Observing how other librarians utilized features such as exporting and copying charts, and filtering and sorting in Excel, and working with multiple editions was extremely helpful.

One of the biggest challenges discussed in the WCA Users Group meetings was the ability to understand exactly what participants were trying to learn from the analysis. It was not possible to establish strict guidelines for everyone because there were too many variations among collections, and different flaws within the product that affected specific subject collections differently. The social science disciplines, for example, were affected less by the FRBR issue than the humanities. Also, science librarians found the project difficult since their collections are mainly serial dependent. Most of their serials come from aggregator databases and do not appear as KU holdings in WorldCat.

Desk-side coaching was also provided for running WCA reports and Excel training. These one-on-one sessions were helpful in assisting users to better understand what they were attempting to extract from the tool. It was also more effective to train people on different features of Excel one-on-one because of the varying needs and skill-levels of individual users. KU subject librarians greatly improved their Excel knowledge and skills while working on this project.

**Results**

WCA reports were run comparing the entire KU Libraries’ collections to several peer groups, including KU’s ARL peers, Big 12 peers, and Kansas State University (see Appendix B), the latter with whom KU has collaborated on several collection development projects. These broad comparisons gave a clear indication that KU has relatively strong (14) collections based on the high number of titles that overlapped with our (14) aspirational peers. Compared to the ARL peers, KU has 77.48% overlap, while KU has 70.86% overlap with Big 12 peers. KU’s overlap with Kansas State was only 35.12%, which is not surprising since KU has a larger collection and the two schools have many differing academic programs.

Forty-five individual WCA reports for specific subjects were submitted by subject librarians. Overall, the results were positive and can be used for several different collection management activities, including retrospective collecting, approval plan adjustments, changing future firm orders, and augmenting collection development policy statements. For example, the Political Science collection analysis found an overall homogeneity between its holdings and the comparison peers (Big 12, ARL Peers, and Political Science Peers). This is a positive result as KU does not strive to have a unique collection in this area, but there were no significant weaknesses when compared with its peers. All subject areas were strong except for the U classification, but KU does not have a military science program so this was not an area identified for adjustment. The Map collection was described in the subject librarian’s report as “on par with the very largest research libraries.” The list of titles not held by KU will be used for retrospective collecting in the areas of history of cartography and environmental sciences. The United States History collection was also found to be strong in the E-F classifications. The only significant weakness observed was in the area of Pacific States and Territories, but this was not cause for concern as this is not a widely studied area at KU. The Journalism collection was found to be
strong, except in the areas that focus on reporting specific issues, e.g., the Iraq War. These findings resulted in ordering titles not previously held by KU, and adjustments in future firm order priorities. Some collections were found to be weak overall. The subject librarian for the Women’s Studies collection stated that the collection “contains the minimum of what should be in a research library collection, and is behind both our ARL Peers and our Women’s Studies Peers.” These results are significant because a PhD program in Women’s Studies has recently been approved for Women’s Studies and the current collection may not support advanced research in this area. Interlibrary loan and circulation data, in conjunction with the list of titles not held by KU, will be used to expand the collection. The African Studies collection analysis also found that the KU collection is far behind its peers in collecting African materials.

Even with the majority of analyses finding positive results, most subject librarians reported varying levels of dissatisfaction because they felt they did not gain significant insight into their collections from the WCA comparisons; rather, the results reaffirmed what they already assumed about their collections or the data was inaccurate. Even though there was dissatisfaction with the product, KU Libraries now has documentation about their collections based on the reports of individual librarians.

- Thirty-nine reports found that KU collections were comparable to aspirational peer libraries based on the number of titles that overlapped with titles in peer libraries and unique titles held by KU.
- Weaknesses in collections were identified in 44% of the reports submitted by librarians based upon low overlap percentages. However, 29% noted that the titles not owned by KU would not be typically selected for our collections because they fall outside of KU’s collecting scope (i.e. commercial publishers, professional literature, subject content, textbooks).
- 42% of the reports reported that books held by peers would be purchased as a result of the peer analyses.
- 16% reported that they would make adjustments to the approval plan in their subject areas. Two librarians stated that they would begin collaborative collection development projects with other libraries based on their findings in the WCA reports.

There were many negative comments in the final reports regarding the difficulty of using the WCA and the lack of usable data collected.

- 16% of librarians reported that they found the WCA reports of no use, because their subject areas are serials dependent.
- 44% of librarians reported problems with the WCA data, including too much duplication due to multiple editions, items not cataloged in WorldCat, and data from WorldCat not being uniform.
- 11% of librarians thought that the WCA was too difficult to use.
- 7% of librarians were disappointed that a particular language was not available in the WCA.
Conclusion

After reflecting on the entire WCA project at KU Libraries, some elements of the implementation and process were considered to have worked well, and others need improvement. When the project was completed, the implementation committee met to discuss the successes and failures of the project. After reading all of the librarians’ reports, the committee members agreed that they would have conducted the process differently had they been aware of the challenges beforehand. Initially the implementation committee thought it would be beneficial for all subject librarians to be involved in the project, but later agreed that a smaller, core group of librarians should have run all of the reports so that there was more consistency in the data that was collected. Even though a template for reporting data and analyzing the reports was designed and shared with the subject librarians, a few librarians did not use it, and even those that did often supplied inconsistent data, making it difficult to analyze the big picture.

One of the benefits of WCA is the ability to download lists of titles that the library does not own into Excel. These lists have several uses, including providing titles for retrospective collection development purchases to fill in gaps and providing information for future comparisons. However, if KU were to subscribe to the WCA again, all of the title lists would be run by a smaller group of participants and stored in a centralized repository for future use.

An additional benefit of the project emerged because librarians were simultaneously writing new subject collection development policies. The WCA results frequently reaffirmed claims made about collections over a number of decades, including the claim that KU has research-level collections that cover all major areas in its academic disciplines. These results then informed the newly written policies. Both the WCA reports and the collection development policies will provide documentation for future collection managers that will enhance their understanding of the history of the collections, and the reasons behind collecting decisions.

One of the major set-backs experienced by KU Libraries during the year-long WCA project was technical difficulties. Subject librarians were consistently reporting problems related to the inability to download reports because WCA would “time out.” The problems were repeatedly reported to OCLC, but they were not resolved immediately. The “time out” issue not only prevented librarians from running reports when they had scheduled time to do so, but also the problems created a significant amount of frustration with the project as a whole. Although many of the technical problems KU encountered throughout the year are not documented elsewhere in the literature, we would recommend that OCLC make the debugging of these problems a priority. As noted by Negrucci (2008) and Orcutt and Powell (2006), the rigid subject conspectus persists and is a problem that OCLC also needs to address. KU also agrees with Orcutt and Powell (2006) and Beals (2008) (36) that OCLC needs to update the WCA on a more frequent basis. A monthly update would be a significant improvement and improvements to FRBR would also make the product much more useful.

The WCA would be a much more useful product if analyzing all of the serials we have access to could be achieved. KU subscribes to many of the largest aggregator databases, but the serials we access through these databases do not display as owned by KU in WorldCat. WorldCat could provide a
knowledgebase of serials titles lists that are accessible from aggregators to add to the serials cataloged in OCLC. Making the subject conspectus less rigid would also make the WCA far more functional than it is now.

Other libraries would be advised to consider the overall benefits and drawbacks of the WCA before implementing any library-wide projects. Libraries will want to identify their goals, and potential problems that might be caused when running WCA reports due to inadequate records in individual library catalogs. They will want to ensure that all those participating understand what the WCA can and cannot do, as well as the technological components of the product. KU found that standardizing the data collected as much as possible is advantageous for all involved in any analysis project and will produce stronger results. Furthermore, assigning a smaller group of librarians to conduct the analyses would also ensure that the results are more consistent among collections, as well as maximizing expertise and minimizing staff time devoted to the project.

Overall, KU Libraries found some value in subscribing to the WorldCat Collection Analysis tool even though there were challenges because of problems with the tool. Many subject librarians concluded that their specific collections are strong. Librarians also identified publishers with whom they were previously unfamiliar. They identified sub-areas of specific collections that are strong, wrote more informed collection development policies, and improved Excel skills. The collection documentation (title lists of monographs not owned by KU and collection policies) that was produced is one of the primary benefits resulting from the project. KU may subscribe to WCA in the future if the current drawbacks to the product are addressed by OCLC. However, if KU Libraries subscribes to WCA in the future, changes will be made in the implementation process.

Collecting data over time to compare the KU collection with its peers would produce useful information to track the changing nature of the KU Libraries collections. Gathering longitudinal data would also assist KU Libraries in developing an awareness of how collections are changing at a national level. Tracking these changes will help KU Libraries’ understand research library is and will be in these quickly changing times.
Appendix A Report Template

WorldCat Collection Analysis Tool Report–FY10

Subject Librarian

Subject Area

Methodology: Explanation of Reports

1. Comparison groups used (e.g., ARL peers, WorldCat, custom peers, standard list)

2. Limits utilized (if applicable):
   a. Subject
   b. Years
   c. Call number range
   d. Holding count
   e. Language
   f. Format

3. How were reports sorted:
   a. Call number
   b. Publisher
   c. Language
   d. Holding count

Results/analysis (please provide supporting data where applicable):

1. Please provide a general description of your collection.

2. Strengths of collection (e.g., unique items, completeness of collections).

3. Weaknesses of collection (e.g., missing call number ranges, publishers, years).

4. Application of results (how will you use the data collected to make decisions about the collection; e.g., approval plan adjustment, budgetary requests, retrospective collecting, accreditation purposes, collection development policies).

5. Difficulties (fallacies that impeded results; e.g., multiple editions, serial-dependent, records not in OCLC, alternative formats)

Additional feedback:

1. Was this analysis of your collections useful? Why not?

2. How successful were you in getting the information you wanted from the WCA reports?

3. What information were you hoping to find from the WCA?

4. How should the WCA tool be changed to make it more useful?
Appendix B KU Libraries Peer Comparison Groups

<table>
<thead>
<tr>
<th>KU Libraries’ ARL Peers</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Colorado, Boulder</td>
</tr>
<tr>
<td>University of Iowa</td>
</tr>
<tr>
<td>University of North Carolina, Chapel Hill</td>
</tr>
<tr>
<td>University of Oklahoma</td>
</tr>
<tr>
<td>University of Oregon</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KU Libraries’ One-to One Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>KU Spencer</td>
</tr>
<tr>
<td>Kansas State Univ</td>
</tr>
<tr>
<td>Univ of Michigan</td>
</tr>
<tr>
<td>Univ of Missouri</td>
</tr>
<tr>
<td>Univ of Nebraska, Lincoln</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KU Libraries Big 12 Peers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baylor</td>
</tr>
<tr>
<td>Iowa State</td>
</tr>
<tr>
<td>University of Missouri</td>
</tr>
<tr>
<td>Oklahoma State</td>
</tr>
<tr>
<td>Texas A&amp;M</td>
</tr>
<tr>
<td>Texas Tech</td>
</tr>
<tr>
<td>University of Colorado</td>
</tr>
<tr>
<td>University of Nebraska</td>
</tr>
<tr>
<td>University of Oklahoma</td>
</tr>
<tr>
<td>University of Texas</td>
</tr>
</tbody>
</table>
References


