All [global] data is local: How academic libraries are enabling discovery and access for institutional data collections

Shawn W. Nicholson
Asst. Director for Digital Information

Terrence B. Bennett
Business / Economics Librarian

| TCNJ THE COLLEGE OF NEW JERSEY |
| MICHIGAN STATE UNIVERSITY |

Not on our mark!

Despite our careful data-collecting strategy, we did not get set for a smooth sprint. Help us leap over the hurdles so that we can go to the finish line.

This pilot analysis attempts to reflect the information-seeking behavior of typical users of research libraries; however, the coding process (and to some extent, the data derived from the coding) raises some concerns about the reliability of this approach.

In her article for the 2010 Encyclopedia of Library and Information Sciences, “Information Searching and Search Models,” Iris Xie delves into the nuanced distinctions of user search tactics and search strategies, all of which are informed by each user’s other knowledge, learned behaviors, usage patterns, and expectations. Can our “first scan, then browse, then search” approach to uncovering the presence of information, specifically institutional-produced non-numeric data collections, suffice as a stand-in for users’ search behaviors?

New means of accessibility and discovery for research data enhances the potential for research progress. They allow the verification of study results and the reuse of data in new contexts. Academic libraries’ role in this environment continues to unfold. Considerations of domain data sharing practices, confidentiality, open access and privacy remain important. Research data whole consumption could be conceived as non-rivalrous might prove to be the entree for libraries to remove barriers and provide unfettered discovery, access, use and analysis protocols and infrastructure. However, our initial attempts to discern if libraries are embracing best practices in this arena have been beset by unexpected hurdles. Can we leap over these hurdles in order to complete a meaningful inquiry?

The library’s role in resource discovery continues to be reimaged. Libraries leverage catalogs, digital repositories, wayfinding across websites, and vended discovery layers to meet demand. While there’s been a renewed focus on metadata and documentation to foster findability, use and reuse, we hypothesize that local research data collections are generally not showcased — or even easily discoverable — from the library. We also attempt to examine how (or if) academic libraries are positioning their local holdings — particularly digital texts, image files, audio archives, and other non-numeric collections — as research datasets (rather than as artifacts of limited local and/or historical interest).

Libraries enable discovery of local non-numeric collections. Are they research data?

Definitional struggles abound. The coding scheme revealed multiple interpretations that led to additional questions. This was particularly true when coding for digital text data collections, where the intention was not to classify every collection of scanned books as “research data,” but to focus specifically on instances of linguistic corpora. However, questions persisted around scaling and potential use; for example, could a large enough set of digitized books be divorced from their original context and reprocessed for linguistic analysis? If so, what criteria are needed to satisfy this transformation?

Methodology

- Generate a sample from ARL and RLUK
- Develop a coding sheet
- Dichotomous schema to reflect information-seeking behavior
- Coder bias is inevitable
- Qualitative intercoder reliability test
- Execute test
- Apply coding schema to sample

Selected Results

Research data is not featured on the library’s home page: Fully 90% of the sample [n=18] offered no signal to users that the library provides access to numeric datasets or other research data.

Browsing for numeric data is generally successful: Browsing the library website for numeric data terms was rarely unsuccessful [n=3]. However, a browse that yielded institutionally-created numeric datasets was almost equally rare [n=4].

The advantage of browse vs. search is difficult to discern: Current trends in website infrastructure make it difficult to assess the library’s role in promoting data discoverability. Nearly one-third of the libraries in the sample [n=7] did not offer a library-specific search function; instead, only an institution-wide search was possible.

Browsing for numeric data

<table>
<thead>
<tr>
<th>1. No mention of datasets, images, etc on library homepage</th>
<th>2. Library homepage prominently features digital datasets search function</th>
<th>3. Library homepage provides link to research data collections</th>
<th>4. Library homepage promotes data discovery (e.g., through a blog)</th>
</tr>
</thead>
</table>

Details from the master coding sheet.