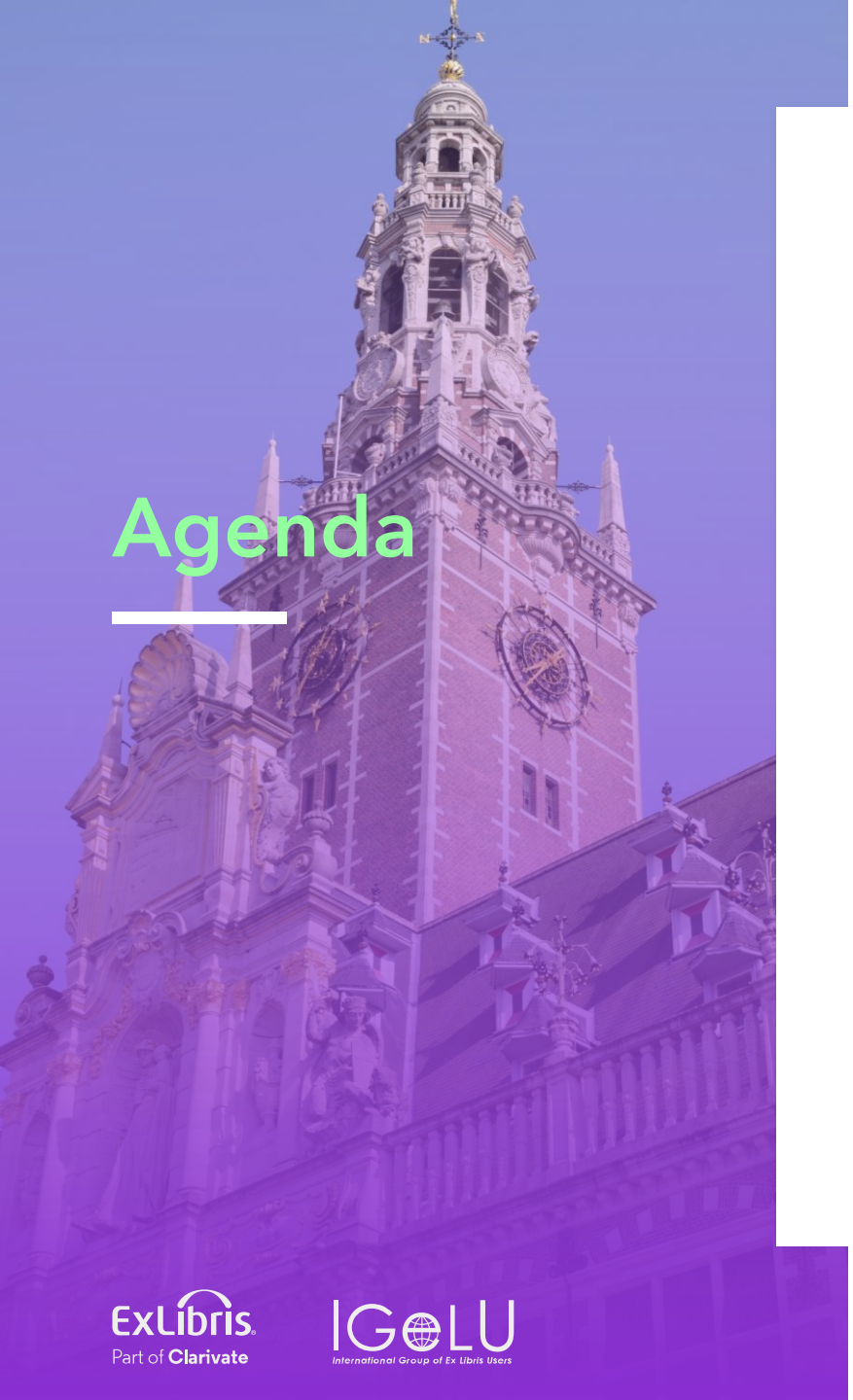
ExLibris
Part of Clarivate
IGeLU
International Group of Ex Libris Users
Achieving Better Content:
Leading Metadata to Higher Standards
Tamar Ganor | Content Product Manager
Christine Stohn | Discovery Product Management



Agenda

- The Content Mountain – a story of growth
- Growth challenges
- Keeping up with the content
- Data improvements in CDI to enhance discoverability and user experience
- Data flows

Ex Libris Part of Clarivate

IGeLU International Group of Ex Libris Users

© 2023 Clarivate

Ex Libris
Part of Clarivate

IGeLU
International Group of Ex Libris Users

The Content Mountain

Depth, Breadth and Perspectives

The Content Mountain

12M

Bib records in Alma

40,000

Collections in KBs

150

Languages with 45 Authority vocabularies

5.2B

Records in CDI:

- 964m articles
- 830m books/eBooks
- 148m patents

© 2023 Clarivate

Let's talk metadata

Highly Activated

44 collections with over 1,000 activations

Local Representations

1,667 local collections in CDI

Great Variety

8,408 collections with fewer that 10 activations

Content Providers

Over 5,000 in CDI

Open Access

333M OA items in CDI

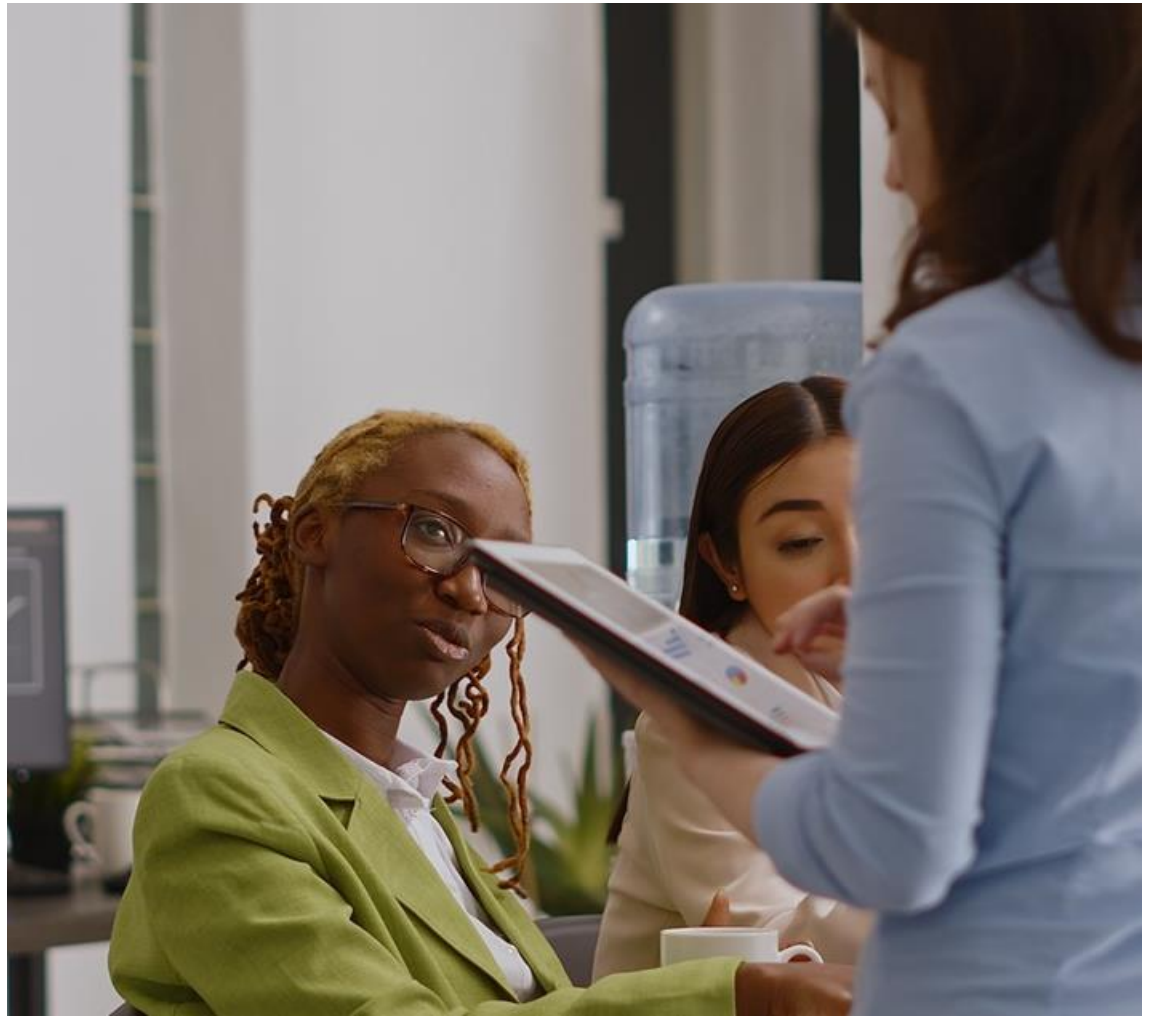
Authority Vocabularies

57.1M Authority records in Alma CZ

© 2023 Clarivate

Growth Originated Challenges

- Staying current
- Metadata quality



© 2023 Clarivate

Ex Libris
Part of Clarivate

IGeLU
International Group of Ex Libris Users

The Way to Data Excellence

Data Excellence

Capacity, Efficiency, Speed and Quality

Quicker Updates
Increase update frequency of Alma

Better Quality
Improve and expand MARC metadata in Alma CZ

Improved Discoverability
Enhance linking and rights for CDI users

Transparent communication and collaboration with community and providers

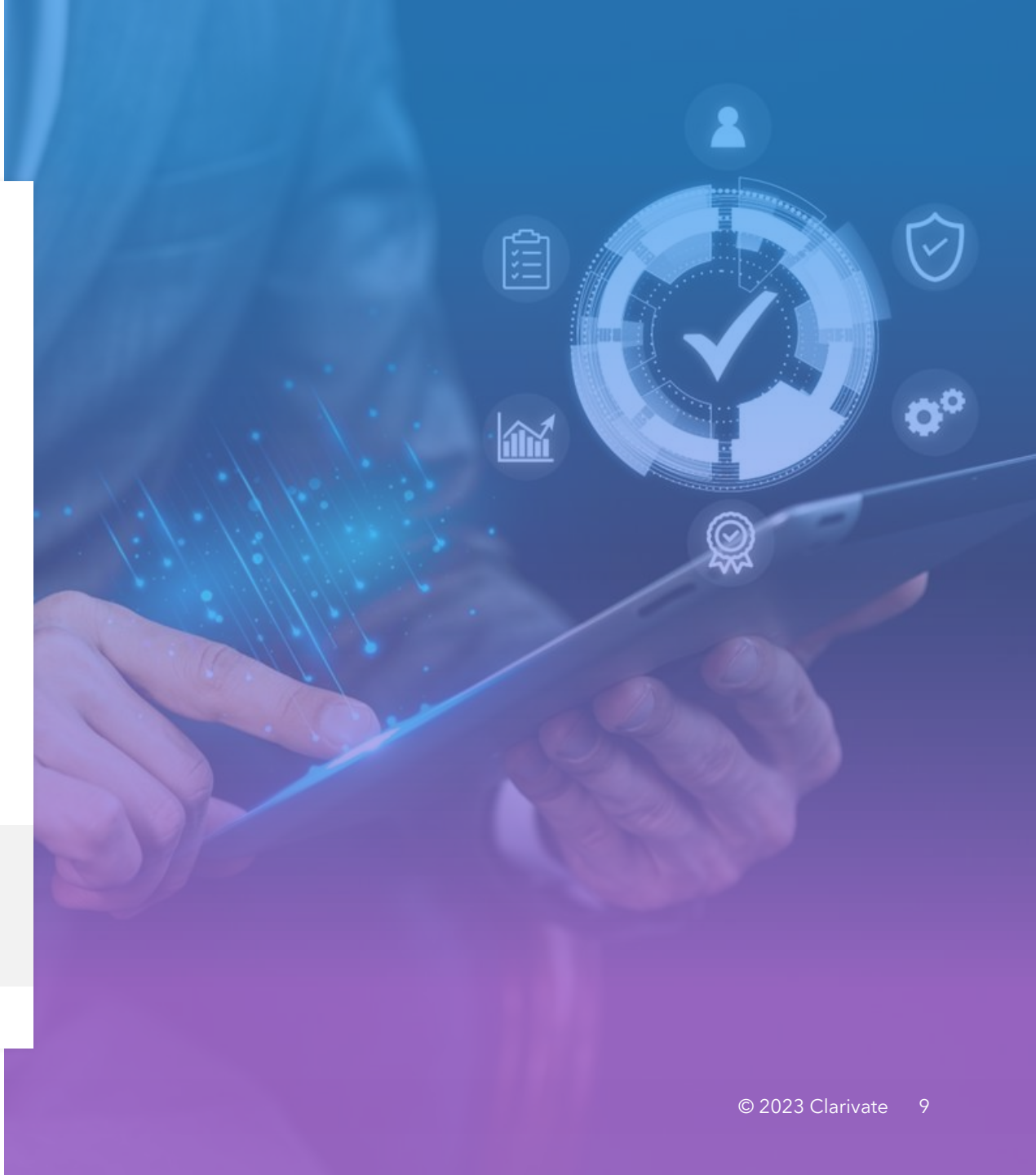
© 2023 Clarivate

Data Excellence: Updates Challenge

Daily Ingest

- Reduce turnaround time to daily updates
- 20 million portfolios affected
- ~165,000 activations
- By end of 2023:
  70% of automated CZ portfolios will be updated daily

Create automation, improve ingest process, clean up collections

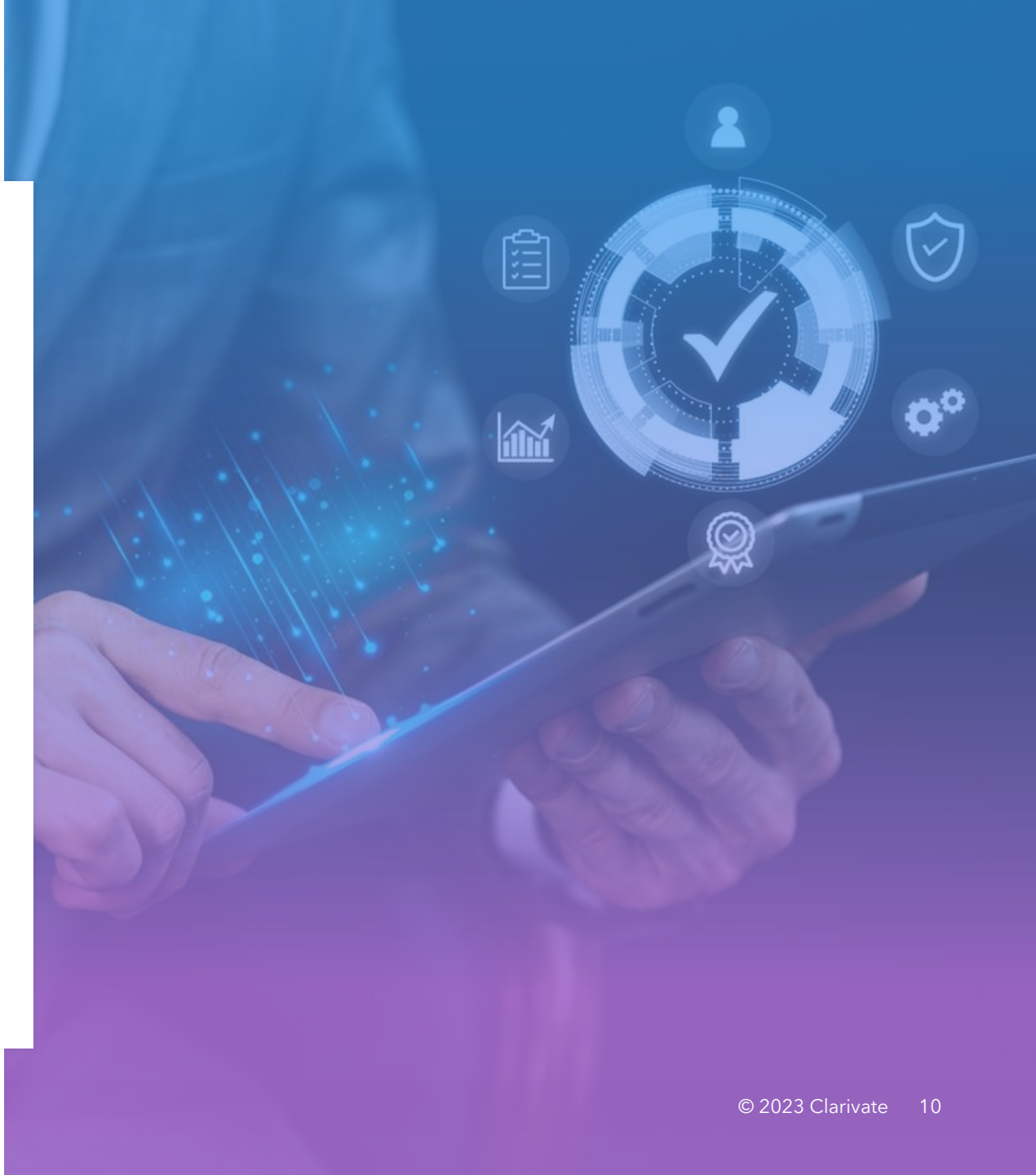


© 2023 Clarivate

Data Excellence: Quality Challenge

Better metadata through various sources

1. Selecting highly activated eBook collections:
   - So far over 153K Bib records enriched
   - Planning to enrich ~200K Bib records by end of 2023
   - Over 20,500 activations



© 2023 Clarivate

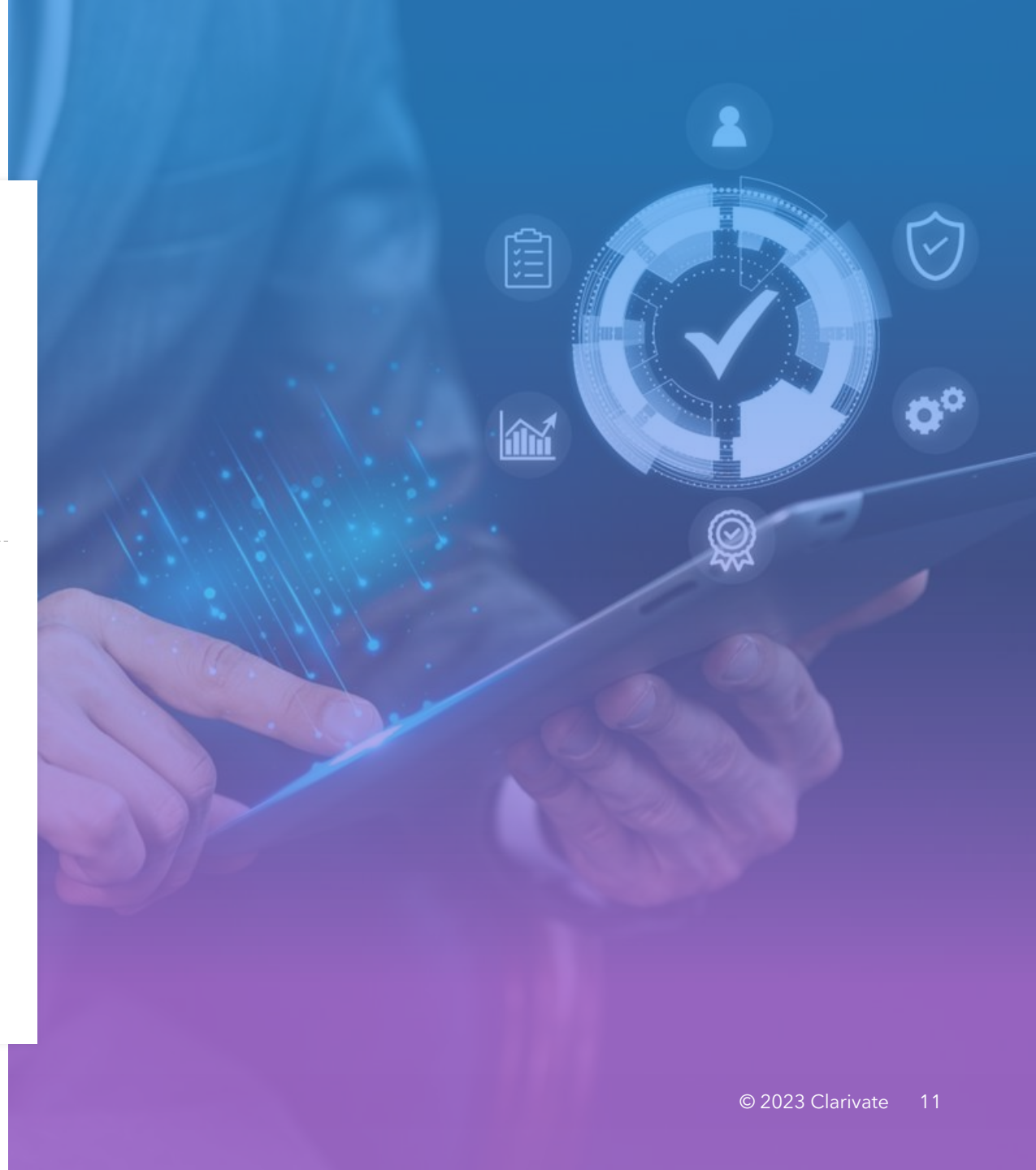
Data Excellence: Quality Challenge

2 Cooperating with providers

- Reviewing available metadata
- Increase capacity of Provider Relations team

3 Alternative sources enrichment

- LoC print records
- Project Gutenberg
- Future prospect - AI



© 2023 Clarivate

Ex Libris
Part of Clarivate

IGeLU
International Group of Ex Libris Users

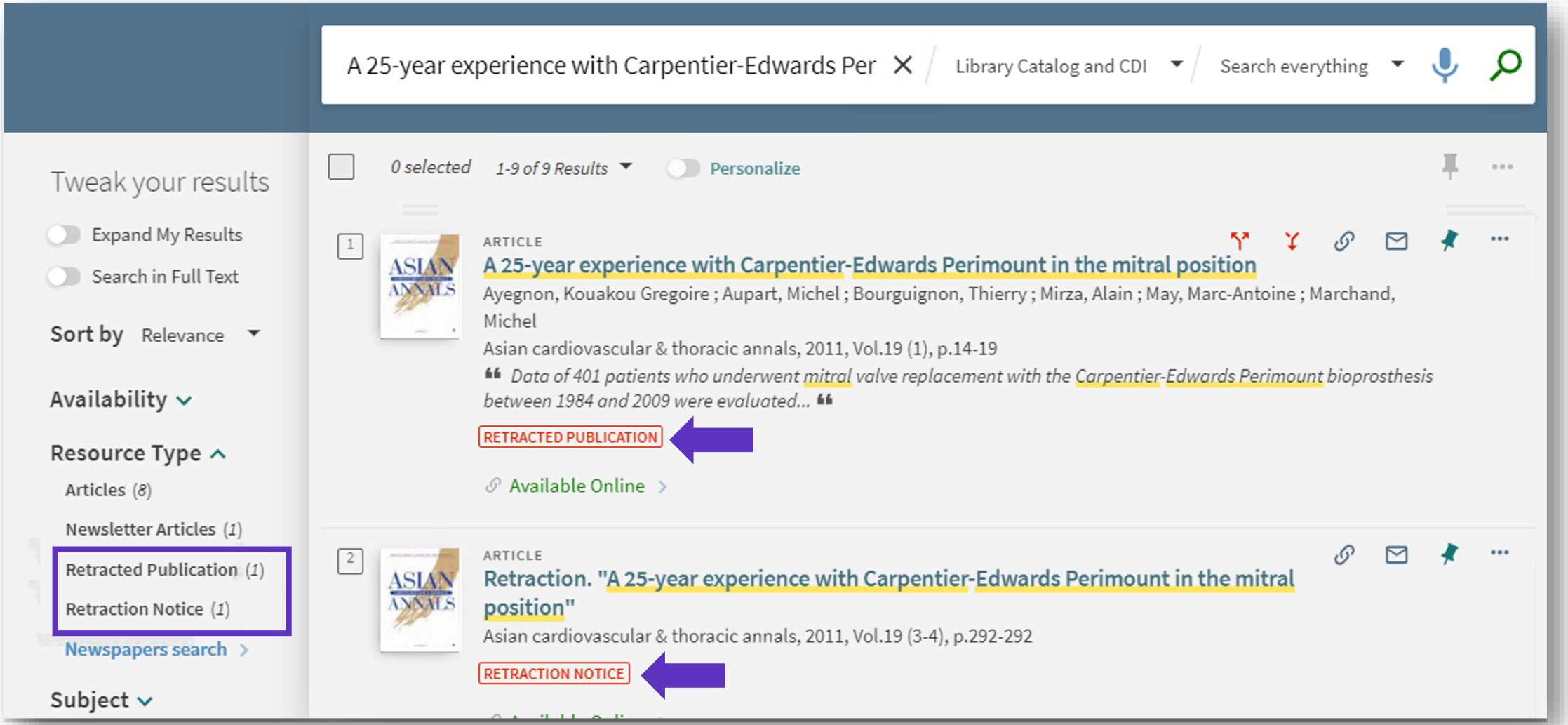
Data Improvements in CDI to Enhance Discoverability and User Experience

Document Attributes (Secondary Resource Types)

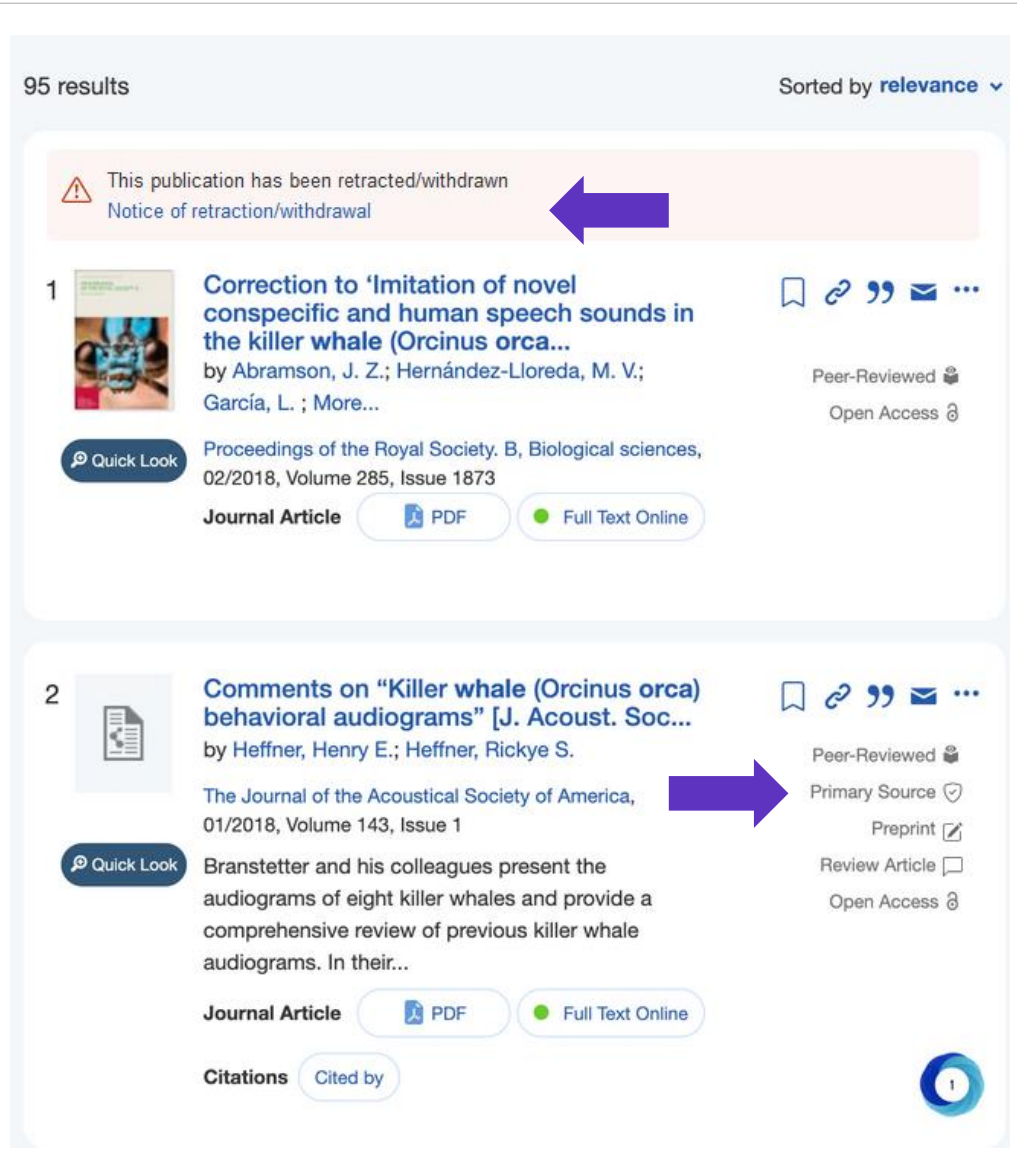
Definition	• **Attributes** associated with a record to provide more information • Does not replace but **adds value to** the existing 'primary' resource type
Purpose	• **Display additional information** about the record/item (display flag) • Used in search and as filters (facets)
Priority 2023	• *Review article* • *Primary source* • *Preprint* • *Retracted publication* • *Retraction notice* • *Withdrawn publication* • *Withdrawal notice/ item withdrawal*

© 2023 Clarivate

© 2023 Clarivate



© 2023 Clarivate



Documentation is available on the Knowledge Center

https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Documentation_and_Training/Documentation_and_Training_(English)/CDI_-_The_Central_Discovery_Index/CDI_Document_Attributes

© 2023 Clarivate

OA Handling for Hybrid (and other) OA Articles

Handling of hybrid Open Access articles in CDI

- DBID level collections can be activated for some publishers (where we have reliable flags on article level), including SAGE, Emerald, OUP, IEEE, Taylor & Francis, Springer (more will be added)

NEW: we are indexing Unpaywall

- Can be activated as DBID level database
- Can be used as Link in Record resource or/and in combination with the API
- **Available in November**

© 2023 Clarivate

Other CDI Data Enhancements in 2023

- **New resource types** (for Primo): magazines, magazine articles, questionnaire

  <add release date and affected collections>

- Improve consistency of **language** and **publication date** fields between Index and display

- Better identification of **book reviews** by a set of rules

© 2023 Clarivate

Subject Heading Normalization Project

Three Fields

- **Subject:** populated from controlled vocabulary, deduplicated and normalized
- **Keyword:** all subjects/keywords that cannot be mapped to the subject field
- **LATER Discipline:** all discipline level subjects, according to a controlled list

Iteration 1

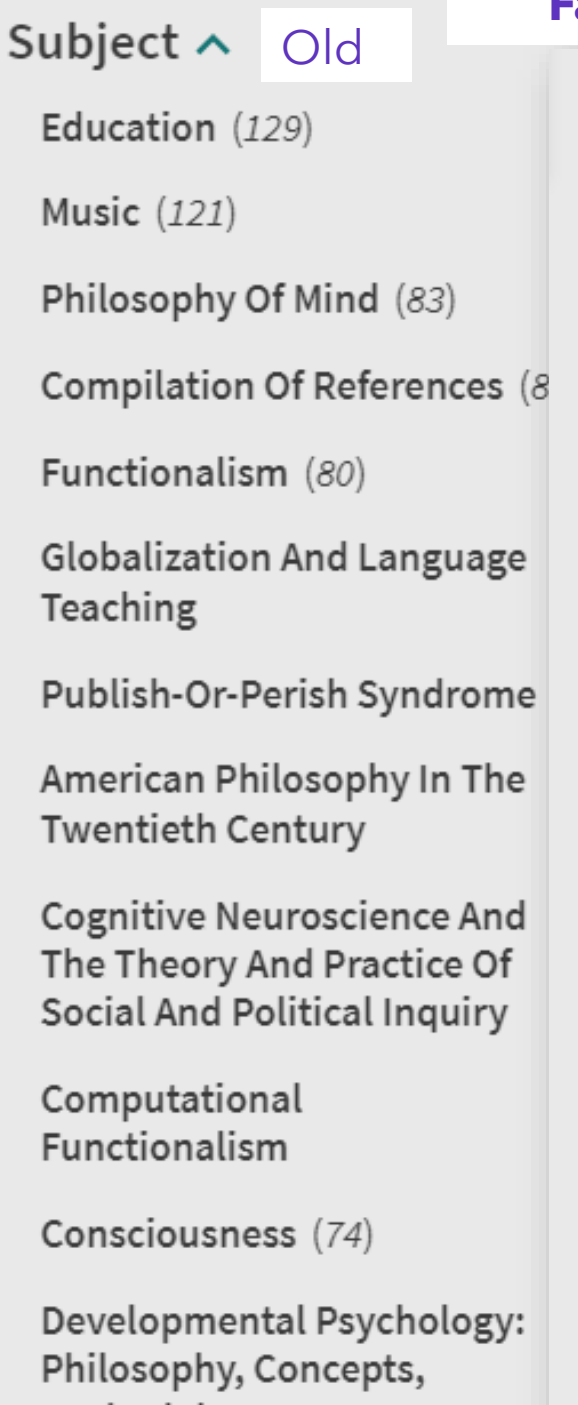
- Subject and keyword field
- LCSH, MeSH, and some additions from PQ thesauri
- Optional: opt-in by activating it – **November** (Summon), **February** (Primo)
- Test option (for the facets) available for Primo customers **&searchCDI=normalizedSubjects**

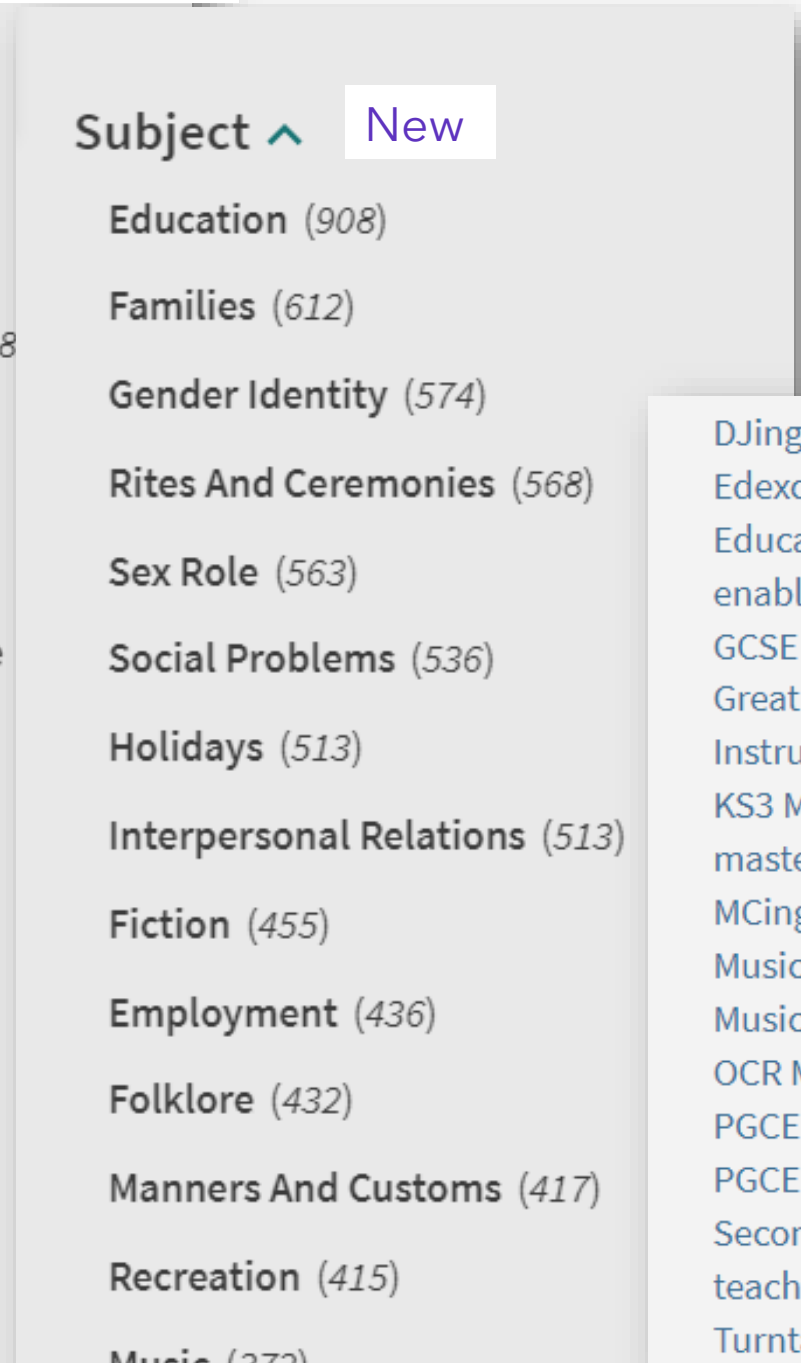
© 2023 Clarivate

Facet

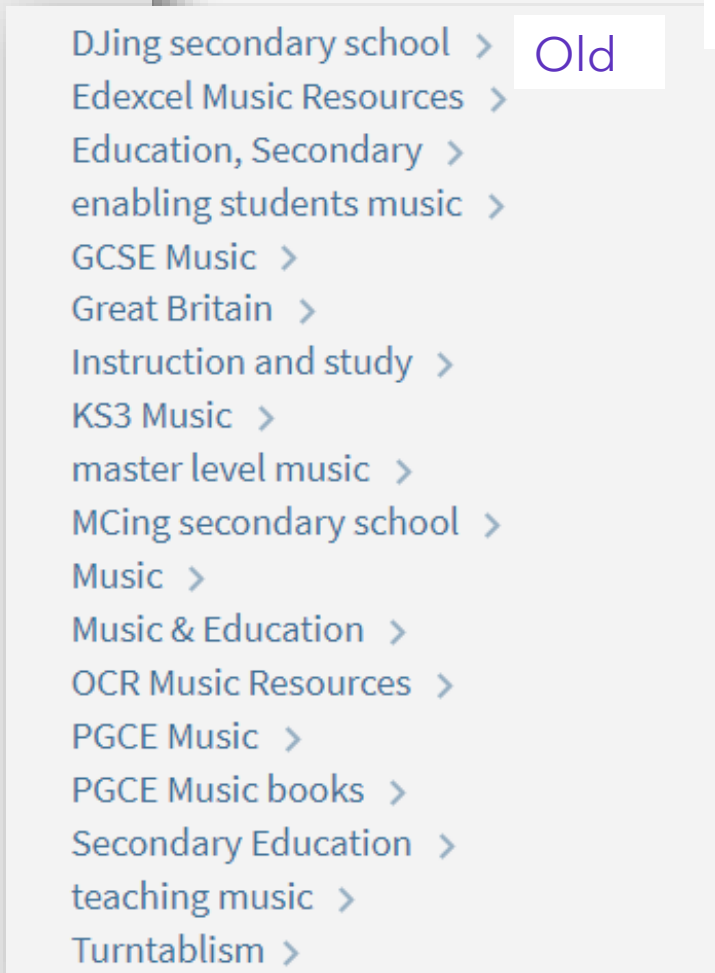
&searchCDI=normalizedSubjects

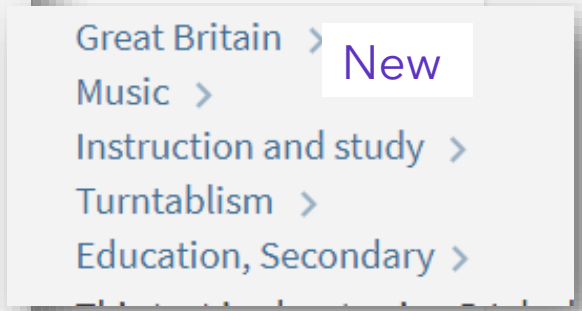
- Test facets and subject field in full record
- No key word field is displayed with the test option





Full view





All other subjects are going into the keyword field (visible in the PNX)

© 2023 Clarivate

Another Important Data Project

Author identification

- Enrich CDI with **ORCIDs** and connect to author names where available, to identify authors better

In future

- Use additional sources for author identification

Both the subject and the author project help us to move towards a Linked Data Network in CDI

© 2023 Clarivate

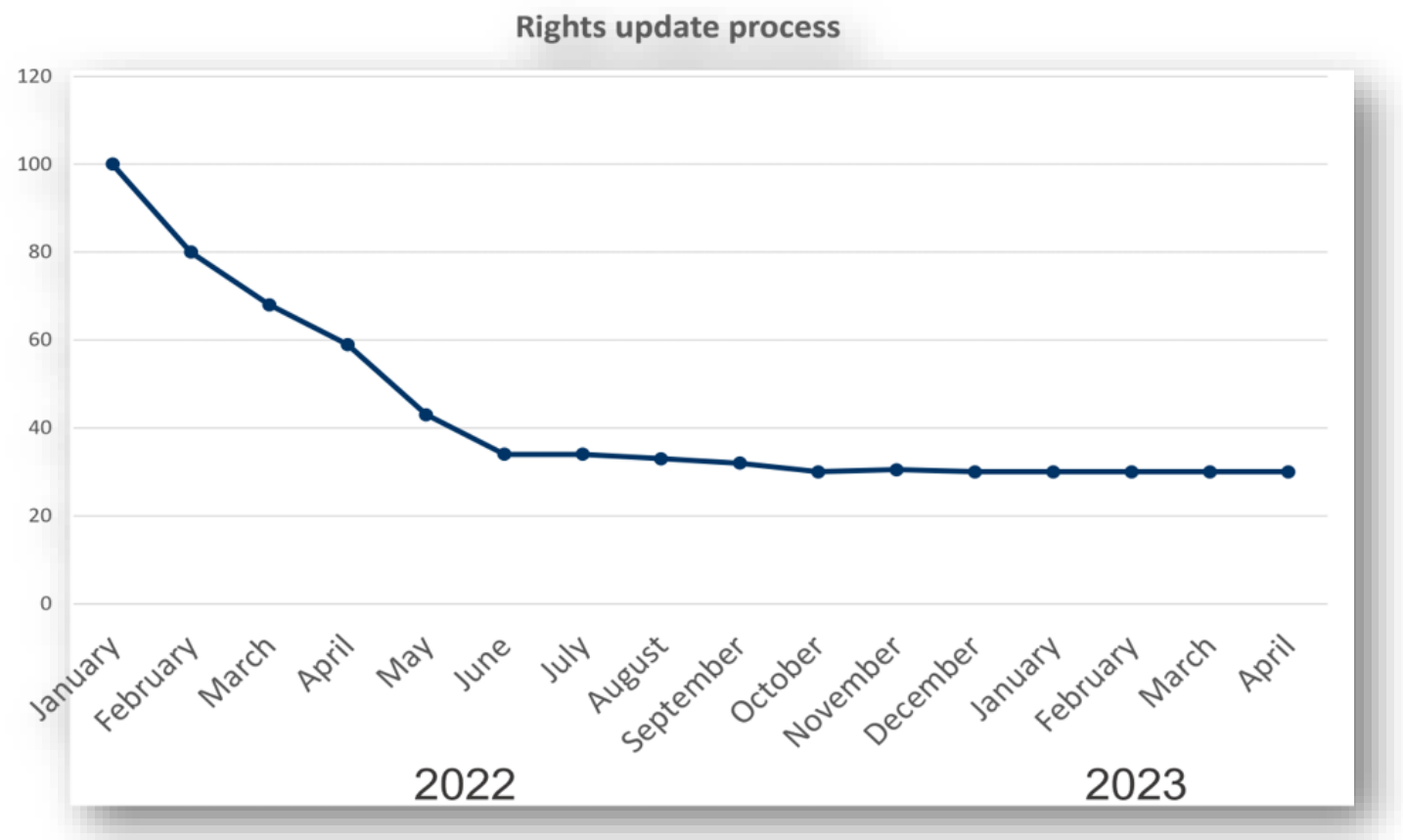
Ex Libris
Part of Clarivate

IGeLU
International Group of Ex Libris Users

Data Excellence for Rights and Access

Quicker Availability with Faster Rights Processing

From 100 hours to an average of 30 hours to process activation to discovery



© 2023 Clarivate

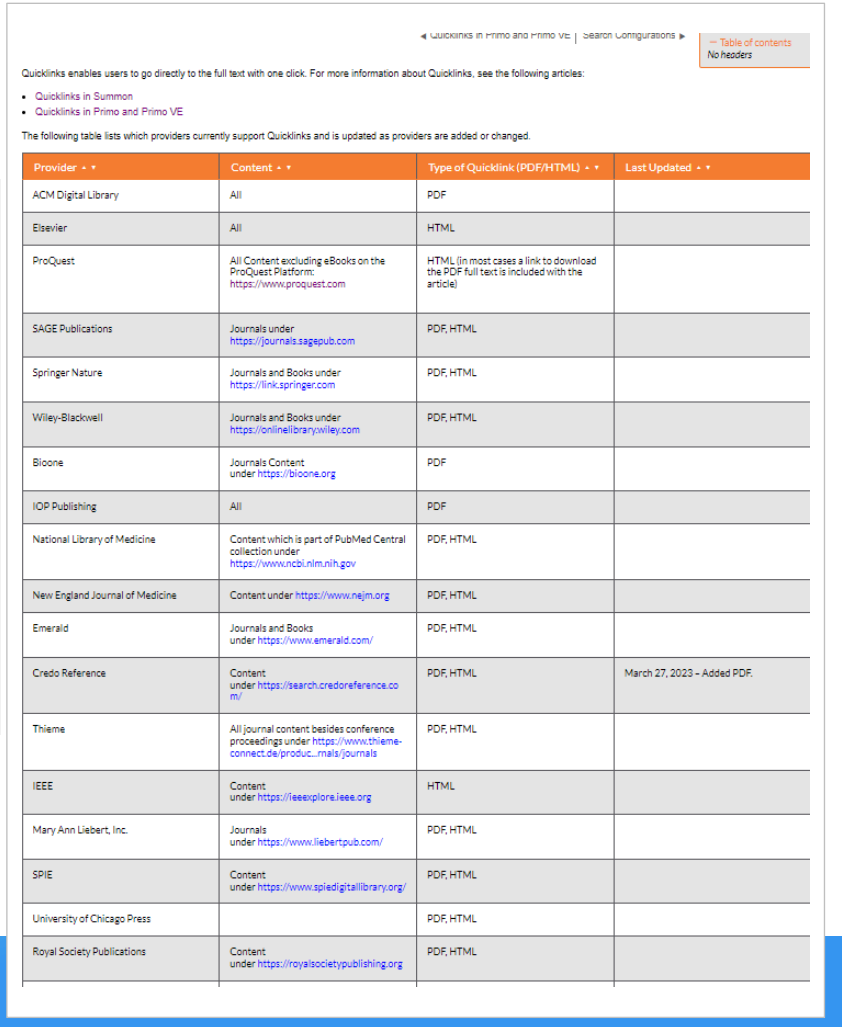
Fulltext Access Enhanced: Linking Improvements

- Increased use of document ID for linking accuracy
- Focus on Quicklinks

- Now with 37 supported providers
- Over **16,304** collections with **1.56 billion** records
- Linking success rates 99%

- Over 400 institutions currently active or testing
- More than 1.3m clicks in March 2023

More providers coming in 2023



◂ Quicklinks in Primo and Primo VE | Search Configurations ▸

Table of contents
No headers

Quicklinks enables users to go directly to the full text with one click. For more information about Quicklinks, see the following articles:

- Quicklinks in Summon
- Quicklinks in Primo and Primo VE

The following table lists which providers currently support Quicklinks and is updated as providers are added or changed.

Provider	Content	Type of Quicklink (PDF/HTML)	Last Updated
ACM Digital Library	All	PDF	
Elsevier	All	HTML	
ProQuest	All Content excluding eBooks on the ProQuest Platform: https://www.proquest.com	HTML (in most cases a link to download the PDF full text is included with the article)	
SAGE Publications	Journals under https://journals.sagepub.com	PDF, HTML	
Springer Nature	Journals and Books under https://link.springer.com	PDF, HTML	
Wiley-Blackwell	Journals and Books under https://onlinelibrary.wiley.com	PDF, HTML	
Bioone	Journals Content under https://bioone.org	PDF	
IOP Publishing	All	PDF	
National Library of Medicine	Content which is part of PubMed Central collection under https://www.ncbi.nlm.nih.gov	PDF, HTML	
New England Journal of Medicine	Content under https://www.nejm.org	PDF, HTML	
Emerald	Journals and Books under https://www.emerald.com/	PDF, HTML	
Credo Reference	Content under https://search.credoreference.com/	PDF, HTML	March 27, 2023 – Added PDF.
Thieme	All journal content besides conference proceedings under https://www.thieme-connect.de/produc...rnals/journals	PDF, HTML	
IEEE	Content under https://ieeexplore.ieee.org	HTML	
Mary Ann Liebert, Inc.	Journals under https://www.liebertpub.com/	PDF, HTML	
SPIE	Content under https://www.spiedigitallibrary.org/	PDF, HTML	
University of Chicago Press		PDF, HTML	
Royal Society Publications	Content under https://royalsocietypublishing.org	PDF, HTML	

© 2023 Clarivate

The Data Flow

Where is the Data Coming from… and Where Does it Go?

Providers

Through the Products

End-User Functionality

© 2023 Clarivate

In Summary

- **Data Excellence** (DX) is the main initiative for Ex Libris to achieve better content
- We are leading metadata to higher standards using **internal and external solutions**
- Pushing selective efforts to improve **quality**, get quicker **updates** and enhance **availability**
- **Community + Providers + Ex Libris**

© 2023 Clarivate

Please join the Q3 content webinar, for a discussion with Sage and Oxford representatives.

Metadata at its Origin –

Publishers Talk about Metadata

Thursday, September 28th

15:00 GMT

*Feel free to send your questions by September 22nd

© 2023 Clarivate

Not yet registered for the content mailing list??

Don’t miss out!!

© 2023 Clarivate

© 2023 Clarivate
Clarivate and its logo, as well as all other trademarks used herein are trademarks of their respective owners and used under license.