E-Resources: Current Issues in Measuring Electronic Use in Libraries

National Seminar on Electronic Resources in Malaysia
21 December 2010 • Penang, Malaysia
Julia C. Blixrud, Association of Research Libraries
A Decade of Challenges

• Demand for libraries to demonstrate outcomes/impacts in areas of importance to institution

• Pressure to maximize use of resources through benchmarking resulting in:
  – Cost savings
  – Reallocation
Electronic Serials Expenditures


Association of Research Libraries  www.arl.org
Average Yearly Increases in Electronic Resources and Total Library Materials Expenditures
ARL New Measures Projects

- Demonstration project for service effectiveness measures (LibQUAL+®)
- Project to define usage measures for electronic information resources (E-metrics)
Statistics and Data Needs for Electronic Information

- Financial Support
- Infrastructure
- Comparisons
- Vendor Negotiation
ARL E-Metrics Project

Three phases:

– Initial Phase (May-October 2000): **What do we know?** Inventory of current practices at ARL libraries as to statistics, measures, processes, and activities that pertain to networked resources and services.

– Second Phase (November 2000-June 2001): **What can we collect?** Identified and field tested an initial draft set of statistics and measures.

– Final phase (July 2001-December 2001): **What difference does this make?** Build linkages to: educational outcomes/impact, research, technical infrastructure.
Recommended Statistics & Measures

- Patron Accessible Electronic Resources
- Use of Networked Resources & Services
- Expenditures for Networked Resources & Related Infrastructure
- Library Digitization Activities
- Performance Measures
Patron Accessible
Electronic Resources

• Number of electronic full-text journals
• Number of electronic reference sources
• Number of electronic books
Use of Networked Resources & Related Infrastructure

- Number of electronic reference transactions
- Number of logins (sessions) to electronic databases
- Number of queries (searches) in electronic databases
- Items requested in electronic databases
- Virtual visits to library’s website and catalog

Association of Research Libraries
Expenditures for Networked Resources & Related Infrastructure

- Cost of electronic full-text journals
- Cost of electronic reference sources
- Cost of electronic books
- Library expenditures for bibliographic utilities, networks & consortia
- External expenditures for bibliographic utilities, networks & consortia
Library Digitization Activities

- Size of library digital collection
- Use of library digital collection
- Cost of digital collection construction & management
Performance Measures

- Percentage of electronic reference transactions of total reference
- Percentage of virtual visits of all library visits
- Percentage of electronic books to all monographs
- *Percentage of electronic journals to serial subscriptions [note: serials now counted by title, rather than subscriptions]*
Project Documents

• Measures for Electronic Resources (E-Metrics)
  Part 1: Project Background and Phase One Report
  Part 2: Phase Two Report
  Part 3: E-Metrics Instructional Module
  Part 4: Data Collection Manual
  Part 5: Library and Institutional Outcomes

• www.arl.org/stats/initiatives/emetrics/index.html
Decade of Learning

• Agreement on what to count is hard
  – Change from serial subscriptions to titles
  – Ebooks
• Complexity of systems and interfaces
• Packages/bundles complicate use counts for specific resources
• Comparisons across institutions difficult when part of consortia
• Digitization projects dependent on environment

Association of Research Libraries  www.arl.org
Launched in March 2002, COUNTER (Counting Online Usage of Networked Electronic Resources) is an international initiative serving librarians, publishers and intermediaries by setting standards that facilitate the recording and reporting of online usage statistics in a consistent, credible and compatible way. The first COUNTER Code of Practice, covering online journals and databases, was published in 2003. COUNTER's coverage was extended further with the launch of the Code of Practice for online books and reference works in 2006. The body of COUNTER compliant usage statistics has steadily grown as more and more vendors have adopted the COUNTER Codes of Practice. This has contributed to the new discipline of usage bibliometrics and a great deal of work is underway to try to establish value metrics associated with usage, in which the COUNTER compliant statistics play an increasingly important role.

COUNTER does more than just set the standards for usage reports; we are co-operating with a number of organizations to develop a range of usage-related research and services. In 2006 COUNTER carried out research, sponsored by JISC (the UK Joint Information Systems Committee) on the effects of publisher platforms on usage and we are currently collaborating with the UK Serials Group on the possible development of a new Journal Usage Factor metric. Summary reports on both these projects can be found on the COUNTER website at http://www.projectcounter.org/news.html. COUNTER has also worked with NISO on SUSHI (Standardised Usage Harvesting Initiative) to develop a protocol to facilitate the automated harvesting and consolidation of usage statistics from different vendors. This protocol may be found on the NISO website at http://www.niso.org/schemas/sushi/index.html#COUNTER

ANSI/NISO Z39.7-2004

An American National Standard
Developed by the National Information Standards Organization

Approved October 6, 2004
by the American National Standards Institute

This standard is continuously maintained, and changes may happen on a periodic basis via the continuous maintenance procedures. This version indicates the most current changes; please visit the change log and comments for additional information about past changes and potential upcoming edits. In addition, the archives retain copies of the full standard prior to regular changes enacted via the continuous maintenance procedures.

- **Table of Contents**
  - Hyperlinked document - smaller files for easy browsing
- **Complete Document**
  - Complete document in one HTML file - larger download
- **Change log**
  - The change log enumerates all the changes made to the document since the previous version.
- **Emetrics Element Listing**
  - All the emetrics data elements listed together, with links to full text within the document.
Standardized Usage Statistics Harvesting Initiative (SUSHI)

SUSHI Standing Committee Quarterly Report Released (August 31, 2010)

- About SUSHI
- Join the SUSHI Developers Email List!
- About COUNTER

About SUSHI


The protocol was designed to be both generalized and extensible, meaning it could be used to retrieve a variety of usage reports. An extension designed specifically to work with COUNTER reports is provided with the standard, as these are expected to be the most frequently retrieved usage reports.

The standard is built on SOAP (Simple Object Access Protocol) for transferring request and response messages. The GetReport method is used for transferring ReportRequest as the input message and returning ReportResponse as the output message.

The standard includes a versioned Web Services Description Language (WSDL), to describe the Web service namespace and operations, and a generalized XML schema with the syntax of the SUSHI protocol. Rules for report naming are outlined and complemented by an external reports registry, which provides for the definition of both COUNTER and non-COUNTER reports.
PIRUS2

PIRUS2, sponsored by JISC (the United Kingdom Joint Information Systems Committee) builds on the outcomes and recommendations of the original PIRUS (Publisher and Institutional Repository Usage Statistics) project, also funded by JISC, which was completed in January 2009. The full report of the original PIRUS project may be found at: http://tinyurl.com/PIRUSreport10.

The original PIRUS project demonstrated that it is technically feasible to create, record and consolidate usage statistics for individual articles using data from repositories and publishers, despite the diversity of organizational and technical environments in which they operate. If this is to be translated into a new, implementable COUNTER standard and protocol, further research and development will be required, specifically in the following areas:

- Technical: further tests, with a wider range of repositories and a larger volume of data, will be required to ensure that the proposed protocols and tracker codes are scalable/extendible and work in the major repository environments.
- Organizational: the nature and mission of the central clearing house/houses proposed in the original project has to be developed, and candidate organizations identified and tested.
- Economic: we need to assess the costs for repositories and publishers of generating the required usage reports, as well as the costs of any central clearing house/houses; investigate how these costs could be allocated between stakeholders.
- Advocacy: the broad support of all the major stakeholder groups (repositories, publishers, authors) will be required.

The objective of PIRUS2 is to address these issues and by doing so specify standards, protocols, an infrastructure and an economic model for the recording, reporting and consolidation of online usage of individual articles hosted by repositories, publishers and other entities.

MESUR: MEtrics from Scholarly Usage of Resources

Project objectives:
The project's major objective is enriching the toolkit used for the assessment of the impact of scholarly communication items, and hence of scholars, with metrics that derive from usage data. The project has created a semantic model of the scholarly communication process, and an associated large-scale semantic store that relates a range of bibliographic, citation and usage data obtained from a variety of sources. After mapping the structure of the scholarly community on the basis of the established reference data set, MESUR will conduct an investigation into the definition and validation of a range of usage-based metrics. The defined metrics will be cross-validated, resulting in the formulation of guidelines and recommendations.

Quick facts:
Funding: The Andrew W. Mellon Foundation
Timeline: October 2006 - October 2008
Principal investigator: Johan Bollen
Institution: Los Alamos National Laboratory
Team: Digital Library Research & Prototyping Team of the LANL Research Library
People: Johan Bollen is the Principal Investigator, Herbert Van de Sompel serves as an architectural consultant, and Aric Hagberg of the LANL Mathematical Modeling and Analysis group serves as modeling consultants. Marko A. Rodriguez, a recent PhD graduate at the University of California Santa Cruz and now LANL post-doc at the LANL Center for Non-Linear Science, has supported the project's research and development. Ryan Chute of the LANL Research Library is now the project’s main developer and database manager.

The MESUR data base:
CLICKSTREAM MAP OF SCIENCE

This is the first map created from large-scale, world-wide, scholarly usage data. It visualizes the collective flow of scientists’ movements from one journal to another in their online navigation behavior.

The MESUR project (see mesur.org) collected a database of nearly 1 billion user requests made by the readership of some of the world’s most significant publishers, aggregators and large university consortia, among them the Web of Science, Elsevier (ScienceDirect), JSTOR, Ingenta, University of Texas (6 campuses, 6 health institutions), and California State University (23 campuses). All usage logs acquired by the MESUR project contain session identifiers that identify the individual clickstreams of individual scientists navigating from one article to the next.

Pairs of journals are connected when they have a high probability of being followed by each other in users’ clickstreams. The colors represent individual journals. A line between two colors indicates that they are strongly connected on either direction. The colors indicate the scientific domain a journal belongs to according to their Dewey Decimal and JCR classification codes that were mapped into the Getty Research Center’s Arts and Architecture Taxonomy (AAT) to allow classifications at various levels of detail. The size of a circle corresponds to the strength (degree centrality) of a journal’s connections in the map. The map is arranged by the Force-directed Graph algorithm that treats connections like springs, connected journals are drawn together, but they are not allowed to get too close.

This map is derived from usage data and therefore also reflects the actions of those who merely browse the literature but rarely publish themselves, e.g. practitioners and learners. As a result, practitioners’-oriented themes, such as nursing, medical, and tourism studies are prominently featured. The natural sciences vs. the social sciences are shown as two distinct domains. Each domain is further divided into branches that are connected via various specific subdisciplines. Most domains are highly interdisciplinary. But this is not the case for the social sciences, where sociology, psychology, and economics show some fragmentation. Mathematics and computer science are not represented as one specific cluster, but spread throughout the map.

Like citation maps, this map is based upon a particular sample of the scientific community, about one that includes non-publishing scientists and non-subscribers, and a much smaller number of publications. From MESUR’s database of 1 billion user events, we collect a matrix of 5 million connections between approximately 100,000 journals. From that matrix, we selected only 12,000 connections with the highest number of observations, ranging from approximately 40,000 to 170 observations. This subset of connections pertained to the 2,367 most used journals. This procedure may introduce some bias, but this seems minor.

This map should therefore not be construed as a final map of scientific knowledge, but as an immensely beautiful and revealing representation of scientific activity from usage data. We hope this methodology will provide useful insights into the real-time scientific activity as it can be observed from scholarly clickstream data.

When we cut the AAT taxonomy at the top level, only two distinctions remain: natural sciences and human sciences, with a further distinction for science and humanities (yellow nodes). Some journals along the spokes of the wheel have classifications (colors) that do not correspond to their location in the wheel. This indicates either that journal in question is highly interdisciplinary, and/or that it is classified as less common or less emphasized by the AA, or that the link to this journal is biased towards how scientists actually use the particular journal.
Welcome to MINES for Libraries®

Measuring the Impact of Networked Electronic Services (MINES for Libraries®) is an online transaction-based survey that collects data on the purpose of use of electronic resources and the demographics of users. As libraries implement access to electronic resources through portals, collaborations, and consortium arrangements, the MINES for Libraries® protocol offers a convenient way to collect information from users in an environment where they no longer need to physically enter the library in order to access resources.

Sixteen libraries in Canada have implemented MINES for Libraries® through a contract between ARL and the Ontario Council of University Libraries (OCUL). The StatsQUAL® portal to MINES for Libraries® presents interactive analysis for the OCUL Scholars Portal by institution.

For more information on MINES for Libraries®, see: http://www.arl.org/stats/newmeas/mines.html
Continuing Challenges

- Definitions
- Vendor systems
- Bundles/packages
- Consortial services
- Freely accessible titles
- Digitization projects
Links

- http://www.projectcounter.org/
- http://www.niso.org/dictionary
- http://www.niso.org/workrooms/sushi
- http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/tiki-index.php
- http://www.mesur.org/MESUR.html
Julia C. Blixrud
Assistant Executive Director,
Scholarly Communication
Association of Research Libraries
21 Dupont Circle, Ste 800
Washington, DC 20036
jbl@arl.org
202-296-2296 ext. 133
202-872-0884 (fax)
202-251-4678 (cell)