OCLC and linked data
OCLC and linked data

No other library organization has created and shared more linked data than OCLC. We know that linked data brings new contextual relationships to WorldCat® with opportunities to modernize and improve both library and patron workflows.

Our strategic approach is threefold.

- Build on the foundation of WorldCat.
- Collaborate and partner with community leaders.
- Make iterative progress rather than wait for perfection.

And, we are actively working with library and technology leaders to create new capabilities that will

- **deliver MARC records with authoritative links** to help improve the quality of cataloging and metadata workflows;
- allow OCLC to **import, store, and enrich data with links** that can then be **published** in a variety of formats, including BIBFRAME and Schema.org, for web accessibility; and
- give libraries the ability to create and cooperatively manage a **global network of data relationships in WorldCat**.
Expanding the impact of WorldCat

As the power of 420+ million WorldCat records, with 2.6 billion holdings, evolves into a global data network of relationships that can be cooperatively managed by a growing list of 16,500+ libraries, the library community is increasingly empowered to do new things that have traditionally been difficult, if not impossible, to achieve.

Weaving library data into the web

Linked data enables libraries to connect their collections with resources out on the web and vice versa. Rather than working in a “walled garden” of metadata as libraries have in the past, they will instead participate in the broader world of information.

Driving library use through web visibility

Search engines, such as Google, are increasingly using linked data via Schema.org to make sense of the web. The best example of this is the knowledge card that often appears at the top of Google search results. As we continue to innovate and make library data on the web available to search engines by adding Schema.org vocabularies (oc.lc/schema) to existing WorldCat records, more users will find their way to their local libraries.
Providing clarity for users

Work clusters link together multiple manifestations of a work to make it easier for library users to conduct an initial search, such as “Hamlet,” without getting overwhelmed by pages and pages of results. By using the FRBR-inspired concept of a work and linked data techniques, new kinds of user interactions become possible. Our work with WorldCat work entities allows us to experiment with these processes while we continue to refine these clustering algorithms.

Making new connections

Linked data techniques create connections between the initial work and subsequent translations. This will make it possible for systems to present the broader global context of a work and its translations in the language that the searcher prefers.

Serving global needs

Linked data offers easier and more effective ways to aggregate, link, and use controlled vocabularies, such as the VIAF® (Virtual International Authority File) service. OCLC is actively working with the VIAF Council to expand the role that VIAF plays in a linked data world.

Modernizing metadata workflows

Linked data methodologies create new opportunities to programmatically transform existing library metadata into data relationships that will simplify the metadata management and discovery workflows of the future.

<table>
<thead>
<tr>
<th>TODAY</th>
<th>TOMORROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy cataloging</td>
<td>Adding relationships</td>
</tr>
<tr>
<td>Original cataloging</td>
<td>Entity and relationship management</td>
</tr>
<tr>
<td>Authorities</td>
<td>Library-sourced vocabularies</td>
</tr>
<tr>
<td>Siloed searching</td>
<td>Expanded web visibility</td>
</tr>
</tbody>
</table>
Continuing our work on behalf of the library community

At OCLC, we have been working with linked data for many years, exploring ways to leverage the power and scale of WorldCat to drive new efficiencies and visibility for our members. We have been pioneers in linked data research, and we have developed prototypes, pilot programs, and business relationships that continue to inform our work as we integrate library linked data into key products and services.

Linked data research

OCLC Research continues to be one of the world’s leading centers devoted exclusively to the challenges of today’s libraries. Our efforts focus on new ways to make the best use of linked data capabilities as libraries begin their transition from MARC.

OCLC Research teams are actively exploring linked data uses from a variety of angles: as a publisher, consumer, applications builder, and project partner as well as through our involvement in linked data work with standards bodies. This work, including the 2018 international linked data survey, part 3 (oc.lc/ld-survey), has shaped OCLC’s thinking and direction with respect to our prototypes, experimental datasets, products, and services.
OCLC is also working with academic, research, public, and national libraries to prototype a suite of services (oc.lc/ld-prototypes) that will show the production value of linked data. We are conducting this work in the context of real library production use cases to inform OCLC’s metadata strategy. The work includes

- a reconciliation service to connect legacy bibliographic information with linked data entities and
- a service to view, create, and edit linked data entity descriptions and relationships.

Visit our OCLC linked data research page at oc.lc/ld-research for a comprehensive list of activities and results.

**WorldCat data enrichment**

OCLC is also working with national libraries and universities as well as the Program for Cooperative Cataloging (PCC) to define transitions and practical workflows that will modernize and improve cataloging and metadata creation.

In June 2017, the PCC Linked Data Advisory Committee published a report (available at oc.lc/pcc-ld-report) describing several unsolved areas of focus that will require community attention in order to create viable linked data implementation and infrastructure models. We agree with this report’s finding that, as libraries convert their local catalog data to linked data, there is a potential for institutions to create entities that are already described on the Semantic Web.

The task of entity creation introduces a need for both automated and manual reconciliation workflows. These topics are being actively discussed in the PCC Task Group on Identities, which includes OCLC representatives Diane Vizine-Goetz and Jean Godby, as well as in the Linked Data for Libraries (LD4) Community Working Group on Reconciliation (oc.lc/ld4-comm-recon), which includes OCLC representative Roy Tennant.

The PCC Task Group on URIs in MARC recently introduced a change requesting that Uniform Resource Identifiers (URIs) be added to MARC bibliographic and authority records. This is a prerequisite for more efficient and faithful conversions to BIBFRAME and other linked data standards.

OCLC provides the option, when exporting bibliographic records, to include $0 subfields containing URIs for authorities from GND, LCNAF, LCSH, and MeSH. This feature, which will be expanded in the future to support more authority files, is available in WorldShare® Collection Manager and WorldShare Record Manager. OCLC is also working with the PCC and other community groups to develop best practices for the use of the newly defined $1 URIs. URIs appropriate for $1 have been available in VIAF for several years.
In 2006, we introduced WorldCat.org to help expose library holdings to the web. In 2012, we added Schema.org descriptions, followed by WorldCat works and WorldCat person entities several years later. This work, combined with partnerships that we’ve formed with more than 50 websites, including Goodreads, Google Books, and Wikipedia, helped drive more than 73 million visits to WorldCat.org in 2017 and more than 10 million visits to local library catalogs around the globe.

WorldCat works

WorldCat work entities are clustered using the OCLC FRBR work set algorithm, which provides the foundation for new services. The algorithm collects bibliographic records into groups based on author and title information from bibliographic and authority records.

Data elements from records within a given work set are aggregated at the work level to form descriptions that are richer and more complete than the descriptions in individual bibliographic records. More than 215 million WorldCat work entities are available (learn more at oc.lc/wc-work). This year, OCLC has been working with the PCC to prototype the construction of newly defined MARC 758 fields where linked data URIs for works and other resources can be recorded.
WorldCat persons

WorldCat person entities connect related information about a person into a brief description that includes various formats of the person’s name, creative works that the person has produced, and biographic sources of information about the person. WorldCat persons include more than 117 million descriptions of authors, directors, musicians, and others, which have been mined directly from WorldCat. These entities were used in a linked data pilot program in which libraries used WorldCat persons in their regular workflows.

Descriptions of works and their translations

As of June 2018, WorldCat records represent works in more than 490 languages. We leveraged this data in our recent study of works and their translations to produce two key outcomes.

- We identified best practices, available at oc.lc/ld-transition, for describing translated works using the most machine-readable features of MARC to ease the automatic conversion from MARC to linked data.
- We are merging descriptions (oc.lc/multilingualism, .pptx) of translated works with corresponding Wikidata descriptions to obtain the language and script of the original work, the source and target languages, and the names of translators. This information may be incorrect or not represented in the source MARC record.

We anticipate that these improvements will make it easier to obtain the preferred version of a work in a WorldCat search.
**VIAF**

VIAF (oc.lc/viaf) makes library authority files less expensive to maintain and more generally useful to the library domain and beyond. To achieve this, VIAF matches and links the authority files of national libraries. All descriptions for a given entity are merged into a cluster that brings together the different names for that entity. By linking disparate descriptions for the same person or organization, VIAF provides a convenient means for a wider community of libraries and other agencies to repurpose authority data produced by libraries serving different communities. VIAF has been available as linked data since 2009 and is now one of the most widely used linked data resources published by the library community.

**FAST**

Faceted Application of Subject Terminology (FAST) is derived from the Library of Congress Subject Headings (LCSH), one of the library domain’s most widely used subject vocabularies (see oc.lc/fast-research). The broad purpose of adapting the LCSH was to create a faceted subject heading scheme that retains the rich vocabulary of LCSH but is easier to understand, control, apply, and use (read the report at oc.lc/fast-report). FAST maintains compatibility with LCSH, and valid LCSHs can be converted to FAST headings. FAST has a linked data design and has been available as linked data since 2011 (oc.lc/fast-ld).
We’ve learned a lot

It’s no secret that we’ve done a lot of experimenting and prototyping as we prepare for the transition from legacy standards and as we design an environment in which metadata is created and managed in BIBFRAME, Schema.org, and a variety of other Semantic Web vocabularies. It’s a lot to think about, and we’ve learned several key lessons along the way.

**Producing linked data requires more than simply converting records**

Linked data demands a fundamental redesign of our data. This redesign will shift us from exchanging records to managing collections of facts—asserting relationships between our resources and the people, places, organizations, and concepts that are important to libraries and their patrons.

**Scale matters**

To be truly effective, linked data statements must be formulated from as large a corpus of data as possible and must draw upon additional data sources, such as Wikidata. Our work in the development of WorldCat work and person entities will provide the bibliographic spine needed in this new environment.
Putting library linked data on the web is important, but it is not a panacea to solve web visibility problems alone

Differentiating between the effects of traditional search engine optimization (SEO) and the use of linked data to improve web visibility is difficult. There is evidence that traditional SEO techniques can be more successful than referrals from library linked data in the current environment, so it may take time before linked data representations become more beneficial than SEO techniques.

One standard does not fit all

The BIBFRAME standard continues to evolve, and OCLC remains committed to working with the Library of Congress to help finalize the official version. Additionally, multiple extensions to and variations of BIBFRAME have emerged, driving OCLC to think more broadly as we build capabilities that support a more complex landscape.

MARC isn’t going away anytime soon

While we believe that linked data representations will eventually become the de facto standard, we also believe that MARC and other record-based systems will continue to be used by the library community for many years to come, resulting in our decision to support MARC, BIBFRAME, and Schema.org.
Conclusion

OCLC has been working with linked data for many years, and we know where the challenges are. We understand that it’s more about progress than perfection, and iteration is essential as linked data standards and vocabularies continue to evolve. That’s why we continue to prototype new services (oc.lc/ld-prototypes) while contributing our expertise on standards committees.

OCLC is committed to serving the needs of library staff members who wish to continue working in traditional ILS systems by enriching WorldCat MARC records with authoritative URIs.

OCLC is also committed to building on the size and scale of WorldCat to further develop a global data network of relationships and new services to help libraries create and manage entities and relationships that can be published in a variety of formats, including BIBFRAME and Schema.org.
Continuing our commitment

No other library organization has created and shared more linked data than OCLC. By building on the foundation of WorldCat, collaborating with community leaders, and making iterative progress rather than waiting for perfection, we will continue to develop new opportunities to modernize and improve library staff and patron workflows.

If you have any questions or comments, please contact us at linkeddata@oclc.org.