

Against the Grain

Volume 25 | Issue 6

Article 43

2013

Curating Collective Collections--Data for Collection Assessment at a More Granular Level: ICON As An Example

Amy Wood

Center for Research Libraries, Wood@crl.edu

Sam Demas

Sam Demas Collaborative Consulting, sdemas03@gmail.com

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

 Part of the [Library and Information Science Commons](#)

Recommended Citation

Wood, Amy and Demas, Sam (2013) "Curating Collective Collections--Data for Collection Assessment at a More Granular Level: ICON As An Example," *Against the Grain*: Vol. 25: Iss. 6, Article 43.

DOI: <https://doi.org/10.7771/2380-176X.7438>

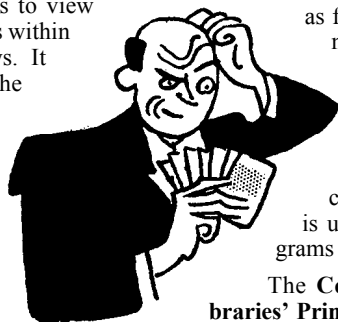
This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

It is fascinating to see that the general pattern of usage remains the same, both before and after the introduction of a resource discovery system. Subject librarians have reported that having access to usage data via UStat is very helpful. One librarian used it to present additional information about library support and resources for their subject department's regular program review of its curricula. They also used the data effectively during this year's journal subscription review. One librarian said, "UStat gives us the type of data we need to show faculty how much or how little a journal is used and to provide cost-benefit analysis for specific journals." Following changes in aggregator database coverage of a journal's articles, this same librarian reinstated a subscription to it based on usage data easily accessible to her in UStat.

Many of **Milner's** subject librarians used UStat to support difficult cancellation decisions during this year's journal review. They could easily determine low- and zero-use titles across all of their subscriptions. One of the more useful features is the ability to see use across platforms, when a title is available through multiple resources. This enables librarians to determine if usage in a database is adequate, making a direct subscription less vital. Viewing usage across platforms or resources informs librarians of how and where patrons seek out and use materials. See *Figure 3*.

Although UStat is useful, it is not perfect. At **Milner Library**, demand for usage statistics for streaming media resources is growing, joining the need for better electronic book usage data. UStat does not yet handle either type of resource. Project COUNTER 4, a major revision, goes into effect in December 2013, with reports already appearing in that format. Although report files can be manipulated into a format that will load into UStat, the chances of data corruption increases as more changes are done by a person. If UStat development lags behind **Milner Library's** needs, the library may have to look for a different usage statistics system. However, the process currently in place for gathering statistics should transfer easily as care was taken to create a process not reliant on a specific system.

Electronic resource usage statistics are no longer a mystery for librarians at **Milner Library**. Data is readily available and accessible at the point of need. Using a system like UStat has enabled librarians to view their collections and the materials within those collections in different ways. It is worth repeating though, that the information in UStat is just numbers. Trends can be identified, but it is still up to the individual librarian to assign value and meaning. Nevertheless, it is now easier to determine the value and meaning of those numbers. 🌱



Curating Collective Collections — Data for Collection Assessment at a More Granular Level: ICON As An Example

by **Amy Wood** (Director of Technical Services, Center for Research Libraries) <wood@crl.edu>

Column Editor: **Sam Demas** (College Librarian Emeritus, Carleton College & Principal, Sam Demas Collaborative Consulting) <sdemas03@gmail.com>

In a previous column, **Richard Fyffe** reflected on risks libraries faced as they move toward collective collections or interconnected print collections. What we think we know about the holdings of other collections has grave impact on print retention decisions made locally. As curatorial trends run their course, decisions made without sufficient information can have long lasting, if not permanent, repercussions. The level of detailed information about print holdings required by the community in a rush to clear space is minimal — often title level is sufficient or existing holdings statements in local catalogs disclosing what we think is on the shelf. This article invites us as a community to reflect on how collective collections and other coordinated curatorial efforts can be improved by investing in better data at the risk of slowing short-term gains. It also offers a glimpse at **ICON**, a tool for comparing holdings of newspapers, as an example of emerging best practices in data for collection assessment.

Ithaca set a stake in the ground a few years ago with its **What to Withdraw** tool. The tool uses detailed information about journals at the issue level to support local collection development and management decisions. The foundation is there for applying quantitative methods of rating value, as in the ratio of images and text. A drawback to the tool is the limited data set it supports — **JSTOR** titles from two dark archives that do not allow access to the print.

The **Association of Southeastern Research Libraries' (ASERL)** has developed a noteworthy tool — **Journal Retention and Needs Listing (JRNL)**. **JRNL** was developed for participating institutions to track journal retention commitments between the **Association of Southeastern Research Libraries (ASERL)** and **Florida State University System (SUS)** partners. It is a tool for individual libraries to track their data and a data repository to aggregate a program's data. One drawback is that data is accepted as formatted and is therefore not always consistently expressed. This makes truly automated aggregation of the data impossible. The bigger drawback, for the wider community, is that the tool is unavailable for other programs to use.

The Center for Research Libraries' Print Archives Preservation

Registry (PAPR) takes the realm of tools for aggregating data about print holdings beyond the local. **PAPR** has roughly 50,000 records for 35,000 titles committed for archiving by twelve separate programs (49 combined institutions within those programs). One of **PAPR's** most useful features is the title, holdings and gap reports by program or search results, which users can download. **PAPR** also has a service allowing users to compare a list of publications from their collections with those in **PAPR**. Developing a means to aggregate issue-level data in an automated way is in the works. **PAPR's** big drawback is similar to **JRNL's**, the free-text data fields allow inconsistency of expression of holdings.

As print archiving or shared collection programs mature, **OCLC** continues to improve the tools it offers for collection analysis, and commercial products are being developed as well.

Despite the growing number of tools, the community lacks focus on creating better data. The challenge is getting data in a format which allows us to use it to make better informed decisions. **CRL** addressed that challenge in improving its **ICON** database to better assess newspaper collections. Of utmost concern was providing a tool for the automated comparison of local library collections with the electronic holdings of commercial newspaper databases.

Projects or programs like the United States Newspaper Program,¹ the National Digital Newspaper Program² and the Florida Digital Newspaper Library³ are important examples of how much coordinated efforts can accomplish with regard to collecting and exposing library newspaper holdings. What they lack is information about the holdings of commercial databases and the tools to compare and assess these collections against libraries' print and microform collections. To make decisions about preserving their own collections and purchasing commercial databases, librarians need to know exact holdings down to the issue level and to have at their disposal tools that automate comparisons at that level between collections.

The ICON Database

CRL's primary goals in developing the **ICON** database were:

- to increase the amount and quality of information on newspapers that are and have been published in the U.S. and abroad;
- to increase transparency of commercially produced collections of digital

continued on page 75

newspapers, particularly information about their source for producing digital versions;

- to enhance librarians' ability to make informed selection, retention and preservation reformatting decisions.

To do this, the **ICON** database focuses on exposing information at the individual issue level. This is important because the same title offered in digital format by various aggregators may not include the same issues or a microfilm set may not offer complete overlap with the print holdings that are being considered for deaccessioning.

Unlike a library catalog or resource sharing system, the focus of **ICON** is not on resource discovery. Instead, success is measured in the finding and use of collection material, in performing collection analysis of holdings at the issue level and in comparing issue-level library holdings data with titles offered by commercial aggregators and publishers of electronic newspaper databases and packages.

Data Sources and Project Partners

Two of the project's participants — the **American Antiquarian Society (AAS)** and **NewsBank, Inc.** — provided issue level data. **AAS** provided metadata for over 10,000 titles and 1.9 million associated issues. **AAS** had developed their database to store issue-level holdings for data management and for patron viewing prior to the need to share data for this project. The **AAS** database, called Clarence,⁴ is a Web-accessible database of **AAS** newspaper holdings. The **AAS** catalog not only provided important issue-level information, it also provided a workable data model to use as a foundation for **CRL's** **ICON** database.

Readex, a division of **NewsBank, Inc.**, contributed metadata for 152 titles and 253,163 associated

issues from the **World Newspaper Archive (WNA)**.⁵ The **World Newspaper Archive** is an online database of digitized historical newspapers developed through a partnership with **CRL**, focusing on newspaper collections held by **CRL** and some of its member libraries. **WNA** issue-level metadata was exposed for **CRL** harvesting through a Web-accessible **LOCKSS** manifest. From the manifest, a title list or table of contents led to an index of issues. Scripts were created to identify and "pull out" the metadata needed from the coding in the **WNA** data. Metadata used from this source included: publication ID, title, and issues. Additional metadata came from **MARC** catalog records for these titles.

Frequency of publication was extrapolated from the digitized issues and from information in the catalog records.

The **Library of Congress** contributed over 140,000 newspaper titles and associated summary holdings, which have been "unpacked" into almost 17 million issues to date. Adding issue-level data continues. Data for these titles and holdings came from **Chronicling America**⁶ and included titles and holdings in the **U.S. Newspaper Registry** and the holdings included in the digitized newspaper database. Bibliographic and holdings metadata was harvested with automated calls to URLs associated with Web pages on the **Library of Congress** Website.

CRL contributed about 15,000 bibliographic records and summary holdings statements, from which 11.5 million issue records were generated. **CRL's** holdings were embedded in the note field of each bibliographic record from its library catalog. **CRL** holdings included shelf location number, number of microform reels (when applicable) and holdings expressed in a locally developed format. **CRL** records were exported from the catalog and run through a custom text processing script to strip out unnecessary characters, normalize holdings and punctuation and unpack the summary holdings statements to generate the issue level holdings.

Generating Issue Level Holdings Records from Summary Statements

Generating issue-level holdings records from summary statements depends on particular data points: beginning and end publication dates, beginning and end holdings dates, frequency of publication (taking into account frequency changes), day of the week of publication (Sundays only, daily except Mondays, etc.). Each of these data points has to map to an individual data field or be

continued on page 76

OECD iLibrary

www.oecd-ilibrary.org

OECD's global knowledge base

To subscribe

call: 202-785-6323 or

email: washington.contact@oecd.org



SIDEBAR — CRL and Newspaper Preservation

Since its inception in 1949, **CRL** has been ensuring long-term availability of all formats of newspaper content. Early efforts were in storage and preservation of an eclectic collection of print newspapers deposited by its members. **CRL** also began to subscribe to many newspaper titles, so that members could discontinue subscriptions of their own. Later **CRL** began to support preservation of newspaper content through area studies microfilming projects and its **International Coalition on Newspapers** program (**ICON**).

The **Area Studies Microfilming Projects (AMPs)** are six independently governed projects whose activities are coordinated and supported, administratively, by **CRL**. These projects identify, preserve and make accessible to scholars and researchers unique, uncommon, and endangered research material, including newspapers. **AMP** participants often work with international partner institutions to safeguard at-risk historical documentation and cultural heritage resources, using traditional preservation techniques and, increasingly, digital technologies.

The **ICON** project, begun in 1999, has addressed all aspects of newspaper preservation, including bibliographic access, copyright, information dissemination and reformatting content. Through **ICON**, **CRL** created and maintained a clearinghouse of information relating to project reports and presentations; technical standards and best practices for selecting, preserving and cataloging newspapers; links to resources for the discovery of information about newspapers; news and developments of current preservation projects. **ICON** also developed a free, Web-accessible database to provide reliable information about where newspapers published outside of the United States were held in U.S. libraries and selected non-U.S. libraries.

Today **ICON** is under **CRL's** Global Resources Program umbrella. It provides a framework for sharing critical information among members of the **CRL** community to support informed, strategic local decisions on investing in digital collections and services and controlling the costs of managing physical collections.

Curating Collective Collections from page 75

clear enough to be called by a text processing command. The most difficult problems faced were: absence of data needed, multiple ways of expressing the same thing and dealing with multiple data elements in a single free-text field. The challenge became how to solve these problems and not let them be game stoppers.

A calendar program was developed that would generate the precise days. Beginning and end dates of summary holdings, frequency and day of the week and the calendar were used to extrapolate individual dates for the holdings. All results were reviewed by staff in a separate program that is interactive, which allowed them to correct errors or to enter the data when data normalization needs prevented the programming from working with the summary statements.

Using ICON

ICON has a public interface that enables users to search for titles by years of publication and particular cities or states in addition to ISSN, OCLC number, Library of Congress Control Number and title. Once a title of interest is retrieved, there are a number

of tools that enable the user to drill down to each separate issue to see which repositories hold it, in which format it is held, what the archiving commitment is and any condition issues disclosed.

Perhaps more useful than searching for individual titles, ICON offers a variety of statistics and graphical representations of the data to make the data within the database work for a wide variety of collection assessment and reporting activities.

- Holding repository statistics lists: contributing repository, its location, number of publications, number of issues and the date range of issues contributed.
- Newspaper [by] country statistics lists: total number of publications, number of issues and date range of issues contributed for each country and U.S. state.
- Issue year statistics lists: number of publications and number of issues per year contributed.
- Issue format lists: number of publications and issues by format. Thirteen different formats are currently listed including some like

photomechanical and microopaque, which may not be the first to come to mind but may help prioritize preservation or reformatting action for the content.

Custom reports can be run on the data by request. Comparison of holdings can be done and reports provided.

Conclusion

This is a rallying call. It is essential that we, as a community committed to responsible preservation and long-term access to scholarly resources, work together to solve the challenges of our existing data rather than develop programs that work around the lack of acceptable data. Data that is issue specific, shared and in tools that are easy to use must be there for us to prevail. 🌱

Rumors from page 43

Did you notice **Rip Van Winkle** (AKA **Anthony Watkinson**) on page one of this issue of *ATG*? It's **Rip's** latest grandchild, **Samuel**, one-day old! Gosh, **Anthony**, how will you keep up with your brood of grandchildren?

Have you renewed your subscription to *Against the Grain* for 2014 yet? We will be sending invoices shortly but, hey, you don't need a reminder, do you?

We already told you that **Charles Lyons** will be guest editing an issue, and we are expecting issues guest edited by **Tom Gilson** (February) **Ramune Kubilius** (April), **Rob Virkar-Yates** (June), **Lindsay Johnston** and her team (September) and November is being guest edited by **Charles Lyons**. **Kristin Eschenfelder** (who gave a great keynote in Charleston a few years ago on licensing) is one of our guest editors for 2015. And why don't YOU YOU YOU get involved with *Against the Grain* — online or print!? Send us your ideas!

Have a good ALA and talk to y'all next issue! 🌱

Core Competencies of Electronic Resources Librarians Adopted as NASIG Policy

The NASIG Board approved and adopted "Core Competencies of Electronic Resources Librarians" as NASIG policy at their June 2013 meeting in Buffalo, New York.

Sarah Sutton, former chair of the **Core Competencies Task Force (CCTF)**, notes that she and the CCTF have high hopes that both library and information professionals and LIS educators will find the document a valuable resource upon which to base their work. **Sarah** writes, "I am so gratified that many practitioners have already used the draft document, which circulated in the professional community over the past few months. It has sparked much interest and use, as evidenced by the wonderful sessions at the recent **NASIG Annual Conference**. I think the document supports **NASIG's** Vision to promote dialogue and professional growth, to provide learning opportunities, to advocate for its constituents, to challenge assumptions and traditions, and to take a leadership role in the information environment."

"Core Competencies of Electronic Resources Librarians" is available in the Continuing Education section of the **NASIG** Website, <http://www.nasig.org>.

Sanjeet Mann and **Sarah Sutton** for the **Core Competencies Task Force**. 🌱

Endnotes

1. "The United States Newspaper Program is a cooperative national effort among the states and the federal government to locate, catalog, and preserve on microfilm newspapers published in the United States from the eighteenth century to the present. Funding is provided by the National Endowment for the Humanities. Technical assistance is furnished by the Library of Congress" — Website <http://www.neh.gov/us-news-paper-program>.
2. "The National Digital Newspaper Program (NDNP), a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC), is a long-term effort to develop an Internet-based, searchable database of U.S. newspapers with descriptive information and select digitization of historic pages." — Website <http://www.loc.gov/ndnp/>.
3. "The Florida Digital Newspaper Library exists to provide access to the news and history of Florida. All of the over 1.5 million pages of historic through current Florida newspapers in the Florida Digital Newspaper Library are openly and freely available with zoomable page images and full text. The Florida Digital Newspaper Library builds on the work done in microfilm within the Florida Newspaper Project." — Website <http://ufdc.ufl.edu/fdn11>.
4. American Antiquarian Society's Clarence database of newspaper holdings url: <http://clarence.mwa.org/Clarence/>.
5. "The World Newspaper Archive is an online database of digitized historical newspapers, created by CRL in partnership with Readex, a division of NewsBank. The initiative has drawn upon the holdings, expertise and resources of CRL and its member libraries to preserve and provide access to historical newspapers from around the globe." — Website <http://www.crl.edu/collaborative-digitization/world-newspaper-archive>.
6. **Chronicling America** is a Website providing access to information about historic newspapers and select digitized newspaper pages and is produced by the National Digital Newspaper Program (NDNP). Description and more information about **Chronicling America**: <http://chronicling-america.loc.gov/about/>.