November 2006

People Profile: Tracey Powell

Editor

Follow this and additional works at: https://docs.lib.purdue.edu/atg

Recommended Citation

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.
subcategories generally exhibited a similar rate of 10-20% under representation in WCA. As we discuss in the next section, there is reason to believe that WorldCat's title numbers in other ways actually over represent titles in some listings. Further exacerbating this difference between WCA numbers and catalog records.

Currency of WCA data is also an issue. The tool relies on an extract from WorldCat taken once per quarter (theoretically, at least; the second quarter 2006 update never occurred). For more current imprints, the frequency of updates precludes tenable comparisons. For example, in Fall 2005, we checked titles of 2005 imprints against our own OPAC and those of the other institutions and found in every case (across comparisons and subject areas) that the majority of unique titles reported in WCA were apparently no longer unique. For large and/or retrospective analyses, currency may be moot, as even tens of thousands of records likely would not radically alter the composition of a large research library's holdings (10,000 out of 1 million records equates to 1%), and few libraries would routinely catalog exceedingly large numbers of older volumes.

Other limitations include sorting by language within an institution (for example, NCSU's Urdu collections cannot be analyzed with this tool), and which subject divisions may be selected for closer inspection. Both of these limitations hamper the tool's ability to assess diverse and interdisciplinary collections. More (or, better yet, user-delimited) categories would better address the needs of research-level institutions.

**Sampling issues**

One of the largest issues we encountered in using WCA is its inability to accommodate sampling methods. A project examining the existence and size of core collections shared by ARL-ranked institutions presented the greatest difficulties for WCA. For this analysis, we identified for comparison the five top- and five bottom-ranked ARL institutions within each of the three categories of "land grant," "public," and "private" universities. We could easily review the size of shared collections using the numerical data by subject area readily accessible in WCA, but encountered difficulties when we attempted to sample titles in six selected subject areas.

Several things called into question the reliability of the data we had gathered. Obtaining workable data required that we gather data within certain limits. First, we needed to limit our analysis to books alone, as other format types presented difficulties, including especially high rates of multiple OCLC accession numbers for like items and apparently lower rates of institutional reporting to OCLC of video and other formats. We also had to exclude titles with imprint dates within the last two years in order to account for differences in cataloging and acquisition rates across institutions. WCA is therefore not the tool for comparing very recent acquisitions.

A much larger problem that detailed our core collections research project consists of a "bug" in WCA that appears in many title lists. In one instance, the title list repeated the same titles with the same OCLC numbers at multiple places in the list. Fortunately, our sampling caught this occurrence and alerted us to the problem, but sampling techniques might not reveal such repetition. Others later reported the same issue via OCLC's listserv for WCA users (OCLC-WCA@OCLC.ORG).

Combined, these problems limit the usefulness of WCA as a research tool. The accuracy of the data and its ability to handle increasingly complex cataloging records and collection decisions in a sophisticated, user-friendly way are necessary if OCLC truly wishes the tool to provide the type of information which could result in more informed decision-making on the part of collection managers.

**Embedded Self-Centeredness**

The implicit institution-centeredness of WCA rears its head in several ways. The tool allows one to create a group of two to five institutions for comparison, or to compare against any one of several predefined groups. Each of our studies utilized individually-selected institutions except for the comparison with the "Google Five," which exists in WCA as a predefined group. The tool only allows for comparison of the user's institution to the aggregate "collection" of the group of institutions selected (or mere analysis of that aggregate). While this might be a sufficient default for simple comparison work, it does not allow for more complex comparisons. For example, in examining the holdings of various ARL-ranked institutions we sought to include NCSU as one of the schools in our group of five so that we could directly examine shared holdings among all of the schools, rather than a comparison between NCSU and the other four. WCA does not allow this flexibility, which necessitated some awkward calculations in order to determine overall overlap and uniqueness in ways readily provided for groups of other institutions. This need to identify one's home institution as the basis for comparison suggests a competitive rather than cooperative stance as the implied norm for WCA comparisons. It also hampers the potential of conducting more general collections research.

Analyzing one's own institution within a context of cooperation, such as our own TRLN consortium, requires a wholly separate (and additionally paid) "group" subscription to WCA. In group accounts, the self-focus (this time, on the group) prevents even comparison against peer consortia, let alone analyses of cooperation across consortia. Unless "group" subscriptions were to include all relevant institutions, one could not analyze a collection's place within, for example, a geographic region, which is but one alternative context which selectors might find compelling.

**Not Learning from the Past**

Most of the problems and unfulfilled potential continued on page 48