The Linked Data Landscape and the CSU System 2025

Report from the Linked Data Task Force[‡] (2022-2025)

Authors: Allison Baillund (San Diego State)
Greta Heng (San Diego State)
Michael Herrick (San Francisco State)

June, 2025

Executive Summary

This white paper aims to provide the California State University's Council of Library Deans (COLD) a succinct summary of the current state of linked data applications in libraries. There are no decision points required of deans at this time. The coming 5 years, though, will likely see:

- Increased potential for linked data-based features to **enhance discovery layers**
- Variations in how key **library service providers of systems and metadata** will support both the data production workflows and the discovery application options they offer
- A need to **support library workers**, in resource management, discovery and other areas in retooling their knowledge bases. The **way we catalog** may change by or around 2030. What is the CSU pace for managing this change?
- A potential re-evaluation of how the ULMS committee structure and Digital Repositories integrate linked data expertise to provide **maximum benefit to students on all campuses** as they explore library, and non-library, resources in an increasingly integrated search and discovery environment

Because the Library of Congress has embarked on an ambitious timeline to convert many of its internal workflows from MARC to BIBFRAME in the coming years, other libraries will be following their lead. This raises the question of if, when and how the CSU consortium will participate in this conversion. With interdependencies in technical services workflows and discovery applications, the CSU will have both cost saving opportunities and it could also face potential obstacles as it charts the coming decision points.

Some general, temporal, guideposts framing these issues for the CSU:

- 2030: Ex Libris/Clarivate **current contract ends**
- 2027-2030: Library of Congress, on the FOLIO platform, might start to drive metadata production and **cooperative cataloging workflow shifts**
- Now: Linked data-powered knowledge graphs are helping to drive the first experiments in **AI enabled**, library-specific, discovery environments. Ex Libris has, simultaneously, started building in modular, and more controllable, linked data enhancements to its traditional **discovery environment**

The key players to follow:

- Clarivate/Ex Libris
- EBSCO/FOLIO
- OCLC
- Share-VDE, Blue Core and other consortial or pod-like alignments

Introduction

In the past decade, linked data has influenced how humans connect socially and it has provided a key piece in the overall development of AI systems. Linked data is now on the brink of transforming library-based discovery, after two decades of steady, collaborative effort by many stakeholders. Bibliographic data models, system infrastructures, and discovery platforms are quickly moving to concrete implementations this decade. Ex Libris has started integrating linked data functionalities in recent years, including BIBFRAME data imports and a native linked data, metadata production and editing tool is coming soon. Simultaneously, non-proprietary initiatives and traditional library service providers, like OCLC, are expanding efforts in linked data creation, storage, and sharing. These advancements offer new opportunities for metadata enrichment and discovery platform integrations with the full internet. But, they also raise critical questions regarding data ownership, interoperability and future system compatibility.

When to make the strategic shift in retraining staff to meet the evolving technological landscape is quickly becoming an immediate, practical, question for us. What will the operational costs be to maintain MARC and BIBFRAME workflows simultaneously? For how long? Several CSU library workers spent the past year reaching out to experts and testing Ex Libris software releases which aim to build robust support for linked data applications in the Alma/Primo environment. Our work has raised questions about the following strategic factors which CSU library deans will need to understand. Some of these questions will touch on significant decision points about systems and discovery:

- Data ownership and transition
    - To what extent do we own our metadata?
    - How much flexibility will we have to hook into resources not explicitly hardwired into the Ex Libris system(s)?

- System compatibility
    - Will the systems and the software, in development by Ex Libris and which will support metadata production and editing, be interoperable with future solutions provided by proprietary or open-source solutions?
    - Will shared and open metadata pools be accessible and easily integrated with local production and publishing environments?
    - Will future “unified” systems be so complex that content, metadata and discovery packages will be inextricably bundled for purchase?

- Cost-benefit timing:
    - What are the financial and operational costs of both adopting Linked Data creation and maintaining MARC?
        - Discovery costs
        - Cataloging costs

    - How will we measure immediate costs (staff training, system upgrades) against long-term rewards (efficiency, discoverability, and sustainability)?

Alma/Primo releases in 2024 and 2025 demonstrate that work with linked data, by systems and metadata librarians, is finally becoming truly hands-on. The ULMS community will need to orient itself so that experts in Discovery can recommend and implement, intelligently, features becoming available. Many decisions about the application of a variety of ontologies[1], vocabularies and other data sources, powering robust discovery and exploration experiences for our students, will need to be made in concert with metadata experts.

Decision points:

- The **way we catalog** may change by or around 2030. What is the CSU pace for managing this change?
- ULMS itself is already evolving from its "migration to Alma" 2017 structure. What role will Chancellor's Office staff play in **System-wide efficiencies** around linked data applications?
- Discovery, UX and Retrieval experts in the CSU will need to become more immersed in linked data technologies to analyze, implement, and interpret **discovery experience changes** successfully for our students. Will a CSU-wide structure need to come into existence to support this?
- Are there linked data-associated factors the CSU should be abreast of fully to understand the **total cost-benefit analysis** when formatting and evaluating responses to RFPs for library systems platforms and services in the coming decade?

Initiatives driving Linked Data development in libraries

- **LD4 Linked Data for** [mostly libraries]: a community of multiple affinity groups experimenting with and building LD products and workflows

[The following two initiatives do not derive directly from the LD4 multi-year, Mellon-funded grant projects and the ensuing community of practice with the same name, but they are strongly associated with them and a GLAM-focused ethos of not-for-profit collaboration in the cost modelling]

[1] An ontology is like a controlled vocabulary but it maps relationships within a specific area of human knowledge or discourse. BIBFRAME is one ontology which defines relationships within the bibliographic domain. MARC is merely a data encoding schema to facilitate both data transfer and discovery or publishing applications.

- ❖ Blue Core: (LD shared data pool): conceived in 2023. A minimal viable product is, currently, planned for release in 2026, spearheaded by a core set of institutions. The cost model and collaborative nature of this endeavor is, as yet, unclear

- ❖ Share-VDE (Virtual Discovery Environment): a beta project to demonstrate an international linked data-powered, library, discovery environment. The Share Family project aims to make its enriched metadata openly available under a CC0 license

- **Vendors**
  - Ex Libris: developing AI-assisted metadata production, ingest and transformation workflows to power its integrated discovery environment
  - The Folio Community, Blacklight and EBSCO: open source solutions with hosting and custom implementations by EBSCO
  - OCLC: working on metadata production, transformation, export and a bibliographic-based knowledge graph

---

What to watch for:

- Will Blue Core develop into a **viable alternative data pool** to OCLC?

- Will the Ex Libris **discovery experience be flexible** and its underlying knowledge graph **rich**?

- Will the vast entity[2] ecosystem which OCLC is developing be interoperable? If so, what will its **access and sharing model** look like?

- More broadly, if libraries start to evolve away from the OCLC-based metadata ownership and sharing model of the last half-century, what will the ensuing **metadata management and discovery application** systems look like and how will the **cost modelling** work?

Discovery and Exploration

A library discovery experience, fully integrated with knowledge sources on the open web, is not far off. Converting library datasets to BIBFRAME-powered discovery and exploration will get our users out of the siloed, card-catalog-type structure which still underlies the basic system architecture of OneSearch.

---

[2] An entity, in data modelling, is any agent (person, corporation, etc), place, thing, event, timespan or concept which can be defined and has possible relationships with another entity. OCLC is working through its entire data set to enrich it with entity-based URIs so that what we used to call description “elements” with a fixed position on cards and record-based screens will become machine crawlable entities with their relationships to one another elucidated through ontologies.

These are some of the features either already available or in the works for discovery. Primo's release, in 2024, of the person entity feature, using information drawn from Wikidata, is the first of many linked-data-powered enhancements which Ex Libris has planned.

- Person Cards/Information: A section in the discovery layer displays author related information. Authority data are fetched from online sources dynamically using URIs in the bib records, stored in either MARC or BIBFRAME

- Other "explore" features:
    - Data models and clustering
    - Graph-based visualizations

As the Ex Libris, library-based discovery ecosystem becomes ever larger, the CSU will be wise to assess, periodically, how the content, metadata and discovery applications interrelate. Is it turning into one impenetrable subscription-based bundle? Or is Ex Libris continuing to build stand alone, modular products which a consortium or individual library can select from and pair with other content and services? Some Alma institutions provide their discovery layers via the open source Blacklight software instead of Primo. This allows them to have greater control over the discovery experience for students and provides them with flexibility in negotiating Ex Libris's out-of-the-box functionality. More detailed information on Discovery options can be found here.

What to watch for:

- Will Ex Libris releases for Primo enable a diverse and flexible **explore and discover experience for our students, faculty and staff**?

- What will the, mostly **open source and open metadata**, alternatives offer in the coming years:
    - as interoperable components with the Ex Libris ecosystem?
    - as competitive, complete, alternatives to Ex Libris and/or OCLC products?

Metadata Creation, Enrichment, Ownership and Interoperability

Decision points facing the CSU in the coming years:

- **Data models**:
    - Confirm that BIBFRAME will be the primary linked data model in the U.S., as seems likely at this point.
    - RDA in RDF could become dominant as it closely follows the official RDA and has some simpler pathways to the implementation of robust library discovery environments.

- **Linked Data editing software**: When choosing software supporting linked data production environments, institutions must weigh the differences between

maintenance costs (for free, open-access tools) and membership or subscription fees (for paid tools).

- **Linked Data Storage**: two institution-neutral options seem to be emerging:
  - Blue Core: open access with current support storage in BIBFRAME
  - OCLC: will continue to support MARC while enabling conversion to and from linked data formats such as BIBFRAME and RDA/RDF from different Linked Data editors.

- **Data Ownership**:
  - MARC: Owned by MARC data creator and OCLC
  - Linked Data data model: open access

- **Data Interoperability**: Variations in data model versions and the additional extended fields implemented by vendor platforms complicate the transfer of linked data between different systems and the transfer from MARC.

- **Ethical Concerns**:
  - Data provenance and accuracy
  - What is too much information for library discovery environments? For instance, full birth dates for authors (from copyright registration)?

More information on metadata can be found here.

What to watch for:

- How will **LD editing software,** bundled into proprietary systems, be used?
- How will **data storage options** be integrated with production and discovery environments? How much **flexibility and interoperability** will there be?
- How much agency will libraries have to influence **ethical concerns**?

Conclusion

Discovery environments in some libraries will change significantly over the coming 5 years. AI-based changes are already influencing research assistance approaches in the classroom. The pace of linked data related developments in both discovery applications and metadata creation and management is still not known precisely. The Library of Congress and all of the major marketplace vendors and players are signalling changes in the period 2027 to 2030 and beyond. However, a large portion of libraries will, for some time, remain wedded both to MARC and to the types of discovery environments we have been using for several decades.

The CSU System and its individual libraries, on all campuses, will likely face some key decision points about managing the pacing of change. Whether the CSU will be able to capitalize on system-wide efficiencies in managing this change will in large part depend

on opportunities in the marketplace. COLD may be making consequential decisions about both major contracts and ULMS-related structures which could have far reaching consequences for students on all campuses. How well we navigate this change on behalf of our students, faculty and staff may hinge on how expert we become in perceiving the relevance and import of key decision points. Metadata creation techniques, ownership and management over time could hold the key to cost savings while also enabling or limiting the richness of exploration and discovery environments for our students.

---

‡The Linked Data Task Force was established by the Council of Library Deans (COLD), with a clear charge in early 2022, as an exploratory effort under the Resource Management Committee to learn about use cases for linked data and to provide targeted knowledge-building, specifically within the Resource Management community. Thirteen formal task force members mounted several surveys, outreach efforts and learning opportunities in 2022, 2023 and 2024. This resulted in engaging dozens more library workers across the CSU. The activities of this action-oriented, engagement and outreach team are recorded in its monthly meeting minutes. Because linked data is not only a Resource Management, nor even just an ULMS-specific, topic, FY24-25 activity focused on embedding linked data awareness and knowledge building into ULMS, Digital Repositories and, increasingly, special collections and archives communities. The structure of future engagement with linked data topics in the CSU is reflected in the ULMS Guide. This report focuses on the ULMS-specific, Resource Management and Discovery issues which might influence future COLD-based decision making. This year, a group of CSU library workers also produced an annotated guide to resources for self-directed learning about linked data extending beyond ULMS.