Smarter Libraries through Technology

Reshaping Academic Library Systems

By Marshall Breeding

The last five years have brought substantial change in the realm of resource management and discovery services for academic libraries. Key drivers to these changes relate to the urgent need to have workflow tools for library personnel more in tune with the shape of their collections, to provide effective tools for discovery, and to provide the infrastructure for supporting library activities beyond traditional areas of service.

The increasing dominance of electronic resources and digital collections created tension with the longstanding integrated library systems (ILSs) in place that were designed to manage print materials and were not especially well suited for managing complex collections of electronic resources. The business processes for licensing packages of content, managing access to open access materials, and other workflows are entirely different from purchasing print materials.

An early wave of change began in around 2009 with the introduction of index-based discovery services able to search and retrieve material within a library's subscriptions at the article level. These products, including Serials Solutions Summon, Ex Libris Primo, WorldCat Local, and EBSCO Discovery service, have become very widely adopted. Although exact statistics are not available, most major academic libraries in the United States, Canada, Europe, Australia, and other developed countries have adopted some form of index-based, article-level discovery. Given the costly investments academic libraries make, it has been viewed as important to provide patron-facing interfaces able to fully exploit their collections of electronic scholarly resources.

These index-based discovery services represent only one flavor of access. Many—if not most—users do not make use of the interfaces provided by the library, but instead conduct their research via Google Scholar, other general web search engines, or discipline-specific tools. This reality drives interest in implementing mechanisms that improve the discoverability of library resources through these web search engines, often based on semantic web or linked data technologies.

Another key strategy for improving access to library collections can be seen in reading list applications that can be embedded in learning management systems. These products, such as Talis Aspire, Rebus:list, EBSCO's Curriculum Builder, and Ex Libris Leganto, enable course instructors to select reading materials within the library's collection or to obtain materials and manage copyright issues for those not already held.

We can expect further development on the discovery front. New products such as Yewno Discover have come out in the last year or so that take advantage of machine learning and other concepts related to artificial intelligence. These technologies may eventually provide comprehensive tools for accessing library collections beyond what has so far been possible through full-text indexing and linked data.

Library services platforms have been on the rise since about 2011 to help library staff manage complex collections spanning electronic, digital, and print formats. Ex Libris and OCLC both launched development efforts at that time, which have resulted in products that have been subsequently implemented by a large portion of academic libraries. Ex Libris Alma has been especially popular among large academic libraries, multi-campus systems, and consortia. Alma was originally designed to work with Primo as its patron interface. However, in recent years, support

has been extended to Summon and open source interfaces like Blacklight and VuFind. OCLC has likewise been quite successful with its WorldShare Management Services, though its customer base is skewed more toward mid-sized libraries than the larger tier. Two initiatives to create library services platforms did not result in completed products—ProQuest Intota and Kuali OLE.

In the six years that Ex Libris Alma and OCLC WorldShare Management Services have been available, they have captured a substantial portion of the academic libraries seeking to migrate from legacy ILSs. The momentum of Alma seems especially strong, winning the majority of deals involving large libraries and consortia. Ex Libris has continued to enhance Alma and expand its capabilities into new areas.

Despite strong momentum, the shift of academic libraries from ILSs to Alma and WorldShare is far from universal. The FOLIO project posits that time remains for the introduction of a new option. FOLIO has been positioned to attract libraries interested in a different approach relative to the prevailing options. It appeals to the interest in open source rather than proprietary software, to modular design rather

than comprehensive platforms, to software development that includes libraries as well as vendors, and a fresh approach to technical infrastructure based on microservices.

The next year or so will be a critical time for FOLIO. The existing library services platforms, Alma and WorldShare, continue to gain in strength and have reached a level of maturity that many libraries wait for rather than join an early adoption cycle. Another segment of academic libraries may remain with their current ILS vendor, anticipating new developments oriented to improve their ability to support academic libraries. Companies like SirsiDynix and Innovative have a lot at stake in retaining their academic library customers. FOLIO likewise has its sights set on the academic library front. Although the clock is advancing rapidly, the FOLIO project has set out on an aggressive agenda and is backed by EBSCO, which brings substantial resources to the project. Time will tell whether FOLIO will disrupt the momentum of Alma and become established as another major option in the library services platform genre. This issue of *Smart Libraries Newsletter* features a new look at FOLIO and the major milestones it has accomplished since our last look at the project.

FOLIO: Advancing from Vision to Software

The open source FOLIO project to create a new library services platform continues to see progress in its development timeline. Since the project was launched in late 2015, the project has crossed a number of significant milestones in its technical development, community-building, outreach, and in support structures. Project documentation posits the end of 2018 for the availability of its initial version able to replace a library's existing ILS. This three-year development effort from project launch to a minimally viable product can be seen as ambitious but reasonably consistent with the timelines of other products. Figure 1 illustrates the development and production timelines for the successful and failed initiatives to develop library services platforms (Available on Library Technology Guides: https://library technology.org/chron/libraryservicesplatforms.pl/).

Index Data was commissioned to perform the development of the core infrastructure components of FOLIO beginning in late 2015. One of the early software development goals involved the creation of the initial version of the microservices infrastructure components (OKAPI) and user-interface toolkit (Stripes). This work was completed by Index Data and made available in August 2016. The completion of these components paved the way for other organizations and individuals to begin

the work of creating the functional modules that would eventually comprise the FOLIO library services platform.

Most of the current efforts focus on creating the functionality that will enable libraries to migrate away from their existing ILSs and electronic resource management systems. This development is taking place through more than a dozen development teams, each working independently on a specific module or infrastructure component. Several commercial software development companies are contributing personnel to these teams. Several libraries are also directly involved in software development, including those associated with the Open Library Environment as well as others that have engaged with FOLIO independently.

Some of the commercial organizations involved in the development of FOLIO include:

- **Index Data**, a well-established, though modestly-sized development firm specializing in open source software components for libraries. The company has previously developed infrastructure components used within commercial and open source library systems to implement standard protocols such as Z39.50 and MARC-oriented
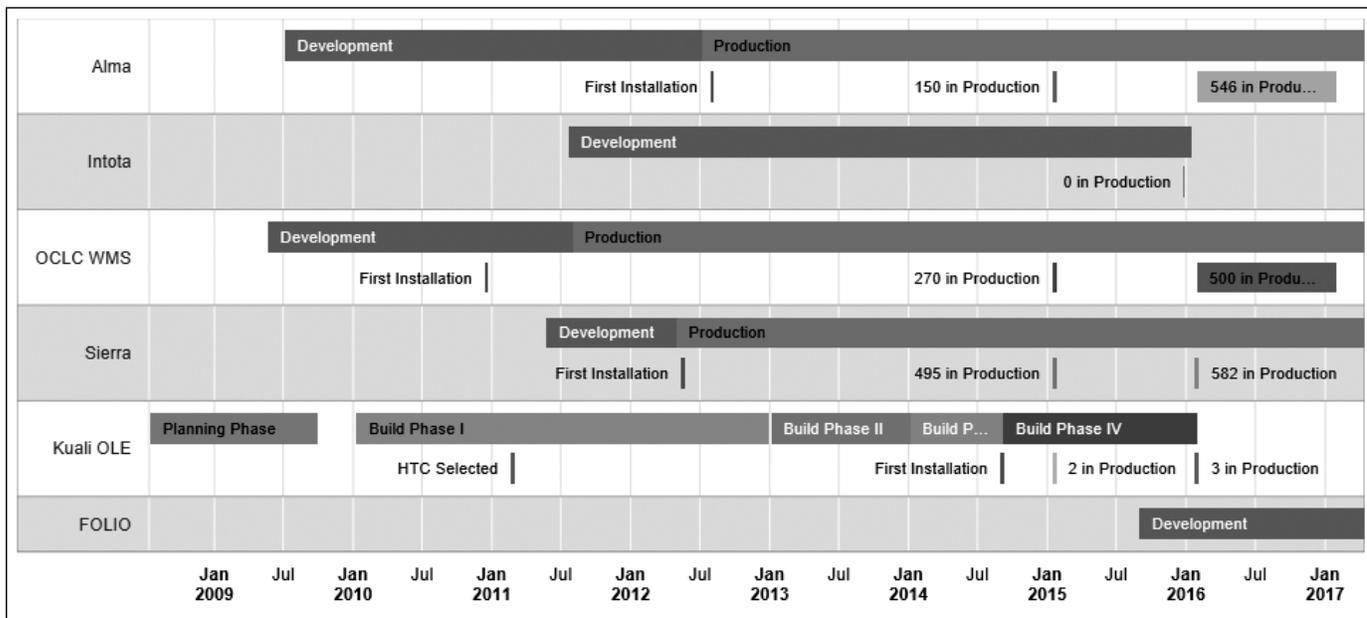
Figure 1.1: Development Timeline for Library Services Platforms

search and retrieval engines. The FOLIO directory of developers currently lists 13 personnel from Index Data, more than any other organization.

- **Qulto**, a company that offers the Monguz ILS product, which has been adopted by libraries in Hungary and other parts of Eastern Europe, currently contributes three personnel to FOLIO development.
- **@Cult Srl**, a technology consulting firm based in Rome, Italy, which offers the OLISuite ILS and is involved with open linked data and semantic web technologies. This firm participates in at least one FOLIO development team.
- **Stacks**, an offshoot of the Hybrid Forge company specializing in custom implementations of Drupal for libraries, has partnered with EBSCO for the development of Stacks as a content management system for libraries (see the November 2016 issue of *Smart Libraries Newsletter*) and is also contributing three personnel to FOLIO.
- **EBSCO Information Services** has four personnel listed on software development teams. In addition to these developers, many other executives, directors, managers, and sales personnel are involved with articulating the vision behind FOLIO and advancing it into the library marketplace.
- **Frontside Software**, a company based in Austin, TX, specializing in user interface technologies, has four personnel participating in the FOLIO development.
- **Samhaeng**, a small software development firm in Copenhagen, contributes to FOLIO development.

In addition to the personnel from commercial software development firms, 13 individuals from libraries involved with the Open Library Environment as well as two from Universidad Nacional Autónoma de México participate in software development teams.

According to FOLIO project documents, as of December 2016, at least 62 persons are actively involved in the development of FOLIO. In addition to this technical work, another 149 persons participate in the Special Interest Groups helping to create requirements, design the functionality of modules, and perform testing and quality assurance.

At this point when the software remains in the development stage, projections regarding when libraries might adopt the software and their implementation dates remain somewhat speculative. A cadre of libraries have dedicated resources to the project and intend to use the software as it becomes viable for their requirements.

Libraries in the United States engaged with FOLIO include Texas A&M Universities, University of Chicago, North Carolina State University, Auburn University, Fenway Libraries Online, Villanova University, Duke University, and Cornell University.

International libraries participating in the project include UNAM, Universidad Nacional Autónoma de México; North Rhine-Westphalian Library Service Center (hbz); Northern Germany Library Service (GBV); Chalmers University; University of Sydney; University of Aberdeen; SOAS, University of London; and China Academic Library and Information System.

Besides the hands-on technical work, EBSCO, the Open Library Foundation, and others have been very active in promoting FOLIO in almost all regions of the globe. In addition to webinars and other virtual events, in-person events have been organized to provide information to those with potential interest in FOLIO and to enable those already involved to share their perspective.

Open Source: A Services-Based Economy

Open source projects allow—and often encourage—organizations to provide fee-based services. Although the software itself must be made available without licensing fees, any organization can charge for hosting instances of the software and providing services for activities such as implementation, conversion, support, and custom development. The Apache 2 license under which FOLIO has been developed offers terms especially conducive for commercial activities.

Deployment Options

As an open source project, the FOLIO software will be available to any organization. Some institutions may opt for a local installation that they support independently. Other organizations may provide commercial hosting and support services. FOLIO has been designed as a multi-tenant platform to enable each instance to support multiple libraries.

This approach contrasts with the existing library services platforms Ex Libris Alma and OCLC WorldShare Management Services, where a single global instance of the platform supports all the libraries using the products on a unified code-base. Neither Alma or WorldShare have yet been offered for local institutional installations.

EBSCO's Commercial FOLIO Services

It's no secret that EBSCO Information Services has been one of the major forces behind FOLIO. The company launched the vision of the project, which it has supported with funding, project management, advocacy, and promotion. EBSCO Information Services has devoted resources to the establishment of the FOLIO project, its technical development, and to promoting the project globally. Since FOLIO is being developed as open source software, EBSCO will not have exclusive access to the software but intends to be one of many organizations supporting instances of the product.

Although the company has no exclusive ownership or control of the FOLIO software, it plans to become one of multiple organizations offering services in support of the software.

EBSCO has implemented an instance of FOLIO using Amazon Web Services cloud infrastructure and will provide commercial hosting and support services. Offering commercial services for FOLIO falls within the company's growing number of software as a service (SaaS) offerings.

In preparation for delivery of FOLIO services, EBSCO has created a new team, initially populated with four new employees. This team will be increased over time as the initial version of the product approaches completion and additional libraries contract for its services.

EBSCO brings considerable assets to the table for its FOLIO services. The company has deep expertise in hosting and infrastructure through its existing global platforms, such as EBSCOhost, EBSCO Discovery Service, Full Text Finder, and a growing array of other SaaS offerings. EBSCO will also be able to provide its FOLIO customers access to is knowledge base and discovery service.

In November 2017, Chalmers University of Technology became the first institution to publicly announce it has engaged with EBSCO as a FOLIO beta partner for its hosting and support services. Based in Gothenburg, Sweden, the library currently uses a Sierra ILS and intends to deploy FOLIO to provide a less cumbersome way to manage its resources and to launch new services. Although other libraries have expressed a commitment to eventually deploy FOLIO, Chalmers University of Technology is the first to partner with EBSCO for its hosting and support services.

EBSCO's Edge: Knowledge Base and Discovery Indexes

FOLIO has been developed primarily for academic libraries, which devote most of their collection budgets to subscriptions to electronic resources. Given the proportions of these investments, FOLIO will need to provide strong functionality for the management of electronic resources and for their discovery and access. The current paradigm of effective electronic resource management depends on a knowledge base of holdings representing the vast number of products and portfolios available to libraries. In the present environment, only three organizations have developed full-fledged e-resource knowledge bases: EBSCO, ProQuest, and OCLC. For both ProQuest and OCLC, the knowledge base is largely tied to their own products. EBSCO has focused more on providing linkages to its knowledge base and discovery services to other resource management products, including FOLIO. EBSCO will be able to distinguish its FOLIO services by integrating knowledge base and discovery services. The Global Open Knowledge Base (GOKb) was initially posited as a likely component for FOLIO, but this project seems to be on

hiatus and has not been populated with data approaching the scale of the three commercial global products.

EBSCO does not make either its e-resource knowledge base or EBSCO Discovery Service available for free but offers them as licensed products. Libraries that have already purchased EBSCO Discovery Service or Full Text Finder can integrate them into their existing ILS or discovery interface. EBSCO has worked with most ILS vendors and open source projects to facilitate the technical integration mechanisms. EBSCO has likewise facilitated integration of its knowledgebase and discovery service into FOLIO. The open and modular architecture of FOLIO would likewise support integration with other knowledge bases and discovery services. The current market dynamics will likely reinforce an affinity between FOLIO and EBSCO Discovery Service despite the technical possibility of other options. Some implementations may involve open source discovery interfaces, such as Blacklight or VuFind, but relying on the EDS index for article-level discovery.

Open Library Foundation

Open Library Foundation was formed as a non-profit organization to facilitate open source projects and to serve as the entity to provide legal governance. Current projects affiliated with the Open Library Foundation include FOLIO and GOKb. The Open Library Foundation was founded by EBSCO Information Services and the Open Library Environment. The Open Library Environment, the group which was originally formed to build the Kuali OLE software with funding from the Andrew W. Mellon Foundation, has shifted its focus to the FOLIO project and is organizationally affiliated with the Open Library Foundation.

The Open Library Foundation issued a statement in November 2017 listing its current board members:

- David Carlson: Dean of Libraries and Provost, Texas A&M University (President)
- Deborah Jakubs: University Librarian & Vice Provost for Library Affairs, Duke University
- Kristin Antelman: University Librarian, California Institute of Technology
- Qiang Zhu: Library Director at Peking University and Deputy Director of CALIS (China)
- Ross MacIntyre: Head of Library Analytics Services at Jisc (U.K.)
- Sam Brooks: Executive Vice President, EBSCO Information Services

See http://www.openlibraryfoundation.org/ for more information on the Open Library Foundation.

Although the software remains in the development phase, the FOLIO project has made substantial progress toward completing an initial version of its software according to its planned development timeline. The project has also jelled organizationally as seen though a network of decentralized though coordinated development teams. The work of the project is documented on its wiki (https://wiki.folio .org/), and the software is shared on github (https://github .com/folio-org).

Smart Libraries Newsletter has followed each of the other library services platforms from the point of the articulation of their product vision through the development of software and eventual adoption. FOLIO has reached the point where the software is beginning to take form but has not yet been completed or implemented in libraries. Upcoming issues of *Smart Libraries Newsletter* will continue to track major milestones or events related to this latest initiative in the library services platform arena.

Harris Computer Systems Acquires ResourceMate

The library automation industry has seen another increment of consolidation, this time in the small library sector.

ResourceMate, an ILS developed by Jaywil Software Development based in Guelph, Ontario, has been acquired by Harris Computer Systems, which is one of the largest software firms in Canada. Harris Computer Systems also does business in the United States and in other international regions. It was founded in 1976 and produces software and technology products for many aspects of public service, local government,

and other business sectors. Jaywil software now operates as part of Harris Computer Systems and will continue to develop ResourceMate.

Harris Computer Systems is owned by Constellation Software Incorporated, a Toronto-based holding company, which also owns and operates several other companies that provide products to a diverse set of business sectors.

ResourceMate has been implemented mostly by small to mid-sized public libraries and school libraries. Harris Computer

Systems states in its announcement that it sees ResourceMate as a complementary offering to its business systems for schools and local governments. Likewise, existing ResourceMate sites may help drive interest in other software products offered by Harris.

For more information see https://www.resourcemate .com/ and http://www.harriscomputer.com/.

Smart Libraries Q&A

Marshall Breeding responds to questions submitted by readers. Email questions to Sam Imburgia at simburgia@ala.org.

Our library is looking to digitize historical photos for patron access as a research tool. There are a lot of companies offering these services. What specs and features should we look for when comparing various database options in terms of user friendly uploading and metadata collection, as well as user friendly on the patron side?

Digitizing historic photos can be an important activity for libraries. Creating digital representations of photographs provides an opportunity to provide broader access and helps preserve photographs for future generations of scholars and researchers. Digitizing historic photographs not only falls within a library's role in helping to preserve cultural heritage, but it also provides an opportunity to deepen its level of engagement with its community.

The standards and technical components needed for a digitizing project will vary according to the scale and scope of the project and the resources available. Such a project also involves several components including the format of the media files created through the digitization of the photographs, the type of digital storage in which the files will be placed, the metadata that describes the content of each photograph, and the applications used to manage and provide access to the digital collection.

The format of the digital image files will include specifications such as the resolution of the digitized files, the tonal resolution, and the type of media file produced. The resolution of the image will vary depending on the size of the original object, the level of quality expected to be achieved, the capabilities of the digitizing equipment, and the quantity of storage available. Only a few years ago, the typical resolution used to digitize image would have been 300 dpi (dots per inch or pixels per inch). The capabilities of digital cameras and the capacity of affordable storage makes it now possible to capture digital images at much higher resolutions. In most cases, a scan of 600 dpi will produce a high-resolution image that captures all of the visual information of the original.

The size and quality of the original material will factor into determining the resolution. 35mm slides, for example, require much higher resolution to gain the same level of quality as photographic prints. Scanning 35mm slides at 2,000–4,000 dpi should yield excellent results.

Large-format items, such as maps or posters, would ideally be scanned at equivalent resolutions as standard-sized prints. Such scans may require significantly more storage and special digitizing equipment. Not all digitizing labs will include large-format scanners or cameras. If you are outsourcing large-format materials, ensure that the vendor has access to the appropriate digitizing equipment and expect to pay significantly more for these materials than for smaller prints or slides.

Digitizing projects will also need to specify the tonal depth for the digital files produced. If the original images are color, the digital files should be produced with at least a 24-bit RGB profile. Black and white images should not be digitized as simple bitonal images, but rather with at least an 8-bit grayscale.

Most projects require two categories of digital images: preservation and access. The preservation images are meant to be the highest quality and resolution. Although in most cases the original prints, slides, or negatives will also be preserved, these digital masters provide an additional layer of protection should the original materials become lost or damaged. Another set of digital files can be derived from the digital masters in a format suitable for presentation through web-based interfaces. The files for access will usually have substantially lower resolution than the masters—typically 72 dpi—and require only a fraction of the digital storage and require less bandwidth for presentation.

The library will also need to specify the technical format of the digital files to be produced. TIFF files have been a mainstay for digital masters and should be provided if possible. If a digital camera is used to capture the images, it is also beneficial to receive the RAW files produced. Both TIFF and RAW files can be quite large, especially for larger items captured at high resolution. These files are well suited for digital preservation but do not work well for presentation via web-based interfaces. More web-friendly formats can be derived from

the digital masters for viewing through access systems, either produced in advance or dynamically within a digital collections management application. JPEG2000 has become widely adopted as a compressed image format since it does not introduce any loss of information. JPEG images are also commonly used but do lose information upon compression.

The technical specifications of digitizing cannot necessarily be determined by a rote formula. They will depend on many factors, such as the conditions of the original materials, the capabilities of the available equipment, as well as the budget available. The curators of a collection will usually work with a technical expert to design specifications that meet the requirements of the project. The final specifications should achieve the highest image quality possible relative to the conditions of the original materials, the budget available, the capacity of available storage, and any requirements of the applications used to manage, preserve, and provide access to the images in the collection.

In order to function as a coherent digital collection, the digitized images need to be associated with metadata that describes what is represented in each image, including names, dates, locations, and other descriptive information. The structure and standards of the metadata represents part of the initial design process of a digital collection. Decisions will need to be made regarding the general metadata standard to be followed, such as Dublin Core, and how each of the available fields will be structured and populated. The ways in which all names and places are recorded will need to be standardized in order for the collection to be easily browsed and searched. In some cases, the metadata design might include use of established vocabularies or ontologies, such as the Art and Architecture Thesaurus (AAT) offered by the Getty Research Institute. Use of these standards will be important especially if the local digital collection will be linked to other regional, state, or national collections. The design of the metadata schema will make a big difference in how well researchers will be able to search or browse the collection and on its interoperability with other information systems.

The creation of metadata for each image will require an investment of time. In most cases, it will take much longer to describe an image than to digitize it. Some metadata elements may be available from documentation associated with the original photographs. Even if it is not entirely complete, this information can provide a good starting point for a more complete description of the photograph based on additional research.

Some libraries and museums have had good results in using crowdsourcing to discover additional information about historic photos. Once the image is digitized and preliminary metadata has been created, the organization can make the images available publicly and invite community members to contribute information on persons, places, and dates that they recognize. Librarians or collection curators can then follow-up to confirm any contributed metadata. Crowdsourcing not only helps the library enrich the metadata for its digital collection, but also represents an opportunity to strengthen its engagement with its community members.

Some type of collections management tool or digital asset management system will need to be implemented to bring together the digitized images and the descriptive metadata to create interfaces for searching, browsing, and viewing the digital images. These products will also provide the library with tools for ingesting the images, importing or entering metadata, and for other aspects of managing the collection.

Many different collection management tools are available for libraries and other types of institutions involved with managing digital images. In the library arena, commercial products such as OCLC's CONTENTdm or Ex Libris Rosetta have been widely implemented. Those interested in open source tools can consider products such as Fedora, Samvera, or Islandora. Several companies and non-profit organizations offer implementation and support services for open source products to libraries that may not have extensive expertise in-house.

All these factors lead us back to the question regarding the features to specify for metadata creation and user-friendly patron interfaces. As we have noted, digitizing and providing access to a collection of photographs involves many layers of standards, workflow, and technical infrastructure. The specific features and capabilities of the technology products will also depend on the complexity and scale of the project. Collections including hundreds of thousands or millions of images will naturally require more industrial-strength technical infrastructure than smaller collections. As with other categories of technology products, we can't expect a one-size-fits-all solution. Finding the best environment for a project such as digitizing historic photographs will involve a thorough review of the expectations and requirements of the proposed collection and should take into consideration a broader digitization program that the library might want to incorporate into its ongoing operations.

January 2018
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**.

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