



Smart LibrariesTM

Formerly Library Systems NewsletterTM

50 East Huron Street, Chicago, Illinois 60611-2795, USA



Smarter Libraries through Technology

Making Technology Smart: Content and Connections

By Marshall Breeding

Libraries face incredible challenges to fulfill their missions of building and providing access to digital and print collections as well as an ever growing array of services to meet their clientele's information needs. So many libraries lack adequate financial resources to accomplish these goals and must continually make difficult decisions on how to allocate their funds and efforts. Technology forms part of the essential infrastructure upon which libraries rely on to carry out every aspect of their work. In the context of these challenges, the technology systems that support libraries must offer much more than traditional functions. They must also be populated with content components and be intrinsically designed to connect and communicate with a complex array of persons and organizations.

The traditional integrated library system might be seen as a mere starting point of what libraries ultimately need for technology infrastructure. An automation system itself, though offering practical functionality, seems but an empty shell until populated with content resources and connected into the fabric of the library's business infrastructure and its patron's personal and social networks. The bibliographic records providing

a basic representation of its collection are a good starting point. Today's libraries benefit from a much richer approach to content and connections, which can be thought of in several categories.

E-Resource Knowledge Bases

Especially for academic libraries, e-resource knowledge bases are an essential content component needed for the management of collections increasingly dominated by electronic materials. The universe of scholarly and professional resources from which libraries select materials for their patrons via subscriptions or open access is vast and dynamic. Rather than requiring each library to track the titles and coverage available within each resource package offered, knowledge bases have been assembled in support of activities such as electronic resource management, OpenURL-based link resolvers, alphabetical lists of journal titles, and other finding aids. These knowledge bases enable portfolio-based resource management, where selecting a content package propagates the details of the titles and coverage of the materials it contains. To be effective, a knowledge base must accurately represent the totality of content packages offered so that libraries can activate the packages to which they subscribe. Naturally such comprehensiveness is but an ideal. Libraries and knowledge base providers must cooperate to fill in gaps and repair errors.

Assembling a knowledge base requires substantial resources, including the technical platform upon which it resides, the technical processes to populate it with data provided by publishers, personnel to manage its quality control, and business processes to negotiate with publishers to contribute title and holdings data. It is not surprising that only a handful of comprehensive e-resource

IN THIS ISSUE

Innovative Launches its New Knowledge Base
PAGE 3

An Update on FOLIO
PAGE 5

Update on Ex Libris Alma
PAGE 6

NISO Establishes New Workgroup to Recommend an API Framework for E-Content Integration
PAGE 7

knowledge bases have been created and that their producers are large and well-resourced organizations: ProQuest, Ex Libris, TDNet, EBSCO, and OCLC. Open access knowledge bases have been envisioned since the early days of electronic resource management, though none have approached the completeness of the commercial offerings. The Global Open Knowledgebase (GOKb) launched in recent years shows promise. New standards such as Knowledge Base And Related Tools (KBART) have significantly lowered the difficulty in compiling knowledge bases in more automated ways from title and coverage lists provided from publishers.

Deep Indexes

Library users bring along their expectations gained from Google and other search engines when they use the tools and interfaces provided by the library. The genre of web-scale or index-based discovery services aims to fulfill this expectation through a central index populated with the individual content items represented in the library's physical and electronic collections. While the traditional online catalog would list the titles of the periodical to which the library subscribes, these index-based discovery services take a giant step forward by addressing the individual articles or chapters. In many cases, these indexes are populated with the full text of the materials in addition to the structured citation terms.

It is important to understand the distinction between e-resource knowledge bases and discovery indexes. Knowledge bases address journal titles and the dates of coverage and may include a few million resource records. Discovery indexes are populated with articles and book chapters, and therefore the number of content items may surpass multiple billions. Discovery indexes ideally represent the totality of scholarly and professional literature. As libraries implement an index-based discovery service they may use their e-resource knowledge base to configure their implementation to address only the materials available via their subscriptions and open access selections.

Not only must these index-based discovery services address an incredibly vast scope of content, but they must also offer a user experience that meets the expectation of increasingly web-savvy users and provide convenient access to content items. Library patrons naturally prefer immediate access for reading or downloading, though patrons may need to visit the library for physical materials.

Even more than e-resource knowledge bases, index-based discovery services involve massive resources to create and

maintain. The acquisition and maintenance of the content addressed in their central index requires highly scalable technology infrastructure and state-of-the-art search and retrieval capabilities and user experience design. Again, only a small number of organizations have the capacity to produce these index-based discovery services: ProQuest, EBSCO, and OCLC. No comprehensive open access discovery index seems in sight.

Competitive and Comparative Intelligence

Librarians need lots of relevant data in order to make well-informed resource allocation decisions. Flat collection budgets and unrelenting subscription costs mean continually having to make difficult decisions regarding what items to cancel or renew and what new resources can be acquired. Traditionally such decision making has been informed via reports generated within the scope of the library's own automation system. It is naturally helpful to understand usage data, budget allocations, and other local data. Libraries benefit further through an expanded universe of data, such as the ability to compare their collection profiles with peer institutions, to have access to impact metrics of journals, and to have access to other relevant data sources.

It's likewise important to have strong analytical tools. As the volume of data increases, the ability to incorporate it in resource decisions becomes difficult without tools to visualize the data and make analytical comparisons. The provision of content and analytical tools for decision support is a component of technical infrastructure of increasing strategic importance.

Business Connections

Libraries have little time for redundant data entry. To the fullest extent possible, library systems must dynamically communicate with all relevant internal and external business partners. Patron records should be automatically populated and synchronized with student records and personnel systems of the parent institution. Financial transactions should be securely transferred to ERP or other financial systems. Orders, invoices, and other procurement transactions should be electronically exchanged with suppliers. Resource fulfillment requests should be disseminated to interlibrary loan or document delivery providers. These are just some examples of the array of business transactions needed in a library's technical infrastructure so that it can function with the least manual intervention possible. In the current phase of library systems, many of these activities take

Discovery indexes are populated with articles and book chapters, and therefore the number of content items may surpass multiple billions.

place through batch transfers of data files. As the API ecosystem matures, we can anticipate that these will increasingly become more transactional and dynamic.

Connections with Content Providers

The ability to seamlessly incorporate access to content provided by external providers ranks as an essential capability. Public libraries offering e-book and audiobook lending services typically take advantage of external providers, such as Overdrive, Bibliotheca, or Baker & Taylor. These providers offer a catalog of titles licensed from a wide array of publishers and have created platforms to deliver digital content on behalf of a library.

In the early phase of e-book lending, libraries would essentially jettison their patrons to the platform of their content provider to search, select, and download titles of interest. In this current phase, libraries increasingly demand that their e-book lending services take place within their own virtual environment, maintaining their branding and the attention of their patrons.

This more library-centered paradigm of digital lending requires a sophisticated infrastructure. Instead of a simple hand-off, library catalogs and discovery interfaces need to engage in a complex set of interactions with the content provider platform to incorporate its titles in search results, present descriptive information, and ultimately check out and download a digital item. Supporting technologies include harvesting metadata into the library's discovery environment and a suite of APIs to manage the presentation and streaming or downloading of content.

Important progress has been made in the integration of e-book lending services into a library's virtual presence. Many of the developers of library catalogs have worked out partnerships with the major digital lending providers to implement mutual APIs to conduct these transactions, though some gaps remain in the overall experience. The process remains

technically challenging with different APIs or other technical mechanisms needed for each combination of discovery interface and digital content provider. The complexity and other shortcomings of the current state of integration technologies drive the need for a more consistent and robust approach. The recently formed NISO workgroup on API framework for e-content integration, covered in this issue of *Smart Libraries Newsletter*, promises to make important progress in this area.

Connections with Patrons

Ultimately, library content and services must flow to the patron. It is important to present the virtual presence of the library in a way that meets expectation of user experience and to deliver content in forms that can be conveniently accessed, stored, and manipulated. Patrons, of course, don't come with predictable technical infrastructure as do the other entities with which libraries interact. The means to deliver content in more meaningful ways will always be highly variable. One of the key challenges for libraries lies in improving the ways that we deliver information resources to mobile devices and to social and professional networks.

It is easy to think of library systems in terms of checklists of functionality that they can perform. Such an approach was adequate in times where libraries were more isolated and worked mostly with collections they owned, which were housed in their own physical facilities. Times have changed dramatically. Library collections are increasingly digital and delivered through a long roster of external providers. Library collections can no longer be managed through the brute force of one-item-at-a-time procedures, but must rely on processes which work at scale, powered by knowledge bases and pre-populated indexes. Libraries today require not only appropriate features and functionality in their technology environment, but an ecosystem of content components and connections into the broad landscape of their partners and providers.

*Library content and services
must flow to the patron.*

Innovative Launches its New Knowledge Base

Innovative Interfaces has launched the Innovative Knowledge Base (release 1.0) to assist academic libraries in the management of electronic resources. The knowledge base includes both technology and content components in support of more efficient management of electronic resources. The technology

components are developed by Innovative as part of its open initiative, while the content components are based on data acquired from other services or providers. The Innovative Knowledge Base can be used to support electronic resource management performed within Sierra or third party tools, OpenURL

resolution via WebBridge LR, and other standard link resolvers and index-based discovery services, such as Encore Duet.

The knowledge base is a multi-tenant cloud-based offering built on the Innovative Open Library Stack platform that Innovative has developed for the deployment of new products and services. This platform, which leverages an infrastructure resident in the Amazon Web Services set of global cloud services, provides a highly scalable, reliable, manageable, and secure modern foundation for Innovative's new applications. The technology components of the knowledge base include tools for working with bibliographic and holdings metadata, integration with Innovative's electronic resource management functionality, and infrastructure upon which the Innovative Knowledge Base resides.

Innovative focuses on the development of technology, but also has experience curating metadata content through their SkyRiver bibliographic utility service. Similarly, the Innovative Knowledge Base will be populated from a variety of proven sources, including partnerships with commercial organizations, as well as open access projects. EBSCO Information Services contributes ongoing updates from their e-resource knowledge base that was developed to support the company's own Full Text Finder, link resolver, and electronic resource management tools. According to EBSCO, the knowledge base includes over 3.2 million unique titles and over 11,500 databases and packages. Libraries using the Innovative Knowledge Base can also take advantage of holdings data from the KB+ open knowledge base that currently describes more than 1,529 packages and 30,601 titles. Innovative states that as the product evolves, libraries will benefit from the inclusion of other commercial and community-supported repositories.

Each organization subscribing to the Innovative Knowledge Base will benefit from its own view of the data, stored in Innovative's Open Library Stack. The product includes the ability for libraries to maintain a profile of local holdings coverage and mechanisms to automatically synchronize the library's local automation system, such as Sierra or Polaris, with the central system of record. Innovative will maintain the accuracy of the knowledge base by publishing updates to the global content every two weeks as part of this initial release, with a goal of increasing frequency over time.

Innovative emphasizes the open character and flexibility of its knowledge base. It is not locked into a single monolithic source but enables the library to work with a variety of sources to assemble their own instance of the knowledge base that best meets their needs.

The new knowledge base supersedes the company's legacy offering, Content Access Service (CASE), a set of coverage metadata in support of the WebBridge link resolver and the

Millennium Electronic Resource Management module, which initially launched in October 2005. CASE was based on metadata licensed from OCLC, which no longer offers this data service. OCLC's data service was based on the knowledge base it gained through its acquisition of Openly Informatics in January 2006. OCLC has since revamped the knowledge base that underlies its WorldShare License Manager and other services reliant on e-resource holdings data.

CASE was based on older technology and workflow models that required each institution to manually download coverage, derived from the CASE database, into their local systems. CASE data was updated monthly from a single third-party source, and any resources owned by a library that were not represented in CASE had to be added and maintained manually.

By contrast, the new Innovative Knowledge Base employs a responsive web-interface with predictive search capabilities and rapid response times based on the elastic nature of the platform. Libraries using the Innovative Knowledge Base have the option to compare local holdings with those of other participating libraries, and they can take advantage of open representational state transfer (REST) APIs that support the integration of multiple content streams, as well as the ability to automate library workflows that engage multiple systems.

E-resource knowledge bases exist to simplify the management and access of electronic resource packages. Most of the electronic resource packages to which libraries subscribe contain many individual titles. It is essential for a library to know what titles are included and what range of dates is included. Rather than each library having to independently track coverage and holdings, a knowledge base provides this data in a continually updated form. A knowledge base should be populated with the largest possible representation of the academic and professional resources published. Libraries subscribing to the knowledge base can then activate the specific packages or portfolios to which they subscribe. This model greatly simplifies the management of electronic resources by requiring libraries only to activate portfolios rather than to track individual titles and holdings manually. The knowledge base must naturally be as comprehensive and accurate as possible so that libraries do not have to make corrections or create their own profiles for packages not already represented.

The concept of using knowledge bases as in support of electronic resource management and linking is well established. Ex Libris introduced SFX in June 2000, based on technology it acquired from Ghent University. Ex Libris was the first company to offer an OpenURL-based link resolver based on a knowledge base of e-resource holdings. Serials Solutions introduced its Article Linker (later 360 Link) in 2002, which was based on a knowledge base of e-resource holdings. Serials

Solutions earned a reputation of providing a very high quality knowledge base. Serials Solutions was acquired by ProQuest in July 2004. ProQuest's acquisition of Ex Libris in December 2015 will lead to the merging of the knowledge base curation processes of the two antecedent companies, though the technical implementations may remain separate. Other commercial providers of knowledge bases include the EBSCO Information Services, whose knowledge base supports their own link resolver and discovery solutions, as well as TDNet.

Proprietary knowledge bases have dominated the realm of electronic resource management and linking. Although some efforts at creating open access knowledge bases have been undertaken, they did not reach a level of completeness and accuracy to compete with the proprietary products. A more promising endeavor, the GOKb, was launched in June

2012 with funding from the Andrew W. Mellon Foundation to create an open access knowledge base managed through a platform based on open source software. GOKb currently describes 47,959 titles and 524 packages. The sister project run by JISC in the UK, KB+, has seen similar results.

Knowledge bases of e-resource holdings represent a critical component of an automation environment for academic libraries. Most academic libraries dedicate the vast majority of their collection budgets to e-resource subscriptions with much smaller proportions allocated to physical materials. A well-accepted approach to managing electronic resources involves workflows based on activating portfolios, with granular holdings provided via a knowledge base. Given that academic libraries represent about half of Innovative's customer base, the provision of a high-quality knowledge base is critical.

An Update on FOLIO

Progress continues on the FOLIO project to create a new open source library services platform for academic libraries. Recent milestones include the initial release of the lower-level platform software needed for the development of functional apps and some new organizational structures.

FOLIO will be based on a microservices architecture, with a lightweight platform providing common infrastructure services to the functional modules or apps. The development of this infrastructure layer was contracted to Index Data, a firm based in Copenhagen, Denmark and Boston, MA that specializes in open source library software. The initial announcements of the FOLIO project anticipated an initial release of the framework in August 2016, which was accomplished on schedule.

Generally consistent with this timeframe, Index Data has completed some of its initial work for FOLIO, with code repositories now openly available on GitHub. It must be emphasized that the software currently available is oriented to developers and does not yet provide higher level functionality. According to Sebastian Hammer, Chief Strategist for Index Data, "the software created at this time should be considered more like an operating system to support apps, which will provide smaller units of functionality which can then be stitched together to form complex systems."

The architectural design of FOLIO organizes the technical infrastructure into several layers. The lowest system layer will be comprised of components mostly oriented to storage for bibliographic and transactional data as well as indexing and configuration services. Another layer, called OKAPI, serves

as an API gateway, providing a variety of services to manage communications among apps. A UI toolkit will also be provided to facilitate the creation of apps with consistent user interfaces. The FOLIO development environment will include Stripes for OKAPI, which builds on the React framework created by Facebook and Instagram. The creation of apps can commence in earnest once the system layer, OKAPI, and UI Toolkit are in place.

A variety of communications channels have also been set up to support developers interested in working on FOLIO. The project envisions individuals and organizations throughout the world contributing their efforts to design, develop, test the software, and create documentation. Robust communications will therefore be required to coordinate these efforts.

Conversations among developers are currently taking place in Slack (folio-project.slack.com) as well as in threaded discussions managed through Discuss (discuss.folio.org). The main page, folio.org is growing rapidly to include or link to all aspects of the project.

The Open Library Foundation was recently established to house the governance of FOLIO and to support other allied open source projects, such as the GOKb and the Open Library Environment (OLE). Open source projects need some type of organization to manage decision making, to hold assets such as copyrights or other intellectual property, and to acquire and distribute any financial resources. Examples can be seen across many open source software projects. These roles are provided by LYRASIS for CollectionSpace and ArchivesSpace;

DuraSpace provides these services for DSpace and Fedora. The Apache Foundation manages many of the large-scale open source projects such as the Apache web service, SOLR, Lucene, and dozens of other core technologies for the web.

The Open Library Foundation was initially established as a legal entity in June 2016 as a first step in becoming an organization able to provide services for open source projects. In September 2016 the Open Library Foundation announced its website (openlibraryfoundation.org) that FOLIO and GOKb had joined. This foundation functions as a legally independent entity.

Most foundations are non-profit corporations. The Open Library Foundation was incorporated in March 2016 in Delaware and is currently working to obtain 501c3 status, which is needed to be recognized by the Internal Revenue Service as a non-profit and not subject to federal taxes. The foundation's initial board of directors includes:

- Michael Winkler, the Managing Director of the OLE
- David Carlson, Dean of Libraries at Texas A&M University
- Sam Brooks, Executive Vice President of EBSCO Information Services

Winkler states that the foundation intends to expand the board to include up to 12 members, representing a broader range of academic institutions and organizations with related interests.

EBSCO Information Services has been deeply involved with FOLIO, both in its initial conception and in financial support. The company, however, does not own the software

produced nor does it have unilateral control of its design or the processes in which it is produced. Rather, EBSCO participates in its governance among other stakeholders. The governance and administrative structures are oriented to a broad community of individuals and organizations interested in open source library software apart from direct commercial ownership.

The OLE was originally formed to produce a new open source resource management system and joined with the Kuali Foundation that supported a variety of large-scale business applications for higher education. The shift of the Kuali projects to a more commercial model, a delayed schedule in delivering the software, and the advent of FOLIO led this group to change its course. Its organization and most member institutions have opted to engage with FOLIO rather than to continue the development of the Kuali OLE software. In the transition from Kuali to FOLIO, OLE has seen the addition of some new partners, such as Texas A&M, Cornell University, and two German consortia (hzb and GBV). Some of the original OLE partners have not engaged with the new direction, including Indiana University and University of Pennsylvania. The Penn Libraries recently announced that they have contracted with Ex Libris for Alma, which will be used with their Blacklight-based open source discovery environment.

Winkler previously served as the Director for Digital Partnerships at the University of Pennsylvania prior to shifting to the full-time role as OLE's Executive Director. This position is currently administratively housed at Cornell University.

Update on Ex Libris Alma

Ex Libris continues to see strong interest in its Alma library services platform. It especially has been successful among large universities, multi-campus university systems, and consortia comprised primarily of academic libraries. The product continues to mature in features and is delivered with an integrated knowledge base. Its development commenced in 2009, with libraries in production beginning in 2011. Deployed as a web-native multi-tenant platform, Alma has monthly updates with new features and bug fixes, which have been continually applied over the life of the product.

The company announced that by September 2016 that there are now over 500 live implementations of its Alma

library services platform. The company reached this milestone with the Renouvaud library network in Switzerland, serving over 100 libraries in the canton of Vaud. The library network switched from Virtua to Alma as their production automation environment.

Alma has been selected by 31 of the 125 members of the Association of Research Libraries. The University of Pennsylvania, formerly one of the Kuali OLE partner libraries, recently announced its selection of Alma. All previous institutions selecting Alma have also selected Primo as its patron-facing discovery service. The Penn Libraries, however, plan to implement Alma with a customized discovery interface

based on the open source Blacklight software. This unbundled implementation will set a precedent. Ex Libris has also committed to supporting Summon as a discovery interface for Alma following its acquisition by ProQuest.

Ex Libris also touts the openness of Alma made possible through a robust suite of APIs. Libraries increasingly expect to access the data and functionality of their systems through

APIs. Ex Libris reported in June 2016 that the Alma platform currently services over 1 million API requests daily.

Alma has also been the launch pad for other products. Its Leganto product for the management of course lists, launched in September 2015, has been purchased by 20 libraries in 9 different countries in its first year of availability.

NISO Establishes New Workgroup to Recommend an API Framework for E-Content Integration

The National Information Standards Organization (NISO) has approved the formation of a new project to facilitate discovery and access to e-books and other digital content via library interfaces. Libraries increasingly require modern mechanisms to enable them to incorporate digital content from external providers into their own services and interfaces. While progress has been made in the realm of integration of e-book lending services provided from external providers into library catalogs and discovery interfaces, no consistent approach is available that works across all providers. Many of the mechanisms currently available are cumbersome and do not follow modern technical conventions.

This new NISO project will produce a new API framework that can be implemented by digital content providers and libraries able to achieve interoperability with a lower threshold of difficulty and with more flexibility. Rather than working out individual interoperability techniques between each set of partners, the project aims to create a single set of APIs that can be

implemented uniformly and ultimately result in more robust services for library patrons than previously possible. The framework will be based on RESTful APIs and Mobile application calls, consistent with modern technology practices.

This workgroup will work under the Discovery to Delivery Topic Committee (D2D), one of NISO's high-level committees that oversee workgroups that create or maintain specific standards or recommended practices. This project was approved by D2D in July 2016 and approved by NISO voting members in August 2016.

The project was initially proposed by the Queens Library, which had developed its own set of API Requirements for e-Content Partners. The APIs specified in this document will serve as the starting point for the workgroup. The workgroup is in charge of specifying an API framework that will be issued as a NISO recommended practice if accepted by D2D and ratified by the NISO voting membership.

For more information, see niso.org.

Questions or suggestions
for topics in future issues?



Contact Marshall Breeding at
marshall.breeding@librarytechnology.org



Smart Libraries Newsletter
American Library Association
50 East Huron Street
Chicago, IL 60611-2795 USA
Address Service Requested

NON PROFIT
US POSTAGE
PAID
PERMIT 4
HANOVER, PA

October 2016 Smarter Libraries through Technology

Smart Libraries Newsletter

Marshall Breeding's expert coverage of the library automation industry.

Editor

Marshall Breeding
marshall.breeding@librarytechnology.org
Twitter: @mbreeding

Managing Editor

Samantha Imburgia
312-280-3244
simburgia@ala.org

Digital Access for Subscribers

journals.ala.org/sln

TO SUBSCRIBE

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

The 2016 subscription price is \$85 in the United States and \$95 internationally.

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

Production and design by the American Library Association
Production Technology Unit.

Smart Libraries Newsletter is published monthly by ALA TechSource, a publishing imprint of the American Library Association.

alatechsource.org

Copyright © American Library Association 2016. All rights reserved.