

2012

New England Technical Services Librarians Spring 2012 Conference: iLibrary: Digital Futures for Libraries (May 3, 2012, Worcester, MA)

Andrée J. Rathemacher
University of Rhode Island, andree@uri.edu

Michael A. Cerbo II
University of Rhode Island, mcerbo@uri.edu

See next page for additional authors

Follow this and additional works at: http://digitalcommons.uri.edu/lib_ts_pubs



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

Citation/Publisher Attribution

Rathemacher, Andrée J.; Cerbo, Michael A. II; and Lovett, Julia A., "New England Technical Services Librarians Spring 2012 Conference: iLibrary: Digital Futures for Libraries (May 3, 2012, Worcester, MA)" (2012). *Technical Services Department Faculty Publications*. Paper 60.
http://digitalcommons.uri.edu/lib_ts_pubs/60

This Article is brought to you for free and open access by the Technical Services at DigitalCommons@URI. It has been accepted for inclusion in Technical Services Department Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

Authors

Andrée J. Rathemacher, Michael A. Cerbo II, and Julia A. Lovett

<p>Andrée J. Rathemacher Head, Acquisitions University Libraries, University of Rhode Island 15 Lippitt Road Kingston, RI 02881-2011 Phone: (401) 874-5096 Fax: (401) 874-4588 E-mail: andree@uri.edu</p>	<p>Julia A. Lovett Digital Initiatives Librarian University Libraries, University of Rhode Island 15 Lippitt Road Kingston, RI 02881-2011 Phone: (401) 874-5079 Fax: (401) 874-4588 E-mail: jalovett@mail.uri.edu</p>	<p>Michael A. Cerbo II Bibliographic Access & Resource Management Librarian University Libraries, University of Rhode Island 15 Lippitt Road Kingston, RI 02881-2011 Phone: (401) 874-5967 Fax: (401) 874-4588 E-mail: mcerbo@uri.edu</p>
--	--	--

New England Technical Services Librarians Spring 2012 Conference: iLibrary: Digital Futures for Libraries (May 3, 2012, Worcester, MA)

Andrée J. Rathemacher, Julia A. Lovett, and Michael A. Cerbo II

Abstract

This report discusses the program of the 2012 New England Technical Services Librarians (NETSL) annual spring conference, held on Thursday, May 3 at the College of the Holy Cross in Worcester, Massachusetts, titled “iLibrary: Digital Futures for Libraries.” NETSL is a section of the New England Library Association and a regional group of the American Library Association.

Big Data: Dig Deal? New Challenges for Scholars & Librarians

John Unsworth (vice provost, Library & Technology Services and chief information officer, Brandeis University) delivered the conference's keynote address. His presentation highlighted large-scale book digitization efforts, barriers to access due to copyright, and issues around mining textual data for computational analysis.

Unsworth first gave a brief history of book digitization efforts, from Project Gutenberg (1971) to the HathiTrust Digital Library (2008) and the Digital Public Library of America (2010). In 2001, the scale of book digitization took a major leap forward, crossing a boundary from thousands to millions of books being scanned. HathiTrust now holds over ten million volumes, most of which were digitized through the Google Books Library Project. These holdings constitute approximately 3.5 billion pages and 458 terabytes of data—the vast majority of which is still in copyright and therefore not accessible to users. Copyright law presents a major stumbling block to making use of digitized books.

To provide background in copyright law, Unsworth gave an overview of major copyright milestones. In 1989, the U.S. joined the Berne Convention, which requires signatories to recognize the copyright of works by authors from other signatory countries and prohibits signatories from requiring formal registration of copyright. Thus the U.S. copyright system switched from an opt-in to an opt-out model. Subsequent milestones in copyright law have closely shadowed—and even “stalked”—digitization efforts.

If you compare dates from the two histories, this close relationship becomes clear:

Digitization milestones

- 1971: Project Gutenberg
- 1987: Perseus Project
- 1990: Library of Congress's American Memory project
- 1994: NSF Digital Libraries Initiative
- 1995: Making of America, JSTOR
- 2001: Million Books project
- 2004: Google Books
- 2005: Open Content Alliance
- 2008: HathiTrust
- 2010: DPLA

Copyright milestones

- 1989: US joins the Berne Convention
- 1996: World Intellectual Property Organization Copyright Treaty
- 1998: Digital Millennium Copyright Act
- 2003: Eldred v. Ashcroft
- 2005: Authors Guild v. Google
- 2008: Kahle v. Gonzalez
- 2011: Google Books settlement rejected; Authors Guild et al. v. HathiTrust et al.

- 2012: Authors Guild v. Google goes to court

Unsworth explored recent copyright history through the examples of Mickey Mouse and T.S. Eliot's *The Waste Land*, specifically how the "imperative to save Mickey" has affected the copyright status of all other works published since around 1923. With each modification to copyright law, copyright has become more retroactive and longer-lasting, always extending far enough back in time to cover Mickey Mouse. Unsworth displayed a graphic¹ to illustrate this dynamic, known as "copyright creep." Eliot's seminal poem, a mashup of prior western cultural texts, probably would not have been publishable had it been subject to the same stringent copyright laws that have prevailed since Mickey's appearance. Unsworth displayed a quote from Wikipedia explaining that *Steamboat Willie*, Disney's 1928 film in which Mickey made his debut, has been "close to entering the public domain in the United States several times. Each time, copyright protection in the United States has been extended. It could have entered public domain in 4 different years; first in 1956, renewed to 1984, then to 2003 by the Copyright Act of 1976, and finally to the current public domain date of 2023 by the Copyright Term Extension Act (also known pejoratively as the Mickey Mouse Protection Act) of 1998."² One twist in the law is that between 1923 and 1963, copyright renewal did not happen automatically. Many of the books published in this time period are believed to be "orphan works"—works that are in copyright (i.e. the copyright was renewed), but for which a rights holder cannot be located. The University of Michigan is currently undertaking a project to identify orphan works in what Unsworth describes as a praiseworthy and practical effort to deal with a difficult problem.

Unsworth gave a detailed overview of *The Waste Land*'s copyright status, to show the complexity and messiness of copyright law. First registered in 1922, *The Waste Land* was originally under copyright for 28 years, through 1950. In 1950, the copyright was renewed in the United States for 28 more years (through 1978), and meanwhile the Copyright Act of 1976 extended the protection through 1997. On January 1, 1998, the work entered the public domain, narrowly missing U.S. Public Law 105-298, which went into effect in October 1998 and extended the renewal of copyrighted items for another twenty years. If Eliot's work had been published only one year later, in 1923, these laws taken together would have put the work in copyright through 2019. To complicate matters further, *The Waste Land* is still under copyright in the United Kingdom, meaning that publicly accessible online copies of the work may be infringing on publisher rights. Faber and Faber, the book's publisher, recently released an app of *The Waste Land*, further strengthening their copyright claim by producing new content based on the work. Overall, Unsworth stated, copyright law "makes tax code look easy."

Unsworth then transitioned into how to provide access to digitized material for the purposes of non-consumptive, or computational, research. He cited the "necessary characteristics of cyberinfrastructure" as defined in 2006 by the *ACLS Report on Cyberinfrastructure for the Humanities and Social Sciences* (2006)³: It will be accessible as a public good; it will be sustainable; it will provide interoperability; it will facilitate collaboration; and, it will support experimentation.

One example of robust cyberinfrastructure is the HathiTrust Research Center (HTRC), a collaborative effort between Indiana University, the University of Illinois, associated research centers, and HathiTrust. The HTRC will be a virtual research center for computational, non-consumptive work in HathiTrust collections. Unlike other textual data resources, the HTRC will be open to researchers who may not know what they are looking for until they get there. The goals of the HTRC are to maintain a repository of text mining algorithms, retrieval tools, derived data sets, and indices available for human and programmatic discovery; to be a user-driven resource, with an active advisory board and a community model that allows users to share tools and results; and to support interoperability across collections and institutions through use of inCommon SAML identity.

The term “non-consumptive research” originated in the Google Books Settlement proposal. By the most restrictive definition, a researcher conducting non-consumptive research can mine textual data but is not allowed to read or understand the data. This concept is problematic because sometimes reading and understanding the text are necessary for conducting research, particularly to program a machine to recognize certain phrases. For example, if the goal is to identify sentimental passages in a text, how can a researcher program a machine to identify these texts without first reading the passages and understanding their characteristics? The HTRC must provide support for non-consumptive research through bringing algorithms to data, exporting results, and providing a secure environment in which researchers can work with copyrighted materials without exporting them. This effort is worth doing because it enables a new “art of information.”

To demonstrate this new art of information, Unsworth presented a video compilation of data visualizations, “Networks in the Humanities” presented by Elijah Meeks, Digital Humanities Specialist at Stanford⁴. This project resulted from collaboration between English faculty and a librarian. One of the network visualizations, “The Nineteenth-Century Literary Genome: An Experiment in Literary Influence” (2011) by Matt Jockers⁵, depicts the stylistic evolution of nineteenth century writers. The network shifts and revolves as several nodes break away from the main group, representing a group of outlier texts beginning to differ stylistically from the majority.

Data, Unsworth claims, enable researchers in the digital humanities to make new kinds of arguments. Though their arguments can now be based on quantitative and/or empirical data, these data still require interpretation. With new types of arguments come new opportunities for error, especially errors resulting from incomplete data or an incomplete understanding of the data. For example, Google’s Ngram Viewer provides graphical visualizations of words and phrases over time; users can enter words and a time period to generate a new graph.

Unfortunately, capitalization proves a problem, as the system’s tokenization considers capitalized words to be different from un-capitalized words (e.g. reason vs. Reason). In most cases, researchers probably would not want to differentiate between capitalized and un-capitalized words, although in some cases it may be very important separate them, if they have different meanings. To avoid making faulty arguments, researchers need to understand the data

and how it has been prepared; they must also have the opportunity to prepare textual data themselves based on specific research questions.

One new kind of argument is Ted Underwood's⁶ effort to understand the changing etymological basis of English diction between 1700 and 1900. His research focuses on understanding the trajectory of Latin-derived versus German-derived words. English scholars have often drawn specious conclusions based on the "Latinate" or "Germanic" character of a writer's style, and the distinction has long been accepted in the academic community. Nevertheless, linguists have found evidence that older words (Germanic) do predominate in informal, spoken English. Underwood's research therefore uses new digital humanities research methods to settle an old argument. He has found that the fundamental distinction is not actually Latinate/Germanic, but date of entry of words into the English language. Breaking down the results by genre, the data show that fiction and nonfiction prose became more formal (less like speech) in the eighteenth century. Drama and poetry changed little, but older, less formal words always predominated in these genres. It is reasonable to think that these historical trends may have continued beyond 1900, but to investigate that question, we need to have access to the data.

Dataverse and Data Management Plans

Following the morning keynote address, four concurrent breakout sessions were offered. In one, Mercè Crosas (director of product development, Institute for Quantitative Social Science (IQSS),

Harvard University) gave an overview of the Dataverse Network Project and data management plans. Dataverse is a virtual archive for data sets that was designed by IQSS beginning in 2006. Crosas also summarized a sample data management plan that was created for the National Science Foundation (NSF) using Dataverse as the data repository.

The intellectual origins of Dataverse were articulated by Gary King, the current director of IQSS, who wrote in 1995, “the replication standard holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author.”⁷ He and Micah Altman further developed this idea in a 2007 article in *D-Lib Magazine* titled “A Proposed Standard for the Scholarly Citation of Quantitative Data.”⁸ The basic principle King and Altman espoused was that to effectively replicate and reuse scholarly work, scholars must have access not only to scholarly articles, but also to the data, metadata, and supporting files behind the work. The formal data citation they proposed includes the following elements: Authors, Year, Title, Persistent Identifier (handle), Universal Numerical Fingerprint (UNF), Distributor, Version, and optional fields.

Replicating and reusing scholarly work is crucial to the validation, enhancement and advancement of science. This kind of reuse and replication, Crosas stated, requires a repository for research data that takes care of long-term preservation and employs good archival practices while allowing researchers to keep control of and receive recognition for their data. The Dataverse Network meets these criteria. Researchers are responsible for depositing data and entering metadata into the repository, getting the data citation (handle, UNF), displaying data on

a personal website, managing data permissions, and updating new versions of the data. The Dataverse centralized repository takes care of backups and replication of data in different locations (LOCKSS), conversion to archival formats, extraction of metadata from data sets, metadata standards (DDI, Dublin Core), and interoperability (OAI, APIs). While the Dataverse Network is a centralized repository, a Dataverse can be implemented for many different types of collections, such as an individual researcher's work, research center output, data management plans, journals, archives, or dissertation data.

Crosas next gave an overview of data management plans (DMPs). The basic research cycle generally starts with an idea and a grant proposal with a data management plan. Data collection is the next stage, followed by publication, data sharing and archiving. Crosas presented a use case, Harvard Professor James Snyder's DMP from an NSF proposal. Crosas went over each element of the generic NSF DMP checklist and explained how Snyder and the Dataverse team worked together to describe the following elements: Data to be collected, existing data, audience, access and sharing, data formats, metadata and documentation, storage, backup, replication and versioning, security, budget, privacy and intellectual property, archiving, preservation and long-term access, and adherence to the DMP.

The first three elements—data description, existing data, and audience—came from Snyder's grant proposal. Access and sharing capabilities are provided by Dataverse: "All data collected or generated will be deposited in the researcher's Dataverse. The Dataverse allows researchers to deposit data in an organized, well curated and citable network... ultimately facilitating access and

sharing.” Dataverse also specifies data formats. Quantitative data will be deposited in SPSS, Stata, CSV, tab delimited, or GraphML format. Images will be deposited in JPEG-2000 or TIFF, and audio in MP3 or WAV. Dataverse also accommodates all formats to allow for the flexibility researchers need. For metadata and documentation, Dataverse provides approximately 100 fields to choose from, plus custom fields. Basic metadata fields are: Title, identifier, year, author, abstract, and keywords. Additional documentation can be uploaded in PDF or plain text formats, as can code. A formal Data Citation will be generated automatically, and metadata will be exported into XML (DDI, DC). To ensure reliable storage, backup, replication, and versioning, Dataverse keeps multiple versions of data and provides system backups daily. Deposited data is never destroyed unless legally required. The Dataverse Network at IQSS also follows “good security computer use practices” set by the Security and Privacy group at Harvard. Dataverse provides a simple answer to the question of DMP budget, as the “Dataverse Network at IQSS is free and open to all social science research data.” Restrictions are 2GB per file, with no limit on the number of files. (In the future, a fee might apply to archive very large collections. This is currently under review.) To ensure that privacy and intellectual property laws are respected, Dataverse does not allow confidential data to be uploaded. The researcher must also agree to the terms of use, which contain statements about copyright and confidentiality protections. The Dataverse Network commits to good archival practices: metadata is exported to XML; data files are reformatted for long-term access; all versions are kept; and both metadata and data are replicated to multiple locations through LOCKSS.

For more information, Crosas pointed to <http://thedata.org>. She noted that there are two main Dataverse Networks at Harvard: the IQSS Dataverse Network which is open to all social science

research data and the Astronomy Dataverse Network which is open to the Harvard-Smithsonian Center for Astrophysics. The software itself is open source and free to install at any institution.

That's Why I Chose Overdrive!

In another morning breakout session, Todd Gilman (librarian for literature in English, comparative literature, and linguistics, Yale University Library) and Mary Danko (director, Hartland Public Library and board president, Green Mountain Library Consortium) discussed their experiences with OverDrive. OverDrive is a distributor of downloadable e-books, audiobooks, music, and other digital content to libraries.

Gilman spoke about the Yale University Library's initiative of offering e-books through OverDrive for a two-year trial period. After Gilman discovered Boston Public Library's OverDrive service, he wondered if OverDrive offered a model for academic libraries. He learned that both McGill University and the University of Pittsburgh were OverDrive customers and that OverDrive offered more than the "vampire romance novel genre." In fact, OverDrive sells e-book content from more than one thousand publishers, including many university presses.

Gilman outlined the OverDrive business model for academic libraries: OverDrive charges a flat fee each year based on the number of full-time equivalent (FTE) students attending the institution. Half of this fee covers Web site development, and the other half covers e-book

content. The library can spend more than this amount on content, but it must spend at least this much. E-books from OverDrive tend to cost about three times the list price of the print edition. In addition to paid-for content, OverDrive offers 34,000 free, public-domain e-books based on content from Project Gutenberg. The library can offer its own digitized content through OverDrive as well, for example digitized images. A strength of OverDrive, according to Gilman, is that multiple e-book formats are available for download, including the EPUB industry standard as well as PDF; OverDrive books are also compatible with the Amazon Kindle.

Because they were offering the OverDrive service on a trial basis, Yale librarians initially decided not to include MARC records for OverDrive e-books in their online catalog. Instead, they settled for the minimum that other OverDrive library partners have: they worked with OverDrive to create a Web site for searching e-book titles that mimicked the look and feel of the Yale University Library front door. Authenticating users turned out to be a problem, however. OverDrive uses barcodes to authenticate borrowers, while Yale uses CAS (the Central Authentication System). The Yale library had to hire a programmer to create click-through authentication, a solution also used by the University of Pittsburgh.

Yale launched its OverDrive service in September 2011 with about 175 titles, including a small number of audio books. OverDrive e-books follow a “one book, one user” model, meaning that only one borrower can read a book at a time. Because OverDrive’s strength is in popular literature, Gilman generated a list of the most heavily circulated books from the Bass Library (Yale’s main library for frequently used materials in the humanities and social sciences) and the

library's collection of popular literature. He compared these titles to what was available from OverDrive. In addition, he asked OverDrive's collection development specialist for the most popular one hundred OverDrive e-book titles at McGill University and the University of Pittsburgh and reviewed these lists for possible titles to purchase. Other subject selectors at Yale were also consulted for title suggestions.

Gilman understood that in order to be successful, Yale's OverDrive service had to be visible. The library advertised the OverDrive service on their large-screen display monitor, and the *Yale Daily News* wrote an article about the service. OverDrive supplied a logo for the library's home page and offered bookmarks and posters custom-branded for the library. After some time, Yale librarians decided that including records for OverDrive e-books in the online catalog was important, and they instituted a procedure to accomplish this. The OverDrive interface highlights "new" and "most popular" titles, which also helps drive usage.

Gilman stated that the OverDrive service at Yale generates about four hundred downloads a month, which is respectable for the number of titles they have available. The library has also initiated a patron-driven acquisitions plan with OverDrive so that if a title has more than three people waiting to read it, the purchase of an additional copy of the title is triggered. In addition, library users are able to suggest titles for purchase through a link on Yale's OverDrive Web site. Gilman hopes the growing popularity of OverDrive will result in the continuation of the service on a permanent basis. He noted that if a library cancels its OverDrive subscription, all content is lost.

In Gilman's experience, troubleshooting with OverDrive is minimal, but the service nonetheless has its drawbacks. For example, Yale discovered that the Harry Potter e-books they purchased expire after five years. To read an OverDrive e-book, the user must create an Adobe account, download the OverDrive console, and then transfer the e-book to an e-reader; this process is more cumbersome than it needs to be. To accommodate reserve reading, Yale requested that OverDrive limit the lending period for specific items to two hours; OverDrive was unwilling to do so. Gilman feels that there are limits to how much OverDrive is willing to accommodate the academic library market.

After six months of providing e-books through OverDrive, the Yale University Library now owns over six hundred titles. Gilman recommends that libraries considering OverDrive have a budget for the project that allows them to continue to add content over time, in order to sustain interest in the service.

Following Gilman, Danko reported on the Green Mountain Library Consortium's experience with OverDrive over the past four years. Approximately fifty small Vermont libraries formed the Green Mountain Library Consortium in order to affordably purchase downloadable audio books from OverDrive after the Vermont State Library decided not to pursue a downloadable audiobook project. In November 2010, the consortium started offering e-books through OverDrive as well. Usage of both audiobooks and e-books grew steadily, and when OverDrive implemented Kindle compatibility in September 2011, usage jumped. Usage spiked yet again in

January 2012 after, Danko presumed, patrons acquired e-reading devices over the holidays. Checkouts for e-books in all formats through OverDrive increased from under five hundred a month in November 2010 to over three thousand a month in March 2012. Danko expects this trend to continue, since she has read that only a small percentage of all the people who own e-readers have ever checked out an e-book from a library. In the near future, Danko expects e-book access through mobile devices to surpass access through personal computers.

Danko remarked that one of the Green Mountain Library Consortium's greatest challenges is providing online content. "People are screaming for content," yet library budgets are under strain and print materials still need to be purchased. She posed the question of how best to budget for online content: Should libraries add a separate budget line for online content? Should a given percentage, say 10 percent, be moved from a library's print budget to a budget for downloadable e-books?

Another challenge is keeping track of the e-book licensing rules of different publishers—for example HarperCollins requires that their e-books be "replaced" after circulating twenty-six times, and Scholastic "expires" purchased Harry Potter e-books after five years. Penguin has removed their titles from OverDrive; libraries retain access to Penguin e-book titles already purchased, but these titles can be downloaded to e-readers via USB port only. Typically, the consortium purchases an additional copy of an e-book for every seven holds placed on the book. However, Random House recently raised the price for popular fiction e-books to between \$85 and \$100, making multiple copies unaffordable.

Such rules necessitate that the consortium continually educate member librarians about pricing and license agreements. Patrons need to be educated as well. For example, patrons have offered to buy e-books from Amazon and donate them to their libraries; librarians need to explain that this is not possible. Librarians must also be able to help patrons use OverDrive with a variety of e-readers; this has been a challenge. Initially, consortium libraries received a large number of support questions regarding OverDrive, but fortunately these have dropped off since Overdrive improved their “Help” pages.

Danko concluded by stating that the Green Mountain Library Consortium is in serious talks with 3M as a possible alternative or addition to OverDrive. Danko is not comfortable “putting all of their eggs in one basket” and dealing only with OverDrive and their limited suite of publishers. 3M offers cloud-based e-book delivery, as well as the option to select and download e-books through kiosks in libraries. The problem with 3M right now, though, is their lack of Kindle compatibility.

Transforming Technical Services in the iLibrary

Following lunch was a panel discussion featuring Alicia Morris (head, technical services, Tisch Library, Tufts University) and Roger Brisson (head, metadata and cataloging, Mugar Memorial Library, Boston University). Morris spoke first on “Rethinking Tech Services at Tisch Library.”

In 2008, the Technical Services Department at the Tisch Library followed very traditional workflows that were almost entirely print-based. E-journals were managed in collections, and it was the systems librarian, not technical services librarians, who loaded MARC records into the catalog for the few e-book collections the library owned. At that time, Technical Services did not work at all with non-MARC metadata. Recognizing the need for change, the department began conversations about trends in e-resource management and non-MARC metadata and how they could apply their skills in this new environment. They created a strategic plan to imagine what new services they might offer and advocated for Technical Services within the library, pointing out roles they could play in managing e-resources and creating collection-level records for manuscript collections, for example. Morris also reached out beyond the library, looking for opportunities to volunteer librarians' expertise in creating and working with metadata.

One such opportunity arose when a portfolio of fifteenth century manuscripts was discovered in the library. The Collections and Technical Services Departments worked together to digitize the manuscripts, catalog them using Dublin core, and created The Miscellany Collection.⁹ The library showcased the site to faculty, and articles in both the *Tufts Daily* and the *Chronicle of Higher Education* profiled the collection. As a result, The Miscellany Collection began to be used by faculty and students on campus. Classics Department students, for example, had the opportunity to engage in primary research by translating the leaves and publishing their notes and commentary on the site.

The Miscellany Collection opened the door to more campus collaborations for the Technical Services Department. For example, the department has provided metadata for a slide digitization project, offered metadata best practices for student-created multimedia, assisted digital humanities students with coding their research using the Text Encoding Initiative (TEI) guidelines, created taxonomies for departmental and program websites, and consulted on data management plans. Morris emphasized that these projects have been undertaken on a small-scale basis, as pilots. Though the department has received a growing number of requests for its services, it will be difficult to fill them all until the department rebalances its workflows and continues to build the skills of its librarians. Technical services will continue in a hybrid print and digital environment for many more years, and project management skills will be essential to balancing the work of the department. Morris concluded by stating that technical services can no longer be seen as a back room operation. Their experience at Tufts has demonstrated that technical services librarians can be full partners in the research and instruction mission of the university if they are open to new ideas and unconventional opportunities.

Brisson echoed Morris' theme, describing how the Boston University Libraries are also transitioning from the concept of cataloging to that of metadata services. Brisson believes that metadata will drive the future library experience as librarians focus on local users and creating simple interfaces that will facilitate discovery. The Mugar Memorial Library is working toward this future through the implementation of the ExLibris Alma library management service and Version 4.0 of the ExLibris Primo discovery platform, both of which will go live in the fall of 2012.

Brisson spoke of a “collections crisis” caused by the dramatic growth on an international scale of scholarly research, both published and unpublished. In response, research libraries have deployed technology and have questioned fundamental principles of collection development. Examples of these responses include the shift toward patron-driven acquisition as opposed to a belief in “collecting the best” and the expansion of the collecting universe to include documentation, evidence, and primary source material. Brisson offered a recent quote by James Michalko (vice president, OCLC Research Library Partnership) in which Michalko stated that it was not the end of the paper model, but the “end of canonicity in cultural life” that “eliminates a pillar of being for the academic library.” Thus “collecting and stewarding the scholarly record is now a very uncertain and, perhaps, undoable task.”¹⁰

Brisson stated that librarians are addressing the collections crisis through technology, and he noted that the intersection of technology and scholarly communication have a long history. Brisson would place the beginning of the collections crisis in the early part of the twentieth century, when scholars Vannevar Bush and Fremont Rider wrote about the struggle to keep up with the growing volume of published information. Earlier technologies such as microformats were a prelude to the computers that today deliver scholarly materials to us on a huge scale. Now we are witnessing a dramatic shift to e-books, as libraries are in the process of adopting e-books as the primary platform for monographs.

Brisson discussed the impact of these shifts on technical services at the Boston University Libraries. The library had purchased large digital collections previously available on microfilm and had cataloged them at the collection level. An inventory of the library's e-books revealed over 500,000 titles with no individual records in the catalog. Now that e-books are becoming the standard, providing analytics of what is in these collections is very important and a high staff priority. As another example of a practice that needs to change, Brisson noted that the library's "A to Z list" of databases contains over one thousand resources. What does this mean for patrons? Brisson feels that libraries have built out their Web sites to the point that they are too complex from the patron point of view; they are untenable. Viewing some library Web sites requires a great deal of cognitive work. Through overly-complicated Web sites with opaque user interfaces, libraries are shifting demands onto the user.

The solution, claimed Brisson, is metadata. Simplicity should guide us. Libraries are complex, but it is our challenge to make them simple for patrons. It is clear that metadata are essential for effective searching, more so than the material described itself. With a metadata-driven user experience, collection development shifts to access management. The local experience gains primacy, as discovery domains are defined and structured through metadata and the Semantic Web is institutionalized. Librarians need to be willing to expand the boundaries of traditional cataloging and MARC. We need to realize that XML is the future and to become comfortable working with it.

The Boston University Libraries have followed this strategy in their implementation of ExLibris Primo. Primo employs a hosted, central metadata index of a library's scholarly content to provide unified discovery and delivery of the library's resources. Primo imports local data sources such as catalogs and repositories to a standardized XML format to allow cross-collection searching; it is essentially a discovery tool sitting on top of multiple data sources. Brisson displayed a prototype of the new Boston University Libraries Web site, which is driven by Primo. A single search box is displayed prominently, allowing the user to search the Web site itself, along with library resources. Since each page of the Web site is cataloged, a search for library hours would be conducted in the same way as a search of the library catalog. The results of a Primo search are displayed in a way that conforms to the look and feel of the library Web site, so patrons are not aware that they are searching a library catalog. The library's Web site essentially becomes one and the same as the library's discovery platform.

The new Boston University Libraries Web site is being built using a WordPress custom implementation. Public services librarians are developing hundreds of reference guides geared to local user needs. All of the guides are metadata-driven, thus librarians do not need to key a description of each resource listed; they simply reference the resource's Primo catalog record number. These reference guide pages are delivered dynamically, meaning that they are rebuilt every time a patron does a search.

In addition to implementing ExLibris Primo as a discovery platform, the Boston University Libraries have adopted the ExLibris Alma library management system. ExLibris Alma "supports

the entire range of library operations—selection, acquisition, metadata management, digitization, and fulfillment—for the full spectrum of library materials, regardless of format or location.”¹¹

The cloud-based Alma breaks down e-resource silos by bringing together into one system the integrated library system, electronic resource management system, link resolver, and discovery platform. Alma is workflow-driven, pushing work to staff as needed. Staff do not need to access external systems to do their work; external systems—for example, WorldCat—are accessed from within Alma. Boston University Libraries are planning on going live in the fall with ExLibris Primo Version 4.0, ExLibris Alma, and their new Web site.

ILS in the Cloud: Promise or Peril?

The second afternoon panel discussion featured Bob Gerrity (associate university librarian, Systems & Information Technology, Boston College Libraries) and Martha Rice Sanders (knowledge management librarian, HELIN Consortium). Their focus was on cloud computing and its impact on library systems. Titled “ILS in the cloud: promise or peril?” the discussion centered around this new technology as it is beginning to affect the library world. Gerrity began with a definition of cloud computing from the National Institute of Standards and Technology. Cloud computing is “...a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management efforts or service provider interaction.” Discussion moved to the three service models of cloud computing, beginning with Infrastructure as a Service (IaaS). This offers managed and scalable processing, store, networks, and other fundamental computing resources,

with examples including Terremark, Equinix, Savvis, Rackspace, and Amazon Web Services. Gerrity described Platform as a Service (PaaS) as “user-created or acquired applications made using programming languages and tools supported by the provider.” Examples include Google App Engine, Windows Azure, and Force.com. Finally, Software as a Service (SaaS) offers the capability of using provider applications running on a cloud infrastructure accessible from the web. Examples here include Google Docs, Microsoft Office 365, and Salesforce CRM.

There are many benefits to cloud computing, particularly for library use. These include reduced costs for technical support, increased storage capacity, flexibility, and enhanced user mobility. Gerrity presented Ex Libris’s “Alma”, OCLC’s “WorldShare Management Services,” and Innovative Interfaces’ “Sierra” as examples of cloud-based ILS platforms that are introducing librarians to cloud technology. He closed with a review of changes in library technology over the last decade. Gerrity noted that, while products and technological changes usually do not live up to their initial hype and promises, cloud computing does offer a way for libraries to utilize shrinking IT budgets and remain a key players in the delivery of information.

Martha Rice Sanders took a different approach in explaining cloud computing and what its effects might be to librarians and the library world. She began with the definition of what a library catalog is and how it has evolved from information on library holdings arranged in a card catalog to an online catalog that provides access to databases, electronic materials and print materials. This history of the library catalog has one constant: metadata. Metadata is a collection

of words and images that assist the user in finding and obtaining the materials that are needed. This metadata is put into a controlled environment by catalogers, systems people and, more recently, the users themselves. The information that is in the catalog is then used by other librarians, users and discovery services. All of this information is housed on a server which can be in the library (local server), hosted somewhere else, or in the cloud. When “in the cloud,” all of a library’s data is intermixed with other data and shared (similar to OCLC’s bibliographic utility). Therefore this cloud concept is not new to the library world. In fact, libraries that participate in a library consortium with a shared catalog are cloud computing. The HELIN consortium has 23 different libraries that share a central server located on the campus of one of the participating members.

There are real concerns in this new environment in terms of what libraries lose in return for all of the benefits of cloud computing. Is there loss of patron privacy, since their information is no longer located in a single library system? Have libraries lost their power to control their data and its quality? When problems do develop, will there be local troubleshooting or will it take place from a distance? In a shared, cloud-based system, there is real concern over who actually owns the data. All of these topics have become major issues for libraries, with no definitive answers. On a positive note, Sanders addressed the fact that the cloud system offers more services and allows more staff time to be devoted to needs other than technology. Development of cloud-based library systems is collaborative a, with a wider spectrum of players.

The panel discussion concluded with general questions from the library audience that asked what the future was for non-cloud based systems and how local libraries were going to deal with these changes. Most agreed that the old library OPAC will soon disappear and that it is important for librarians to be key players in what will replace them.

Notes

1. Bell, Tom. "File:Copyright term.svg," 2008, Wikipedia, http://en.wikipedia.org/wiki/File:Copyright_term.svg (accessed June 25, 2012).
2. "Steamboat Willie," Wikipedia, http://en.wikipedia.org/wiki/Steamboat_Willie (accessed June 25, 2012).
3. American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. *ACLS Report on Cyberinfrastructure for the Humanities and Social Sciences*, 2006
http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf
(accessed June 25, 2012).
4. Stanford University Libraries and Academic Information Resources. "Digital Humanities Specialist," <http://dhs.stanford.edu/> (accessed June 25, 2012).

5. Jockers, Matthew L. "The Nineteenth-Century Literary Genome: An Experiment in Literary Influence," 2011 <http://www.youtube.com/watch?v=4eiLrjNpwRk> (accessed June 25, 2012).

6. Underwood, Ted. The Stone and the Shell, <http://tedunderwood.wordpress.com/> (accessed June 25, 2012).

7. King, Gary. "Replication, Replication." *PS: Political Science and Politics* 28 (1995): 443-499.)

8. Altman, Micah and Gary King. "A Proposed Standard for the Scholarly Citation of Quantitative Data." *D-Lib Magazine* 13, no. 3/4 (March/April 2007), <http://www.dlib.org/dlib/march07/altman/03altman.html> (accessed June 25, 2012).

9. "The Miscellany at Tisch Library." http://www.library.tufts.edu/tisch/ematlocalstorage/miscellany_collection/home.html (accessed June 25, 2012).

10. OCLC Research. *Above the Fold: A Weekly Newsletter for the Changing World of Libraries, Archives and Museums* 5, no. 14 (April 6, 2012),

<http://www.oclc.org/research/publications/newsletters/abovethefold/2012-04-06.htm> (accessed June 25, 2012).

11. Ex Libris. “Ex Libris Alma: The Next-Generation Library Services Framework.”

<http://www.exlibrisgroup.com/category/AlmaOverview> (accessed June 9, 2012).