

OCLC and linked data: Getting the best results for you

At OCLC, we are deeply engaged with library and technology leaders as we pursue a linked data vision that will drive more effective uses of library data while making it possible for the library community to do things that have traditionally been difficult or 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 search engine optimisation:** 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 make library data on the web available to search engines (as we’ve done by [adding Schema.org vocabularies](#) to existing WorldCat® records), more users will find their way to their local library.
- **Providing clarity for users:** Manifestations of a work will be linked together in a work cluster to make it easier for library users to conduct an initial search, such as “Hamlet,” without getting pages and pages of results with no good way to focus in on the version they want. 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 and to discover where our clustering algorithms can be improved.
- **Making new connections:** Linked data techniques will be used to 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 through the VIAF® (Virtual International Authority File) service. OCLC is actively working with the VIAF Council to explore the role that VIAF will play in a linked data world.
- **Modernising metadata workflows:** Linked data methodologies create new opportunities to programmatically [transform existing library metadata](#) (.pptx) into data relationships that will enable new, simplified metadata workflows.

Understanding the challenges

As OCLC prepares for an environment in which metadata is created and managed in BIBFRAME, Schema.org, and other legacy standards, we continue to face challenges, and we are committed to working with the community to find solutions.

- **Finding library data on the web.** There is some evidence that traditional search engine optimisation (SEO) techniques appear to be more successful than referrals from linked data in the current environment. It may take some time for linked data representations to become more useful than SEO techniques with search engines.
- **Producing linked data is more than simply converting records.** Linked data is a fundamental re-design of our data from records to collections of facts about the important people, places, organisations, and concepts—or “entities”—that are important to libraries and their patrons. 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.
- **The BIBFRAME standard continues to evolve.** OCLC remains committed to working with the Library of Congress to help finalise the official version of BIBFRAME. Multiple variations of BIBFRAME have recently emerged, driving OCLC to 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 will continue to be used by the library community for many years to come. We do plan to support MARC, BIBFRAME, and Schema.org.

Continuing our work on behalf of the library community

At OCLC, we’ve been leading the way with research and experimentation to overcome these challenges. In 2006, we introduced WorldCat.org™ to help expose library holdings to the web. We then formed more than 50 partnerships with popular websites, including Goodreads, Google Books, and Wikipedia. In 2016, this work helped OCLC drive more than 74 million visits to local library catalogues around the globe.

But our work hasn’t stopped, and we continue to research ways to leverage the power of WorldCat to modernise and improve cataloguing and metadata workflows for libraries that want to move to new systems as well as for libraries that wish to keep working in MARC.

We will continue to lead the transformation with a multi-faceted approach to linked data.

- **Linked data research:** OCLC Research continues to be one of the world’s leading centres 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 the library community prepares for the eventual transition from MARC.

OCLC Research teams continue to explore 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 results of two international linked data surveys, is shaping and informing OCLC’s thinking and direction with respect to our prototypes, experimental datasets, products, and services. Visit our [OCLC linked data research page](#) 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 modernise and improve cataloguing and metadata creation. For example, in June 2017, the **PCC Linked Data Advisory Committee** [published a report](#) describing several unsolved areas of focus that will require community attention in order to create viable linked data implementation and infrastructure models.
 - The **PCC Task Group on URIs in MARC** recently introduced a change requesting that URIs be added to MARC bibliographic and authority records. When these requirements are complete, OCLC will add URIs to \$0 and \$1 MARC subfields in WorldCat master records as part of our ongoing role as stewards of quality information.
 - We also agree with this report's finding that, as libraries convert their local catalogue data to linked data, there is a potential for institutions to create entities that are already described on the semantic web. This could necessitate a need for both automated and manual reconciliation workflows, which are actively being 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](#) (which includes OCLC representative Roy Tennant).

- **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. Author names and titles are normalised to construct a key. All records with the same key are grouped together in a work set.

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. As of July 2017, more than 215 million [WorldCat work entities](#) are available.

- **WorldCat persons:** WorldCat person entities connect related information about specific people 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. As of July 2017, 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 programme in which libraries used WorldCat persons in their regular workflows. See an example of a [WorldCat person entity](#).
- **Descriptions of works and their translations:** As of June 2017, 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](#) 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](#) (.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](#) 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. It then groups all authority records for a given entity into a merged “super” authority record that brings together the different names for that entity. By linking disparate names for the same person or organisation, VIAF provides a convenient means for a wider community of libraries and other agencies to repurpose bibliographic data produced by libraries serving different language 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) and has a linked data design that makes it one of the library domain’s most widely used subject vocabularies. The broad purpose of adapting the LCSH was to create a faceted subject heading scheme that retains the rich vocabulary of LCSH and is easier to understand, control, apply, and use ([read the report](#)). FAST maintains compatibility with LCSH, and valid LCSHs can be converted to FAST headings.

Conclusion

No other library organisation has created and shared more linked data than OCLC. We will continue our work with library and technology leaders to modernise cataloguing and metadata workflows by leveraging our state-of-the-art technology and global data resources (including WorldCat, VIAF, FAST, the Dewey Decimal Classification system, and more) to create new capabilities that will

- deliver MARC records with authoritative URIs to help improve the quality of cataloguing and metadata workflows;
- help OCLC import, store, and enrich data with links that can then be published in a variety of formats for web accessibility; and
- give libraries the ability to create and cooperatively manage a network of data relationships.

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