1-1-2014

The Future of Cloud-Based Library Systems

Steven K. Bowers
DALNET, sbowers@wayne.edu

Elliot J. Polak
Wayne State University, elliot.polak@wayne.edu

Recommended Citation

Bowers, S. K. and Polak, E. J. (2014). 'The Future of Cloud-Based Library Systems.' In K. J. Varnum (Ed.), The Top Technologies Every Librarian Needs to Know: A LITA Guide (pp. 43-55). Chicago, IL: American Library Association.
Available at: http://digitalcommons.wayne.edu/libsp/78

The Future of Cloud-Based Library Systems

Steven K. Bowers and Elliot J. Polak

The current and future trends of library services and the software for delivering those services are intertwined. It may be debatable which comes first, but the two are inseparable and necessarily affect each other. It is definite, however, that the limits of the now-aging integrated library system (ILS) software model are constraining the presence of the library within the new reality of the Internet and interconnected World Wide Web.

Utilities for library system cooperation, such as OCLC, have served as a mechanism for facilitating resource sharing, built around a core of shared record keeping for tracking holdings within member libraries. Whether they are called cloud-based ILS or library services platforms, developing and future systems will move beyond shared library resources on shelves to establishing a shared technological infrastructure for supporting all that libraries do.

Library platform systems are under development or in early adoption, and while it is not necessary to name specific companies and their product offerings, it is certainly worth examining common platform features and designs that are planned or have already been implemented. Cloud-based systems have a hosted technological infrastructure, which also allows for staff and public access via a web browser. A platform solution will provide means for management of all library systems, including circulation, cataloging, acquisitions, serials, electronic resources, authentication, the public interface, and analytics for data in the system.

New platform systems currently in use, as well as those under development, are intended to be built on a somewhat common suite of services. As with older

ILS, these platforms already manage bibliographic and holdings data for library materials, supporting cataloging and circulation functions. In various stages of development are associated acquisitions modules for all library materials, print and electronic. As new and developing platforms are designed to manage all library materials, they also include electronic resource management (ERM) services. With all library services as part of one system, platforms will be able to provide analytics for data in the system, including usage statistics for print and electronic materials, as well as reporting features for each of the services in the platform. Several new systems already include analytical dashboards for snapshot views of system statistics.

Discovery catalogs will also be an integral service of future library platforms. Already adopted by many libraries, these new public interfaces often function with external link resolvers and A-to-Z serial title lists for taking users from the catalog to full-text resources. Future systems will have all these components as part of the unified platform. In the platform, these public interface components will integrate with all other services, including authentication services governing access rights.

A single platform for all systems will allow for redesigned and combined workflows. Platform systems will be open so that one platform can use services from another if needed or desired, making use of web services and application programming interfaces (APIs). Systems will be built on service-oriented architecture (SOA), with individual yet connected functions, so that systems are scalable, or can be used by large and small operations to perform complex or simple services. Shared infrastructures will also incorporate shared and linked data sets and tools for analyzing system information and statistics. All this will allow for greater cooperation among libraries and a strengthening of patron-driven print and digital services, thus getting the information to the user and fulfilling her information needs.

CLOUD COMPUTING

In 1962 Ross Perot left IBM and started Electronic Data Systems (EDS), based around the idea that organizations with large mainframe computing hardware needed these systems only during business hours, and could lease afterhours processing time to other organizations. Though not widely credited as a founder of cloud computing, Perot's success demonstrates the idea that the computing needs of organizations are often lower than the capacity of the hardware they

purchase. It is this inefficiency that has led to the recent explosion of cloud-based systems.

Since the mid-2000s the term *cloud* has been used to describe a wide range of services, from hosting providers who allow organizations to offload information technology (IT) infrastructure, to storage providers encouraging consumers to preserve data virtually. *The cloud* has come to mean a shared hardware environment with an optional software component.

While these services may describe cloud computing, they do not accurately describe the current state of cloud-based integrated library systems. Given the noncompetitive nature of most library environments, libraries have the unique opportunity to benefit from not only a shared hardware/software environment but also a shared data environment. A cloud-based ILS may best be understood as a shared system where all users benefit from the data contributions of participants.

These contributions can manifest themselves in a number of ways, from shared Machine-Readable Cataloging (MARC) records, to a shared online public access catalog (OPAC), where searches from all participating libraries enhance the relevancy ranking of results. It is this distinction that separates a cloud-based ILS from simply being a cloud-hosted ILS.

As a precursor to cloud computing, for several years there have been services providing access to software installations. At one time these services were referred to as the application service provider (ASP) model, now more commonly known as Software as a Service (SaaS). ASP/SaaS models have provided opportunities for organizations—including libraries—to access computing services and technology perhaps beyond the means of systems run in-house. Some organizations, although quite capable of running such infrastructures on premises, have been able to realize the cost savings of ASP/SaaS services: reduced infrastructures for technology, staffing, or both. Many organizations have become accustomed to relying on these provided services and have been able to shift resources to other areas of concentration within their work environment.

The use of cloud computing builds on the ASP/SaaS model and expands upon the opportunities for realized savings on reduced infrastructures. Cloud computing offers not only a hosted software and hardware infrastructure, but it provides means for accessing the software or related services online without any local installation of the software in question. The general Internet user is now accustomed to using services and software within the cloud by means of access via a standard web browser, although it may have taken some longer to develop a formal understanding of the term *cloud computing*.

The discontinuance of the need for local software installation—and for separate software instances—is a substantial advantage of true cloud computing over the ASP/SaaS model. The ASP/SaaS model generally requires installation of software locally for access to a system. The infrastructure for the ASP/SaaS model is often still the same as a locally hosted system, but it is handled remotely by the service provider. Cloud-based computing is such that a single instance of a software system can be installed and run for access by many shared users, giving rise to a multitenant infrastructure, unlike the separate systems run in ASP/SaaS models. A shared-system environment running a single instance of a particular application is the basis for the creation of a platform. A platform is the foundation for a software environment with multiple related services and computing capabilities accessible within the cloud.

The multitenant shared infrastructure of a platform allows for quantifiable advantages. The service provider can spend time on development and quick implementation of new features and enhancements, as there is only one instance of the software to maintain. The reduced overhead for separate systems also allows the service provider to realize significant cost savings from not maintaining multiple separate infrastructures. This cost reduction must ultimately be at least partially realized by the libraries or other organizations using the provided software services.

It can be argued that open-source software has encouraged proprietary software vendors to create more open systems. This argument does not need to be made or supported. It is enough to note that now, and in the future, systems must be more open as the functionality of adopted and evolving web standards will require this. Open systems are relevant to cloud computing and platforms because not only are they contemporary developments, but they are essentially all made to work together. Web services provide means for extracting and using data from open systems, and the new platforms provided by cloud services are certainly systems from which data is sought.

Big data, linked data, and the analytics of this content are some of the advantages forthcoming from cloud computing. Libraries are finally realizing access to their own information that has for years been locked away in locally hosted, or at least discretely hosted, systems. Not only will libraries have better access to their data and statistics, the platforms of the cloud-based environment promise to offer better means for analyzing and using that information for creation of new business intelligence and enhanced service delivery.

SHARED TECHNOLOGICAL INFRASTRUCTURE

As can be expected, the model for future library systems has been developed in a broader realm than libraries, as this model is now prevalent among general IT operations. If cloud-based platforms are adopted in libraries, then multitenant architecture software—and the expectations for the provision of platform services in most current computing environments in general—will allow libraries to spend more time delivering services and less time running the systems they employ.

A platform will offer these systems as a package of cloud-based solutions. A shared cloud infrastructure has certain quantifiable and unquantifiable advantages over locally hosted and even individually external hosted systems. The cost savings of outsourcing software and hardware maintenance are measurable. It is in some ways immeasurable what libraries will gain by sharing future platform infrastructures, enabling the analysis of large-scale data sets (big data), along with individual library data that has previously been locked away in local systems, or not maintained by systems at all.

New library platforms will organize and maintain metadata as library systems have always done. By using shared-platform infrastructures, libraries will be able to take advantage of shared metadata within these systems. Bibliographic information will be open for shared use and enhancement, which will reduce duplicative work by eliminating the need to create local records for popular or widely held titles. This shared production environment will be enhanced by partnerships between libraries and other entities. Publishers and vendors are eager to share in the creation and enhancement of data sets, and libraries must become comfortable with allowing partnerships from varied organizations, including those from for-profit companies.

OPEN SYSTEMS AND LINKED DATA

Software systems within libraries and in other computing platforms are shifting from single-function, or single-system, designs, to service-oriented architecture (SOA). Computing services are individually running software functions that are linked to form a complex integrated system that can be seen as a whole consisting of interconnected separate operations. Developing and future library systems will use this newer architecture to connect all of the functions of library software to build the library platform.

A platform built on SOA is a scalable system— meaning that the functions of the system can be used by large-scale operations, or they can be scaled down to smaller implementations that maintain functionality and are meaningful in either environment. A large computing platform is designed to be scaled to need, and can grow to provide continuity in supporting operations. A powerful platform can also be used by a library with small or minimal system needs, as the infrastructure of the system is in the platform and need not be maintained by the individual libraries using the shared infrastructure.

Some libraries will opt to use a cloud-based system built with SOA but will not participate in a shared-platform environment. These systems, although in the cloud, will not be cloud-based in the sense of having a shared implementation; they will be open systems allowing for information exchange from one separate implementation to another. These connected systems will have the potential to connect services and make use of linked data.

The advent of linked data, as with linked systems, will allow for libraries to capitalize on the shared infrastructure of the actual data that is used to access information. In an environment where libraries are sharing bibliographic data, efforts can be made elsewhere, as duplication of work is minimized. Expanded access to bibliographic data as part of the Semantic Web will in fact allow libraries to have access to a much broader set of data, including data from any other entity that has linked data on the Internet. Of equal or greater importance, non-library entities will also be able to use library metadata if that metadata is part of the Semantic Web. This is essential as libraries work to become an included part of information retrieval on the Internet rather than isolated collections of locally accessible materials.

All this new interconnectedness for systems will enhance discovery of information materials, and will mean greater access to information for the user. Discovery systems, or next-generation OPAC replacements, are already using library metadata in new ways that allows the user to "discover" data through single-search interfaces and point-and-click access for refining searches. These search-refining parameters, or facets, are built from organized metadata that has been stored in the more closed integrated library systems of the past. Facets can provide information from multiple sources, including bibliographic records, item or holding records, and even vended data sources. In the future, these systems will become even more powerful as they incorporate linked data from other libraries and non-library resources alike.

NEW BIBLIOGRAPHIC FRAMEWORK

To sustain broader partnerships—and to be seen in the non-library specific realm of the Internet—metadata in future library systems will undoubtedly take on new and varied forms. It is essential that future library metadata be understood and open to general formats and technology standards that are used universally. Libraries should still define what data is gathered and what is essential for resource use, keeping in mind the specific needs of information access and discovery. However, the means of storage and structure for this metadata must not be proprietary to library systems. Use of the MARC standard format has locked down library bibliographic information. The format was useful in stand-alone systems for retrieval of holdings in separate libraries, but future library systems will employ non-library-specific formats enabling the discovery of library information by any other system desiring to access the information. We can expect library systems to ingest non-MARC formats such as Dublin Core; likewise, we can expect library discovery interfaces to expose metadata in formats such as Microdata and other Semantic Web formats that can be indexed by search engines.

Adoption of open cloud-based systems will allow library data and metadata to be accessible to non-library entities without special arrangements. Libraries spent decades creating and storing information that was only accessible, for the most part, to others within the same profession. Libraries have begun to make partnerships with other non-library entities to share metadata in formats that can be useful to those entities. OCLC has worked on partnerships with Google for programs such as Google Books, where provided library metadata can direct users back to libraries. ONIX for Books, the international standard for electronic distribution of publisher bibliographic data, has opened the exchange of metadata between publishers and libraries for the enhancements of records on both sides of the partnership. To have a presence in the web of information available on the Internet is the only means by which any data organization will survive in the future. Information access is increasingly done online, whether via computer, tablet, or mobile device. If library metadata does not exist where users are—on the Internet—then libraries do not exist to those users. Exchanging metadata with non-library entities on the Internet will allow libraries to be seen and used. In addition to adopting open systems, libraries will be able to collectively work on implementation of a planned new bibliographic framework when using library platforms. This new framework will be based on standards relevant to the web of linked data rather than standards proprietary to libraries.

The Library of Congress, with other partners, continues to work on a new bibliographic framework (BIBFRAME). This framework will be an open-storage format based on newer technology, such as XML. A framework is merely a holder of content, and a more open framework will allow for easier access to stored metadata. While resource description and access (RDA) is a movement to rewrite cataloging rules, BIBFRAME is a movement to develop a new storage medium. The new storage framework may still use RDA as a means of describing content metadata, but it will move storage away from MARC to a new format based on standardized non-library technology.

This new framework will encompass several important characteristics. It will transition storage of library metadata to an open format that is accessible for use by external systems, using standard technology employed outside of libraries. This will allow for libraries to share metadata with each other and with the rest of the Semantic Web. The new framework will also allow for the storage of both old and new metadata formats so that libraries may move forward without reworking existing records. Finally, the new framework will make use of formal metadata structure, as the benefit of named metadata fields has more power for search and discovery than the simple keyword searching employed by much of the Internet. Library metadata will become more important once its organized fields of information can be accessed by any standard non-library system.

Embracing a new storage format for bibliographic metadata is much like adoption of a new computer storage format, such as moving your data storage from CD-ROM to an external USB hard drive; the metadata that libraries have created for decades will not be lost but will be converted to a new, more accessible, storage format, sustaining access to the information. Although these benefits may be seen by some, it can be expected that there may be resistance to changes in format as well. It will be no small undertaking to define how libraries will move forward and to then provide means for libraries to transition to new formats. Whatever transitions may be adopted, it will be important that libraries not abandon a structured metadata entry form in lieu of complete keyword formatting.

Much of the Internet has been built on keyword searching for information retrieval, but there is continued value in defining what information is stored in specified metadata areas so that keyword searches can be better targeted. Areas or fields of metadata can be used for providing relevancy ranking for search results or specified searches (e.g., Author or Title), using a ranking system similar to other information sources such as websites. Using formats that are understood by those outside of libraries will make libraries more visible—sustaining libraries'

relevance and increasing, enhancing, and enabling future development of new library services for the user.

ELECTRONIC RESOURCE MANAGEMENT AND EXPANDED ANALYTICS

Many libraries have already begun to acquire electronic materials into their collections. It is no longer relevant to the patron whether these resources are owned, leased, borrowed, or simply open access. Libraries focused on patron fulfillment—that is, meeting the information needs of the user—will continue to emphasize access to information in all its formats. Systems currently under development, and some already in use, are early attempts to incorporate management of all these types of resources. Systems will move from print-based management to new systems built with electronic resource management fully integrated with all functions of library operations.

Current electronic resource management (ERM) systems are often separate components of an integrated library system, and often have little interaction with other modules in an ILS. Although existing systems are not always part of an ILS, they have been built on standards including SUSHI (Standardized Usage Statistics Harvesting Initiative) and COUNTER (Counting Online Usage of Networked Electronic Resources). As future systems will fully incorporate electronic resources and their management, the usage statistics for those materials will be part of the growing data of information available in library platform systems.

Platforms will allow libraries to have a shared knowledge base of information. This shared knowledge base will allow for a shift in workflows. Electronic resources will be acquired and "turned on" from within services in the library software platform once a purchase decision is made. These materials will be instantly available in the system, which will already have defined access rights, holdings data, and other pertinent information maintained in the shared platform.

The knowledge base of a shared system will also allow for enhanced demand-driven acquisitions (DDA). Libraries will be able to offer access to materials by allowing the patron to make the purchase decision. Currently some systems allow for parameters or thresholds to be set so that a purchase is made for the user, or access to information is granted to the user on an at-need basis, instantly when the patron is searching. Systems can allow for brief access without a fee, or a purchase can be triggered after prolonged access or multiple views. Developing and future

systems will allow for access to an unlimited collection of information that will grow as the user selects new materials for access. Access will then be limited only by the data contained in the knowledge base.

Along with DDA, cooperative collection development will be greater enabled through shared knowledge of what libraries already have access to. Libraries will be able to identify unique materials and avoid unnecessary collection overlap. Holdings information will be accessible in shared library platforms, as will be the usage information for those holdings, in both print and electronic formats.

Web services and APIs provide integrated library system users a means to access data over the Web using web-programming languages. This ability to extract and use library data in external systems has been described as open architecture, which is similar to open source. While *open source* indicates open access to source code, *open architecture* indicates open access to a system's data or functionality. Many libraries are already using APIs from discovery layer systems from various vendors to integrate article metadata into custom-created web applications. APIs for web-based systems will allow libraries to access library data for discovery, and will also enable functionality such as circulation, cataloging, electronic resource management, and statistical reporting.

Libraries will be empowered by library computing platforms with increased access to analytical data for holdings, usage statistics, and potentially user and other data in library platform systems. Each individual library not only will have greater access to its own information but will potentially be able to benchmark or compare data across multiple institutions. Platform systems, along with APIs and web services, will allow libraries to access their own information that was previously not accessible in closed integrated library systems. Most early platforms have limited reporting features, but future development is likely in this area as platform systems see greater use and new adopters define new reporting needs. Delivered analytical dashboards are already being designed to allow for custom display; these will include the option of custom reporting at a much greater level as systems continue to develop. Future reporting will include analysis of purchases, collections, and all other platform functions. Systems will be designed to meet the changing needs for analytical reporting, to keep pace with delivery of changing library services.

FOCUS ON FULFILLMENT AND ACCESS

For years, integrated library system models have served as disconnected, separate repositories of bibliographic metadata, in which both records and the work to create them have been often duplicated from library to library. The software and hardware for this model were separately maintained and financially sustained by each library using these systems. Consortia have arisen for sharing both physical and financial resources, but libraries are still disconnected in their daily operations, even within shared systems. New models will run on outsourced infrastructure, allowing efforts and finances to shift toward resource sharing and delivery of library user services.

If the original purpose of library science was to provide curated and organized access to information, the original purpose of integrated library systems grew out of the need to track a controlled inventory. The foundation of library services now—and in the future—is still to provide this access to information. System needs and designs, however, have shifted from inventory control systems to systems designed with the end purpose in mind: delivery of accessible information and fulfillment of the user's information access needs. This natural progression is imminent as libraries shift from curating owned, physical materials to providing access to print, electronic, and other format of materials from multiple collections that may or may not be owned or licensed by the library. Libraries are now being called to provide access to any information resource needed by the end user.

Integrated library systems developed around the premise of fulfillment will facilitate both traditional circulation and OpenURL access to electronic material. Cloud-based ILS will include metadata for journals, articles, and aggregators, along with a built-in link resolver providing a means to connect users to these resources. Libraries can then manage and assess electronic and physical usage within a single system.

Platforms will employ a multitenant architecture, allowing multiple libraries to use shared systems and services. This will, in turn, allow for the expanded provision of access to owned, leased, and borrowed information. Most patrons are not concerned with where the library gets materials, as long as the library is able to deliver access in a timely manner. The connected systems will allow for this expedited access to information, which the patron may already expect to be instantaneous. Security management for access will be part of some cloud-based ILS or linked from a combination of systems including the library platform.

WHAT LIES AHEAD?

In the next few years, there will be a proliferation of open architecture systems. By exposing library data through web protocols, types of access to library data will be endless; this increased portability will further encourage libraries to develop in-house discovery interfaces. These interfaces could then blend traditional MARC-born data with non-MARC metadata, merging traditional library resources with digital collections, institutional repositories, and research data sets. Juxtaposed, some cloud-based ILS may also focus heavily on discovery and will be built to ingest MARC and non-MARC metadata, allowing libraries to manage multiple types of data in a single interface.

The most promising aspect of this newfound portability will be the ability to crosswalk library data into linked data formats that search engines can then ingest and index. Whether developed in-house or by discovery-layer providers, new search engine optimized (SEO) ready search tools will bring users to library resources at the search engine level. SEO is the preparation of a website for search engines to index which then allows the content of a website to be findable within a search engine. Given the geolocation aspect of search engine relevancy—and regional nature of the library environment—matching a user to a resource via search engines will play a large role in future library systems.

Libraries embracing these technologies will look to create a new bibliographic framework focused on compatibility with linked data standards. These standards will enhance both search engine optimization and semantic searching. This new bibliographic format will open the library to new markets and has the potential to bring more users to library resources than ever before. Increased traffic from search engines may even curb downturns in circulation and increase lending between libraries. Increased lending and borrowing between libraries will have a cost, but this cost may actually be less than the cost of purchasing, processing, and storing new materials. Libraries will need to monitor changes in material usage, and to perform a cost/benefit analysis when entering these new environments.

This increased discovery of library resources will lead to an inevitable refocusing of library reference and instruction from navigating databases to evaluating resources. Libraries that continue to see declining reference statistics will likely realign staff into emerging fields such as digital publishing, digital preservation, data visualization, assessment, and scholarly communication. Ultimately, library patrons will benefit from enhanced discovery resulting from enhanced metadata.

The efficiencies brought about by cloud-based ILS will undoubtedly have their greatest impact on acquisitions and cataloging departments. Since these systems typically look to eliminate copy cataloging, libraries with special collections may look to reorganize technical services staff around original cataloging or digitizing collections. Libraries may also look to move catalogers into acquisitions or electronic resource management as demand dictates.

Library information technology operations adopting cloud-based ILS will also see changes in staffing needs. Since cloud-based ILS providers will handle hardware and software upgrades, library staff previously responsible for these tasks will have more time to participate in research and development of projects. These projects may involve integrating cloud-based ILS web services into current library web interfaces and assisting with assessment projects such as business intelligence applications.

Of all the changes resulting from migration to cloud-based ILS, the most significant will be libraries' expanded ability to develop new and innovative services. Libraries will have the opportunity to help solve more complex problems than ever before—and thus, patrons will be served with more library resources than ever before.