

# Edition Clustering in WorldCat Discovery and WorldCat Local – July 2018

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. As a result, we've made changes to rules for determining the default edition, and made plans for other changes as well. This document explains our current edition clustering approach inspired by FRBR, which we use for print monographs and other physical formats. It also 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 our plans for improvements.

Journal articles are handled differently, and not discussed here.

## 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, often including 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, but also that few would recognize if we offered an earlier edition in its place. Therefore, we have changed the default edition displayed in search results from the most widely-held (often an older one) to the most recent locally-held edition, 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, identifying 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.

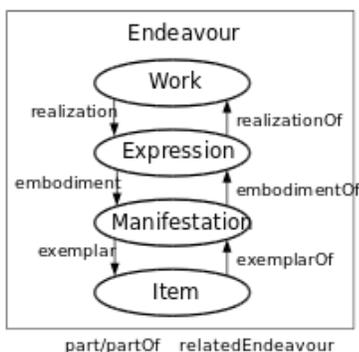
For subject or author searches, where users select between works, edition clustering prevents search results from being swamped with many editions of the same work.

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.*

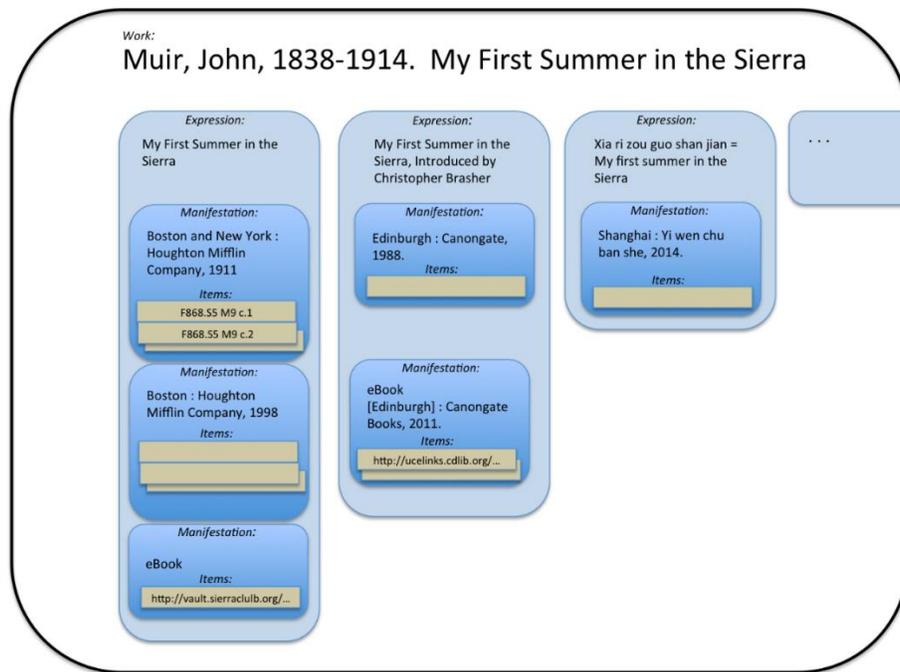
## The FRBR model

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. Following a clustering change in 2017, if more than one edition in your library matches all the search criteria, the edition your library holds that is most recent will represent the cluster. Our testing with users confirms that this change makes the most recent and desirable editions a library holds most visible, even when older editions have more holdings.

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 enabled the “Group related editions” option.

The screenshot shows the University of Maryland Libraries search interface. The search term is "my first summer in the sierra". The "Group related editions" filter is selected and highlighted with a red box. The results are sorted by "Library".

**Search Tools**

- Group related editions
- Library**
  - Libraries Worldwide
  - Big Ten Academic Alliance
  - USMAI Libraries
  - University of Maryland, College Park
    - McKeldin Library
    - Architecture Library[Show More](#)
- Content**
  - Full Text
  - Peer Reviewed (451)
- Format**

Top 6 Formats

  - All Formats
  - Article/Chapter (819)
    - Article (667)
    - Chapter (105)
  - Book (164)
    - Print book (98)
  - Archival material (94)[Show More](#)
- Databases**
  - All Databases
  - WorldCat.org (1069)
- Author**
  - All Authors
  - Muir, John (62)
  - Khachatryan V (49)
  - Sirunyan A M (42)
  - Chatrchyan S (20)
  - Londt Jason G H (5)[Show More](#)
- Year**
  - All Years
  - Last 5 Years
  - Last 10 Years
  - Last 25 Years
  - Custom Year Range:  
from  through[Change Year](#)
- Language**
  - All Languages
  - English (783)

1,098 results in Libraries Worldwide Sorted by: Library

- 1. My first summer in the Sierra**  
by John Muir  
  
Print book 1988 | [View all editions & formats](#)  
Held by: [University of Maryland, College Park](#)  
[View eBook](#)  
16/Aug/18 UMCP McKeldin Library QH31.M9 A3 1988  
[Cite](#) [Link](#) [Email](#) [Save](#)
- 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 | [View all editions & formats](#)  
"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](#)  
Held by: [University of Maryland, College Park](#)  
[View eBook](#)  
On Shelf UMCP McKeldin Library QH31.M9 A3 1997  
[Cite](#) [Link](#) [Email](#) [Save](#)
- 3. Gentle wilderness; the Sierra Nevada.**  
by John Muir, Richard Kauffman, David Brower, Sierra Club.  
  
Print book [©1967] | [View all editions & formats](#)  
Held by: [University of Maryland, College Park](#)  
On Shelf UMCP McKeldin Library F868.S5 M88 1967  
[Cite](#) [Link](#) [Email](#) [Save](#)
- 4. The world in which we occur : John Dewey, pragmatist ecology, and American ecological writing in the twentieth century**  
by Neil W. Browne  
  
eBook ©2007 | [View all editions & formats](#)  
American philosopher John Dewey considered all human endeavors to be one with the natural world. In his writings, particularly Art as Experience (1934), Dewey insists on the primacy of the e... [Read More](#)  
Held by: [University of Maryland, College Park](#)  
[View eBook](#)  
[Cite](#) [Link](#) [Email](#) [Save](#)

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

Detail view



my first summer in the sierra



Advanced Search

Library Links

Staff Sign In

View Filters

My Items (0)

1. My first summer in the Sierra



by John Muir

Microform 1911 | [View all editions & formats](#)

Held by: USMAI Libraries

[View eBook](#)

2. My first summer in the Sierra



by John Muir, Herbert Wendell Gleason

Print book 1911 | [View all editions & formats](#)

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](#)

Held by: [University of Maryland, College Park](#)

[View eBook](#)

✓ Non-Circulating/On Shelf  
UMCP HBK Maryland Room  
QH31.M9 A34 1911

3. My first summer in the Sierra



by John Muir

Print book [1990] | [View all editions & formats](#)

Expresses the transcendent joy, the insatiable curiosity, the nascent sense of mission of a man discovering his passion and his destiny.

Held by: Big Ten Academic Alliance

[View eBook](#)

4. My first summer in the Sierra



by John Muir

Print book 1987 | [View all editions & formats](#)

Held by: Big Ten Academic Alliance

[View eBook](#)

5. My first summer in the Sierra

by John Muir

My first summer in the Sierra

Cite Link Email Save



by John Muir (Author.), Herbert Wendell Gleason (Photographer.)

Print book 1911

Held by: [University of Maryland, College Park](#)

[View eBook](#)

Description

Editions & Formats

118 Found

Availability

Access Online

1 Link

	Publication	Database / Coverage
<a href="#">View eBook</a>	My First Summer in the Sierra	Project Gutenberg E-Books

[Request from Special Collections](#)

University of Maryland, College Park

1 available

[Request Item](#)

UMCP HBK Maryland Room Special Collections Stacks

✓ Available (1)

Call Number	Status
QH31.M9 A34 1911	✓ Non-Circulating/On Shelf

USMAI Libraries

1 available

Loyola Notre Dame Library Stacks

✓ Available (1)

Call Number	Status
F868.S5 M9	✓ On Shelf

Libraries Worldwide

417 Libraries

[Request Item through Interlibrary Loan](#)

## Editions and Formats list



my first summer in the sierra

### 1. My first summer in the Sierra

by [John Muir](#)

 Microform 1911 | [View all editions & formats](#)

Held by: USMAI Libraries

### 2. My first summer in the Sierra

by [John Muir](#), [Herbert Wendell Gleason](#)

 Print book 1911 | [View all editions & formats](#)

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](#)

Held by: [University of Maryland, College Park](#)

✓ Non-Circulating/On Shelf  
UMCP HBK Maryland Room  
QH31.M9 A34 1911

### 3. My first summer in the Sierra

by [John Muir](#)

 Print book [1990] | [View all editions & formats](#)

Expresses the transcendent joy, the insatiable curiosity, the nascent sense of mission of a man discovering his passion and his destiny.

Held by: Big Ten Academic Alliance

### 4. My first summer in the Sierra

by [John Muir](#)

 Print book 1987 | [View all editions & formats](#)

Held by: Big Ten Academic Alliance

### 5. My first summer in the Sierra

by [John Muir](#)

### My first summer in the Sierra

1. [My first summer in the Sierra](#)  
by [John Muir](#)

 Print book 1988 Yolla Bolly Press limited ed., xii, 162 pages : illustrations ; 33 cm English, San Francisco, CA : Sierra Club Books, 1988.

Held by:  
University of Maryland, College Park and 122 Libraries Worldwide

2. [My first summer in the Sierra](#)  
by [John Muir](#)

 Print book 1979, ©1916  
x, 271 pages, [23] leaves of plates : illustrations ; 21 cm English, Boston : Houghton Mifflin, 1979, ©1916.

Held by:  
University of Maryland, College Park and 231 Libraries Worldwide

3. [My first summer in the Sierra](#)  
by [John Muir](#)

 eBook 1972 [©1911] 1 online resource (353 pages) illustrations English, Dunwoody, Ga., N.S. Berg, 1972 [©1911]

Held by:  
University of Maryland, College Park and 15 Libraries Worldwide

4. [My first summer in the Sierra](#)  
by [John Muir](#)

 eBook ©1916 Manuscript ed., 1 online resource (x, 271 pages, [22] leaves of plates) : illustrations, map. English, Boston : Houghton Mifflin, ©1916.

Held by:  
University of Maryland, College Park and 18 Libraries Worldwide

5. [My first summer in the Sierra](#)  
by [John Muir](#), [Herbert Wendell Gleason](#)

 Print book 1911  
vii, 353 pages, 12 unnumbered leaves of plates : illustrations ; 21 cm English,  
Boston and New York : Houghton Mifflin Company ; Cambridge : The Riverside Press, 1911.

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 screenshot shows the OCLC interface for a search of "Alice in Wonderland". The search results are filtered to "All Editions" (16 Found) for the Pepperdine University library. A callout box highlights a note: "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
<a href="#">Alice in Wonderland</a> 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
<a href="#">Alice's adventures in Wonderland ; and, Through the looking-glass and what Alice found there</a> 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
<a href="#">Alice's adventures in Wonderland ; and, Through the looking-glass and what Alice found there</a> 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
<a href="#">Alice's adventures in Wonderland ; and, Through the looking-glass and what Alice found there</a> 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.

Recent changes have improved the appropriateness of the default edition and the visibility of other editions:

- The default edition is now the most recent locally-held edition
- The “View all editions & formats” link now appears by default.
- A “Group related editions” check box enables searchers to see every edition immediately in the search results.

Still, some design decisions that remain can impair the experience of some common cases:

- 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 [customer support](#) in your region.