

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



The library's role in resource discovery continues to be reimagined. 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).



METHODOLOGY

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

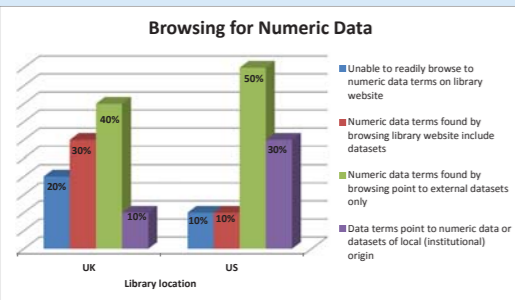
HOME PAGE	SOUND	IMAGES
1.....No mention of datasets, images, etc. on library homepage	4.....Unable to readily browse to voice (sound) collections on library website	8.....Could not readily browse to image collections, and a search of the library website for image collections was unsuccessful
1.1.....Library homepage mentions numeric data or datasets	4.1.....Voice (sound) collections found by browsing library website; no other source of voice (sound) collections was identified	8.1.....Could not readily browse to image collections, but a search of the library website for image collections was successful
1.11.....Numeric data or datasets mentioned on library homepage are not browsable	4.2.....Some voice (sound) collections are of local (institutional) origin	8.2.....Image collections were found by browsing the library website, but a search of the library website for image collections was unsuccessful
1.12.....Numeric data or datasets mentioned on library homepage include links to external sources		8.3.....Image collections were found by browsing the library website, but a search of the library website for image collections was successful
1.2.....Library homepage mentions image collections		8.4.....No library website search is available

Details from the master coding sheet.

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.

Libraries enable discovery of local non-numeric collections.

Are they research data?

Sound Collections

At 65% of library websites [n=13], it was easy to browse to sound/voice recordings (12 of these were local collections).

Digital Text

At 50% of library websites [n=10], digital text was discovered by browsing (9 of these were local collections)

Image Collections

At 90% of libraries [n=18], image collections were readily browsable.

TCNJ THE COLLEGE OF NEW JERSEY

Terrence B. Bennett
Business / Economics Librarian

MICHIGAN STATE UNIVERSITY

Shawn W. Nicholson
Asst. Director for Digital Information

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.

How should libraries respond?

Does a library (or its larger institution) need to describe local collections of images, sound files, digital text, or other objects or artifacts as "data" or "research data" in order to promote such collections' potential for use in support of research?

Do our preliminary findings clearly support any specific, best-practice recommendations for libraries to better position, and showcase, these holdings?

What are [re]searchers' intentions?

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?

What are data collections?

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 repurposed for linguistic analysis? If so, what criteria are needed to satisfy this transformation?

DISCUSSION

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 whose consumption could be conceived as non-rivalrous might prove to be the entrée 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?