

Edition Clustering in WorldCat Discovery and WorldCat Local

Throughout the design, development and adoption of WorldCat Local and WorldCat Discovery, we have heard many thoughtful questions and concerns about our approach to edition clustering—about risks and benefits, about design decisions and about clustering errors. We have created this document in response to member feedback to explain our edition clustering approach inspired by FRBR, which we use for print monographs and other physical formats. Journal articles are handled differently, and not discussed here.

This document describes the details of our approach that extend beyond the FRBR model, including what specific details we choose to display at different stages in a user's path. It identifies the kinds of uses we believe we are currently supporting well, those we are aware we're supporting less well and what our plans for improvements are.

Important user journeys

When assessing the effects of edition clustering, it is useful to group user journeys under these broad headings:

- Many users sometimes *trust their library to offer them whatever is best*.
- Many users sometimes *know exactly what edition they want* and how to describe it.
- Many users sometimes *want to explore*, and can't specify in advance the exact attributes of the things they will be interested in.

Past and ongoing research and testing, and comments and feedback through various channels, reveal some patterns within each of these journeys.

1. Many users sometimes trust their library to offer whatever is best.

Many users, including many college students, expect their library to deliver the best edition of a work without having to figure out for themselves which one that is. Our testing shows that many college students understand that the most recent edition is generally preferable; few recognize they're being offered an older edition instead. But when we display the most widely-held edition or format by default, it will too often be an older one. *In response, we plan to make the most recently published locally-held edition or format the default that displays in search results, unless only a different edition matches the search terms.*

2. Many users sometimes know exactly what item they want and how to describe it.

A college student assigned an item from a reading list or a researcher tracing a citation is aware that he or she is looking for a specific edition, and knows something specific about that edition (date, publisher, perhaps a particular editor or translator). In our usability tests, college students looking for a specific edition often search by including edition-specific information in the keyword search box. This type of specific search retrieves an edition that matches all the search criteria, rather than the most widely-held

edition of the work. For example, a search with a Russian-language title displays a Russian-language edition, even if it is not the most-widely held locally-owned edition. We've seen in extensive testing that the current implementation works quite well in cases where searchers include edition-specific information like dates, as they often do. However, getting a specific edition can be difficult for users who do not specify something distinctive about the edition they want in their search.

3. Many users sometimes want to explore, but can't specify in advance the exact attributes of interest.

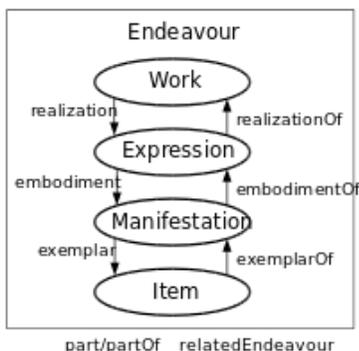
For subject or author searches, where users select between works, edition clustering prevents search results from being swamped with many editions of the same work.

When exploration focuses on different editions of the same work (when a user is not looking for a particular edition, but just wants to know what there is: Is there a Russian edition? What's the earliest edition?) or when an advanced researcher wants to learn about all the editions of a work that have been published with their dates and languages, the researcher will be interested in a list of all editions of the work. In our current implementation, reviewing such a list requires unreasonable time and patience. *Proposed improvements to the list of editions for each work, including filtering for this list, should make exploring the range of different editions of a work easier.*

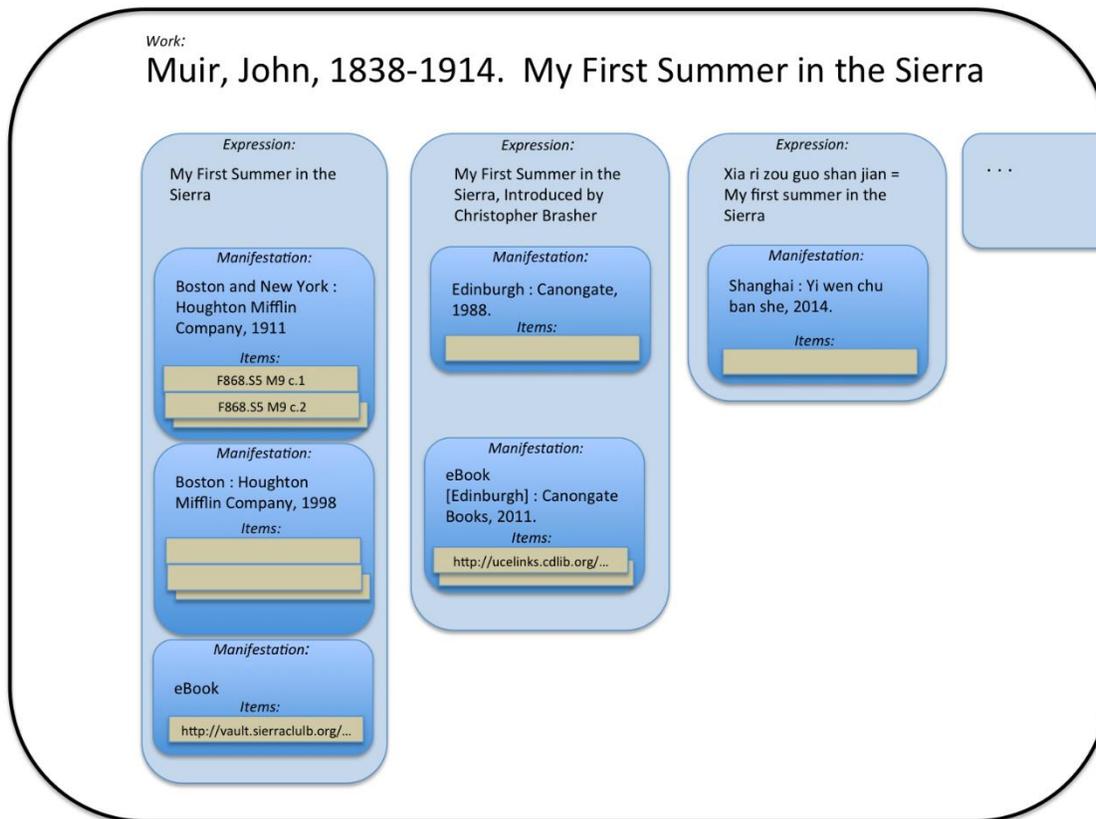
The FRBR model

Our approach of bringing together related editions under the concept of an abstract work—but not the details of how we display those editions or which editions we show under specific circumstances—is inspired by [IFLA's FRBR model](#). The FRBR reference model for bibliographic data defines, in a broad way, the entities, attributes and relationships that are important to represent in the records that libraries and library users typically exchange. FRBR specifies that intellectual or artistic products include the following types of entities:

- **The work**, a distinct intellectual or artistic creation.
- **The expression**, the intellectual or artistic realization of a work.
- **The manifestation**, the physical embodiment of an expression of a work.
- **The item**, a single exemplar of a manifestation.



Source: <http://www.oclc.org/research/activities/frbr.html>



OCLC's implementation of edition clustering inspired by FRBR

As records enter the WorldCat database, OCLC's edition clustering algorithms work in the background to identify and cluster records for all manifestations and expressions of a given intellectual work, including different languages, different translations in the same language, and editions with different dates and publishers.

When a user does a search in WorldCat Discovery or WorldCat Local, the initial search results page shows just one edition from each cluster, a default edition. The edition that displays matches all the search criteria, so if a user includes a date, publisher, editor or a language limit, the search result shows an edition that matches those criteria. If more than one edition in your library matches all the search criteria, the edition your library holds that is most widely-held across all WorldCat libraries will represent the cluster. This design decision makes the most recent and desirable editions a library holds less visible when older editions have more holdings. *We intend to change the criteria and instead show the most recent matching edition your library holds as the default edition on search results.*

We tested search results that summarize each work cluster (showing how many formats, how many languages and what range of dates the cluster included) with academic users. In this testing, we found that those users did not expect or prefer to see an abstract "work" record on search results—the complexity of the different editions was something they cared about later, if at all.

In a departure from FRBR, we refer to manifestations as editions. In another departure, within the list of

editions and formats for a work, we don't present some more closely-related manifestations (different editions of the same translation, for example) as a distinct expression within the work. Consequently, all editions appear together in a single list under the Editions and Formats heading, without editions of the same translation being grouped more closely together within the list.

Display of edition clusters in WorldCat Discovery

Search result: This library has not switched on the "View all editions" link.

The screenshot shows the University of Maryland Libraries' WorldCat Discovery interface. The search bar contains the text "my first summer in the sierra". The results page displays three items:

- 1. My first summer in the Sierra**
by John Muir, Herbert Wendell Gleason
Print book 1911
Held by: University of Maryland, College Park
Famed naturalist John Muir (1838-1914) came to Wisconsin as a boy and studied at the University of Wisconsin. He first came to California in 1868 and devoted six years to the study of the Yo... [Read More](#)
[Access online](#)
Non-Circulating/On Shelf UMCP HBK Maryland Room QH31.M9 A34 1911
- 2. Nature writings : the story of my boyhood and youth : My first summer in the Sierra : The mountains of California : Stickeen : Selected essays**
by John Muir
Print book ©1997
Held by: University of Maryland, College Park
"In a lifetime of exploration, writing, and passionate political activism, John Muir made himself America's most eloquent spokesman for the mystery and majesty of the wilderness. A... [Read More](#)
[Access online](#)
On Shelf UMCP McKeldin Library QH31.M9 A3 1997
- 3. Gentle wilderness: the Sierra Nevada.**
by John Muir, Richard Kauffman, David Brower, Sierra Club.
Print book [©1967]
Held by: University of Maryland, College Park
On Shelf UMCP McKeldin Library F868.S5 M88 1967

The left sidebar contains filters for "Sort by" (Library and Relevance), "Library" (Libraries Worldwide, Big Ten Academic Alliance, USMAI Libraries, University of Maryland, College Park, McKeldin Library, Architecture Library), "Content" (Full Text, Peer Reviewed (452)), and "Format" (All Formats, Article/Chapter (673), Article (556), Chapter (95), Book (140)).

Item 1 represents multiple editions, but there's no clue that this is the case because there is no "View all editions" link.

Search result: This library has enabled the “View all editions” link.

919 Results

my first summer in the sierra

Advanced Search

Sign In

Chat with a librarian

My List

Sort by: Library and Relevance

Library: Libraries Worldwide, Mennonite Historical Library, Libraries in Indiana, Good Library, Goshen College Harold and Wilma Good Library, Good Library, Chemistry Reading Room

Content: Full Text, Peer Reviewed (452)

Format: All Formats, Article/Chapter (673), Article (556), Chapter (95), Book (140)

- 1. My first summer in the Sierra**
by John Muir
Print book 1988
Held by: [Good Library, Goshen College Harold and Wilma Good Library](#)
[Access online](#)
Available Good Library 2nd Floor -- Oversized Materials QH31.M9 A3 1988
[View all editions](#)
- 2. Nature writings : the story of my boyhood and youth : My first summer in the Sierra : The mountains of California : Stickeen : Selected essays**
by John Muir
Print book ©1997
Held by: [Good Library, Goshen College Harold and Wilma Good Library](#)
"In a lifetime of exploration, writing, and passionate political activism, John Muir made himself America's most eloquent spokesman for the mystery and majesty of the wilderness. A...
[Read More](#)
[Access online](#)
Available Merry Lea Learning Center Library QH31.M9 A3 1997
[View all editions](#)
- 3. Gentle Wilderness : the Sierra Nevada**
by John Muir, Richard Kauffman, David Brower, Sierra Club.
Print book ©1964
Held by: [Good Library, Goshen College Harold and Wilma Good Library](#)
Available Merry Lea Learning Center Library F868.S5 M88
[View all editions](#)

Send Feedback

Item 1 represents multiple editions. The “View all editions” has proven in testing to raise awareness and provide an effective path to other editions.

Detail view

UNIVERSITY OF MARYLAND LIBRARIES

my first summer in the sierra

Advanced Search

Library Links

Staff Sign In

View Filters

919 Results

My List

- 1. My first summer in the Sierra**
by John Muir, Herbert Wendell Gleason
Print book 1911
Held by: University of Maryland, College Park
Famed naturalist John Muir (1838-1914) came to Wisconsin as a boy and studied at the University of Wisconsin. He first came to California in 1968 and devoted six years to the study of the Yo...
[Read More](#)
[Access online](#)
Non-Circulating/On Shelf
UMCP HBK Maryland Room
QH31.M9 A34 1911
- 2. Nature writings : the story of my boyhood and youth : My first summer in the Sierra : The mountains of California : Stickeen : Selected essays**
by John Muir
Print book ©1997
Held by: University of Maryland, College Park
"In a lifetime of exploration, writing, and passionate political activism, John Muir made himself America's most eloquent spokesman for the mystery and majesty of the wilderness. A...
[Read More](#)
[Access online](#)

Open "lib.umd.edu" in a new tab

Send Feedback

My first summer in the Sierra

Close Item Detail

Access online

MARC

Description

Editions & Formats 117 Found

Availability / Holdings

Access Full Text Online 2 Links

View eBook

My First Summer in the Sierra, Project Gutenberg E-Books

hdl.loc.gov

This item does not circulate at University of Maryland, College Park.

The collapsed Editions & Formats section is the only clue there are other editions.

Editions and Formats list

The current interface shows a search for 'no:19457813' with 1 result. The main content area displays a list of four editions of 'My first summer in the Sierra' by John Muir. Each entry includes a thumbnail, author information, publication details, and holding information. A 'Send Feedback' button is located at the bottom right of the list.

Planned enhancements to the Editions and Formats list will make the list easier to manipulate through filtering.

Proposed new design for Editions and Formats list

The proposed design features a more structured layout for the 'Editions and Formats list'. It includes a search bar with 'alice in wonderland', a navigation bar with 'Library Links', 'Emma Enduser', and 'My Lists'. The main content area shows the book details for 'Alice in Wonderland' by Lewis Carroll, including a 'Share' button, 'Cite', and 'Save' options. A 'Description' section is visible, with a note: 'Relates to Facet enhancements work. PENDING USER TESTING.' Below this, there are facet filters for 'Library' (Pepperdine University), 'Year', and 'Language' (English). A table lists the editions found, with 16 total results.

4/22/2016
 Other Editions should:
 - Honor appropriate facet selections (format, language, year, availability/holdings)
 - Provide ability to change this filtering only within this section (do not persist changes to these drop downs)

Title / Author	Language	Edition	Availability / Holdings
Alice in Wonderland by Lewis Carroll, Donald J. Gray	English	[1971] [1st Ed.], xi, 434 pages illustrations 22 cm. New York, WW Norton [1971]	Pepperdine University Libraries and 12 Libraries Worldwide
Alice's adventures in Wonderland : and, Through the looking-glass and what Alice found there by Lewis Carroll, Roger Lancelyn, Green, John Tenniel	English	2008 xxxiii, 278 pages : illustrations ; 20 cm. Oxford ; New York : Oxford University Press, 2008.	Pepperdine University Libraries and 12 Libraries Worldwide
Alice's adventures in Wonderland : and, Through the looking-glass and what Alice found there by Lewis Carroll, Roger Lancelyn, Green, John Tenniel	English	1998 xxxiii, 278 p. ; 20 cm. Oxford ; New York : Oxford University Press, 1998.	Pepperdine University Libraries and 12 Libraries Worldwide
Alice's adventures in Wonderland : and, Through the looking-glass and What Alice found there by Lewis Carroll, John Tenniel, Roger Lancelyn Green	English	1971 xxxiii, 277 pages : illustrations ; 21 cm. London ; New York : Oxford University Press 1971	Pepperdine University Libraries and 12 Libraries Worldwide

How we build and maintain edition clusters

OCLC's clustering algorithms primarily use names, title and format to cluster editions. Translations are retrieved with the original work, based on information mined from WorldCat when relationships between a translation and the original work can be identified.

MARC data does not always include the information needed or does not record it in a usable way. Sometimes cataloging is too varied or unreliable. Examples include some TV series, remakes of movies and book series. Collected works can cause problems because the collected titles may or may not reference the included titles, which might have been cataloged individually. Some relationships are built manually to solve specific problems.

Differentiating between an expression and a new derived work can be subjective and difficult even with the objects in hand—sometimes certainty is impossible based on metadata alone.

Clusters are built from scratch each month. New data, plus fixes to algorithms, can result in seven to ten percent of clusters changing on a monthly basis.

How we discover edition clustering problems and fix them

OCLC staff routinely receive reports of the largest clusters (based on the number of editions or the number of holdings) and of a random sample of changed clusters. We examine these for clustering problems. Reports from users to OCLC Customer Support also reveal clustering problems. We investigate to find patterns, determine the role clustering algorithms or data quality play, and develop solutions. We install fixes monthly. Improvements ordinarily affect many more records than those that are reported, because the corrections are based on patterns of problems.

Conclusions

Edition clustering is one of the tools that help make the world's largest library catalog useful to individual libraries for revealing their own collections and the broader range of group and global resources.

For many common cases, edition clustering is demonstrably beneficial:

- For topical searches, author searches and other exploratory searches where users choose between works, edition clustering prevents many very similar editions of the same work from swamping search results—making selection between different works easier.
- For searches where a user is looking for a specific edition and they include something explicit about that edition in the search, edition clustering doesn't hide matching editions because we select the matching edition to represent the whole work cluster in search results.

However, we have made some design decisions that can impair the experience of some common cases:

- Showing the widely-held, locally-owned edition as the default edition in search results (unless another edition matches the search more closely) sometimes eclipses a more desirable more recent edition that is locally owned. *We plan to change this to display the most recent locally-held edition.*
- When we made the “View all editions” link on search results a configurable option (defaulting to off), the effect, for libraries that didn’t choose to display this link, was loss of a valuable indication that other editions are available locally. Library staff can change this by adjusting their profile in [OCLC Service Configuration](#).
- The list of editions and formats of a work can be difficult to navigate when it is long. *We are developing a design that provides better tools for exploring the list of editions within each work.*

Improving discovery of the desired editions and formats is a high priority for us. This effort is a collaborative process, and we appreciate your input. If you have experienced a clustering issue or have feedback for the OCLC team, please post your comment in the [Community Center](#) or contact us at worldcatpm@oclc.org.