Smart Libraries Newsletter

News and Analysis in Library Technology Developments

ALA TechSource
alatechsource.org

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

Smarter Libraries through Technology

Evolution or Revolution?

By Marshall Breeding

The current slate of technology products and services has come about through the developments of a diverse set of companies, non-profit organizations, and open source initiatives. The library technology industry has seen an overarching trend of incremental, evolutionary development. There have also been some episodes where new products have been introduced that depart substantially from the prevailing path through a more revolutionary vein. Sometimes the revolution has been a minor deviation; in other cases, it has led the way to establishing new trends, which then become the basis for a new wave of evolutionary development.

The general model of the integrated library system (ILS) was cast in the 1970s. Prior to that time, there were systems to automate specific areas of library operations, such as circulation, cataloging, and acquisitions. Over time these systems coalesced into a unified set of modules that shared a set of common databases. The advent of the ILS eliminated many of the duplicate workflows and technical overhead involved when operating multiple separate programs. For most libraries, the initial foray into automation came once the ILS had become the established model. ILSs enabled libraries to save enormous time and gain efficiencies compared to the manual procedures previously in place. The initial phase of ILSs came at a time when library collections were comprised of print materials and when computing mostly took place on mainframes or midrange computers.

A very long phase of evolutionary development followed the initial establishment of the ILS. Although dozens of ILS products were developed and implemented, there was far more resemblance among them than differences. Although not well differentiated, there was a vigorous competition among the players based on the maturity of features, quality of support services, and reputation. This lack of substantive differences among a crowded field of players led to multiple rounds of mergers and acquisitions, gradually consolidating a previously highly fragmented industry.

Some of the survivors of this evolutionary development pattern can be seen in products still in use today:

- The current SirsiDynix Symphony ILS traces its origin to the Sirsi Unicorn introduced in 1982.
- Innovative Interfaces ILS products have seen a series of incremental transitions: INNOVACQ (1982), INNOVAC (1989), Millennium (1997), and Sierra (2011). Each of these products inherited major components from its predecessor, though reengineered to current technology architecture and new feature enhancements.
- Polaris, which was introduced in 1997, has seen continuous incremental enhancement throughout its product history.
- Ex Libris Aleph was launched in 1980 and evolved through multiple versions through Aleph 500, which remains in use today.
- Library.Solution from The Library Corporation has seen continual enhancement, including major transitions to web-based interfaces. CARL.X has evolved from one of the pioneering ILS products developed in the early 1980s and has been thoroughly but gradually reengineered from its original proprietary Tandem-based hardware and software platform to a more open and standard environment.

IN THIS ISSUE

ISSN 1541-8820 May 2019 Volume XXXIX Number 5

- OCLC has recently launched a new product branded as Wise for public libraries in the United States positioned as a patron engagement system. This product has a long evolutionary history, descending from the bicat ILS developed by HKA, which OCLC acquired in 2013.

These are just some examples that illustrate the overarching pattern of incremental development, often spanning multiple decades. This evolution falls into the brownfield development, where new products and versions are based on previously established products or software.

The library has also seen some examples of more revolutionary product development, following the greenfield development pattern of creating a new product independent of any previous product. This revolutionary path usually means substantially longer development time to create a new product since it means essentially starting over rather than extending an existing codebase. This path also involves substantially higher risk since the initial versions may not be as rich in features and may enter the market after critical points of opportunity.

There are some greenfield revolutionary product initiatives in the library technology industry that didn't achieve success. Examples include:

- Taos, a new automation system developed by Data Research Associates based on an object-oriented database (ObjectStore from Object Design) in addition to other modernized technical components. Its initial development phase began in about 1995. Despite being selected by a number of major libraries, its development languished, and it saw a relatively low level of sales and implementations. Its final demise came with the merger of DRA into Sirsi Corporation and subsequent acquisition by Vista Equity Partners. Despite some remaining promise of viability, the product was terminated.
- Virtua was likewise positioned as a new system based on forward-looking technologies when it was launched by VTLS in 1995. Its development took longer than expected, missing some important market opportunities. The product did make it into production use, seeing mixed results in the market. There were some high-profile failures, such as aborted implementations in Oxford University and New York University as well as major successes in Queens Borough Public Library and Hong Kong Public Library. Absent major market momentum, the development of Virtua ceased following the acquisition of VTLS by Innovative Interfaces. Many implementations of Virtua remain and support continues, though Innovative no longer offers the product for sale.
- The open source Kuali OLE project, despite initial enthusiasm, ultimately did not result in a completed product. Launched at the onset of the new interest in developing library services platform offering comprehensive resource management, the development of Kuali OLE software took far longer than anticipated. During this interval, the underlying service bus framework became obsolete, and the commercialization of the broader Kuali initiative left the project without a viable route forward. The organization behind the project have since shifted their efforts to the FOLIO project.
- ProQuest, through its Serials Solutions subsidiary, also entered the library services platform movement. It launched a development initiative to create a new product branded as Intota in 2011. This product also fell victim to long development cycles and business transitions. Following ProQuest's acquisition of Ex Libris in 2015, its attention was focused on Alma, which was already well established.

Moving beyond the risks that didn't pay off, there have been a handful of revolutionary, greenfield initiatives that have seen considerable success and have established new directions in the industry. These include:

- Summon, the initial entry into the web-scale index-based discovery, was launched by the Serials Solutions division of ProQuest in 2009. This product was based on a massive index of article-level scholarly content in addition to the library's local resources. Summon provided a significantly better model for academic library search and discovery compared to the federated search technologies prevalent at the time. Other products soon joined the competition for index-based discovery, including Primo with Primo Central, WorldCat Local, and EBSCO Discovery Service. These four products continue to prevail in this product sector.
- Ex Libris launched a new initiative to develop its unified resource management platform, later branded as Alma in 2009. This product proved to be transformative for Ex Libris and for the library technology industry at large. Instead of reworking either of its successful, though aging, ILS products, Ex Libris opted to create an entirely new product, not only based on current technology components and architectures but also on new assumptions and workflows for managing collections in academic libraries dominated by electronic resources. Alma has gone on to see almost unprecedented success.
- OCLC WorldShare Management Services was also launched at about the same time as Alma in 2009. This

initiative built on the WorldCat bibliographic database through an entirely new platform based on current technology and architecture. OCLC has not seen the same level of success with WorldShare Management Services compared to Alma but should be recognized as a visionary product that helped drive the movement of academic libraries to library services platforms.

Summon, Alma, and WorldShare Management Services illustrate the possibilities for revolutionary products gaining success. But these are the ones that survived and prospered, and just as many projects ultimately did not see completion or success in the market.

In more recent times, there are some additional initiatives that fall into the higher-risk category of greenfield development. These would include the Axiell Quria, a new library services platform for public libraries, now in use in a handful of libraries in Europe. Systematic Cicero currently supports all the public and school libraries in Denmark in a remarkably large-scale implementation following an ambitious development initiative from the Danish company Systematic.

In this issue of *Smart Libraries Newsletter*, we take an early look at Innovative's new Inspire Discovery product. Innovative Interfaces has previously been the champion of evolutionary development, following a consistent pattern of building new products on the shoulders of its prior offerings. In a sharp departure from this tradition, Innovative has developed an entirely new platform, branded as Inspire, with Inspire Discovery as its initial product. Following this introduction to the product and Innovative's new development strategy, look to future issues of *Smart Libraries Newsletter* as we track its ongoing development and adoption cycles.

Innovative Launches Inspire Discovery

Innovative has launched a new patron-facing search interface, branded as Inspire Discovery, positioned as its next generation discovery technology. The company officially launched the product at the ACRL conference in Cleveland, OH on April 10, 2019. Inspire Discovery is the initial product built on a new technology platform from Innovative that has been in the design and development stage since early 2017.

Inspire: Innovative's New Multi-tenant Platform

Innovative is developing a new technical platform, branded as Inspire, which will provide the foundation for its new generation products. The Inspire platform is a new multi-tenant environment based on the microservices architecture. Some of the core components of the platform include an Application programming interface (API) gateway, an API registration and discovery service, and a messaging bus. The platform is deployed through Amazon Web Services.

Inspire provides support for multiple data storage and indexing services for applications operating on the platform. It includes PostgreSQL for transactional data, Elasticsearch for keyword indexing and relevancy, and the Dgraph graph database for semantic relationships.

The platform will interoperate with a variety of external systems and services. A set of connectors, for example, will enable data exchange and functional operations with Innovative's ILS products, Sierra and Polaris. APIs will be developed as needed to interoperate with any relevant external systems and services.

The Inspire Discovery service is the first product Innovative has deployed via this new platform. The Inspire platform will support additional modules and services as Innovative builds out its suite of next-generation products.

Inspire Discovery

Innovative designed its new Inspire Discovery service to provide a unique set of capabilities distinct from other products on the market. Its interface includes the usual search box that returns lists of results ranked according to relevancy. But it also goes beyond record lists, providing a context-based visual experience for exploring concepts and relationships in resources explore pathways for discovery that might not be apparent through established conventions such as facets.

Scope of Discovery

The body of linked data addressed by Inspire Discovery currently includes the resources of the library represented in BIBFRAME format. Innovative harvests the MARC records from the library's ILS and transforms them into BIBFRAME, which emerged out of the Library of Congress's bibliographic transformation initiative. The library can also opt to ingest other types of local resources, such as their institutional repository or digital collections.

Innovative has also built a central repository of article-level data so that electronic resources will be discoverable along with the library's collection materials. The primary source of the article-level content is the Crossref database, which currently contains over 104 million records, including 77 million journal articles and 15 million book chapters. The Crossref data is transformed into linked data as it is ingested and loaded into the its internal graph database and indexed by Elasticsearch. The discovery service does not index the full text of article. Rather, as explained by Kathryn Harnish, Senior Vice President of Product Strategy, "Inspire Discovery enables precision searching without the overwhelming number of results that are returned by full-text searches, results that the user must figure out how to contextualize and process."[1] Innovative is working with a third-party partner that acquires the full text of the articles and uses a set of semantic enrichment algorithms to generate concept terms associated with the content of each page of the articles. About six to eight concepts are generated for each page of full-text content. These concepts are incorporated into the keyword indexes and graph database and linked with other resources, concepts, and persons in the Context Engine.

The Crossref database currently contains over 104 million records, including 77 million journal articles and 15 million book chapters.

Driven by Linked Data

The Inspire Discovery service relies on bibliographic data based on BIBFRAME data model with an emphasis on leveraging relationships and linkages among resources, concepts, and people. The resources represented primarily include the library's collection items, including print and electronic materials. Concepts are based on subject and other elements in the BIBFRAME entities. Linkages relating to people include the individuals and organizations involved in creating content.

This linked data model enables the Inspire Discovery to create result sets based on multiple elements of context. In addition to initial search results produced through a comprehensive Elasticsearch index, Inspire Discovery also taps into a graph database that enables users to explore connections among resources expressed in the underlying linked data entities.

A New Discovery Interface

The Inspire Discovery interface has many familiar components. Its initial page offers a search box, which includes a drop-down option enabling the user to change from the default *Everything* to other search options. Once the user enters a search, the interface returns a list of resources, following many of the expected conventions, including a set of facets on the left and the main pane displaying records for each of the search results. Each item view includes cover art when available or a placeholder as well as indicators for resource type and any available descriptive metadata. A *Get It* button is available on each record, which facilitates requests for physical items, e-books, and articles through the library's link resolver, presenting options on how to view or request the item. Inspire Discovery can work with Innovative's own WebBridge, or it can be configured to use any of the other link resolvers.

The interface for Inspire Discovery is based on the Angular web framework, originally developed at Google. The use of this framework ensures that the interface is fully responsive and works well across all types of devices and will adhere to accessibility standards.

This aspect of the search and retrieval of the product is based on Elasticsearch technology. Results to queries will encompass the library's own resources and the indexed article-level electronic resources.

Context Wheel: Graphical Interaction with Results

Inspire Discovery aims to offer a distinctive experience for researchers in the way that it provides access to library collections. Each resource record returned will also present an *Explore More* button, which then launches a visual tool called the *Context Wheel*. This exploration tool provides a visual view of the relationships of resources, concepts, and people represented within the resources and the search context, but it also enables users to explore those relationships and generate new result sets. The user, for example, can select any of the presented concepts and generate a list of the corresponding resource records. Likewise, multiple concepts can be combined for listing results.

The interface includes a Workbench that enable a user to save or *pin* records or searches to use later.

Target Audience

The initial version of Inspire Discovery has been designed for academic libraries. Harnish indicates that work is underway to create features that will be of interest to public libraries.

Expected Availability

Innovative reports that it is working with 70 libraries as strategic partners for Inspire Discovery. Development partners include Hillsdale College in Michigan, Cairn University in Pennsylvania, and Oral Roberts University in Oklahoma.

Innovative states that Inspire Discovery is available for academic libraries effective of its launch in April 2019 and that the product is now available for production implementations.

Next Generation Product Strategy

Inspire Discovery can be seen as the first phase of Innovative's strategy to develop a suite of new-generation products to complement its current slate of products.

Innovative initially launched its Encore discovery interface in May 2006. Although the product has seen ongoing development and enhancement throughout its history, it continues to rely on proprietary keyword indexing and relevancy technology. Inspire Discovery aims to provide a significantly more powerful and intuitive environment than Encore and other discovery tools used in the academic library arena.

Following the release of Inspire Discovery, Innovative will continue to develop new applications on this platform to complement or eventually provide forward migration paths for some of its established products. No specific time lines have been announced for these future products.

Perspective

Innovative Interfaces finds itself in a very challenging business environment. The company has a broad customer base, including public libraries, academic libraries, law libraries, and other types. The competition in the academic sector has been especially fierce, with vendors such as Ex Libris and OCLC developing specialized products that have a long head start and strengthening maturity. The immanent implementation of FOLIO by some early adopters may also mean additional competition in the academic library front. Some of the gains by these products have come at the expense of Innovative, SirsiDynix, and other ILS companies that have more generalized products oriented to multiple library types.

From its earliest days, Innovative has generally followed strategy of evolutionary product development. There was a continuous path in its ILS product as it morphed from INNOPAC to Millennium and eventually Sierra. Each of these products preserved core software components and functionality, though deploying new interfaces and reengineering architectures to then current standards. While this strategy meant that each iteration of the product inherits rich functionality from its predecessor, it also missed opportunities to make an entirely fresh start and embrace new models of functionality.

In the long term, a completed suite of products based on the Inspire platform could add a new alternative to the mix as libraries seek to replace their aging ILS products.

The development of the Inspire platform represents a major departure from Innovative's traditionally evolutionary development strategy. The creation of an entirely new product suite is a long-term project. Efforts have taken at least three years to create a comprehensive library services platform. The development timeline of Inspire Discovery has been ambitious, especially considering that it also involves the development of the platform infrastructure intended as the foundation for a broad suite of products.

Innovative's design of the technology architecture of the Inspire platform seems consistent with current enterprise software trends. The microservices architecture has been well accepted in the broader high-tech industry, including Google, Amazon, Netflix, Uber, and almost all other globally deployed services. In the library space, FOLIO is being developed using a microservices framework. The use of graph databases and linked data is also a forward-looking approach, including the internal use of BIBFRAME as a library-specific implementation of linked data concepts.

It will be interesting to observe the impact that this new product strategy has for Innovative and the library community. The key issue in the short term is whether these new products will strengthen Innovative's position with academic libraries in the context of formidable competition. Will a more compelling discovery product make libraries more interested in staying with or moving to Sierra? In the long term, a completed suite of products based on the Inspire platform could add a new alternative to the mix as libraries seek to replace their aging ILS products.

Smart Libraries Q&A

Each issue Marshall Breeding responds to questions submitted by readers. Have a question that you want answered? Email to Samantha Imburgia, Managing Editor for ALA TechSource, at simburgia@ala.org.

When our library decides to put its data in the cloud for backup storage and emergency access, what encryption, level of security, and access do we need to consider for non-personal information data versus personal data that might include payroll information and such? Is there a going rate that is reasonable? Are there other considerations and issues to watch out for when considering contracting with a tech company providing these services?

Cloud-based storage services can now be considered a routine part the technical infrastructure for almost any organization. Businesses, government agencies, and non-profits regularly rely on public, private, or hybrid cloud deployments for storage of critical data.

It's important for libraries to develop multi-tier strategies that govern how their data is managed. These strategies will ensure that all data is properly controlled at all stages. The level of security and encryption will depend more on the type of data and its business environment rather than the technology employed. Data stored locally should meet the same security and encryption requirements as any cloud-based options.

Sensitive data should be encrypted throughout its lifecycle.

It is often the case that data will be more vulnerable when housed on local devices and servers. Laptop and desktop computers can be seen as having the highest risk. The users of these computers may not always follow the most stringent security practices. Any data files stored on the built-in storage devices have a high level of risk due to possibilities of loss or theft as well as equipment failures. Data files are much safer when directly stored on institutional network drives or to cloud storage services, such as Microsoft OneDrive, Google Drive, DropBox, Box, or similar services. Most of these services enable functional access to files when the computer is offline with automated synchronization when reconnected to the internet.

The weakest link in the security of data are the people involved. It has been my experience that in many organizations, individual staff members treat their work-related files according to their own preferences. These files may be stored on personal cloud services, flash drives, or other media.

Data associated with general productivity applications should have at least three levels of protection. In most cases, the layers would include copies cached on the local computer, a copy on the institutional file server or cloud-based storage service, and an offline archive on the institutional tape library or a cloud service, such as Amazon Glacier. Some organizations may use multiple cloud services to distribute copies of critical data files or at least storage services deployed in different geographic regions. In most cases, the synchronization of the local and cloud copies of data files can take place in real time, with periodic scheduled transfer to offline storage options.

Files with personally identifiable information or other sensitive data requires additional levels of security and protection. Sensitive data should be encrypted throughout its lifecycle. The procedures described above for multi-level storage options guard primarily against the loss of data through technical failures or human error. Additional layers are required to ensure that in the event of any type of intrusion or security incident, any exposure of files will not enable an unauthorized intruder to gain access to their contents. Sensitive files should therefore be encrypted as records are saved onto storage devices and when transmitted through local networks or through the internet.

The encryption of data files will often be controlled by the business applications involved. In the library context, the ILS manages patron records and related usage data. Personnel and financial data may be created through spreadsheets, enterprise resource planning, or other business systems. These applications should be configured to follow security protocols that enforce encryption for all data that the library considers as sensitive. Although sensitive information can be encrypted as it is transferred to backup drives or other storage locations, significant vulnerabilities remain when business applications do not apply such protections when working with live data.

Authentication credentials, especially passwords, also require special treatment. Many security intrusions involve exposure of username and password files, which can then be used for subsequent attacks. The industry-standard practice of using one-way hashing algorithms to store passwords provides a high level of security. Hashed passwords enable the authentication service to verify that the password entered matches the one stored but does not enable that password to be viewed or

extracted. Any product or service where a system administrator or staff member can view a password has not enabled password hashing and would not be regarded as secure enough to store sensitive data.

In summary:

- Cloud based storage is increasingly the standard operating environment for most organizations.
- Additional precautions should be taken for sensitive data to ensure its protection during active use, storage, transmission, as well as in backup or archival copies.
- Cloud storage offers protections beyond that of local devices and servers.
- Desktop and laptops are usually the weakest option for data storage.
- Distribution of data across multiple local and cloud storage options provides the most protection.
- Sensitive data must be encrypted throughout its operational lifecycle, beginning with the applications that record the data and must remain encrypted as it moves through networks to each tier of storage and backup.
- Passwords should be encrypted through one-way hashing algorithms.
- Organizations should have explicit data storage policies that should be followed by the business applications and in the procedures followed by its personnel.
- Treatment of data should not be at the discretion of individuals in the organization but should be managed within institutional infrastructure, policies, and procedures.

Note

1. Kathryn Harnish (Senior Vice President of Product Strategy, Innovative Interfaces), email interview by Marshall Breeding, April 2019.

Questions or suggestions for topics in future issues?

Contact Samantha Imburgia at **simburgia@ala.org**

ALA TechSource
alatechsource.org

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

NON PROFIT
US POSTAGE
PAID
PERMIT 4
HANOVER, PA

May 2019
Smarter Libraries through Technology

Smart Libraries Newsletter

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

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

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

Digital Access for Subscribers
journals.ala.org/sln

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

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

alatechsource.org