

Poster Session Descriptions

Databib: An Online Bibliography of Research Data Repositories

Michael Witt, Mike Giarlo

A number of academic and research libraries are beginning to take a more active role in research data curation on their campuses, applying library science principles to help address the data deluge. An expanded role for research libraries in digital data stewardship was forecasted by an Association of Research Libraries (ARL) workshop report to the NSF in 2006. This forecast was substantiated in August 2010 by a survey of 57 ARL libraries, of which 21 libraries reported that they currently provide infrastructure or support services for e-Science, and an additional 23 libraries are in planning stages.

These services encompass a wide range of activities such as helping researchers formulate funder-required data plans, adapting library practice to help organize and describe research datasets, developing data collections and data repositories, and digital preservation. Some librarians are adapting instruction and reference approaches to directly address data needs, for example, by offering data literacy and data reference—helping patrons find data and integrate it into their learning, teaching, and research.

Librarians are in a good position to provide these services; unfortunately, there is currently no framework in place to support the organization and discovery of data repositories. Many funding agencies are requiring their sponsored researchers to submit their data to repositories without giving further instructions to them. What repositories are appropriate for a researcher to submit his or her data to? How do potential users find appropriate data repositories and discover datasets that meet their needs? How can librarians help patrons who are looking for data find and integrate it into the patrons' research, learning, or teaching?

To help meet these needs, the Institute of Library and Museum Services awarded a Sparks! Innovation Grant to the Purdue and Penn State Universities to develop an online, community-driven, annotated bibliography of research data repositories. This tool, called Databib (<http://databib.lib.purdue.edu>), seeks to become an important resource to data librarians, users, producers, and funders. The Databib platform attempts to challenge the traditional concept of a bibliography by employing the latest technologies to instantiate bibliographic content and integrate it in Web 1.0, Web 2.0, and Web 3.0 environments. One technology in particular, Linked Data, shows a great deal of promise for delivering a “web of data” and giving librarians a new toolkit for describing and classifying data by relating it to other data.

Exploring Methods of Using Public Data on Twitter for Library Assessment

David Meincke

Performing a thorough, quality assessment of library services can be difficult, the results ambiguous, and the overall process time-consuming. As many measures of overall library satisfaction assessment involve considerable costs—including performing surveys, institutional review boards, providing incentives for the survey-takers—libraries are often faced with the choice to either forgo a full assessment of services, or choose alternative methods.

Beginning in January 2011 the Digital Services Librarian at Johnson & Wales University (JWU) Library started to explore the possibility of using public data to assess patrons' satisfaction with library services. For a year, the librarian archived tweets that matched criteria that would indicate the tweet was directly or indirectly related to the JWU Library. Around 400 tweets were collected using four Twitter searches (using keywords and location-based metadata. When the data was cleaned up and various 'noisy' tweets (not pertinent to the JWU library) removed, around 200 tweets remained.

A variety of methods were initially used to analyze the data. First, the library attempted to use sentiment analysis software, but found that it was relatively imprecise in gauging the true tone of the tweets.

A grounded theory approach proved to be the best way to handle the data, with each tweet assigned a code family and a code. The code families included: 'location' (the patron's presence and activity in the library), technology (software, website, and hardware), facilities (the library itself), and service (books, databases, staff, reference. Sentiment was also encoded into the dataset. The dataset then allowed for an analysis of satisfaction on various topics.

Since a single tweet generally only carries an expression of sentiment of only one aspect (or code) relating to library service, the medium of Twitter proved to be remarkably effective in a way a more complex dataset (e.g. an IM transcript) would be.

Many areas for further study and discussion were identified, including privacy issues, consent, and whether or not an IRB request should be sought after all. Also, if a future study is conducted via Twitter sentiment analysis a concurrent study by another method (e.g. transcript analysis, formal survey) may prove useful.

From the field to the desktop: how online access to historical botanical collections, data and literature creates a world of information at your fingertips

Larry Schmidt

Herbarium specimens are important to scientists looking at environmental change over time, new taxonomic features and the classification of new species. Early collectors in North America have supplied herbaria throughout the world with their material. This content has moved online and the data behind these collections can now be accessed on the Internet. Building these collections takes a significant amount of knowledge of metadata schema, scientific nomenclature as well as technical knowledge. The creation of a digital collection of herbarium specimens from one National park now includes type specimens, multiple national park collections, field notebooks of early collectors and geospatial

information through Google Maps. Creating the links between the herbarium specimens (with images), field book entries, historical maps and the current maps with location information is now possible. This presentation will show how libraries have partnered with botanists to create a holistic interface that can be used by scientists, historians and other individuals interested in botany, collections and collecting in the American West.

Implementing DMPTool [Data Management Planning Tool]

Hong Ma

Research Libraries are facing many new opportunities and challenges to define their role in the emerging E-Science arena. Researchers are required to supply detailed, cost-effective plans for managing research data by a number of U.S. funding agencies such as National Science Foundation (NSF), National Institute of Health (NIH), etc. Libraries are seeking strategies to develop possible online data management services to support researchers to establish their Data Management Plans.

DMPTool can help researchers meet the data requirements from funding agencies by providing step-by-step instruction and guidance for developing a data management plan. The DMPTool [Data Management Planning Tool] open source software was developed by a group of research institutions partners: The California Digital Library, DataONE, Digital Curation Centre, Smithsonian Institution, UC Los Angeles, UC San Diego, University of Illinois, and University of Virginia. In this poster presentation, we will share our experience being a contributing institution for the DMPTool software. We added local information to lead local researchers to related resources, such as additional information from our institution and library system. We customized the Data Management plan template to meet our local needs and implemented full Shibboleth authentication to lead researchers to our localized templates by using their campus credentials. The details of implementation, customization, promotion and training strategies will be covered.

Kindred Works: developing a content-based recommendation service for WorldCat

Diane Vizine-Goetz, JD Shipengrover, Roger Thompson, Harry Wagner

The prototype consists of a user interface, Kindred Works, and a machine service.

This project applies principles of the FRBR model to aggregate library metadata into work-level summaries for resources described in WorldCat. The summaries, which include classification numbers, subject terms, and genre characteristics, are used to determine relatedness.

The machine service accepts a query, such as OCLC number or ISBN, and returns a list of related resources. An individual OCLC member library may customize the machine service to return related resources from its collection.

The presentation will provide overviews of the metadata aggregation and recommendation logic, demonstrations of the user interface and machine service, and plan for future improvements.

LibX 2.0

Annette Bailey

In this presentation, the newly released, IMLS funded LibX 2.0 infrastructure will be covered. LibX 2.0 is a new technology infrastructure which allows librarians to easily create new web applications for their patrons. Their patrons can then use the LibX plug-in to install and experience the web with customized applications that librarians create. Users are then able to further customize LibX. We will describe the underlying open source technology and encourage audience participation in creating library applications during the session.

"Pop-up" Your Data!

Andrea Leonard

Do you have trouble understanding where the communications and data exchange breakdowns are happening between your OpenURL link resolver, databases, and your serials management system? Do you find it challenging to verbalize exactly what is happening effectively to technical support so that the problem can be truly solved? Are you a kinesthetic and visual learner in a sometimes untouchable virtual world?

Do you love pop-up books?

Just because you are the E-resources Librarian doesn't mean that your learning style lends itself to the often very flat data and I.T.-driven online world. This poster session will share a way to create a simple paper pop-up of your e-resources systems so that you can actually see, touch, and follow the exchange of data. You'll not only begin to be able to verbalize problems better, but to state more clearly why the development of rigorous metadata standards is so crucial when you "see", for example, that your link resolver needs data for a dissertation, but is sent the title of the database as a 'Book Title' instead.

A "pop-up" of your e-resources systems may not only facilitate trouble-shooting and your own understanding, but could demystify your e-resources systems for your colleagues in the library and others who want to understand. And you don't have to be artistic!

Research Data Management Training for Graduate Students: An Iterative Approach

Rebecca Reznik-Zellen

The University of Massachusetts Amherst Libraries' Data Working Group supports faculty and graduate students who need to meet funder expectations for data management, analyze existing data management practices, or create new practices that best fit the needs of their research projects. This support includes education and training services. With 6,000 graduate students in 51 doctoral and 73 master's degree programs, the graduate student population at the University of Massachusetts is a significant one. Since the fall of 2011, the Data Working Group has offered a series of workshops specifically for graduate students on Data Management Basics.

In conducting these workshops, the Data Working Group observed a significant portion of graduate students who had no prior experience with smart data practices or useful data management resources. At the same time, workshop interactions and evaluations indicate that graduate students work with increasingly large and complex datasets and seek practical, technical solutions to their data management problems. In addition, the diversity of disciplinary data needs across this population creates challenges in presenting basic information in a relevant way. The Data Working Group has continually modified the workshop outreach, content, and format to try to address these issues. This ongoing process has identified a clear need for wider, more intensive education for graduate students on data practices and the data management requirements of national funding agencies.

This poster will describe the processes, challenges, and outcomes of the University of Massachusetts Amherst Libraries' Data Working Group's series of data management training workshops for graduate students.

Robust 'Altmetrics' as a Framework for Measuring Item Usage and Researcher Impact in Institutional Repositories

Stacy Konkiel

Select academic journal publishers and subject repositories have implemented tools that allow authors to see their "impact" at an article level. These usage indicators, called "altmetrics," measure article citations, download counts, pageviews, bookmarks on academic social reference manager sites, and mentions on Facebook and Twitter. Some altmetrics have been shown to predict the likelihood of papers receiving more traditional measures of impact such as citations (Eysenbach, 2011), and a growing number of scholars are calling for altmetrics' recognition as a supplementary measure of an article's influence (Groth & Gurney, 2010; Neylon & Wu, 2009; Priem et al, 2010; Priem & Hemminger, 2010; Taraborielli, 2008). Should institutional repositories offer altmetrics for the research they hold? If so, which metrics? Which tools are best suited to aid in implementation? What barriers to participation exist for repository managers? Using examples from the research literature, academic publishers, and subject and institutional repositories, we explore these questions.

Taking Risks: Improving Productivity, Collaboration, Accountability, and Transparency by Adopting New Project Management Techniques

Mark P. Baggett

Properly handling projects from the initiation phase to closure can be challenging for library technology departments. Often, these departments are viewed as ivory towers that have no regard for others within the library. As a support unit, it is critical that these departments be systematic in their decision-making process regarding project selection and be transparent and accountable throughout the project lifecycle.

Recently, the Systems Department at the University of Tennessee changed the way that project requests are handled. Traditionally, the department attempted to satisfy other library groups by agreeing to take

on any and all requests, despite limited resources. Unfortunately, this method led to projects going largely unfinished or projects being ignored. Most importantly, it often resulted in focusing on the simplest tasks, regardless of importance. In response, the department adopted a strategic plan, a technical reference model, and a formal request process in order to make decisions based on the best interests of the library and to properly allocate resources. Furthermore, the department adopted a project management tool, ActiveCollab, to help manage projects after selection and throughout the project lifecycle. With this tool, tasks can be assigned to specific groups or individuals collaborating on a project while allowing others within the library to see what work is being done and who is responsible. This poster will describe in detail how this tool works and how combining it with a formal decision-making process can improve interdepartmental transparency, accountability, productivity, and collaboration. Adopting these tools was not without risk, however, as it could have been seen as intimidating or too complex to other library staff.

Using the Data Curation Profiles Toolkit to Inform the Development of a Research Data Registry

Gail Steinhart, Leslie McIntosh, Kathy Chiang, Dianne Dietrich, Wendy Kozlowski, Huda Khan, Sarah Wright

DataStaR is a semantic web application in development at Cornell University, which is intended to serve as a data set registry to promote discovery of research data sets by making basic metadata available as linked open data. The application will support the basic description of research data, links to data sets in any location, and publication of data sets to selected repositories supporting the SWORD (Simple Web-service Offering Repository Deposit) protocol. We know from a survey of Cornell researchers that the basic data set registry function planned for DataStaR is a service they are interested in using to demonstrate compliance with funders' requirements for data sharing. To ensure that more specific development plans are based on researchers' actual needs and behaviors, we adapted Purdue University Libraries' Data Curation Profile Toolkit to inform our software development plans. We made minor modifications to the toolkit's interview instrume

nt, completed eight data curation profiles with the assistance of researchers at Cornell University and Washington University in St. Louis, and analyzed the results to develop a set of recommendations to guide further development of the DataStaR platform. We offer some reflections on our specific findings as well as some general comments on the utility of the Data Curation Profiles Toolkit for this particular purpose. The DataStaR project is supported by a grant from the Institute for Museum and Library Services.