April 2004

Resource Sharing Data: Mining for Selection & Meaning in Collection Development

Cyril Oberlander

Mary Oberlander

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

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

Recommended Citation

Oberlander, Cyril and Oberlander, Mary (2004) "Resource Sharing Data: Mining for Selection & Meaning in Collection Development," *Against the Grain*: Vol. 16: Iss. 2, Article 13.
DOI: <https://doi.org/10.7771/2380-176X.4291>

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.

**Majors' Mission is to provide Health Science Libraries
the best possible SERVICE - with a personal touch.**

Majors' Suite of Services Includes:

- Online ordering
- Customized Approval Plans and Continuations/Standing Order Plans
- New and Noteworthy Title Alerts
- Weekly New Book Receipts
- In-hand Book Reviews with Bibliographic Summaries
- Flexible Invoicing
- In-stock Items Shipped within 24 Hours
- Rush Orders Shipped the Same Day
- MARCIVE and PromptCat Cataloging Services
- Personal Account Specialists and Sales Representatives

Let's Do Business!

800-633-1851
www.majors.com

Resource Sharing Data: Mining for Selection & Meaning in Collection Development

by Cyril Oberlander and Mary Oberlander

We often imagine data mining to be work destined for Sisyphus; our efforts can seem just as monotonous and inconsequential as the hopeless labor of ceaselessly rolling a rock to the top of a mountain. However, just as we have seen advancements in technology change every facet of our profession, great strides have been made in transforming the process of extracting and analyzing data from **Interlibrary Loan/Document Delivery (ILL/DD)** transactions for collection development decisions. Although the process was once severely time consuming and complicated, with the help of several new data mining tools, the work has become surprisingly creative, precise, and valuable. In this article, we provide a brief background of ILL/DD data mining for collection development in the library literature, describe several new data mining tools including methods of analysis, and

lastly, provide some techniques to consider for comparative analysis and cooperative collection development.

Background

ILL/DD data has long been viewed as a potentially powerful tool in supporting collection development decisions. As **Roberts and Cameron** (1984) aptly explain, "the justification for analyzing interlibrary loans as an aid to collection development has never been open to question: one of the few incontrovertible conclusions of the previous studies was that material acquired on ILL reflected an 'unmet demand' within the requesting institution." The primary challenge then is not the justification of the connection between ILL/DD data and collection development decisions, but rather, the great difficulty of consistently producing relevant data needed for such a connection to exist.

Throughout the literature, one may find excellent reviews of previous attempts to create inno-

vative methods to extract and analyze ILL/DD data (**Bartolo** 1989, **Lahmon** 1991, **Wilson** 1999). Often noted in these reviews is the challenge of libraries to implement the methods because the processes frequently appear prohibitively labor intensive, complicated, tedious, and produce results that are often difficult to interpret. As **Bartolo** (1989), points out "one is impressed with the thoroughness and tenacity required to undertake these projects," but "one sees that time and labor were the major impediments to the regular inclusion of ILL activity analysis in collection development decisions." Although many articles describe methods that appeared successful in limited situations, lacking is a description of methods that can be used consistently over long periods of time and can respond to a variety of data inquiries on demand.

Timeliness has also been a serious issue. **Wilson** (1999) points out that "compiling and analyzing...records was a laborious and time consuming process that did not permit ready input from ILL when important collection development decisions were needed." As **Bartolo** (1989) explains, "while earlier studies indisput-

continued on page 78

Resource Sharing Data: from page 77

ably attested to the validity of using ILL as one type of litmus test for collection development, they also underscored the fact that until analysis of ILL requests could be easily incorporated into ILL workflow and be quickly disseminated to subject selectors, ILL would not readily be included in collection development." Lastly, the data gathered in these studies was often seen as "complex and difficult to interpret." (Byrd et al. 1982)

For the ILL/DD data to be used effectively in collection development decisions, librarians need information that is easy to access, specific, multi-dimensional, and individualized to the librarians' selection or de-selection decision at hand. This is the crux of the matter—getting relevant data in a timely manner and having the tools to effectively analyze the data. The following data mining tools directly address this issue. We begin our discussion of tools with the relatively basic and progress to those that are more complex and versatile in their capabilities.

Title Analysis

Interlibrary Loan departments have long provided lists of borrowed and requested items for collection development, often conveniently derived from copyright reporting lists with varying degrees of bibliographic data. At **Portland State University**, we create title lists from OCLC Management Statistics data. This data is managed and reports are generated using STATCAT, an MS-Access Database created at Portland State University by **Cyril Oberlander** and **John Clarke** and available from OCLC. The borrowing title lists are easy to use, Web accessible MS-Excel spreadsheets. The following example illustrates how a subject librarian can use an Internet browser to sort and limit a title list to a range of Library of Congress call numbers using the auto filters.

Total Filled Requests	TITLE	Year	OCLC NUMBER	Standard Number	LC Call Number
29	FOREST ECOLOGY AND MANAGEMENT	1976	3131294	0078-1127	J51 F57
22	PERSONALITY AND INDIVIDUAL DIFFERENCES	1980	4862018	0191-8869	BF69 A1 F49
47	MOLECULAR BIOLOGY	1992	25447413	0963-1083	Q6341 J5 M63 B64
43	JOURNAL OF CHROMATOGRAPHY. A	1993			
42	JOURNAL OF THE INTERNATIONAL P				
42	JOURNAL OF THE INTERNATIONAL P				
41	PSYCHIATRIC SERVICES: A JOURNAL OF THE AMERICAN	1995			
39	INTERNATIONAL JOURNAL OF TECHNOLOGY MANAGEMENT	1996			
39	JOURNAL INTERNATIONAL DE LA GES	1980			
38	INTERNATIONAL JOURNAL OF REMOTE SENSING				
37	CANADIAN MODERN LANGUAGE REVIEW - LA REVUE	1944			
37	CANADIAN MODERN LANGUAGE REVIEW - LA REVUE	1979			
37	CHILDREN AND YOUTH SERVICES REVIEW	1986			
36	RESIDENTIAL TREATMENT FOR CHILDREN & YOUTH	1985			
36	JOURNAL OF MEMORY AND LANGUAGE	1989			
34	AIDS CARE	1993			
33	JOURNAL OF EMOTIONAL AND BEHAVIORAL DISORDERS	1993			
33	NEUROLOGY	1951			
32	JOURNAL OF PEDIATRIC PSYCHOLOGY	1976			
31	INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY	1996			
31	INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY	1996			
30	JOURNAL OF OCCUPATIONAL HEALTH PSYCHOLOGY	1996			
30	JOURNAL OF SOCIAL DISTRESS AND THE HOMELESS	1992	22447873	1052-0789	RM1 J28
29	ELECTRONIC WORLD	1993	2421457	0013-4457	TX1 J25
29	FOR THE LEARNING OF MATHEMATICS: AN INTERNATIONAL				
29	JOURNAL OF MATHEMATICS EDUCAT	1980	7862828	0028-0671	QA11 A3 P67
29	JOURNAL OF ADOLESCENCE	1978	4041779	0148-1971	RM99 A1 J274
29	JOURNAL OF GERONTOLOGICAL NURSING	1975	2243369	0098-9134	RC954 J67
29	JOURNAL OF GERONTOLOGICAL NURSING	1976	1767752	0197-1497	RC954 J67

STATCAT presents subject selectors with highly used titles in their subject area and allows them to weigh the cost of subscribing to or purchasing a particular title. A great advantage to this tool is its accessibility. Using STATCAT, librarians can access and edit current, individualized data reports from virtually anywhere via the Web. The following is a sample cost benefit analysis of frequently requested titles using reports from STATCAT.

Total Filled Requests	Title	LC#	Borrowing Charges	Cost to Subscribe
28	Canadian Modern Language Review	PB5 .C36	\$ 467.50	\$ 53
28	Journal Of Memory And Language	BF455.A1 J6	\$ 1,169.00	\$ 575
22	Journal Of Gerontological Nursing	RC954 .J67	\$ 362.50	\$ 60

User Analysis

User analysis offers a variety of opportunities for collection development assessment. Public libraries, for instance, can use zip codes to determine geographic use patterns, while academic libraries can use department affiliation and status to analyze the re-allocation of collection funds and evaluate the effectiveness of approval plans. The following abbreviated list of PSU ILL users and number of requests comes from a modified Web report from ILLiad, an ILL request management system. Selectors can observe user behavior details that help isolate use patterns in subject areas. They can also be alerted to changes in faculty interest areas or new programs that may warrant adjustments in approval plan profiles or subscriptions.

Department	Distance Ed	Faculty	Graduate
Accounting	0	1	1
Anthropology	0	14	12
Applied Linguistics	0	13	41
Architecture	0	1	0
History	0	23	37
Psychology	1	8	47
Social Work	9	15	86
Avg.		7.25	

Title-User Analysis

Correlating the title list with user data greatly enhances a selector's ability to evaluate the relationship between collection, ILL-use and a specific user group. In the following example, a report by discipline is a simple query report that combines ILL/DD borrowing data from OCLC Management Statistics to obtain Library of Congress call number with user affiliation and status from ILLiad. (see illustration p.79)

The accompanying title-user list is useful in identifying title priorities of distinct user groups, targeting approval plan profiles to department needs and department funds, and illustrating the breadth and scope of interest based on an institution's discipline. Each year, subject librarians at PSU review these reports of frequently requested titles by their department or discipline. Librarians can also share these title-user lists with departments or teaching faculty to assess perceived values of frequently requested materials so as to better determine the necessity of adding titles to the collection.

Hybrid Lists, Graphs and Analysis

Hybrid lists and graphs involve correlations of data sets to help librarians better conceptualize data and understand the "big picture" using comparative and parallel analysis. These analysis techniques typically provide trend analysis, useful in determining budget allocations and alerting librarians to areas of new research interests.

At **Portland State University**, simply using three data elements reported from our ILS system and STATCAT, specifically LC number, number of checkouts, and number of ILL requests, we created use factor charts to better conceptualize how the collection is used. The chart plots collection size and ratios of Circulation and ILL/DD use as a percentage of the collection size. Selectors view the charts and data tables through an Internet browser, easily picturing the collection size against relative use. The higher the circulation and ILL/DD relative use factor, the higher the need for building a collection in that aggregate LC call number.

continued on page 79

Department: Mechanical Engineering

Request Type

Frequently Requested Call Numbers

29 - ?
 12 - TN693.T7 S96
 10 - TA401 .M382
 8 - RK652.5 .D44
 7 - CLASSED SEPARATELY
 5 - TK7881 .E46
 4 - TP490 .R48
 3 - TA462 .P76
 3 - TA401 .J68
 3 - TS227 .A282
 3 - TS670.A1 E43

Unknown LC#

Status Summary

?	Article	37009014	Science and technology of welding and joining.	Graduate
?	Article		International symposium "Principles and processes of c	Faculty
?	Article	32081466	Workshop on the Thermophysical properties of molten	Faculty
?	Article	44487358	ECOSS-18 : proceedings of the 18th European Confer	Faculty
?	Article		In: High-Productivity Joining Processes. Proceedings, I	Graduate
?	Loan		Acoustic Emission Monitoring of Pinch Welds	Graduate
?	Article	35407117		Graduate
?	Article	8401972	Proceedings of the thirteenth Symposium on Nondestr	Graduate
?	Article	1757173	Metal construction and British welding journal.	Graduate
?	Loan	40499426	Mathematical modelling of weld phenomena 4 /	Graduate

Resource Sharing Data:

from page 78

(Total Number of ILL/DD's in LC#) / (Total Number of ILL/DD's)

$$\text{ILL Use Factor} = \frac{\text{(Total Number of ILL/DD's in LC\#)}}{\text{(Total Number of ILL/DD's)}}$$

(Total Titles in LC#) / (Total Number of Titles)

Example of PSU Use Factor and Collection Data for LC# range "P-PZ"

REMEMBER WHEN?

Subscription Agencies were *service* businesses and your account was as important as any other account your agency handled?

Basch Subscriptions, Inc. provides serial *service* that saves you and your library both time and money. Any agency will take your order, only **Basch Subscriptions, Inc.** will prove its commitment every day with the personal service you deserve.

Services for Libraries and Publishers

Basch Subscriptions, Inc.

88 N. Main Street
 Concord, NH 03301

Phone: (603) 229-0662

Fax: (603) 226-9443

E-Mail: Subs@Basch.com

Shared Hybrid Analysis

When conducting hybrid analysis, there are even greater rewards in sharing library data. The simplest example of a shared hybrid analysis is a cost benefit strat-

egy with ILL/DD borrowing data from more than one library. The following sample chart is created from approximately one year of borrowing data from four libraries and compares the total borrowing fees (a sum of copyright and lending charges) with the cost to subscribe. This type of analysis offers consortia libraries, with reciprocal agreements, a potential for significant cost savings. Libraries can reduce their cost of borrowing resources while purchasing new titles that are highly used at their institution. This cooperative collection development strategy is a win-win situation for participating libraries; the library that most often requests articles from a title has the greatest benefit in purchasing it, thereby reducing copyright royalties. Each library in a group picks a title that benefits the group, and in effect, builds a cooperative access selection strategy that doesn't reduce resources for existing collection allocations.

continued on page 80

Resource Sharing Data: from page 79

In the example above, LibStatCAT graphs journal supply and demand by LC call number. Each vertical line represents a journal

sions. As **Sherrer** aptly explains:

"As resource sharing continues to grow, institutional lines blur... as do the proprietary rights of collections... It is time for the library community to investigate the impact of expanded resource sharing in the context of collection development and technological enhancements." (**Sherrer** 1998)

New technology has indeed transformed the collection development process, allowing us to mine resource-sharing data for more

Title	Total request count	OHS Lib.	ORE Lib.	ORU Lib.	ORZ Lib.	Total Borrowing Fees	Cost to Subscribe
British Journal of General Practice	22	6	8		8	\$304.00	\$264.00
Child & Adolescent Psychiatric Clinics of North America	36	2	7	25	2	\$997.70	\$242.00
International Journal of Geriatric Psychiatry	47	3	30		14	\$1,535.00	\$1295.00
Journal of Cancer Education	18	11	3		4	\$292.00	\$205.00
					Total	\$3,883.70	\$3,731.00

Benefits of cooperative access selection strategies, as well as cooperative collection analysis greatly improve with more data and data elements in the data pool. For example, combining extensive bibliographic data with use and cost data enables selectors from groups of libraries to cooperate in selection and de-selection based on comparisons of supply and demand of a particular title or a subject range. To extend our capabilities for comparative and parallel analysis of data sets, **Cyril Oberlander** and **Dan Streeter** of **Portland State University** developed **LibStatCAT**, a data managing graphical tool to analyze library data sets from one library and functions as a data farm for groups of libraries. **LibStatCAT** is a prototype library statistical collection assessment tool that enables selectors to perform comprehensive analysis of their collections using a stacked histogram that helps librarians visualize queries of the data tables (**Oberlander** 2003).

title, making this chart a virtual shelf. On this shelf, the height of the line is relevant to both use and holdings, specifically, the length below the x-axis represents the number of libraries in the Oregon region holding that title, and the line above represents how many times an Oregon library requested an article from that title from three of the large academic libraries in Oregon: **Oregon State Univ.** (ORE), **Portland State Univ.** (ORZ) & **Univ. of Oregon** (ORU). Lastly, we added one more data characteristic; **LibStatCAT** colors the line red if **Portland State University** has an electronic subscription to that particular title.

As we are better able to understand and manage our ILL/DD data, we find ourselves looking beyond our own collection, drawing the parallel that data sharing as well as resource sharing has the potential to transform how we view collection development deci-

effective selection and greater meaning.

The data mining tools described in this paper help support the complicated considerations of serious collection development decisions, producing data that is more precise, timely, multi-dimensional, in-depth, and individualized than ever before. The potential uses are countless: from evaluating the cancellation of a single journal title, to creating regional cooperative collection development strategies; from observing the efficacy of approval plans, to identifying trends in department research interests. The methods for mining for ILL/DD data may change, but the desire is the same; create meaningful and understandable reports of data and analysis that support the collection development process, leading ultimately to more user satisfaction.

Bibliography

Allen, G.G., "The economic efficiencies of alternatives to purchase and loan of library resources," *Iatut Quarterly*, vol. 2 (4), 1988, p. 209-214.

Anderson, Kristine; Robert Freeman, Jean-Pierre Herubel, Lawrence Mykytiuk, Judith Nixon, Suzanne Ward, "Buy, Don't Borrow: Bibliographers' Analysis of Academic Library Collection Development Through Interlibrary Loan Requests," *Collection Management*, vol. 27 (3/4), 2002, p. 1-11.

Bartolo, Laura M., "Automated ILL Analysis and Collection Development: A Hi-Tech Marriage of Convenience," *Library Acquisitions: Practice & Theory*, vol. 13, 1989, p. 361-369.

Burch, Brian; Arthur Davies, "Acquisitions and Interlibrary Loans: A Correlation?" *Interlending and Document Supply*, vol. 15 (3), 1987, p. 84-87.

Byrd, Gary; D.A. Thomas; Katherine Hughes, "Collection Development Using Interlibrary Loan Borrowing and Acquisitions Statistics," *Bulletin of the Medical Library Association*, vol.

LibStatCAT Journal Analysis (Library of Congress LC#: L)

continued on page 81

Resource Sharing Data: from page 80

70 (1), Jan. 1982, p. 1-9.

Graziano, Eugene, E., "Interlibrary Loan Analysis: Diagnostic for Scientific Serials Backfile Acquisitions," *Special Libraries*, vol. 53, 1962, p. 251-257.

Khalil, Mounir, "Applications of an Automated ILL/DD Statistical Analysis as a Collection Development Tool," *Journal of Interlibrary Loan, Document Delivery & Information Supply*, vol. 4 (1) 1993, p. 45-54.

Lahmon, Jo Ann, "Using Interlibrary Loan Data in Collection Development," *OCLC Micro*, vol. 7, Oct. 1991, p. 19-22.

Mellendorf, S.A., "A practical method for using ILL data to assist Librarians with Collection Development," *OCLC Systems & Services*, vol. 9, Summer 1993, p. 45-48.

Nitecki, Danuta, "Document Delivery and the Rise of the Automated Midwife," *Resource Sharing and Information Networks*, vol. 1 (3/

4), Spring/Summer 1984, p. 83-101.

Oberlander, Cyril; Dan Streeter, "LibStatCAT: a library statistical collection assessment tool for individual libraries & cooperative collection development," *Library Collections, Acquisitions, & Technical Services*, vol. 27, 2003, p. 493-506.

Ochola, John, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *Collection Management*, vol. 27 (1) 2002, p. 1-12.

Paustian, Robert P., "Collection Size and Interlibrary Loan in large academic libraries," *Library Research*, vol. 3, 1981, p. 393-400.

Roberts, Michael; Kenneth Cameron, "A Barometer of 'Unmet Demand'," *Library Acquisitions: Practice and Theory*, vol. 8, 1984, p. 31-42.

Rottmann, F.K., "To Buy or to Borrow: Studies of the Impact of Interlibrary Loan on Collection Development in the Academic Library," *Journal of Interlibrary Loan & Information Supply*, vol. 1 (3), 1991, p. 17-27.

Sherrer, Johannah, "Resource Sharing, Collection Development and Technology," *Technical Services Quarterly*, vol. 16 (2) 1998, p. 25-33.

Ward, Suzanne, et. Al. "Collection development based on patron requests: collaboration between interlibrary loan and acquisitions," *Library Collection, Acquisition and Technical Services*, vol. 27, 2003, p. 203-213.

Wender, Ruth W., "Analysis of Loans in the Behavioral Sciences," *Special Libraries*, vol. 60, 1969, p. 510-513.

Wilson, Mary Dabney; Whitney Alexander, "Automated Interlibrary Loan/ Document Delivery Data," *Serials Review*, vol. 25 (4), 1999 p. 11-19.

Wilson, T.D., "Follow-up on Interlibrary Loan Analysis," *Special Libraries*, vol. 53, Oct. 1962, pl. 493-494.

More information and examples of tools and charts used at Portland State University are available at: <http://www.lib.pdx.edu/services/ill/stats/librystats.html>.

STATCAT is available free from OCLC at: <http://www.oclc.org/support/tips/ill/tip15.htm>.

International Dateline — Which is Right? Florence or Firenze?

Issues In International Authority Control

by Claire-Lise Bénéaud and Sever Bordeianu (University of New Mexico)

Introduction

How to name people, places, organizations, and concepts when dealing with multiple languages, cultures, and scripts is no small feat. Through authority control, catalogers have tackled this issue locally forever but have seldom been able to transcend national boundaries. Authority control sets up the correct form of a name and provides the cross-reference structure that gives validity and useability to library catalogs. In the US, under the leadership of the **Library of Congress**, authority control has been an integral part of cataloging. American librarians have created rich and complex authority files that are shared nationally. But, with all its complexity, applying authority control in the US is relatively simple compared to the problems encountered in other countries. Countries outside of North America have longer publishing histories and have applied authority control in various ways over the centuries. East Asian countries have additional layers of complexity imposed by their intricate scripts.

Diversity in Europe and Asia is multifaceted. Languages, scripts, formats, library organizational structures, political systems, geography, resources, and cultural sensitivities all play a role in how authority control is viewed, developed, and implemented. And yet, as technology is making records available world-wide, and libraries are increasingly making efforts to share their collections, the need for authority control is more acute than ever.

For this purpose, the **University of Florence** organized the "International Conference on

Authority Control: Definition and International Experiences." The conference was held in the fourteenth century Convitto della Calza, about a fifteen minute walk from downtown Florence, from February 10th to 12th, 2003. Some 500 librarians attended. The majority were Italian — looking very stylish as a group — and the rest came from Western Europe (Holland, Sweden, France, England, Germany, and Switzerland) with some Americans, Canadians, and a few participants from East Asia. Most of the presentations revolved around the idea of going from local authority control (and by local, sometimes it meant national, and sometimes it meant regional) to global authority control. The aim of the conference was to offer an occasion for rethinking, comparing, and reporting on ongoing projects. It focused on how to name individuals and families. Naming corporations, events, and topics is more complicated and was scarcely addressed. The lack of international cataloging rules, the different authority control models used throughout the world, linguistic, cultural and political considerations all have an impact on the development of a world-wide structure.

Why Authority Control Matters

There was little disagreement among the participants that authority control is essential. **Mauro Guerrini**, the conference organizer, and **Michael Gorman**, the keynote speaker presented the arguments for the continuous need for authority control. Effective cataloging requires consistency of the form of access in order to identify, select, and obtain the correct in-

formation. Access points need to be standardized. Without a standard form of a name, catalogers reach different conclusions even when presented with the same evidence. An authoritative name is the most common form of a name within a cultural setting. **Guerrini** set the tone of the conference by stating that plurality is not only linguistic but cultural. Authority control should respect this plurality and therefore, a worldwide adoption of the same form of a heading by all is not tenable. The issue is how to share authority records in different languages and scripts rather than establish universal headings.

Coupled with a lack of universal availability of headings is an even more worrisome trend, that is, not applying any kind of authority control. **Gorman** mentioned the **Dublin Core** and the Web as examples of either minimum or non-existent authority control. Web searches are ineffective and retrieve very large volumes of information with many aberrant results. As a test, **Gorman** did a search on his own name in **Google** and retrieved over 7,700 entries, most having nothing to do with him or his works. He characterized the Dublin Core as naïve, simplistic, and unsuited for reflecting the complexity of what needs to be cataloged. With his trademark wit, he concluded that a complicated world needs complicated rules.

Challenges

Lack of International Cataloging Rules

The world has no global cataloging rules, no global classification system, and no global

continued on page 82