Smarter Libraries through Technology: Striving for Neutral and Objective Coverage

By Marshall Breeding

In my role as a journalist and analyst for the library technology industry, I aim to provide objective coverage of all the major stakeholders. Maintaining objectivity means taking a non-promotional, or neutral, position relative to the organizations, products, technologies, and even development models.

My readers include library staff that use technology products and services and the people who create and support them. I aim to help libraries understand the range of options available, providing enough information, context, and perspective to help decision-makers shape technology strategies and assemble their technical infrastructure. The newsletter's readership also includes future library or information professionals. I'm delighted when I meet new librarians and students who mention reading my essays and articles as part of their education about technology in libraries. People on the supply side of library technology products and services represent another important segment. Not only does each organization pay close attention to how they are covered, but they naturally want to thoroughly understand their competition and all alternatives available to libraries.

In my coverage of the library technology industry, I aim to be neutral relative to each of the organizations involved and the various organizational and development models. I have valued relationships with individuals at all levels of the organizations—often including the CEO—of each of the major companies in the industry. These relationships provide important opportunities to gather information about the products, roadmaps, and strategies of each of the companies that inform specific articles or reports

Get Your Smart Libraries Newsletter Online in PDF!

Subscribers to ALA TechSource's *Smart Libraries Newsletter* or *Library Technology Reports* can read digital versions through the scholarly content host, MetaPress. Subscribers also have access to an archive of past issues. Visit alatechsource.metapress.com to begin reading. Beginning with the April 2015 issue, *Library Technology Reports* and *Smart Libraries Newsletter* will leave the MetaPress platform. Subscriber will have ongoing access to the current issue, archives, and other American Library Association publications at journals.ala.org/ojs. Thank you for subscribing! Your purchase funds advocacy, awareness, and accreditation programs for library professionals worldwide.

I will not publicly promote any given organization above another and . . .
I will cover each objectively, including both strengths and weaknesses

that I write or that provide background context. In all these relationships is an explicit understanding that I will not publicly promote any given organization above another and that I will cover each objectively, including both strengths and weaknesses. My ability to continue these relationships and to gather information about the companies and products depends on maintaining confidences and by ensuring the accuracy and balance of my reports and public presentations.

I often write about the business side of the industry, including the mergers, acquisitions, and ownership status of organizations. The industry comprises a diverse range of organizations—for-profits and nonprofits; companies owned by founders or their families others by private equity or smaller investors; and even publicly traded. I appreciate the benefits of each of these ownership models as well as the potential pitfalls. I generally observe the strategies and personalities influence outcomes for libraries more than the ownership arrangement or organizational models.

The library technology industry includes a mix of open source and proprietary offerings. Here also, I hold a neutral stance. I have been involved in open source projects and use open source software and see many positive qualities in this model of software development. Companies that produce products and services not released as open source dominate in most segments of the industry, providing products that benefit libraries. But even these proprietary products are increasingly expected to provide robust and well-document APIs that open their data and functionality to libraries.

In the current state of the industry, proprietary and open source products compete vigorously in many sectors. Open source ILS products, such as Koha and Evergreen, now considered mainstream options, compete with the proprietary offerings based on their merits in functionality, design, architecture, support services. With both the open source and proprietary products, I also aim to provide as much information as I can about the organizations behind their development support. In an industry of mostly private companies, I am generally pleased with the level of disclosure by these organizations. For my various reports they provide figures regarding many aspects of their businesses, such as sales performance, numbers and allocation of personnel, and often customer lists. In the finite universe of libraries, it is usually possible figure out what systems

are acquired by what libraries in any given year through other means, but getting that information directly enhances my ability to track and analyze the trends in the industry.

In describing the organizations that produce technology products, I am constantly reminded of the difficulty in comparing the companies that produce proprietary products with the communities involved in creating open source software. Among the many characteristics of these organizations, I consider their capacity for development and support one of the important factors. When these functions fall within the same organization, it is relatively easy to quantify them. To the extent that organizations disclose the numbers of personnel allocated to these activities, it is possible to at least make rough comparisons regarding their relative capabilities.

In developing the open source Koha ILS story in this issue, I wanted to provide perspective on the relative size of its development community compared with that of the organizations that product proprietary software. My initial strategy involved attempting to aggregate the numbers of developers employed by each of the firms involved with Koha, as well as those working individually. Outside of the support firms, many individuals working in libraries voluntarily spend part of their time contributing to open source projects, some with no specific affiliations. Throw in similar scenarios, involving hundreds of individuals distributed throughout the world with no specific organizational structure, and the calculation of effort turns out to be a quite complex task. This ethos of open source development does not necessarily lend itself to the tracking of time. Projects such as Koha, when managed through GitHub or similar repositories provide another set of measures, especially the number of lines of code committed by each developer. The statistics describing the Koha codebase are truly impressive. So while I have not yet been able to derive a general comparison between the Koha development community and the development resources available to the proprietary products, I have gained more of an appreciation for the quantity of efforts contributed and for the fundamental differences in these two approaches to software development.

Neutrality does not imply lack of critical assessment. It means approaching each product, company, or development model without an initial bias. Absent that bias, a multitude of comparisons can be made to identify the relative merits and weaknesses of each alternative. It's also important to recognize the differences in the requirements of libraries where any given characteristic may be well-suited for one type of library and a liability for another. Creating data and points of reference that help to objectively compare and analyze the various alternatives have long been the basis of how I cover this incredibly interesting and complex library technology industry.

EBSCO Provides Major Grant to Open Source Koha ILS Project

EBSCO Information Services has made a significant grant to the open source Koha ILS project in support of an ambitious set of enhancements and extensions. The grant was awarded to Koha Gruppo Italianio, an organization devoted to the promotion of Koha in Italy. The development work of the enhancements covered in the grant will be carried out through other Koha support firms, including Catalyst, based in Wellington, New Zealand and ByWater Solutions. This grant signals strong support toward open source library software that complements EBSCO's partnerships with companies offering proprietary ILS products.

Koha Gruppo Italiano initiated a grant request to EBSCO in May 2014; it was granted on Feb 11, 2015. The specific monetary amount of the grant was not disclosed, but the tasks funded by the memorandum of understanding were itemized in the press announcement and described below. The initial proposal requested funding for one from a selection of enhancements. According to Senior Vice President Scott Bernier: "We considered their proposal and decided that we wanted to support Koha in an even more significant way and decided to fund all of their enhancements."

Koha finds use in all regions of the world, implemented by an estimated 7,000 to 9,000 libraries and rising. As open source software that has been implemented by individual libraries and is promoted or implemented by regional or national library authorities in many areas of the world, tracking globally the exact number of libraries using Koha is difficult. In the United States, at least 855 library systems, representing 1,568 individual branches, have implemented Koha. Of this number, 41 (5.6%) have collections larger than 200,000 items; 353 (50.9%)have collections between 20,000 and 200,000; and 299 (43.1%) have collections smaller than 20,000 items. Many different types of libraries are represented, including 507 public, 104 academic, and 84 in K-12 schools. These numbers, taken from the libraries.org directory, may not be entirely comprehensive.

The April 2014 issue of *Smart Libraries Newsletter* provided coverage of Koha, including "History and Background of Koha."

The Koha Community

Koha is developed through a globally distributed community of developers. OpenHUB, a resource that analyzes open source software projects, characterizes Koha as having a "very large, active development team" with 88 individuals contributing new programming code in the last year; and 322 developers since the initiation of the project in 1999. OpenHub reports the Koha development team is in the largest 2 percent of projects it tracks.

Many libraries, especially in the developing world, implement Koha through the efforts of their own personnel, often assisted by members of the broader community via e-mail discussion lists and the Koha IRC channel. Dozens of commercial and nonprofit organizations around the world provide services to support libraries implementing Koha and that contribute to its development. One of the original developers of Koha, Chris Cormack, continues work with Koha through Wellington, New Zealand-based Catalyst, a large technology services company involved with open source software in many business sectors. Cormack reports that Catalyst employs more than 235 personnel, including 182 developers, though only a small number are devoted to Koha. In the United States, ByWater Solutions ranks as the leading support provider for Koha with at least 538 clients spanning 892 library facilities. Equinox Software, though primarily involved in services related to the open source Evergreen ILS, also provides support for Koha. BibLibre provides support services to libraries in France and has been one of the companies most involved in software development. PTFS Europe provides services related to Koha in the United Kingdom and in Europe. Koha has become a one of the preferred automation system in public and school libraries in Argentina and is gaining adoption in academic libraries. The Universidad Nacional de Córdoba has been the focal point support, advocacy, and development of Koha in Argentina. Dozens, if not hundreds, of other companies in many countries provide services in some way related to Koha.

The Koha Gruppo Italiano was formed in February 2012 to promote and support Koha throughout Italy. The organization focuses on advocacy, awareness, and fundraising more than technical support or development. With funding gained, it works with other Koha support firms, including ByWater Solutions, Catalyst, and BibLibre for the execution of technical projects. The organization has organized a number of events to introduce libraries in Italy to open source software, including VuFind and DSpace in addition to Koha. Individuals involved in the organization include:

- Franziska Wallner of the American University of Rome, which implemented Koha in 2006,
- Stefano Bargioni of the Pontificia Università della Santa Croce (Koha since 2011),

- Diego Ramirez of Pontificia Università della Santa Croce,
- Sebastian Hierl of the Arthur and Janet C. Ross Library of the American Academy in Rome, which implemented Koha in 2013 and relies on hosting and support services from ByWater Solutions.

Koha Gruppo Italiano has been active in raising interest and resources for the extension of Koha to include Elasticsearch since February 2014.

Koha Technology Stack

The Koha ILS has been evolving in functionality and architecture since its initial version created for the Horowhenua Library Trust, a relatively small three-branch library system north of Wellington, New Zealand. The original version was a relatively basic database-driven automation system. In the 15 years since its introduction, Koha has been developed into an integrated library system with features and functionality at a comparable level with many of the proprietary products. Koha continues to evolve in its technical architecture to sustain use by larger and more complex libraries.

The codebase underlying Koha is large. According to the GitHub stats for its public repository (http://git.koha-community .org/), the project currently totals 9,398,192 total lines of code, distributed over 6,196 program files.

Koha manages its data through the MySQL open source relational database management system, is programmed in Perl, relies on the Apache web service, and runs on the major implementations of Linux. Koha currently includes support for Zebra, an open source search engine developed by Index Data. Zebra includes native support for library-oriented standards including MARC and Z39.50. It is a relatively lightweight infrastructure module that easily co-exists within the same server as the other components of Koha.

The earliest versions of Koha relied entirely on MySQL to deliver search results. LibLime, one of the early commercial support and development firms for Koha, began work in December 2005 to incorporate Zebra from Index Data to support the MARC database and search components of Koha. (LibLime has since been acquired by PTFS and focuses on its own LibLime Koha ILS, which forked from Koha in about 2009.) As Koha reaches into larger and more complex libraries, there has been considerable interest in an alternative search technology for greater scalability and additional features.

Outside the scope of the grant, architectural improvements are being developed to improve Koha's performance in processing transactions. As noted, Koha is written in Perl, which interprets each script into a computer's native machine language in real time rather than being compiled in advance. Each script is also normally executed as a new process, adding system overhead. In an implementation with a high transaction load, these factors can lead to slower performance. To improve performance, an additional layer, called Plack, is being introduced that significantly reduces processing overhead and increases performance. Plack support has already been implemented selectively. Some support and hosting providers have already incorporated it in the online catalogs of their production systems. Delivering the staff interface via Plack is also underway with completion expected in the coming weeks.

New Search Architecture Based on Elasticsearch

One of the most ambitious tasks supported by the EBSCO grant will be the extension of Koha to incorporate Elasticsearch (http://www.elasticsearch.org/) as a new option for search and retrieval. Zebra has been a very pragmatic indexing engine and search component for Koha, but does not necessarily offer the performance levels and features of alternatives such as Apache SOLR or Elasticsearch, which were developed for very large scale applications across many different industry sectors. Zebra, which requires fewer hardware resources to operate, was originally developed for library-specific applications, with direct support for MARC record formats.

Elasticsearch and Apache Solr are the two most popular indexing and search servers for large-scale applications that involve retrieval of information from large content repositories. Both offer a powerful set of features and are offered as open source software. Elasticsearch and Apache Solr rely on Apache Lucene as part of its indexing and search infrastructure. Solr is used by index-based discovery services, including Ex Libris Primo and ProQuest Summon.

The use of Elasticsearch with Koha will also provide more capabilities related to the creation of facets to narrow search results and in improved relevancy in the presentation of search results. Elasticsearch finds use in many large-scale projects, such as for The Guardian, processing more than 40 million documents per day; GitHub, indexing for all the repositories; and many others (see http://www.elasticsearch.org/case-studies/).

The enhancement of Koha to use Elasticsearch will be implemented as an optional installation configuration. Many smaller libraries may not require the additional capabilities of Elasticsearch and may prefer not to have to manage a more complex set of components that come with its use. Koha with the Zebra search component can be managed easily on a single server.

Work on the implementation for Elasticsearch in Koha was already underway prior to the award of this grant. Currently

one developer is working on the programming to accomplish the integration, with to others assisting in testing, according to the Universidad Nacional de Córdoba's Tomás Cohen Arazi, an active participant in the Koha community of developers and the Release Manager for Koha 3.20.

Other New Features

The grant will also fund the development of the ability for patrons to browse the contents of the library's collection according to author, title, subject, or call number. Browsing according to these categories has been a standard feature in major ILS products. Koha has previously offered the ability in its advanced search to limit queries by specified fields, but has not offered structured browsing.

Other enhancements covered include the development of a "MARC to RDF crosswalk" and the ability to support other forms of metadata besides MARC21 to describe resources. This capability will be facilitated through the implementation of Elasticsearch, according to Sebastian Hierl, member of Koha Gruppo Italiano and librarian for the American Academy in Rome.

Extending the Koha Patron API

The scope of the grant includes enhancing the functionality of the Koha APIs related to patron-oriented tasks, which benefits Koha in a variety of ways. EBSCO will benefit by improved integration of its EBSCO Discovery Service in libraries using Koha. EBSCO works with almost all of the major ILS products, including Koha, to integrate its EBSCO Discovery Service. This integration can take the form of either a full replacement for the online catalog of the system or as a supplementary article-level index queried through the interface of the online catalog pro-vided with the ILS. EBSCO has not developed its own ILS, but has worked to bolster the exposure of EBSCO Discovery Service through partnerships with a broad range of ILS providers, including both proprietary and open source products.

One example of many libraries that have used EBSCO Discovery Service in conjunction with Koha include the Hammermill Library of Mercyhurst University. EBSCO reports that there hundreds of libraries using Koha that also subscribe to EBSCO Discovery Service, many of which are fully integrated.

EBSCO has also provided support for the Kuali OLE project. The company became a Kuali Commercial Affiliate in April 2013. The scope of Kuali OLE does not include the provision of a discovery interface. In addition to its general support, EBSCO facilitates the integration of Kuali OLE with EBSCO Discovery Service as one of the interface options available. As seen in its partnerships with proprietary providers and open source projects, EBSCO supports a library technology ecosystem or infrastructure that gives libraries options to choose different discovery interfaces with any given integrated library system. This separation between resource management products and discovery services opens options for libraries that may have different choices for these two different domains. The ability to mix and match discovery and management products depends on a robust set of APIs to enable interoperability among diverse systems.

This grant by EBSCO to the Koha project reflects an interesting set of dynamics between one of the largest companies in the industry, involved in a wide range business activities related to providing content, services, and technology to libraries, and a broad-based open source project. As Koha continues to find use in a growing and diverse set of libraries, EBSCO gains recognition within that global community, which includes libraries that may also be current or potential subscribers to its products and services.

Reading List Product Category Grows

There has been considerable interest internationally for a genre of products that manage lists of materials assigned by instructors for the classes offered in a college or university. Two of these reading list management products compete in the UK, including Talis Aspire Reading Lists and Rebus:List introduced in July 2012 by PTFS Europe. While these products may be populated primarily by content provided by the library, they also must handle materials outside of its collections. Reading lists generally have a closer connection with the learning management systems used by the campus than the resource management system or discovery service of the library.

Talis recently announced that 77 higher educational institutions have selected its product, primarily in the UK, but also including universities in Australia, New Zealand, Norway, Cyprus, Malaysia, and China. PTFS Europe reports 21 universities using its Rebus:list, all in the United Kingdom.

Until recently reading list products have not seen strong interest in the United States. US-based institutions have not

subscribed to the products offered by PTFS Europe or Talis, and until recently technology providers active in the US have not offered these products.

EBSCO Information Services entered the fray in August of 2014 with a related product called Curriculum Builder, which was covered in the August 2014 issue of *Smart Libraries Newletter*. Curriculum Builder leverages the content and functionality of the library's implementation of EBSCO Discovery Service to provide a plug-in to the institution's learning management system. Using the industry-standard Learning Tools Interoperability (LTI), course instructors can build resource lists, including articles or other materials from the library's subscriptions or from their own personal copies of materials. Curriculum Builder operates entirely through the LMS course page and does not require that the instructor directly invoke EBSCO Discovery Service via the library's site.

Ex Libris has now launched an initiative to develop its reading list product in partnership with a group of five universities, an internationally diverse set of Ex Libris Alma customers: KU Leuven in Belgium; Imperial College and Kingston University in the United Kingdom; the University of Oklahoma; and University of New South Wales in Australia. The product is in the early development phase and will be designed with the input of

Perceptions 2014: An International Survey of Library Automation

This eighth annual Library Automation Perceptions Report provides evaluative ratings submitted by individuals representing over three thousand libraries from 80 countries describing experiences with 154 different automation products, including both proprietary and open source systems.

these development partners. This yet unnamed product, will interoperate with the institution's learning management system and will be based on new functionality delivered through Ex Libris' Alma platform. Readers can expect a more in-depth treatment of this product once more information is available.

Reading lists products have the potential to help educational institutions make better use of the resources acquired and managed by the library. They provide a locus of discovery that operates through the learning management system in addition to the catalogs or discovery services offered by the library. In many libraries, electronic reserves programs and the separate systems used to manage them have waned as course materials shift to being managed by learning management systems. While this shift have relieved libraries of some of the burden involved in identifying, acquiring, or even digitizing course-related materials, it has also meant a further degree of separation of libraries from involvement with the content consumed in the classroom. This emerging genre of reading list management gives libraries an opportunity to ensure that their collection materials are easily accessed for inclusion in course reading assignments, but also provides statistics and analytics to help acquire them acquire content that is well-aligned with the curriculum.

Library Technology News

ProQuest transforms research with digitization of rare, historical works from Bibliothèque nationale de France

ANN ARBOR, Mich., February 12, 2015—Researchers can now explore early European history and culture as it happened with ProQuest's release of the first 2 million pages in its massive digitization project with Bibliothèque nationale de France (BnF) in Paris. When complete, the project will give researchers cover to cover access to more than 28,000 rare European books printed from 1400 to 1700—10 million pages—in crisp, fully searchable images. Approximately 5,800 titles are now available in Collections 4 and 6 of ProQuest's Early European Books, enabling

researchers around the world to benefit from the Library's centuries of acquisition, curation and preservation.

Explore pages from books in this collection here: http://bit .ly/BnFimages.

The BnF's collection is vast and wide-ranging, including 3,000 incunabula (works printed before 1501) and many rare and valuable yet obscure texts. Subjects range from literature and history to science and engineering, from law to aesthetics and art criticism, from politics to philosophy and theology. The books themselves come in many forms, with popular chapbooks (which were widely distributed but rarely preserved) at one end of the spectrum and luxury editions aimed at a wealthy, courtly audience at the other.

OCLC and Library of Congress linked data initiatives compared and contrasted in new white paper

29 January 2015—OCLC and the Library of Congress have jointly released a white paper that compares and contrasts the compatible linked data initiatives at both institutions. The paper, "Common Ground: Exploring Compatibilities between the Linked Data Models of the Library of Congress and OCLC," is an executive summary of a more detailed technical analysis that will be published later this year.

The white paper summarizes the recent activity of the Bibliographic Framework Initiative at the Library of Congress which proposes a data model for future data interchange in the linked data environment that takes into account interactions with search engines and current developments in bibliographic description.

This summary also provides an overview of OCLC's efforts to refine the technical infrastructure and data architecture for at-scale publication of linked data for library resources in the broader Web. In addition, it investigates the promise of Schema.org as a common ground between the language of the information-seeking public and professional stewards of bibliographic description.

Access the white paper at: oc.lc/CommonGround.

ProQuest brands new integrated ebook platform: ProQuest Ebook Central

ANN ARBOR, Mich., January 30, 2015—ProQuest unveiled the name of its anticipated new ebook platform: ProQuest Ebook Central. Launching in mid-2015, ProQuest Ebook Central will integrate key elements from both ebrary and EBL - Ebook Library, along with all-new functionality.

The patron interface for ProQuest Ebook Central and the ProQuest Ebook Central Readerwill be based on ebrary's recently re-designed reader.

The librarian portal—ProQuest LibCentral—will be based on EBL's popular, newly rebuilt LibCentral administrative module. It will support discovery, selection, acquisition, customizable management and analytics of ebooks.

ProQuest will announce in early 2015 the launch of a beta program for ProQuest Ebook Central. A stepped rollout of the fully integrated platform will follow.

Questions or suggestions for topics in future issues?

YOUR COMMENT COUNTS!

Contact Marshall Breeding at
marshall.breeding@librarytechnology.org

March 2015
Smarter Libraries through Technology

Smart Libraries Newsletter

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

TO SUBSCRIBE

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

The 2015 subscription price is $85 in the United States and $95 internationally.

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