



Conference Paper

Business Intelligence Infrastructure for Academic Libraries

Joe Zucca

Director for Planning and Organizational Analysis

University of Pennsylvania Libraries

Philadelphia, Pennsylvania, United States of America

Email: zucca@upenn.edu

© 2013 Zucca. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 2.5 Canada (<http://creativecommons.org/licenses/by-nc-sa/2.5/ca/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objective – To describe the rationale for and development of MetriDoc, an information technology infrastructure that facilitates the collection, transport, and use of library activity data.

Methods – With the help of the Institute for Museum and Library Services, the University of Pennsylvania Libraries have been working on creating a decision support system for library activity data. MetriDoc is a means of “lighting up” an array of data sources to build a comprehensive repository of quantitative information about services and user behavior. A data source can be a database, text file, Extensible Markup Language (XML), or any binary object that contains data and has business value. MetriDoc provides simple tools to extract useful information from various data sources; transform, resolve, and consolidate that data; and finally store them in a repository.

Results – The Penn Libraries completed five reference projects to prove basic concepts of the MetriDoc framework and make available a set of applications that other institutions could test in a deployment of the MetriDoc core. These reference projects are written as configurable plugins to the core framework and can be used to parse and store EZ-Proxy log data, COUNTER data, interlibrary loan transactional data from ILLIAD, fund expenditure data from the Voyager integrated library system, and transactional data from the Relais platform, which supports the BorrowDirect and EZBorrow resource sharing consortiums. The MetriDoc framework is currently undergoing test

implementations at the University of Chicago and North Carolina State University, and the Kuali-OLE project is actively considering it as the basis of an analytics module.

Conclusion – If libraries decide that a business intelligence infrastructure is strategically important, deep collaboration will be essential to progress, given the costs and complexity of the challenge.

Introduction

Since the late 1990s, the academic library community has held a wide-ranging discussion on library metrics for the digital age. Beginning in 1998, this conversation took on formal dimensions with two noteworthy developments: first, the guidelines for measuring the use of electronic resources issued by the International Coalition of Library Consortia (ICOLC); and second, the emergence in Europe of Equinox, a project to create performance indicators for the “hybrid” library (International Coalition of Library Consortia, 2006). Soon after, the Association of Research Libraries (ARL) (2001) identified electronic use statistics as a key priority for its Statistics Program, and launched the E-Metrics Project.

The ARL effort eventually broadened into an attempt to restructure the canon of statistics that describes and tracks the value of library services in the 21st century. It has long been recognized that the traditional ARL statistical corpus – holdings, expenditure, and staff size – cannot adequately represent library contributions to academic outcomes, or engagement with the strategic interests of the academic community, such as library support for collaborative methods of teaching and learning, e-science and e-research, and the globalization of higher education.

Even as the search for more relevant metrics has unfolded, academic libraries have been buffeted by paradigm-altering events. They have seen their purchasing power erode, their budgets constrict, and their audiences shift to powerful new commercial information services, such as

Google and Amazon. In their planning, libraries have had to tackle difficult questions about their very nature and purpose in the academy. To quote one study: “Unless libraries take action...they risk being left with responsibility for low-margin services that no one else (including the commercial world) wants to provide” (“A continuing discussion”, 2008, p. 4).

Academic libraries, regardless of Carnegie designation, share a common mission to support the teaching and learning enterprise, and the fulfillment of that mission amid today’s pressures is increasingly linked to intelligence about resource consumption, service quality, and the library’s impact on research and student learning. Clearly, libraries have entered a period where measurement and mission are inextricably linked, where effective management is evidence based management (Wilson, 2008).

The challenges of the past decade have sparked a keen interest in assessment and an even sharper focus on accountability and the elusive questions of what to measure and how (Luce, 2008). ARL’s commendable reevaluation of the statistical canon notwithstanding, only nominal progress has been made on new metrics or on the critical problem of assembling data for effective, cost-efficient, and sustainable assessment. Further, some of the most promising work has originated outside the ARL community, for example, in the Los Alamos Digital Library’s MESUR initiative and Project COUNTER. JISC is another source of good recent work that sheds light on tools, methods, and developmental pathways for business intelligence in libraries (Kay & Van Harmelen, 2012).

ARL has had notable success at building a nascent community of practice around library assessment, elevating quantitative methods employed within the community through LibQUAL® and other initiatives. But if libraries are to link evidence to management and planning effectively, the assessment effort will require additional focus, leadership, tools, and technical infrastructure. The thrust toward evidence based management has been particularly hobbled by the problem of gathering and mining information from data—vast amounts of data arising from service and user interaction with librarians. Until data can be quickly and routinely harvested and made ready for study, the evolving community of practice, along with effective leadership in assessment, will struggle to coalesce.

This situation seems paradoxical given that nearly every library service leaves some kind of data trail to mine, from circulation records to e-journal logs to emails about research questions. Enormous in size and potential, these trails of evidence are as inaccessible as they are ubiquitous; they are locked up in silos that bar retrieval and thwart investigation; they are expensive and complicated to render usable. At the present time, assessment's most critical assets are, in effect, the detritus of library systems—traces in the clickstream captured by some log or millions of transaction records stored in an esoteric database table.

Libraries are not wanting for analytical methods, even if the data they need are hard to reach. A variety of protocols has been developed in recent years including: a means to analyze the depth of reference services, to measure the impact of networked electronic resources, and to estimate return on investment (ROI) in academic libraries (Gerlich & Berard, n.d.; Franklin & Plum, 2008; Kaufman, 2008). But in each case, the commodity most critical to sustained, productive use of these methods is also the hardest resource to muster. Liberating an institution's data and converting them into knowledge which informs budgetary decisions,

staff allocations, new service models, and a sophisticated understanding of research output and scholarly workflows is fundamentally important to evidence based practice and, by extension, to the course of libraries and the universities they serve. Duderstadt argues that the evolution of the library in the digital age prefigures the evolution of the university: "In a sense the university library may be the most important observation post for studying how students really learn. If the core competency of the university is the capacity to build collaborative spaces, both real and intellectual, then the changing nature of the library may be a paradigm for the changing nature of the university itself." (2009, p. 220) This reasoning underscores the critical need for an improved understanding of how scholars interact with and use the services that libraries provide.

Meaning and Scope of a Decision Support System

As an enterprise approach to systematic decision support, the University of Pennsylvania Libraries (Penn Libraries) is developing MetriDoc to provide an information technology (IT) infrastructure that facilitates the collection and transport of data. As such, our goal is to address the assessment challenge cited above, specifically to unlock the vast and rich data reserves that libraries possess and to tap them for planning and decision-making.

MetriDoc constitutes several layers of a tiered Decision Support System (DSS). In the literature, the concept of DSS has many connotations, which encompass technology but also speak to the non-technical facets of data administration and evidence based management. For present purposes, I follow Turban, Leidner, McLean, and Wetherbe (2004) in describing a DSS as:

"a computer-based information system that combines models and data in an attempt to solve semi-structured and some unstructured problems with extensive user involvement." (pp. 550)

Again, following Turban et al. the MetriDoc approach to a DSS possesses these features:

- 1) Data Management Layer: the range of data that originate from disparate sources and are targeted for harvest into a database or repository layer of the DSS. (As Turban points out, extract into a database is not a prerequisite of the DSS, but that is the method we employ with MetriDoc.).
- 2) Model Management and Data Governance: structural components of data that form the building blocks of DSS applications and require continuous coordination with the production systems that generate transactional data.
- 3) Data Warehouse: repository of refined, normalized data from raw sources.
- 4) User Interface Layer: a discovery interface that aids users in identifying and isolating relevant data, performs basic aggregation and analysis, and outputs results to dashboards, feeds (RSS and/or Atom), structured reports, or even integrates with third-party applications such as Excel, SAS, R, or Software Environment for the Advancement of Scholarly Research SEASR.

As Lakos and Phipps (2004) have noted, the management of library services employs multiple data sources that often have overlapping relationships, such as the linkages between expenditure and use, or the more complex interconnections between user populations and resource consumption. For this reason, a single, integrated DSS should be developed that supports sophisticated use of both descriptive and inferential statistics. The DSS should make quantitative information readily available and easy to access by all levels of staff. Data should be routinely harvested, modeled, updated, and archived. A management structure should be in place with sufficient staffing and executive support to deal

with data governance issues and manage the flow of quantitative information throughout the organization.

Options for Developing DSS Capabilities

The case for developing decision support systems for libraries dates back to at least the 1980s (McClure, 1980). By the late 1990s, the idea had found a prolific champion in Amos Lakos (1998), whose work with Shelley Phipps (Lakos & Phipps, 2004) gives a prominent place to the DSS in furthering what is commonly termed the culture of assessment in libraries.

Though the need for such systems is well established in the literature, there has been little institutional investment in their creation. Lakos cites automated DSS systems in some stage of development at only a handful of universities, including the Penn Libraries' Data Farm project, which we discuss in more detail below (1998).

The rarity of DSS projects in the academic library community, particularly given the need to clarify mission, optimize finances, and cultivate new services and management methods, testifies to the difficulty and expense of the endeavor.

The Commercial Development Sphere

For the majority of library administrators, keeping pace with mission-critical technologies, such as their Integrated Library Systems (ILS) and web applications, absorbs most of the staff and technical expertise available to them. As a result, the appeal of vendor support in this realm is especially strong.

All ILS vendors provide some level of report writing, but these capabilities are deeply integrated into the architecture of proprietary systems and thus fail to provide the flexibility or richness of data analysis that libraries need. OCLC's WorldCat Collection Analysis tool is yet another of these "blackbox" solutions. Regardless of their strengths or flaws, both the

ILS and OCLC provide business intelligence primarily about print collections; gathering and processing data on other aspects of library services would involve a multiplicity of systems, which works against the need for economy and integration in a DSS solution. The DSS space is also occupied by commercial firms active in the university market. Here the need is for enterprise level data warehousing to provide metrics related to admissions, student performance, retention, and the like. Firms in data warehousing have not made a foothold in libraries due to the expense of implementation and support.

Whether the commercial sphere is prepared to engage with libraries and the complicated mix of data sources they handle is unclear. Libraries need to integrate budgetary data, bibliographic measures, web analytics, personnel information, courseware measures, and a wide range of usage data from local and licensed sources. While library-oriented data warehousing systems have appeared from vendors, they require substantial contributions and start-up costs involving a range of library staff to implement. The ongoing costs for a commercial solution are uncertain, but clearly, libraries will have no control or proprietary stake in the products they are helping vendors to design and market. In the end, a proprietary solution will struggle to satisfy the scope of library needs, but it will add extraordinary new costs and slow deployment of DSS technology. The commercial option is also apt to inhibit prospects for multi-institutional collaboration around metrics, just as the commercial ILS inhibits cooperative efforts by hardening the silos around data and systems architecture.

Community Development Model

A development role in DSS, under an open or community source model, would be advantageous to the library community, specifically enabling:

- maximization of local data reserves,
- effective use and development of domain expertise,
- financial and functional sustainability, and
- infrastructure required for collaborative research and development.

Community-sourcing does not exclude commercial interests, but changes the fundamental dynamics of the library market, allowing vendors and libraries to forge new relationships around the support of software and the extension of that intellectual property for the best interests of the community. Open development of a metrics framework insulates libraries from a destabilizing reliance on vendors for product development and support, while also building a knowledge base that strengthens intra- and inter-institutional cooperation around strategic problems. Open development can also spur competency-building within the library community, encouraging the acquisition of statistical skills and creating professional opportunities around data modeling, metadata design, and data governance, in addition to statistical methods and presentation.

MetriDoc: A System Overview

With the help of the Institute for Museum and Libraries Services (IMLS), the Penn Libraries have been working on the feasibility of creating a DSS for library activity data, and have developed a deployable, extensible technology, MetriDoc, that other libraries can use to broach the challenge. MetriDoc is a means of “lighting up” an array of data sources to build a comprehensive repository of quantitative information about services and user behavior. A data source can be a database, text file, XML, or any binary object that contains data that has business value. MetriDoc provides simple tools to extract useful information from various data sources, transform, resolve and consolidate that data, and finally store them in a repository. The

repository is comprised of various storage mechanisms to make it easy to extract data for reports and statistical processes. With this in mind, the Penn Libraries are designing MetriDoc to meet the following requirements:

- create a simple framework that handles the complexities of extracting, resolving and storing data
- provide hooks into the framework so non-enterprise programmers can use Metridoc with a combination of scripting languages, XML and project schemas
- create reusable solutions specific to the library space, such as extracting data from popular ILS systems, handling COUNTER data, resolving EZproxy logs, etc.
- follow best practices when storing and curating data in the repository to enable the widest possible distribution of decision-support information so that data analysis can become a routine and continuous facet of organizational administration and culture.

MetriDoc must be understood within the context of the Penn Libraries' Data Farm initiative. The Data Farm website (<http://metridoc.library.upenn.edu/>) has authentication controls, but this page suggests features available to staff. That said, a number of Data Farm functions deliver data on schedules directly to managers and do not require interaction with the web. In addition, Penn Libraries Management Information Services provide considerable ad hoc analyses from Data Farm sources.

A program that began in 2000, the Data Farm represents a substantial institutional investment in assessment. In brief, the Data Farm is a "collection" of DDS functions that run on a common Oracle instance and output to the web or Excel (Cullen, 2005; Zucca, 2003). The underlying data come from a variety of sources, for example: the Voyager ILS system, Apache

web server logs, a local database that powers segments of the Penn Libraries website for metrics on e-resource usage, COUNTER data from vendors (this includes a Penn-designed SUSHI harvester which we deploy in MetriDoc), and input from public services staff who consult with students and do bibliographic instruction. The Data Farm is also the reporting utility for the BorrowDirect and EZ-Borrow programs (two large-scale resource sharing cooperatives in the Northeast). The Data Farm is used heavily by more than 70 members of these cooperatives, as well as Penn Libraries bibliographers, public service managers, and Strategic Planning Team. But for all of that, in certain fundamental respects the Data Farm is a prototype for study and experimentation.

MetriDoc represents a more rigorous phase of Data Farm development, and leverages the knowledge the Penn Libraries have gained since 2000. The key points of distinction between Data Farm and MetriDoc are represented in Table 1.

The four service layers comprising MetriDoc support the following functions:

- 1) Extraction of raw data sources. Routines within MetriDoc are designed to "recognize" specific data structures and extract what is of primary interest to measurement, for example, relevant information from a log or database.
- 2) Transformation of the raw extract into normalized, decoded information (such as the resolving of ISSN numbers into a serial title, or an SFX object identifier into citation elements). Transformation is a complex but critical process that sets the stage for the third function,
- 3) the loading of normalized and anonymized data into a query-able data repository. The fourth MetriDoc tier sits above the other three (ETL) service layers and allows for the integration of the data repository with statistical analysis and visualization tools, or the

distribution of flat files for use with statistical programs.

The MetriDoc service layers are more fully described below and illustrated in Figure 1.

1. Extraction Service – The extraction API, or application programming interface, can be accessed directly with code via scripts. This process creates the payload for ingestion by the MetriDoc repository – in most cases a data construct that defines a database table and rules for validation.

2. Transformation Service – Data elements within a log stream often include encoded or identity information. Encoded data must be resolved to capture the meaningful information for analysis and reporting. For example, Digital Object Identifiers (DOI) or ISSN numbers are commonly used to identify specific instances of articles or journal titles. Identity information provides useful demographic class descriptions about a user's department, status, and rank. The MetriDoc Resolution Service consists of processes that tap external data sources, such as national bibliographic utilities or the university data warehouse, and query for matching content

from these sources. Once deployed, these resolvers can be linked in order to resolve data points iteratively within a log or other data source. The MetriDoc document is returned to the messaging channel with enriched data about the bibliographic and demographic components of service events.

3. MetriDoc Repository Service – MetriDoc provides a repository service that houses MetriDoc event data processed from source files and exposes that data for user query and retrieval. This service abstracts the actual data store to provide scalability and flexibility, and can comprise a wide variety of repositories, from relational databases such as Oracle or MySQL to a mere file system. Additionally, abstraction allows storage to be distributed across physical locations for improved resiliency and fault tolerance.

4. Data Farm Service Layer – The MetriDoc architecture abstracts user interaction from the ETL components of the framework. In the Penn Libraries context, interactive services are supported by the Data Farm Service layer, which can be developed using a variety of

Table 1
Data Farm and MetriDoc Structural Features

Data Farm Structural Features	MetriDoc Structural Features
Builds a specific extraction and ingestion tool for each type of data source.	Abstracts the ingestion process and delegates specific extraction to small pieces of code.
Builds source-specific data structures in an Oracle tablespace.	Generalizes each log transaction into an abstract representation of an “event.”
Resolves identity and bibliographic data after ingestion.	Resolves identity data on the fly from a rich and diverse set of resolution sources.
Exposes a single discovery interface, tightly coupled with the end-user tool.	Isolates discovery of datasets and provides workflow tools to combine, refine, analyze, and augment data, and then expose it through a multifaceted delivery service layer.
Comprises a single technology stack.	Composed of loosely coupled service layers consisting of four distinct services that are integrated through easy-to-use, RESTful interfaces.

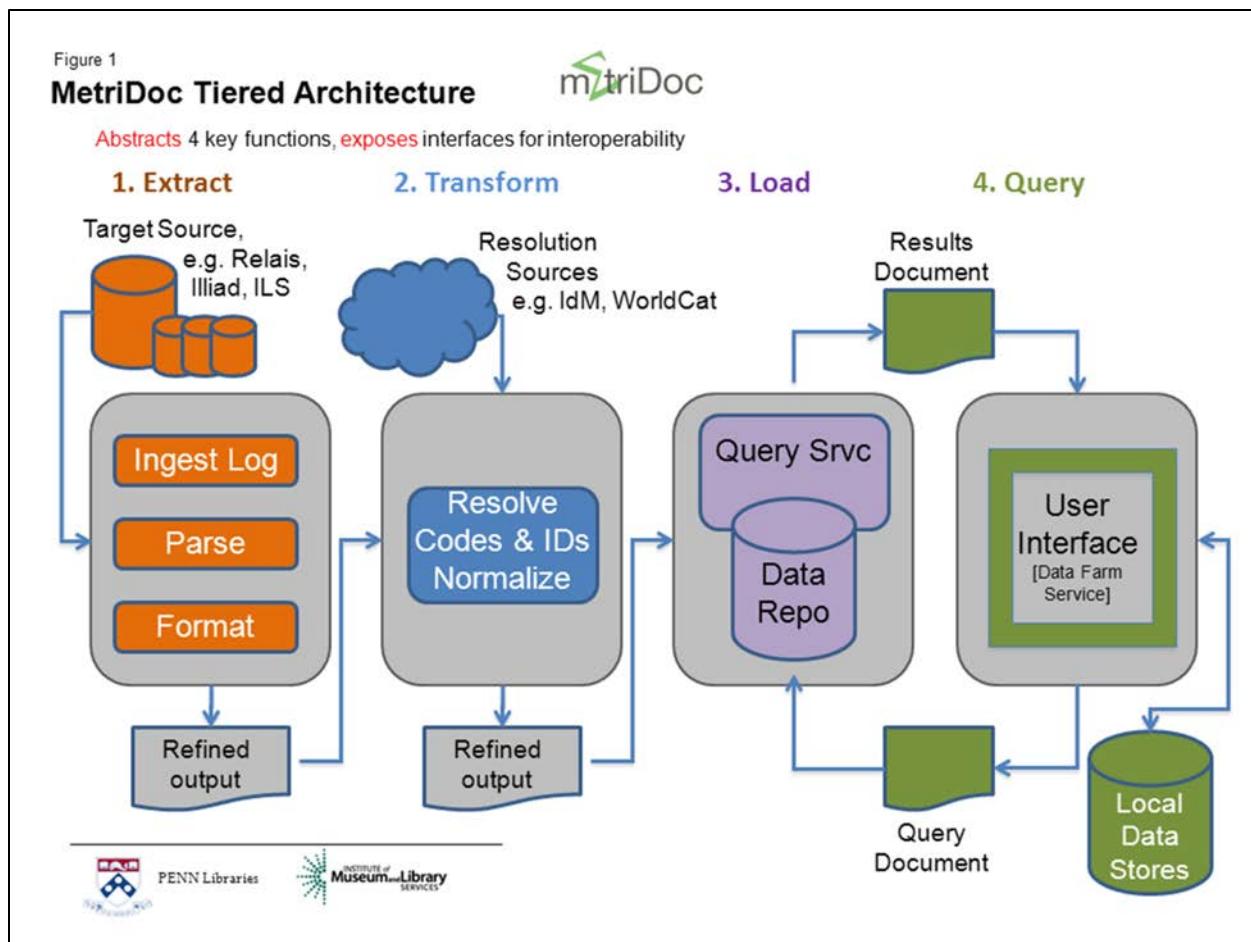


Figure 1
MetriDoc tiered architecture

commercial tools or locally designed solutions. By design, the MetriDoc repository can be exposed to report-building applications via a RESTful interface, or to scripts that generate dashboard pages, datasets in Excel format for download, or comma delimited files for ingestion into a third-party analytics repository such as eThority. In this last scenario, the Data Farm Service can contain an extensible repository with a library of datasets and data visualizations, and the ability to create refined datasets for analysis, using a statistical language such as R or SAS. This service can support analysis tools that are shared across domains to assist in comparison, reporting, and analysis.

The four MetriDoc service layers are an orchestrated chain of services that ingest, resolve, normalize, store, index, query, deliver and transform event data regardless of their native structures. It is designed to provide flexibility, extensibility, and consistency to data flows. The technologies used are common in enterprise applications including Spring, Hibernate, Java, and Grails.

Current Development

With funding from the IMLS received in 2010/11, the Penn Libraries completed five reference projects to prove basic concepts of the

MetriDoc framework and make available a set of applications that other institutions could test in a deployment of the MetriDoc core. These reference projects are written as configurable plugins to the core framework and can be used to parse and store EZ-Proxy log data, COUNTER data (with and accompanying plugin for data harvest with SUSHI), ILL transactional data from ILLIAD, fund expenditure data from the Voyager ILS, and transactional data from the Relais platform which supports the BorrowDirect and EZBorrow resource sharing consortiums. The projects represent a range of challenges and repository concepts that a DDS will encounter in a library setting.

As of this writing the Penn Libraries are also developing a MetriDoc module for data related to research consultations and bibliographic instruction services. The MetriDoc framework is currently undergoing test implementations at the University of Chicago and North Carolina State University, and the Kuali-OLE (Open Library Environment) project is actively considering it as the basis of an analytics module that will ship with OLE.

Benefits of Collaboration

The purpose of MetriDoc is to make available vast, unutilized quantitative information in support of library strategic planning and decision-making. Success in this endeavor opens a range of partnership opportunities. Deployed in a collective environment, a MetriDoc-like framework can:

- provide libraries a tool for conducting the foundational research leading to new performance metrics;
- aid cross-institutional study of collections, which advances collaborative collection development;
- be deployed in resource-sharing initiatives which will help partners identify best practices and optimize the distribution of physical materials;

- increase an institution's knowledge of local research interests and patterns through the demographic analysis of transaction records;
- expose metadata based on resource use to discovery systems for improved resource access and research intelligence;
- enable the integration of usage and expenditure data to identify cost efficiencies and help libraries apportion budgets more effectively across communities;
- gather electronic use data on both locally created and licensed digital resources; and
- provide a platform for relating usage information to customer satisfaction and other parametric measures of quality.

Conclusion

Powerful new tools for visualizing and distributing data are available to the assessment community. Measurement standards for library performance and the potential for creating a robust canon of library metrics are also within reach. The challenge remaining is posed by the data: by the complex and ornery problem of harvesting, structuring, and storing the vast troves of activity data resting dormant in the systems libraries all use to conduct business. MetriDoc, and ETL solutions generally provide an answer to this problem.

The academic library community faces some tough decisions with regard to business intelligence. First, this is not an assessment project, but a matter of technical and staff infrastructure, on the level of our commitments to ILS technology and similar IT supported functions. It is, additionally, an area requiring development resources, as there are no shrink-wrap solutions for our particular challenges. Infrastructure creation and development are expensive activities and will test the importance of business intelligence in the spectrum of this community's strategic priorities.

In the end, libraries will or will not rank this as strategically important. If it is, deep collaboration will be essential to progress, given the costs and complexity of the challenge. A community effort on business intelligence infrastructure can expedite innovation and instigate new relationships among academic institutions and between the academy and commercial sector. But how will this deep collaboration come about? One wonders if this is an area where ARL can be an effective broker, providing a space for potential partners to begin addressing the challenge of creating and governing a critical new infrastructure for managing library services. Such an effort is afoot in the U.K. where, under JISC sponsorship, the focus by libraries on activity data is picking up steam and maturing faster than here in the States. The MetriDoc effort has joined that conversation even as it looks for development partners closer to home (Zucca, 2012).

References

- A continuing discussion on research libraries in the 21st century. (2008). In *No brief candle: Reconceiving research libraries for the 21st century* (pp. 1-12). Washington, DC: Council on Library and Information Resources (CLIR).
- Association of Research Libraries (ARL). (2001). E-Metrics: Measures for Electronic Resources. In *Association of Research Libraries*. Retrieved January 2010. Retrieved 18 May 2013 from <http://www.arl.org/stats/initiatives/emerics/index.shtml>
- Cullen, K. (2005). Delving into data: businesses have used data mining for years, now libraries are getting into the act. *Library Journal*, 130 (13), 30-32.
- Duderstadt, J. J. (2009). Possible futures for the research library in the 21st century.
- Journal of Library Administration, 49(3), 217-225. doi:10.1080/01930820902784770
- Franklin, B., & Plum, T. (2008). Assessing the value and impact of digital content. *Journal of Library Administration*, 48(1), 41-47. doi:10.1080/01930820802029334
- Gentleman, R., & Ihaka R., et al. (n.d.) *R project for statistical computing*. Retrieved 18 May 2013 from <http://www.r-project.org/>
- Gerlich, B. K., & Berard, G. L. (n.d) *Reference Effort Assessment Data (READ) scale*. Retrieved 18 May 2013 from <http://readscale.org/>
- International Coalition of Library Consortia. (2006). Revised guidelines for statistical measures of usage of Web-based information resources. In *International Coalition of Library Consortia (ICOLC)*. Retrieved 18 May 2013 from <http://iclc.net/statement/revised-guidelines-statistical-measures-usage-web-based-information-resources>
- Kaufman, P. T. (2008, Jan.). *University investment in the library: What's the return? A case study at the University of Illinois at Urbana-Champaign*. American Library Association MidWinter Meeting, Philadelphia, PA, USA. Retrieved 18 May 2013 from <http://hdl.handle.net/2142/3587>
- Kay, D., & Van Harmelen, M. (2012). Activity data: Delivering benefits from the data deluge. In *Jisc*. Retrieved 18 May 2013 from <http://www.jisc.ac.uk/publications/reports/2012/activity-data-delivering-benefits.aspx>
- Lakos, A. (1998). Building a culture of assessment in academic libraries: Obstacles and possibilities. *Living the*

- Future II, Tucson, AZ, USA. Retrieved 18 May 2013 from
<http://www.library.arizona.edu/conferences/ltf/1998/proceeds.htm>
- Lakos, A., & Phipps, S. (2004). Creating a culture of assessment: A catalyst for organizational change. In *portal: Libraries & the Academy*, 4(3), 345-361. Retrieved 18 May 2013 from
http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v004/4.3lakos.html
- Luce, R. (2008). Raising the assessment bar: A challenge to our community. In S. Hiller, K. Justh, M. Kyrialladou, & J. Self (Eds.), *Proceedings of the 2008 Library Assessment Conference* (pp. 7-11). Seattle: Association of Research Libraries. Retrieved 18 May 2013 from
<http://libraryassessment.org/bm~doc/proceedings-lac-2008.pdf>
- McClure, C. (1980). *Information for academic library decision making: The case for organizational information management*. Westport, CT: Greenwood Press.
- Turban, E., Leidner, D., McLean, E., & Wetherbe, J. (2004). *Information technology for management: Transforming organizations in the digital economy*. Hoboken, NJ: J. Wiley.
- Wilson, B. (2008). Accelerating Relevance. In S. Hiller, K. Justh, M. Kyrialladou, & J. Self (Eds.), *Proceedings of the 2008 Library Assessment Conference* (p. 14). Seattle: Association of Research Libraries. Retrieved 18 May 2013 from
<http://libraryassessment.org/bm~doc/proceedings-lac-2008.pdf>
- Zucca, J. (2003). Traces in the clickstream: Early work on a management information repository at the University of Pennsylvania. *Information Technology & Libraries*, 22(4), 175-179.
- Zucca, J. (2012). Metridoc is an extensible framework that supports library assessment and analytics, using a wide variety of activity data collected from heterogeneous sources. In *Jisc*. Retrieved 18 May 2013 from
<http://www.jisc.ac.uk/publications/reports/2012/activity-data-delivering-benefits.aspx#penn>