Newsletter

ALA TechSource
alatechsource.org

Smart Libraries™

Formerly Library Systems Newsletter™

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

Smarter Libraries through Technology: Expanding the Models for Library Management

By Marshall Breeding

Over decades, it's been interesting to watch the change in conceptual and technical models of infrastructure supporting libraries—both in how automation systems have organized functionality and the technology platforms and architectures in which they have been deployed. In recent years, I have observed considerable divergence in models of technical infrastructure. Today multiple approaches coexist, each oriented to the distinct operational needs present in different types of libraries.

The integrated library system (ILS) stood as the dominant form of automation for many decades. In about the early 1980s the mold of the ILS was cast as a system centered on common databases, with modules oriented to cataloging, circulation, acquisitions, serials management, with an online catalog for patron access. The ILS was a step forward in consolidating the previous set of standalone circulation, cataloging, or acquisitions products. As I have written many times, the ILS emerged at a time when library collections comprised primarily in print materials, and it developed functionality to manage them with great sophistication. The character of the integrated library system has turned out to be enduring and persistent.

Libraries have changed considerably over time, but not in a homogenous way. Academic libraries have seen their collections shift toward a predominance of electronic resources with a smaller emphasis on print materials. Public libraries continue to purchase and circulate ever more print materials and e-book lending programs have proven to be extremely popular. Corporate and other special libraries diverge even more, with an emphasis on enterprise knowledge management, dealing mostly with business and legal content sources.

The integrated library system continues to persist most strongly in the public library sphere. The ongoing vigorous circulation of physical materials falls nicely within the capabilities of these products, especially as they have matured and evolved over the years. Many have sprouted new public-facing discovery interfaces or catalogs that offer more a more modern user experience and that better incorporate diverse types of materials. Most importantly, these interfaces now integrate well with e-book lending platforms to provide an unified presentation of print and e-books and with drastically simplified mechanisms for enabling patrons to download and read e-books.

The affinity between the online catalog or discovery interface and the integrated library system has proven to be quite strong, with most public libraries relying on products from the same vendor. The main exception is BiblioCommons, which has emerged as a user

IN THIS ISSUE

ISSN 1541-8820 August 2015 Volume XXXV Number 8

interface platform that many public libraries have chosen to use instead of that provided with their ILS.

The academic library arena has seen the most dramatic changes in the models of technical support infrastructure. The first wave of change came in the form of index-based discovery services that provided article-level search spanning all the resources in a library's subscriptions. The initial phase of change saw libraries use these index-based discovery services in tandem with their existing ILS products, often supplemented by an electronic resource management system. The second wave of change came with the emergence of library services platforms. They diverged considerably from the model of the integrated library system, managing print and electronic resources through a more unified, Web-native multitenant platform. OCLC's WorldShare Management Services and Ex Libris's Alma follow this model closely, while others including Kuali OLE and Innovative's Sierra incorporate many of its characteristics.

Library services platforms diverged considerably from the model of the integrated library system.

Neil Block from EBSCO Information Services recently suggested that the the patron-facing Discovery Services Platform serves as the foundation for library technology infrastructure. This view infers that the discovery service isn't an add-on to the internal resource management systems, but that it might work the other way around. In this model, the discovery service becomes the locus of integration for integrated library systems, electronic resource management, learning management systems, and other components needed to exchange data or deliver services. The concept of the discovery services platform aligns with EBSCO's strategy of offering EBSCO Discovery Service with APIs for interoperability with components of the library's choosing, rather than developing its own integrated library system.

In this month's issue of *Smart Libraries Newsletter*, we take a look at yet another possible model for library technology infrastructure. Libraries of research institutions have a somewhat different set of priorities than large academic institutions. They may play more of a role in the curation and dissemination of the research output of the institution, where the institutional repository may be the most critical component of the infrastructure. In this vein, the Invenio software developed at CERN as the basis for an institutional repository has become the foundation for providing support to other areas of the library, including digital asset management and preservation, curation of research data, a publishing platform for scholarly journals, and to support the management and circulation of print and e-books. The Invenio software is now available to institutions outside of the CERN community through the professional services of a new company named TIND Technologies. The Caltech library has engaged TIND in an ambitious project to implement Invenio over the coming months to replace its existing integrated library system and to eventually extend its capabilities for electronic resource management. It will be interesting to observe this project and to see whether this new model gains traction in the academic and research library sector.

TIND Technologies and Invenio: A New Model of Automation for Research Libraries

Invenio, developed under the auspices of CERN is open source software created to address the major areas of involvement for libraries associated with research organizations. The software can be used to support institutional repositories, manage digital assets or multimedia content, provide digital preservation, manage data sets associated with research projects, and to manage library materials. Invenio has been most widely implemented to support institutional and disciplinary repositories but is seeing adoption to support these other activities. As interest in Invenio expanded beyond the community of institutions closely affiliated with CERN, it established TIND Technologies as a new commercial company to promote and provide professional services.

Many libraries serving research institutions focus the majority of their efforts on their collections of electronic resources, documents, and digital objects, with print materials representing a decreasing level of interest. Although physical materials continue as an essential component of these library collections, they may not necessarily rank as the center of its existence. The traditional integrated library system was initially conceived to manage print collections and has evolved over time to also handle other types of materials. Invenio

offers a new approach of building a library management system based on a repository platform in contrast to the established integrated library systems.

Functionality

Invenio was originally developed to provide an institutional repository of scientific papers for CERN and other institutions but has expanded in scope of functionality. Designed to be flexible and adaptable to many resource management needs, the software has been the foundation for additional modules to addressing other types of use. Core functionality includes the classification, indexing, curation, and access to documents. Core functionality centers on the management and dissemination of documents, with its initial implementation as the CERN Document Server, one of the key sources of pre-prints and research reports for high-energy physics. It includes modules for the ingestion, classification, description, search, and access for documents. Support for OAI-PMH enables Invenio to efficiently harvest documents from other repositories and to disseminate its contents to other services. Invenio supports more than 23 different languages.

Serving the interest in content beyond documents, Invenio can manage images, video, and other forms of digital media. Through customized interfaces and plug-ins, users can search and view or play multimedia.

Invenio also provides support for digital preservation, providing workflows and features that follow the OAIS (Open Archival Information System) reference model. The system is able to calculate and validate checksums associated with each digital object and provide multiple layers of security for short- and long-term storage.

In addition to individual bibliographic items or digital objects, Invenio also includes functionality to help organizations manage the data sets associated with research projects. Providing services in support of the management of research data has been a key area of interest for academic and research libraries in recent years.

Beginning in 2011, developers extended Invenio to provide the basic capabilities of an integrated library system for the circulation and management of physical library materials. The original functionality of the product included providing an online catalog in tandem with a separate ILS. The BibCirculation module enables libraries such as CERN to shift away from the management of these materials in a separate application.

Invenio uses MARC21 as its native record structure and includes an editor for working with this format. The software can export records in many other formats, including Dublin Core, MARCXML, JSON, or any type of delimited file. Bibliographic records can be produced and modified individually, or they can be imported in batch.

Invenio does not yet have specialized functionality for electronic resource management, link resolution, or an article-level discovery service. The simplified library automation model may also not include some of the detailed functionality present in integrated library systems, especially in the areas of acquisitions, serials management, or circulation and resource sharing for institutions with large collections, multiple campuses and services points, and complex loan policies. Invenio has primarily been implemented for library management in institutions that serve researchers and not in those with large student populations.

Development History

Invenio was originally developed in support of the library at CERN, a major research organization specializing in high-energy physics. CERN may be known best in recent years for its work to build and perform research through the Large Hadron Collider. The CERN library provides support to the researchers affiliated with CERN and in the dissemination of the documents and reports produced by CERN and related organizations.

CERN initially automated its library using a commercially provided integrated library system. In 1983 the library became one of the earliest implementers outside of Israel of the Aleph ILS from Ex Libris. Given that the World Wide Web was invented at CERN by Tim Berners-Lee, the library was interested in making resources available in this emerging environment. In 1993 the IT group supporting the CERN library created a Web-based interface for Aleph called weblib. This technology was extended in 2002 to create a full document repository, called the CDS Document Server, to provide a more efficient management and distribution mechanism for the documents created at CERN and the many requests it received for copies. The software was originally known as CDS Ware (CERN Document Server), and was later renamed to Invenio, reflecting its development and use by a community of institutions beyond CERN.

As time progressed, Invenio found use to support a broader set of activities at the CERN library, leaving Aleph in place primarily to manage the book collection. Beginning in 2011, Invenio was extended with library management features to more fully integrate the library's technical infrastructure and to relieve it from having to maintain Aleph. This move both saved the expense to the library in fees paid to Ex Libris, but it also enabled a simplified and more efficient environment for management and access of library materials.

Extending Invenio to manage library materials was accomplished through the creation of a new module called BibCirculation. This development was accomplished through an 18-month project carried out by Joaquim Jorge Rodrigues Silvestre, working in the Information Technology Department of CERN. A complete description of the context, the development process, and the capabilities of the module were described by Silvestre's Master's Thesis (*An Integrated Library System on the CERN Document Server*. April 2010. Master's Thesis in Computer Science Engineering. Universidade de Evora).

Below is the timeline of CERN and the development of Invenio.

- **1954:** CERN established (European Center for Nuclear Research)
- **1983:** CERN Library implements Aleph to manage its print collections. One of the first European libraries to implement Aleph from Ex Libris
- **1989:** Tim Berners-Lee invents the Web at CERN
- **1993:** CERN Preprint Server made available on the Web as an institutional repository
- **1996:** CERN Library Server (weblib) launched to also include books, periodicals and other materials, functions as a front-end interface for CERN Aleph installation
- **2000:** CERN Document Server (CDS) launched for multimedia material and internal documents, building on weblib
- **2002:** CDSware released as open source for use by other institutions
- **2006:** CDS Ware renamed to CDS Invenio (Release 0.9x)
- **2010:** Release of Invenio 1.0
- **2011:** library management features added to Invenio.
- **2013, May:** TIND Technologies established
- **2014:** Release of Invenio 2.0

Technical Components and Architecture

Invenio has been in development for more than a decade and has seen some evolution in the technical architecture and components. The software comprises more than 40 modules, which can be selectively implemented depending on the scope of functionality required. Invenio provides all public and staff-oriented functions through Web-based interfaces.

The application makes use of a relational database to manage bibliographic records and operational data. Databases supported include MySQL, PostgreSQL, SQLite, and MongoDB. The application is primarily programmed in Python. SQLAlchemy provides object-relational mapping and other services between the underlying database and Python.

Version 1.0 and earlier of Invenio use MARC21 as the underlying bibliographic record format. Records will be stored in JSON with version 2.0. The user interface for Invenio makes extensive use of JavaScript and the jQuery libraries, with Bootstrap planned for future versions. Other technical components include HAProxy for load balancing and the Apache Web server. Invenio can optionally be configured to use SOLR for indexing and retrieval. A change from SOLR to Elasticsearch is also planned for version 2.0. The current production version of Invenio is 1.2.

Invenio has been created as a server-based application where each instance requires its own copy of the codebase. It is not a multi-tenant platform, as would be expected for new software services created today. Each instance of Invenio can be expected to scale to support large collections with millions of records. The modules of Invenio include RESTful APIs to communicate with external applications or scripts.

TIND Technologies has developed its own hosting infrastructure for efficient deployment and maintenance of Invenio. The company is able to test and validate fixes and software updates and push approved changes to the software to all of its clients simultaneously.

TIND Technologies Launched to Provide Invenio Services

CERN operates a Knowledge Transfer office to identify opportunities and to create spin-off companies to advance software or other intellectual products created by the organization. Knowledge Transfer partners with an entrepreneurship program of the Norwegian University of Science and Technology. CERN spin-offs seek investors to develop a company and share a portion of the revenues generated.

As external organizations began making use of the Invenio software, CERN continually received requests for information, help with implementation, and other types of assistance. CERN IT planned to continue to develop the software, but was not in a position to devote resources to provide external support. This apparent demand for services surrounding Invenio drove Knowledge Transfer to consider the creation of a spin-off company to fill this niche.

TIND Technologies AS was formed as a commercial CERN spin-off company in 2013, with Kenneth Hole and Alexander Nietzold as its co-founders. The new company works closely with the Invenio development team at CERN, with the Knowledge Transfer group, and with the CERN Library.

CERN has established TIND as the exclusive organization for providing professional services for Invenio. It offers TIND access to the development team for Invenio, gaining important expertise. TIND primarily supports the software as developed at CERN, but has in-house technical personnel,

CERN has established TIND as the exclusive organization for providing professional services for Invenio.

several of which came to TIND from CERN IT, for creating and maintaining its hosting platform, for conversion and installation, customization, configuration, and other services. Any individual or organization can download, modify, use, or redistribute Invenio as it is offered under a GNU open source license, but only TIND has direct access to the Invenio development team, beyond the mailing lists, documentation, and other resources generally available to anyone. TIND is licensed to use the CERN brand and logo and returns a portion of revenues as royalties to CERN.

Because Invenio is available as open source software, the primary business model for TIND lies not in license fees, but for support services. TIND has developed a robust hosting infrastructure able to efficiently deploy and maintain instances of Invenio through software as a service (SaaS). The company provides hosting services for any type of implementation using Invenio, with libraries representing the largest portion.

Principals of TIND include:

- Alexander Nietzold, Managing Director
- Kenneth Hole, Product Manager and Project Manager
- Fredrick Carlsen, Software Architect
- Audun Bjørkøy, Technical Director

The company employs a total workforce of eight individuals. It is based in Trondheim, Norway and has offices in Geneva and Paris. TIND Technologies is owned by its employees and does not rely on capital from external investors. TIND has also received some government grants, primarily from Norway.

Current Implementations

TIND is a relatively new company, but has established itself as a provider of services to a variety of organizations. Institutions that have implemented the library management system from TIND include:

- United Nations Office of the High Commissioner for Human Rights
- United Nations Office in Vienna

Those working with TIND for repository or research data management include:

- The Institute of Applied Mechanics of the Czech Academy of Science
- University of Applied Sciences in Western Switzerland

For more information:

- TIND Technologies: tind.io
- Invenio: invenio-software.org
- CERN: home.web.cern.ch

Caltech Sets Tech Strategy Based on Invenio

The library of the California Institute of Technology (Caltech) has recently selected the Invenio platform supported by TIND Technologies as the basis of its library management environment. Caltech ranks as a leading research and education institution dedicated to advancing science and engineering. In addition to its research activities, Caltech serves a relatively small student body, with 2014–15 enrollment reported of 977 undergraduates and 1,204 graduate students.

According to University Librarian Kristin Antelman, the Caltech library intends to implement Invenio with hosting and support through TIND Technologies and integrated with EBSCO Discovery Service for article-level discovery. The Caltech library currently uses the Millennium ILS from Innovative, initially implemented in 1989.

Caltech will work with TIND Technologies to expand the capabilities of Invenio for the management and access of electronic resources. Antelman came to Caltech from North Carolina State University, which is a partner in the Kuali OLE project and the lead institution for the Global Open Knowledgebase (GOKb). Antelman anticipates leveraging the design features of Kuali OLE to extend the capabilities of Invenio to meet academic libraries' needs for managing large collections of electronic resources. GOKb will provide the knowledgebase to support electronic resource management in Invenio. GOKb, though closely allied with Kuali OLE, has from its inception been intended for use with any other project that can benefit from an open access knowledgebase populated through the efforts of librarians.

Caltech shares some characteristics of CERN and other institutions that have implemented Invenio. They share an orientation to research in science and engineering with the need to provide repository services and to mange research data in addition the management of library materials. But as a university library serving a modest-sized student body, Caltech brings some requirements that differ from others that have implemented Invenio. Antelman, however, sees Invenio as a stronger foundation for supporting the libraries repository and digital asset management needs, extending beyond the traditional ILS model. Caltech will be the first library in North America to implement Invenio.

The Caltech library anticipates an implementation schedule with concerted work over the summer of 2015, with migration from its current Millennium ILS by October 2015.

Library Technology News

OCLC's work with library linked data detailed in new book

DUBLIN, OH, 19 June 2015 – OCLC's work to help increase the visibility of library collections on the Web through the creation of library linked data—moving from a web of documents to a web of data—is described in a new book, *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*.

Written by OCLC Research staff members Carol Jean Godby, Shenghui Wang and Jeffrey K. Mixter, the book focuses on the conceptual and technical challenges involved in publishing linked data derived from traditional library metadata. This transformation is urgent, the book maintains, because it is common knowledge that most searches for information start not in a library, or even in a Web-accessible library catalog, but elsewhere on the Internet. Modeling data in a form that the broader Web understands may help keep libraries relevant in the network environment.

In the book, the authors explain how the new Web is a growing "cloud" of interconnected resources that identify the people, places, things and concepts that people want to know about when they approach the Internet with an information need. They also explain why linked data is an appropriate architecture for the description of library resources.

EnvisionWare launch RFID tablet tag

ATLANTA, GA, July 6, 2015 – EnvisionWare, a provider of self-service and library efficiency solutions, announced its new RFID Tablet Tag is designed to enable libraries to circulate mobile devices securely, including iPads, Androids, and other tablets, as well as e-readers.

The Tablet Tag works like a traditional RFID tag used in books and other print materials and can be used for both security and circulation.

ProQuest Ebook Central enters beta

MOUNTAIN VIEW, CA, June 24, 2015 – ProQuest launched the beta program for its new ebook platform. ProQuest Ebook Central integrates key elements from both ebrary and EBL—Ebook Library, along with new functionality.

Beta users include libraries at Oxford University, University of Michigan, Fordham Law Library and University of Wollongong. ProQuest Ebook Central combines the strengths of the EBL and ebrary platforms. The platform's administrative portal simplifies the ebook workflow.

Ebook Central's patron experience encompasses ebrary's new Reader. Designed to work on a variety of devices, the Reader's text and page quality boosts readability while new options, including direct chapter downloads from results and detail pages, give users more control of their reading experience. The beta also debuts a variety of other new features, such as improved search and library branding.

ProQuest expects to conclude beta and go live with ProQuest Ebook Central later this year. Learn more at www.proquest.com/libraries/academic/ebooks.

Total Boox and Califa Library Group announce agreement

NEW YORK, TEL AVIV, June, 17, 2015 – Total Boox, an ebook service for libraries, and Califa, the largest library network in California, announced an agreement, making Total Boox's collection of over 40,000 ebooks available to 32 libraries in

California, including San Francisco, San Jose, Oakland, and Santa Clara Public Libraries.

According to Susan Hildreth, Califa Executive Director, the focus will be on supporting technical and entrepreneurial communities, but patrons of those libraries will be able to instantly access and read ebooks in a wide range of subjects and genres.

SirsiDynix and Zepheira announce strategic partnership

LEHI, UT, June 26, 2015 – SirsiDynix and Zepheira announce their intent to develop a product that will make libraries' collections visible and searchable on the Web.

"Currently, only a fraction of all searches begin at the library," said Bill Davison, CEO of SirsiDynix. "Our goal is to take the mystery and complexity out of Linked Data and deliver to our customers a product that is plug-and-play. We want libraries to easily transform their MARC data into robust, Web-searchable, geo-locatable Linked Data—ready for the world to find."

Libraries who are interested are encouraged to contact SirsiDynix as soon as possible to assess their library's readiness to participate.

Zepheira announces Libhub Initiative Early Adopter Program for academic libraries

June 10, 2015 – http://zepheira.com – Zepheira provides Web visibility to libraries through Linked Data, and now offers a new program for academic libraries to reveal their resources and collections to the Web. The program builds on the successful Libhub Initiative Early Adopter Public Library Program launch with its Founding 12 Partners.

The Early Adopter Program for Academic Libraries builds awareness of the power of Linked Data through education, training and exploration. Linked Data can be used to better reveal a library's materials, which may include ebook, print, special, archival, music, rare books, and/or other collections of value to alumni, students, faculty, researchers and administration.

The current ILS and PAC or discovery layer are not affected, as this is a parallel activity. Should users find library material via the Web, they are linked back to the library's catalog for further action.

The program includes:

- Attendance to Zepheira's Practical Practitioner Linked Data & BIBFRAME Essentials online course (from 1–10 participants)
- Linked Data Readiness Assessment:
  - Data analysis of up to 100,000 MARC records
  - Vendor analysis of partners and systems providers
  - Change management analysis and preparation
- Linked Data Pilot of transforming up to 1 million MARC records using Zepheira's BIBFRAME transformation tools and publishing them to the Web
- A community space for actively sharing information and experiences
- Shared access to early BIBFRAME tools as they are developed

For more information on this and other program offerings, please contact John (JR) Richardson at solutions@zepheira.com.

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

August 2015
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
Patrick Hogan
312-280-3240
phogan@ala.org

Digital Access for Subscribers
journals.ala.org/sln

TO SUBSCRIBE

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

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

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

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

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

alatechsource.org