Putting Linked Data at the Service of Libraries
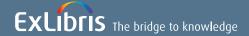
The Ex Libris Vision and Roadmap

Disclaimer

The information in this document will be subject to periodic change and updating. Please confirm that you have the most current documentation. There are no warranties of any kind, express or implied, provided in this documentation, other than those expressly agreed upon in the applicable Ex Libris contract. This information is provided AS IS. Unless otherwise agreed, Ex Libris shall not be liable for any damages for use of this document, including, without limitation, consequential, punitive, indirect or direct damages.

Any references in this document to third-party material (including third-party web sites) are provided for convenience only and do not in any manner serve as an endorsement of that third-party material or those Web sites. The third-party materials are not part of the materials for this Ex Libris product and Ex Libris has no liability for such materials.

Contents

Introduction

"It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data." Tim Berners-Lee[1]

The concept of linked data has gained ground in recent years as a means of fundamentally redefining the way libraries manage data and connecting libraries' data to the greater expanse of online knowledge. The potential advantages of linked data include exposing libraries' data to the web, enhancing data accessibility, enriching cataloging and technical services, and developing new services that will benefit end users and library staff alike. However, the road to harnessing the power of linked data is still paved with multiple challenges.

This document describes the Ex Libris vision for linked data as it applies to library resource management and to discovery and delivery. The key principles that guide Ex Libris in developing linked data functionality and services are outlined, along with the role that linked data is expected to play in the Ex Libris product roadmap.

Linked Data

Linked data is data that is "published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets."[2] Built on standard web technologies such as HTTP and URIs, linked data can be read not only by humans but also by computers.
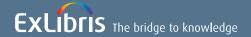
The linked data infrastructure lends itself to the development of numerous types of user services. In their research, patrons access a wide variety of data sources; through linked data, patrons are presented with enriched data in the appropriate context regardless of the interface in which they are conducting their search. In addition, linked data can be exploited to enrich the library catalog, which other applications can use to enrich their own data.

BIBFRAME

The Bibliographic Framework (BIBFRAME) Initiative is a Library of Congress project for defining a bibliographic data model. Based on linked data principles, BIBFRAME has been designed to replace the MARC standards and to make bibliographic data more useful both within and outside the library community.

---

[1] Tim Berners-Lee, "Linked Data," June 18, 2009,
https://www.w3.org/DesignIssues/LinkedData.html,

[2] Christian Bizer, Tom Heath, and Tim Berners-Lee, "Linked Data—The Story So Far,"
International Journal on Semantic Web and Information Systems 5 (2009): 2.

BIBFRAME is expressed in Resource Description Framework (RDF) format, which is based on the idea of making statements about resources (particularly web resources) in the form of subject-predicate-object expressions. These expressions are referred to as triples.

Ex Libris and Linked Data

As a vendor that is deeply engaged with the global library community and benefits from collaborative and forward-thinking customers and user groups, Ex Libris is at the forefront of discussions about linked data. One of the key driving forces behind this initiative is the company's close collaboration with the IGeLU-ELUNA Linked Open Data Special Interest Working Group (LOD SIWG), which has been contributing its experience and ideas for developing "essential linked open data features in all Ex Libris products where appropriate, both from the data publishing, the data consuming, and the data integration perspective"[3] for over four years.

The combination of the Ex Libris Alma® resource management service with the Primo® and Summon® discovery solutions enables Ex Libris to leverage the power of linked data to the benefit of libraries and end users and to support end-to-end services that are based on and enriched by linked data. The merging of services supplied by Primo and Summon with data supplied by Alma will empower discovery-system users as well as library staff with new and exciting possibilities, including richer metadata, enhanced workflows for technical services, improved search results, and new ways to explore content. In addition, third-party tools supporting linked data will consume linked data supplied by Alma and Primo.
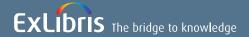
Library and End User Benefits

In light of linked data's potential to make libraries an active part of the web information infrastructure and to render their contributions to the web more visible and attractive to end users, Ex Libris has assumed a role of leadership in linked data initiatives. More and more Ex Libris customers are, indeed, indicating an interest in linked data technology as the right technology for new applications that enrich the discovery experience. Ex Libris is also aware that large providers of metadata are looking into publishing library data in linked data structures.

Ex Libris sees the clear advantages of linked data for library catalogers, other library staff members, and library patrons.

- For catalogers:
  - Authority control via linked data will be online and will use a variety of international sources.
  - Catalogers will be able to add metadata during cataloging when they encounter additional enriched data that might complement the cataloging process.
  - Data supplied by catalogers and additional services related to suppliers and publishers will become an integral part of library workflows.

---

[3] "Linked Open Data SIWG | IGeLU," accessed August 25, 2016 http://igelu.org/special-interests/lod.

- For end users:
  - Users' searches can be automatically enhanced through linked data, displaying results that users were not aware of previously.
  - Related information that is stored externally can be made available to users.
- For the library: The library's collection gains exposure by allowing applications that consume linked data to use and create valuable links to library information.

Key Elements of Linked Data for Ex Libris Roadmaps

The following principles related to linked data have helped shape the roadmap of the Alma resource management solution:
- The use of linked-data format for loading and publishing bibliographic records.
- URI support for cataloging and technical services: identifying "things" based on URIs instead of simple identifiers.
- Access to linked data to enrich data displayed to staff in routine workflows.
- Support for the BIBFRAME model and ontology as they mature.

The following principles have helped shape the roadmap of the Primo discovery and delivery solution:
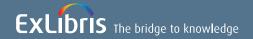- Discovery of the underlying metadata and access to it via URIs.
- The use of linked data by non-library applications.
- The discovery system as the key interface to make data accessible to people and computers (including availability to web search engines).
- The use of RESTful APIs to provide support for applications based on linked data.

Status of Ex Libris Linked Data Projects

Ex Libris is involved in multiple linked data projects, including the Europeana cultural portal, the European Digital Library project, and the IGeLU-ELUNA LOD SIWG. Experience has revealed the following challenges:

- On-the-fly linking of triples in distributed data stores is rather slow and hinders sophisticated discovery. Harvesting is necessary to enable a search engine to use the triples.
- Keeping RDF triples up to date in a central index is problematic. Maintaining triples is a matter of scale, and even medium-size institutions cannot surmount the problems.
- Most of the current metadata sources do not provide RDF triples, and the ontology is not standardized. The metadata has to undergo conversion.

Alma supports a wide variety of RESTful web services, such as services for the retrieval of bibliographic records, holding records, and purchase orders. Retrieved data may be in either XML or JSON format. The RESTful nature of these web services means that the Alma responses include URIs of related entities.

Recognizing the importance of up-to-date URIs that are part of BIB records, along with the large number of linking-based services that can be provided through such URIs, Ex Libris has released a RESTful API in Alma for retrieving any record in a library's catalog in linked data format.

When this API is used, links will be created as embedded URIs or will be based on existing IDs that can be processed to generate full URIs. Alma will make as much use as possible of existing data sources and APIs to generate full URIs. Third-party applications and databases for which URIs will be created include:
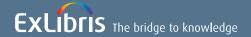- Library of Congress.
- Virtual International Authority File (VIAF®), which links name authority files from national libraries and agencies into a single OCLC-hosted name authority service.
- Integrated Authority File, also referred to as GND (from the German *Gemeinsame Normdatei*), which is managed by the German National Library for the purpose of removing ambiguity in personal names, subject headings, and the names of corporate bodies.

Primo supports a variety of RESTful web services for generating searches, retrieving full records, and retrieving patrons' e-shelf contents. The Primo APIs include embedded URIs for Primo records and patrons' e-shelf contents. The Primo web service responses are in JSON-LD format, containing URIs pointing to records; any application that consumes linked data can embed these URIs to create valuable links to bibliographic records indexed in Primo. The inclusion of URIs and JSON-LD-formatted data in the returned results supports the streamlined consumption of Primo data already in the form of linked data.

2016 Linked Data Collaboration Program

Ex Libris is actively turning linked data into reality through the 2016 Linked Data Collaboration Program, working with customers to roll out the first set of linked data functionalities into Alma and Primo. Among the directions being explored by the program are:
- Alma Technical Services
  - In light of the Ex Libris vision of linked data services as a major element in the facilitation of cataloging, the Alma Metadata Editor and Staff Search screens will show linked data services embedded in a frame or displayed on a menu. In this way, linked data services will become a seamless part of an empowered cataloging process.
  - In order to both introduce the ability to use linked data sources for cataloging, and to leverage the existing MARC based data, Alma's development plans include workflows for catalogers to be able to use well-known linked data sources while cataloging MARC-based bibliographic records, and embed URIs directly.
  - Catalogers will also be able to find and store URIs in the $0 of relevant tags. This data will be available via Alma's existing APIs and publishing mechanisms to Primo and other third-party systems.
- BIBFRAME – publishing library bibliographic data in BIBFRAME format. This is just the first step in the road that leads to native BIBFRAME support in Alma.
- Discovery – downstream use of URIs in Primo to dynamically access and display data to users within the new Primo user interface.

Product Roadmaps

Ex Libris continues to develop comprehensive end-to-end services based on the capabilities of both Alma and Primo. The linking of Alma enhancements to Primo and Summon enhancements will enable patrons who access Alma records through Primo or Summon to enjoy much richer services.

Alma Resource Management

Linked Data
The Alma RESTful API for retrieving catalog records in linked data format will be enhanced to include support for the RDA/RDF format. This will complement the already existing JSON-LD based API available in Alma.

Alma has been designed to serve as the metadata platform providing the infrastructure for other tools that supply services based on linked data. In addition, Alma will be the platform over which such services are delivered. Thus, Alma will help enrich the user experience and support streamlined and more powerful back-office processes for librarians, who will leverage services based on linked data inside Alma.

To this end, Alma will include a linked data services menu on the Repository Search Results screen, indicating that linked data services are associated with the record. Librarians who use Alma records for any process, be it related to acquisitions, fulfillment, or cataloging, will be able to access linked data services for a richer and more efficient experience.
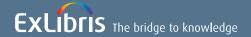
In light of the Ex Libris vision of linked data services as a major element in the facilitation of cataloging, the Alma Metadata Editor will also show linked data services highlighted in a frame or displayed in a menu. In this way, such services will become a seamless part of an empowered cataloging process.

Alma today already has the capability to publish the institution's catalog incrementally with Primo, in a format enriched with URIs. This will be enhanced to also support publishing linked data enriched records to other third-party systems. Publishing in this way works the same way as Alma's RESTful APIs.

In order to implement the use of linked data sources for cataloging and leverage the existing MARC-based data, the development plans for Alma include workflows enabling catalogers to use well-known linked data sources while cataloging MARC-based bibliographic records, and embed URIs directly in these records. The records will be available to Primo and third-party systems via the Alma APIs and publishing mechanisms.

BIBFRAME
Given the expected role of BIBFRAME as a major bibliographic description data model, and in light of the cataloging-agnostic design of Alma, Ex Libris has set a clear roadmap for implementing functionalities in Alma that will improve its integration with BIBFRAME records.

Alma will support both the export and the import of catalog records in BIBFRAME format. Thus Alma records will be part of BIBFRAME-based record workflows outside Alma. A new option will be added to the title-level export job, so existing MARC-based bibliographic records will be exportable in BIBFRAME format. Similarly, imported catalog records in BIBFRAME format will seamlessly become part of the Alma catalog, regardless of the format in which the catalog is managed. Alma will use the metadata import framework with BIBFRAME as a source format.

Discovery and Delivery

Currently Available: Linking to Information in Primo
As a discovery tool, Primo can supply useful URIs that support enriched, linked data services outside the context of a discovery interface. Ex Libris thereby allows applications to use Primo record information within their local interfaces. More Primo data (for example, relationships calculated by Primo) will be exposed as accessible URIs.

Primo will continue to make metadata that has been enriched with linked data available via web services. Existing linked data APIs will be enriched and enhanced, and new web services will be provided.
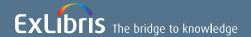
Linking from Primo and Summon Records to Related Resources
As vehicles for supplying linked data services as part of the discovery experience, Primo and Summon will provide such services to researchers who use either interface for discovery. As such, Primo and Summon will display links, data, and services based on harvested URIs and will enrich the discovered content to make use of linked data. Some of these services will rely on the maturity of publishers' harvested linked data information.

Primo and Summon will leverage harvested linked data from source systems that contain URIs into dedicated fields within their records. Patron discovery will be enhanced by connecting the discovery interface and external linked data sources. For example, Primo and Summon will harvest linked data URIs that are calculated by Alma. They will then be able to link to third-party applications and databases, such as the Library of Congress, VIAF, GND, and GeoNames.[4]

In addition, Primo and Summon will seamlessly enrich the user interface with information related to discovered records without forcing users to follow a link out to a third-party system. This capability will enable the library to leverage information held externally and present users with rich metadata and information that lie beyond the library catalog. Related to aspects such as location, subject, or author, such information could prove valuable to a user's research, extending the scope of research to additional works.

---

[4] See *"Status of Ex Libris Linked Data Projects,"* above, page 3.

Figures 1 and 2 present two examples that demonstrate how linked data can enhance the discovery experience. In Figure 1, Primo displays information retrieved from the Library of Congress after linked data URIs calculated by Alma were harvested into Primo and then used to enrich the results. Figure 2 presents a mock-up of a Primo window that shows information retrieved from VIAF about selected co-authors and other works by the same author.
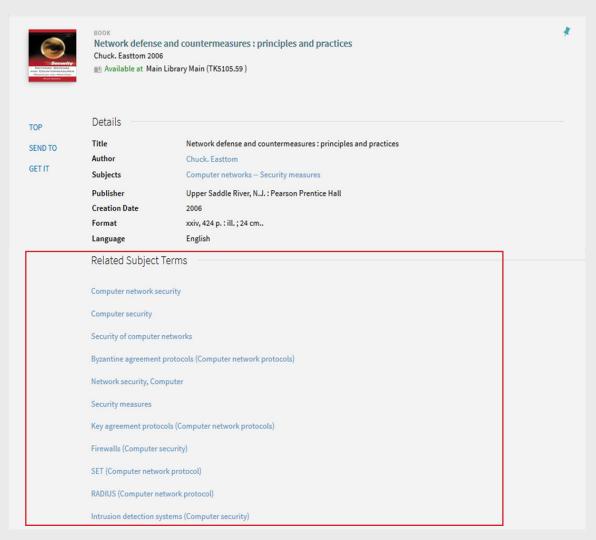

Figure 1. Information retrieved from the Library of Congress (outlined in red) showing related subject terms, displayed within the Primo interface
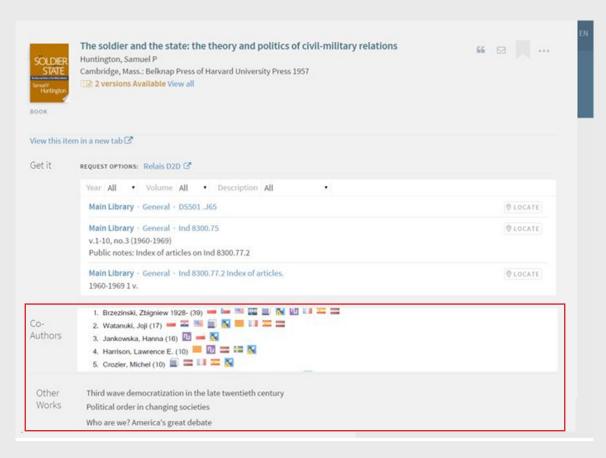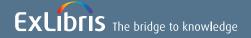
Figure 2. Information retrieved from VIAF (outlined in red) about selected co-authors and other works by the same author, displayed within the Primo interface

Looking Ahead

While there is a shared understanding that the use of linked data will have many benefits in the form of new services for both library staff and end users, the precise nature of the possibilities is still a matter of discussion and debate. Ex Libris is working closely with libraries around the world to identify the scenarios and use cases that are expected to yield the greatest value to libraries and patrons, and is actively planning and implementing linked data services as part of the Alma resource management solution, as well as the Primo and Summon discovery solutions.

About Ex Libris

Ex Libris, a ProQuest company, is a leading global provider of cloud-based solutions for higher education. Offering SaaS solutions for the management and discovery of the full spectrum of library and scholarly materials, as well as mobile campus solutions driving student engagement and success, Ex Libris serves thousands of customers in 90 countries.

For more information about Ex Libris Group, see our website and join us on Twitter, Facebook, YouTube, and LinkedIn.