Smarter Libraries through Technology:

Open and Linked Data

By Marshall Breeding

Despite a proclivity toward openness, libraries have historically been involved with data and content under various levels of restrictions imposed by copyright and proprietary business arrangements. Libraries have also dealt with data primarily through self-contained applications, mostly managed within relational databases. While these modes of operation persist today in the mainstream production systems of interest to libraries, a strong movement is underway toward the principles of the Semantic Web and more open approaches to sharing data. I see libraries now at an interesting juncture between the realm of proprietary, restricted, and isolated data and a new era with possibilities for more openness in sharing data and putting it to work through Semantic Web technologies.

The library community seems to be ready to push boundaries and more liberally share previously restricted data. To what extent can metadata related to collections be shared with other organizations or exposed freely for the entire world? When can the actual content of collections be shared? Copyright, licensing terms, and business contracts impose specific restrictions, but libraries have been working toward arrangements that allow more sharing.

The open access movement, for example, has been underway for a decade or so, aiming to transform scholarly publishing. Yet, it seems that the progress achieved has yet to significantly relieve the financial burden that libraries face in gaining access to scholarly publications. Still, there is reason for optimism as increasing numbers of important academic institutions, such as Harvard University, become strategically committed to open access to the works created by their researchers.

We see more aggressive movement toward liberal models of access in the metadata that describes library collections or bodies of published content. In this issue of *Smart Libraries Newsletter,* we explore developments toward more freedom in sharing metadata stores of interest to libraries, such as Dewey Decimal Classification, the Virtual International Authority File, the bibliographic records from libraries such as Harvard University, and ultimately OCLC's World-Cat database. International efforts such as Europeana have set a high standard for openness that other organizations will do well to follow.

In addition to sharing data, the library community has also become increasingly involved in Semantic Web technologies that enable more powerful use of that data. One of the technologies associated with the Semantic Web is linked data, which exposes data elements with built-in interconnectivity. Rather than using proprietary formats, linked data makes use of W3C standards such as RDF (Resource Description Framework); exposes each element through a

unique URI (Uniform Resource Identifier); and provides ways to query the linked data sets through the URI. This approach results in a global network—the Semantic Web—where data are truly machine readable, meaningfully structured, and interconnected, enabling the creation of new systems or applications that bring together relevant information from the ever-expanding universe of linked data.

So far, library automation has not yet seen widespread adoption of Semantic Web technologies. The business systems common in libraries make use of data enclosed within relational databases. The current slate of discovery services, although incorporating ever larger volumes of metadata and content, continue to operate on the basis of pre-built indexes within search engines rather than through linked data. The current technology trends toward cloud technologies, such as software-as-a-service and shared knowledge bases, do not generally cross over into the realm of the Semantic Web.

While the number of library activities and projects involving Semantic Web technologies continues to grow, it seems to me that they currently are in the research and development stage and have yet to reach strategic production-level systems. They hold great potential for libraries, and I will be closely following these technologies, looking for significant new developments.

OCLC Navigates the Realm of Open Linked Data

OCLC's stance on linked data and open data has become a key topic of interest for libraries. OCLC has been involved with the Semantic Web and linked data as research projects for several years. It has released open data licenses for some of its ancillary products. Where the massive WorldCat database, OCLC's prime strategic asset, fits into plans for open data remains a pressing concern for librarians. Through a series of actions, OCLC has demonstrated a progressively more open approach to the information assets under its stewardship.

OCLC Steps into the Realm of Linked Data

Linked data and the degree to which it should be openly available is not an entirely new area of interest for OCLC. Since about 2009, the organization has made a number of forays into exposing information assets as linked data, with such initiatives as dewey.info, FAST, and VIAF.

Dewey.info was launched as an experimental linked data service in August 2009, applying the principles of linked data to the Dewey Decimal Classification scheme. OCLC acquired Dewey Decimal Classification from the Lake Placid Education Foundation in 1988. The dewey.info service exposes each of the Dewey concepts as RDF triples available through an URI and as Web pages. OCLC makes the data underlying dewey.info available under the Creative Commons Attribution-Non-Commercial-NoDerivs 3.0 license, meaning that reuse is allowed with attribution for non-commercial purposes and that no derivative works can be created based on the data (see: http://creativecommons.org/licenses/by-nc-nd/3.0/).

OCLC has demonstrated a progressively more open approach to the information assets under its stewardship.

In December 2011, OCLC released the **Faceted Application of Subject Terminology** (FAST) as an experimental linked data service under the Open Data Commons Attribution License. The FAST authority file was created through collaboration between OCLC and the Library of Congress. FAST derives from the Library of Congress Subject Headings, breaking the long pre-coordinated strings into simpler terms better suited for discovery services and other interfaces that employ faceted navigation.

The **Virtual International Authority File**, or VIAF, recently became a service under OCLC's sole stewardship under an open data arrangement. The VIAF aggregates a variety of international name authority files and makes it available through a Web-based service with an underlying linked data model. VIAF has been underway since April 1998 as an experimental project, initially including the Library of Congress, the German National Library (Deutsche Nationalbibliothek), and OCLC. These organizations formally joined as the VIAF Consortium in 2003, which has since expanded to include National Library of France (Bibliothèque nationale de France). More than 22 organizations currently contribute to the project. In September 2009, through an initiative of OCLC Research VIAF was released as linked data.

In April 2012, VIAF transitioned to become a production service of OCLC, with the understanding that it would

continue to be open and freely available under the Open Data Commons Attribution License (ODC-BY), which includes the freedom for others, including commercial organizations, to share, create, and adapt the data with the provision that the original source be attributed (see http://opendatacommons.org/category/odc-by/). With this transition, OCLC's operational activities in VIAF moved from its Research division to the groups within the organization responsible for production-level services.

WorldCat Rights and Responsibilities Guidelines and Open Data Licenses

Currently, WorldCat, the organization's top strategic asset, remains governed by the "Rights and Responsibilities for the OCLC Cooperative" that outlines the policies regarding the use of records in WorldCat. OCLC's policies regarding the ability to share WorldCat records have long been controversial. OCLC constituted a Record Use Policy Council in September 2009. The council created the current Rights and Responsibilities guidelines through a process that included opportunities for the broader OCLC community to provide input. Among other issues, the Rights and Responsibilities document, which took effect in August 2010, aims to advise libraries on sharing records outside of the OCLC membership. OCLC positions the Rights and Responsibilities document as reflecting the norms for its membership rather than as binding contractual terms. Yet, in a time when the library community resonates with the principles of open data, the guidelines of use stated in the Rights and Responsibilities policies remain a source of creative tension (http://www.oclc.org/worldcat/recorduse/policy/default.htm).

One action not necessarily endorsed under the Rights and Responsibilities policies is representing a library's collection as a public release of WorldCat records that it did not create. A library has claim to original records it produces directly. However, WorldCat records that represent its holdings but were created by other OCLC members are subject to the Rights and Responsibilities guidelines.

The National Library of Sweden Chooses Not to Join OCLC

In December 2011, the National Library of Sweden announced that it would not participate in WorldCat, specifically citing the constraints implied through the Rights and Responsibilities policies as in conflict with its strategy to openly share data. Libraries that participate in Libris, the Swedish National Union Catalog, expect to be free to take their records out and place them in other systems, a scenario that might be problematic if those records were derived from WorldCat. Participation in Europeana, for example, requires the CC0 (Creative Commons zero) public domain license inconsistent with OCLC's recommended policies. OCLC reports that there are continuing discussions with the National Library of Sweden on this issue.

Jim Michalko, Vice President, OCLC Research Library Partnership, wrote a subsequent blog post addressing this issue, stating that while the free public release of records would be inconsistent with the guidelines of the WorldCat Rights and Responsibilities, libraries ultimately can follow their own discretion in how they release their records to other organizations. It mentioned that licensing options that OCLC has followed for FAST, the Open Data Commons Attribution License, could be applied to how libraries release their own cataloging data that involves WorldCat. According to a comment on a blog posting by Eric van Lubeek, OCLC's Managing Director EMEA: "…on 18 April the OCLC Global Council in order to advise the OCLC Board of Trustees passed

a resolution endorsing the Open Data Commons Attribution License (ODC-BY) as consistent with the WorldCat Rights and Responsibilities for the OCLC Cooperative."

Harvard University Releases Complete Bibliographic Database

In April 2012, Harvard University exercised its discretion and released its entire bibliographic database, including over 12 million catalog records from its 73 libraries under the CC0 public domain license. This action, consistent with Harvard's mandates regarding open access to its creative output, goes beyond the Open Data Commons Attribution License that OCLC mentions as consistent with WorldCat Rights and Responsibilities. Harvard did work with OCLC as they considered the terms in which the data would be released and agreed to recommend to those that use the data to give attribution and to follow the community norms consistent with OCLC's guidelines, even though the CC0 does not require those actions. Harvard released its bibliographic records as a simple repository of MARC records, not as linked data.

EBSCO Makes Commercial Use of Harvard University Bibliographic Records

Less than two weeks after Harvard's release of its catalog records, in May 2012, EBSCO announced that these records had been incorporated into the base index of EBSCO Discovery Service. Libraries that subscribe to EDS would have access to these records, an example of the commercial use allowed under the CC0 public domain license.

Ex Libris to Incorporate Harvard records into Alma

Ex Libris has also announced its intentions to make commercial use of Harvard's bibliographic records. The company will incorporate these records into the Alma Community Catalog, a repository of data shared by all of the organizations that use the company's next-generation automation platform. Alma follows a hybrid data model. Its Community Zone includes resources such as the Alma Community Catalog, where participants can associate their holdings with shared records. Its Local Zone is for resources managed separately by an institution.

Ex Libris will also index Harvard's Digital Access to Scholarship (dash.harvard.edu), an open access repository of articles based on the institution's research. Harvard University makes the metadata associated with DASH available under CC0, though it restricts the articles themselves to non-commercial personal, teaching, and research use. Harvard requests, but does not legally require, attribution for use of the DASH metadata. It also requests that any derivative works be made available under the same terms.

A Linked Data Evangelist at OCLC

Richard Wallis, a longtime proponent of the Semantic Web and linked data recently joined OCLC in April 2012 in a position titled as Technology Evangelist. The arrival of Wallis comes at a time when OCLC has been increasingly involved in the exploration of linked data relative to its strategic information assets and as interest in the Semantic Web and linked data becomes more prevalent in the broader library arena. Wallis is expected to help communicate OCLC's activities in these issues both internally and to the library community and to help the organization explore and navigate its strategic path in this area.

Wallis departed Talis in January 2012 to establish a consulting company named Data Liberate. He will continue to be based in Birmingham in the United Kingdom.

— *Marshall Breeding*

Talis: Making a Business through the Semantic Web

At least some of the awareness of linked data in the library community can be attributed to Talis, a UK-based company that has promoted the topic internationally. Richard Wallis, who recently joined OCLC following a 20-year involvement with Talis, was its Technology Evangelist. In recent years Talis has been in transition from a company offering traditional library automation products to one focused on business activities surrounding the Semantic Web. In March 2011, Talis sold its library automation division, which supported the Alto library management system and the Talis online catalog, to Capita Group, choosing to focus solely on products and services related to Semantic Web and linked data.

Following the divestment, Talis was a company focused not so much on libraries as on its platform, based on Seman-

tic Web technologies, and on related consulting services. One product that proved to be especially successful is Aspire, a tool for creating resource lists to support faculty and teaching in educational institutions. Talis also spun off a subsidiary called Kasabi that offered services for hosting data sets in a linked data model with APIs to facilitate their integration into other products or ser-

vices. Kasabi positions itself as "an online marketplace that brings developers and data publishers together to enable new business models for consumers and producers of data at all scales" (http://kasabi.com/about).

In addition to Wallis, other key staff have left Talis in recent months. Ian Davis, chief technology officer and director for Talis, and chief execu-

tive officer for the Kasabi subsidiary, departed in March 2012. Rob Styles, chief technical consultant departed in April 2012. The core management team at Talis remains intact, including Dave Errington, its chief executive officer, and Justin Levesley, chief strategy officer.

— *Marshall Breeding*

SirsiDynix Development Strategies

In recent years, SirsiDynix has been executing a development agenda that provides libraries with tools and interfaces to improve access to their collections and services. The company continues to make improvements in the functionality of its strategic integrated library systems, Symphony and Horizon, concentrating especially on the development of patron-facing tools. These interface products address the different methods with which patrons might want to access the library: BookMyne and BookMyne+ for iOS and Android based mobile devices, Enterprise and Portfolio for the Web, and the Social Library for Facebook. SirsiDynix has also launched eResource Central for access and management of e-books, complementing the capabilities of its Symphony and Horizon ILS products in managing a library's print collection.

Ongoing Development and Support for Symphony and Horizon

Despite any past announcements or recent rumors, SirsiDynix asserts full commitment to support both its Symphony and Horizon ILS products going forward. In remarks at the recent Customers of SirsiDynix Users Group, Inc. Annual Conference in Orlando, Florida, Talin Bingham, SirsiDynix chief technology officer, said that the new interface products will support both Symphony and Horizon in both locally installed and SaaS deployments. Both Symphony and Horizon have many installations in libraries, making ongoing support and development a sound business strategy. SirsiDynix continues to make incremental improvements in Horizon even as it enhances Symphony more aggressively. A development project underway for Symphony, for example, will produce Web-based cataloging clients , offering a more flexible approach to

the creation of bibliographic records than the current Java-based WorkFlows.

Enterprise and Portfolio

SirsiDynix continues to develop Enterprise as its strategic Web-based discovery product. An enhanced version named Portfolio also includes the ability to manage digital content outside the ILS. SirsiDynix Enterprise was initially launched in June 2008 with relevancy-based search based on GlobalBrain, a search technology licensed from BrainWare.

SirsiDynix Enterprise and Portfolio was re-engineered in Version 4.1, released in June 2011 to make use of the open source SOLR search engine, which has become the most widely used search platform by library-oriented discovery products. By May 2012, 251 library organizations had implemented Enterprise. The most recent 4.2 release of Enterprise/Portfolio brings a tighter integration with the patron functionality of the underlying Symphony or Horizon ILS to provide more complete "My Account" features, such as sign-in, renewing items, managing holds and requests, and making payments. Previous versions offered more limited patron account features. Other enhancements include integration with eResource Central and the ability to easily create citations through RefWorks or Zotero.

eResource Central

With increasing investments in e-books, public libraries want to provide access that is integrated into existing end-user discovery interfaces and to effectively and efficiently manage e-books in their collections. The initial model of public

library e-book lending has largely been based on linking patrons to third-party e-book providers, especially OverDrive. In recent months, interest has grown in integrating e-book discovery and lending to existing library interfaces. The May issue of *Smart Libraries Newsletter* featured the integrated environment from Polaris in partnership with 3M and also BiblioCommons' project for integrating e-book lending. SirsiDynix has created a new product called eResource Central to allow its customer libraries to dramatically improve their e-book services to their patrons.

SirsiDynix initially announced eResource Central in January 2012 as a product offering a radically simplified user experience for access to e-books, audio books, and other digital content. It aims to be a single point of search and discovery of the library's collection, including both the print materials managed within the Symphony or Horizon ILS as well as e-book collections. Patrons will be able to access available copies of e-books through a single click. Through eResource Central, appropriate services will be presented according to the type of material: requests or holds for physical materials; or the download and viewing of e-books. No separate authentication is needed to access e-books that of the library interface, such as Enterprise/Portfolio, BookMyne, or Social Library.

The product also includes functionality to help libraries manage their e-book collections. The staff of eResource Central will provide libraries with tools to select and acquire e-books from commercial suppliers as well as from freely available sources. The libraries' e-book holdings will be indexed and made available through eResource Central in a way that seamlessly integrates with their print collections. The product will also include analytics that will assist the library's collection development processes for the selection of new materials.

SirsiDynix initially announced eResource Central in January 2012 as a product offering a radically simplified user experience for access to e-books, audio books, and other digital content.

The functionality of eResource Central contrasts with the process that many libraries currently have in place to manage their e-book collections, such as loading the MARC records into their ILS and providing links to external e-book providers. When libraries load the e-book records into their ILS, it is not necessarily possible to include features such as their availability in the initial result lists or to provide buttons for patrons to download the item or request it without being sent off to the platform provided by the e-book supplier.

The structure of eResource Central is cloud-based. SirsiDynix hosts both the software and the indexes, eliminating the need for locally installed software in the library. SirsiDynix's primary patron-facing interfaces, including Enterprise/Portfolio, BookMyne, and Social Library, deliver the end-user functionality of eResource Central. Customer libraries using legacy online catalog interfaces, such as iLink, IPS, or Horizon Information Portal (HIP), can use eResource Central through a dedicated Web-based client, currently known as eRC PAC, which will be rebranded in the near future. These older catalog products do not have the capability to make use of the Web Services layer and the search technologies that eResource Central holds in common with Enterprise/Portfolio, BookMyne, and Social Library.

Patrons of libraries using one of the current SirsiDynix interfaces will be able to search both holdings managed in Horizon or Symphony and e-books managed in eResource Central from a single interface. Users will see unified relevancy-ranked search results spanning print and digital materials with facets to narrow results according to the usual topical categories and by immediately available e-book collections.

The discovery component of eResource Central relies on an index maintained by SirsiDynix of the major e-book collections, profiled according to those held by each library that makes use of the product. This index will include those items offered in publicly available collections, such as Project Gutenberg and Open Library as well as those from SirsiDynix's commercial partners.

SirsiDynix is currently working with Baker & Taylor to integrate e-books from its Axis 360 service through eResource Central. This initial effort results in the ability for library patrons to download e-books into the Blio reader with a single click after discovering an item of interest, provided that they have already signed in. The library will also be able to purchase e-book titles through the staff interface of eResource Central. The arrangement with Baker & Taylor is not exclusive. SirsiDynix reports that it expects to form similar arrangements with other suppliers of e-books to libraries.

SirsiDynix expects the initial release of eResource Central to be in the fourth quarter of 2012.

While SirsiDynix describes eResource Central as a tool for access and management of electronic resources, the current features focus on e-books and audio books and the loan models seen in public libraries. It does not currently provide functionality similar to the electronic resource management products used in academic libraries to manage complex collections of

scholarly resources. The capabilities needed for the management and access of e-resource packages comprised of e-journals, book chapters, abstracting and indexing products, and aggregated content platform differs considerably from that needed to support e-book lending. E-book integration services such as the SirsiDynix eResource Central and electronic resource management systems such as Verde from Ex Libris, Innovative's Electronic Resource Management, or OCLC's WorldShare License Manager constitute different product categories and highlight the increasingly divergent automation requirements between public and academic libraries. SirsiDynix indicates that they plan support for journals, serials, and scholarly reference materials in the next major version of eResource Central expected in 2013.

BookShelf

SirsiDynix has created a personal space, called Bookshelf, for patrons to place items of interest. BookShelf will be imple-

mented at no additional cost across its major public-facing products, including BookMyne, Social Library, and Enterprise/Portfolio. It provides a graphical interface, using the cover art of materials when available, as a portable container for patrons to place items of interest, including the ability to share those items with others.

SirsiDynix Social Library

In January 2012, SirsiDynix announced the launch of Social Library, a native Facebook application that adds catalog search and other features to its Facebook page. Social Library operates through the Web Services layer available for both Symphony and Horizon. Recent developments to the SirsiDynix Social Library include the integration of BookShelf. As of May 2012 SirsiDynix reports that around 25 libraries have licensed Social Library.

— *Marshall Breeding*

Library Technology News in Brief

Udini by ProQuest innovates with on-demand premium information service for individual researchers

ANN ARBOR, Mich., May 7, 2012 – ProQuest is launching an inventive new research service that provides individuals with access to premium content. Instant and on-demand, Udini bundles peer-reviewed and trade journal articles, dissertations, international newswires, newspapers, magazines and more from thousands of publishers in a comprehensive cloud-based workflow management tool designed for individual users. The service is aimed at knowledge workers without access to research libraries. For publishers with already-strong academic distribution, Udini opens a new channel to reach users.

"Research libraries play a critical role in our knowledge economy, but not everyone who needs serious content is connected to a scholarly library. Research for these unaffiliated users is confusing and inefficient unless they know exactly what they're looking for. Premium information—when it's accessible at all—is distributed behind many different

paywalls all over the Web," said Rich LaFauci, Senior Vice President and General Manager, ProQuest Research Solutions. "Udini curates and licenses high-quality content and makes it incredibly easy to discover, acquire and use. The entire service is crafted from the end-user's perspective—from the content to the tools to the commerce model. It's simple, easy and flexible."

Proprietary content from market-leading publishers such as Springer, Nature Publishing Group, the Economist, The New York Times, The Washington Post, the World Health Organization, Cambridge University Press and 3,800 others is already slated to be available through Udini. The fast-growing service currently encompasses some 150 million full-text articles as well as the highly regarded ProQuest dissertations archive, the world's most consulted collection of intellectual property and emerging research from universities. Udini combines the kind of information resources offered by scholarly libraries with the intuitive search and content management of a cloud-based web service.

Udini users can search and then add desired content to the Udini cloud-based project organization and management tool, which also welcomes information from personal libraries and the open web, enabling them to capture their projects' whole research file in one always-accessible space.

June 2012
Open and Linked Data

Smart Libraries Newsletter

TO SUBSCRIBE

To reserve your subscription, contact the Customer
Service Center at **800-545-2433, press 5 for assistance**,
or visit **alatechsource.org**.

The 2012 subscription price is $85 in the United
States and $95 internationally.

ALA Techsource purchases fund advocacy, awareness,
and accreditation programs for library professionals
worldwide.