

# Share-VDE and the Share Family Advancements towards production

BIBFRAME Workshop in Europe 2023 19th september 2023

Jim Hahn (University of Pennsylvania Libraries) Tiziana Possemato (@Cult and Casalini Libri) Anna Lionetti (Casalini Libri)



# Share Family principles, processes and stepping stones



# Share Family - Linked Data Ecosystem: Principles



### **CO-OPERATIVE**

Developed and driven by libraries, for libraries, the Share Family is a growing international community built on collaboration. Participating institutions play an active role in defining the vision, aims and progress of the Share Family and its tools.

The Share Family opens the door to a flexible, sustainable, interoperable and co-operative approach to bibliographic data, with time, expertise and costs shared across the community for the benefit of all members.





### **INTEROPERABLE**

By implementing the BIBFRAME data model and facilitating interoperability with different data models and data pools, bibliographic information can transformed into Linked Data, increasing the visibility of research and encouraging greater engagement with library, archive and museum collections.

We strive to encourage open access to data, and support diversity by freely sharing information. We apply and support open metadata policies as part of our commitment to enhancing the discovery of library and cultural heritage resources.





### **FLEXIBLE**

Enriched and structured data can be re-used in local and external systems, across library types and ILS/LSPs, enabling each institution to maintain control of its own catalogue data.

The quality of data is guaranteed both through advanced technical processes and through collaborative data modeling, enrichment and sharing, handled collectively by member organizations.





# Share Family - Linked Data Ecosystem: Processes

# DATA INPUT FROM INSTITUTIONS

MARC21 bib. and holding

MARC21 aut.

**UNIMARC** 

RDF/BIBFRAME

Other formats (eg. FOLIO)

#### LOD PLATFORM SERVICES

### **Mapping**

**URI Enrichment** 

Reconciliation

Creation of Linked Data Entities

Conversion to RDF/BIBFRAME

#### **RESULTS**

### Data Publication

- End user discovery portals for each initiative of the Share Family
- Search APIs for data consumption (GraphQL, REST, SPARQL)

### + Data Distribution

- API / protocols for third parties integration (eg. local LSPs and data editor such as Wikidata, Sinopia BIBFRAME editor, FOLIO etc.)
- Downloadable datasets (BIBFRAME/RDF, MARC enriched records)



### Shared Data Management

- Editing of Share Family entities with JCricket
- Reuse of Share Family BIBFRAME data in local library systems
- Reuse of Share Family MARC representations in local library systems

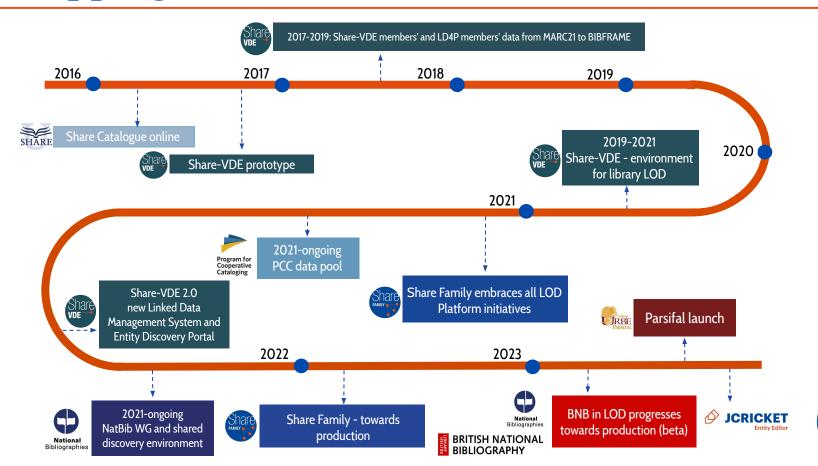


### Additional Services

- Authority control in MARC and BIBFRAME-based workflows
- SFI Share Family Index: registry of entity URIs



# Stepping stones





# The Share Family Linked Data Ecosystem



The Share Family is a global community built on collaboration that brings together libraries, archives, museums, consortia and Library Service Platforms (LSP) and joins their knowledge in an ever-widening network of interconnected bibliographic data.

# Share-VDE - Virtual Discovery Environment



https://www.svde.org

**Berkeley Law Library** 

**Duke University** 

Library of Congress

National Library of Finland

National Library of Norway

**New York University** 

Smithsonian Libraries and Archives

**Stanford University** 

University of Alberta / NEOS Library Consortium

University of Chicago

University of Michigan Ann Arbor

University of Pennsylvania

Vanderbilt University

Yale University



### Share Catalogue: Scholarly Heritage and Access to Research



Share Catalogue discovery portal

Università degli Studi di Napoli Federico II

Università degli Studi della Basilicata

Università degli Studi del Sannio

Università degli Studi di Salerno

Università degli Studi di Napoli Parthenope

Università degli Studi del Salento

Università degli Studi di Napoli L'Orientale

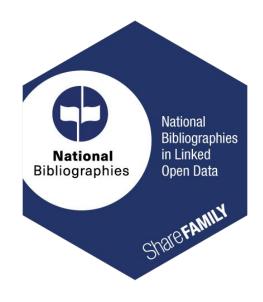
Università degli studi della Campania Luigi Vanvitelli

Università degli Studi Suor Orsola Benincasa

Università degli Studi di Cassino



# National Bibliographies in Linked Open Data



https://natbib-lod.org/

The aggregation of data from National Bibliographies in a shared entity discovery environment; the first of these is the BNB - British National Bibliography, soon to go into production.

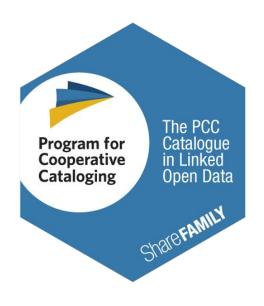


The preview of the BNB beta website is available at

https://bl.natbib-lod.org/



# PCC Catalogue in Linked Open Data



The Share Family hosts a dedicated tenant for the data of the PCC - Program for Cooperative Cataloging, to provide PCC-quality BIBFRAME data housed in an ad hoc data pool

https://pcc-lod.org/



# Parsifal - Integrated Catalogue in Linked Open Data



https://parsifal.urbe.it/parsifal/?l=en

Accademia Alfonsiana

Centro Pro Unione

Pontificia Facoltà di Scienze dell'Educazione "Auxilium"

Pontificia Facoltà Teologica "Marianum"

Pontificia Università Antonianum

Pontificia Università della Santa Croce

Pontificia Università di San Tommaso d'Aquino (Angelicum)

Pontificia Università Gregoriana

Pontificia Università Lateranense

Pontificia Università Urbaniana

Pontificio Ateneo Sant'Anselmo

Pontificio Istituto Biblico

Pontificio Istituto Orientale

Pontificio Istituto Teologico "Giovanni Paolo II" per le Scienze del Matrimonio e della Famiglia

Pontificium Institutum Patristicum Augustinianum

Università Pontificia Salesiana



# Share Art, Share Music, Share MIA



Three pilot projects for shared Linked Open Data environments in the domains of Art, Music and Manuscripts, Incunabula and Ancient books

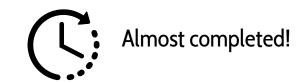


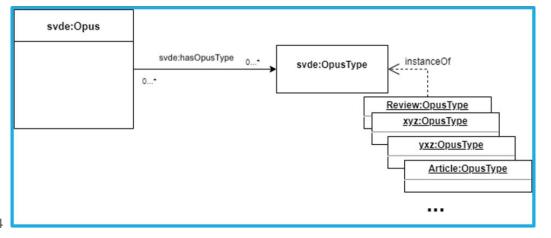
# The Share-VDE ontology



# The SVDE ontology - a BIBFRAME extension

The Share-VDE Ontology supports the Share Family of initiatives (based in federated linked data discovery environments) and is developed as an **extension to BIBFRAME**.







# Translate conceptual state into OWL

Share Family is using the BIBFRAME ontology with the extensions. As the Share-VDE initiative evolved the discovery environment, the entities were in a conceptual state. Translating the conceptual state into the ontology web language allowed the project to unambiguously define core concepts. Conceptual models are limited in defining core concepts with such precision.



# More in Jim Hahn's presentation



# Towards an operational environment



# An integrated and hybrid environment

The mutual exchanges in the BIBFRAME / linked data community are bringing the Share Family towards:

- an integrated, "hybrid" operational environment...
- ...based on a variety of tools and diverse data sources...
- ...including traditional tools (eg. new authority services for MARC workflows) as well as advanced models for data exchange



# An integrated and hybrid environment

DONE ONGOING STUDY PHASE

New authority services for MARC-based workflows - designed with SVDE AIMS working group and Stanford's input

Third parties integration with ILS/LSP - local library services (eg. Alma circulation APIs)

Finer granularity level of the CKB - Cluster Knowledge Base to make it format-agnostic and extend input data capabilities (MARC21, UNIMARC, BF/RDF eg. Sinopia profile etc.)

Third parties integration with ILS/LSP - Sinopia

Third parties integration with ILS/LSP - FOLIO

Third parties integration with authority sources:

- LD4P Questioning Authority
- Wikidata (initial specs by SVDE working groups)
- ISNI (initial specs by SVDE working groups)

Application of Share Family tech to other domains (Art, Music)



# Focus on Authority services

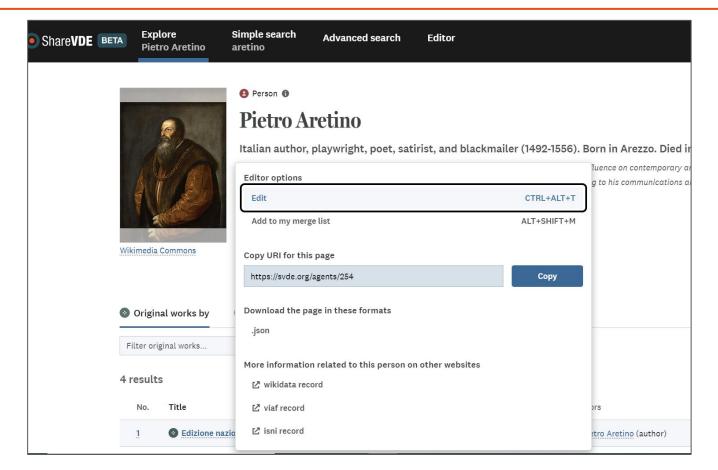
Automatic services for Share Family libraries, piloted by Stanford University:

- validation of MARC 21 bibliographic records (correction of MARC 21 fields and obsolete forms, update of tags and subfields etc.);
- enrichment of MARC 21 fields with SVDE original URIs and URIs from external sources according to ad hoc profiling, including LCNAF, VIAF, ISNI;
- matching processes on external authority files (LCNAF, LCSH, LCGFT, FAST);
- import of authority records from external authority files (LCNAF, LCSH, LCGFT, FAST);
- reporting features providing complete details of the validation and corrections done to the records.

Next step developments: Authority Services fully integrated in the Linked Open Data environments.

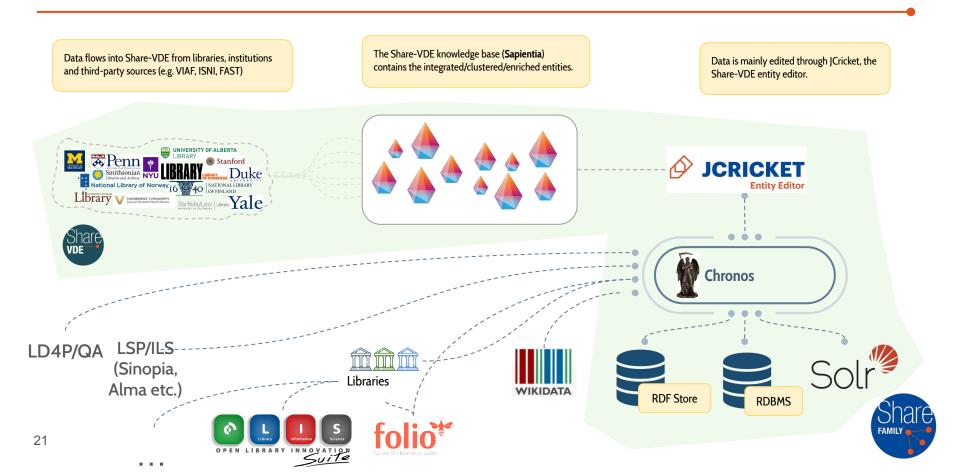


# Enhance data quality and authority control

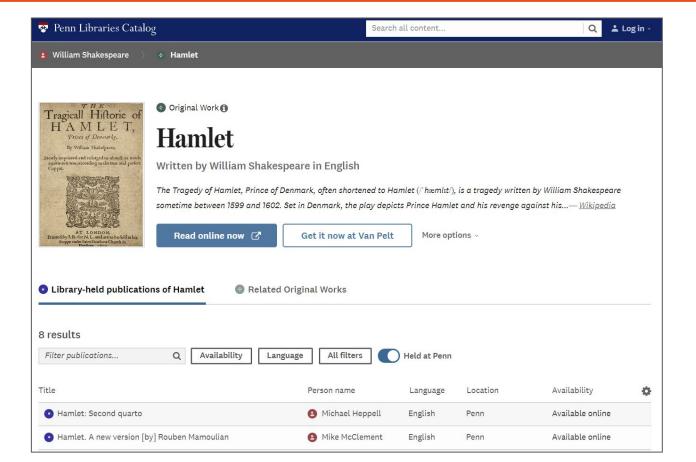




### Third party integration - Outbound Connectors Architecture



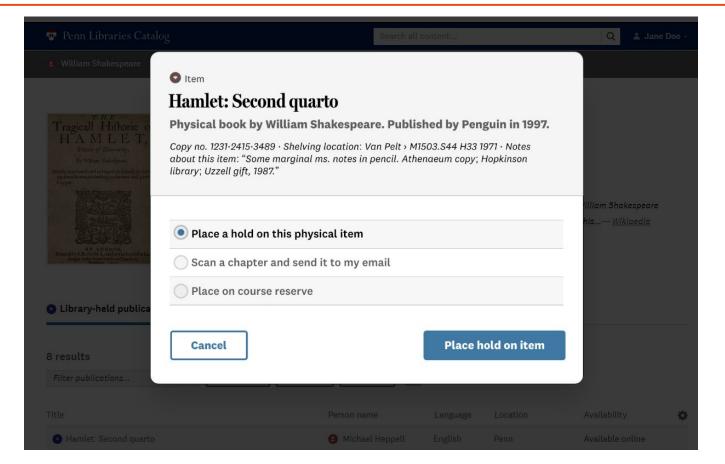
# Integration with local services - circulation info



Integration with local services, e.g. connection to Alma APIs for Penn circulation services



# Integration with local services - lending





# SINOPIA integration: high-level milestones

- set-up the connector to fetch data from Sinopia
- ingested subset of Sinopia data from Stanford
- now creating the parser so that RDF data coming from Sinopia can be clustered by Share-VDE processes
- ot the end of this process, Sinopia data will be included in the Share-VDE CKB Cluster Knowledge Base



# folio integration: high-level milestones

### Level 1: Instance correlation

- Folio inventory instances are retained in dedicated faces of Share-VDE prisms
- The inbound connector receives FOLIO data (instances) and feeds the Cluster Knowledge Base (CKB)
- The outbound connector communicates back data changes to FOLIO

### Level 2a: Agents (and works) correlation

Same interaction as above, but using authority records (agents, works), instead.

### Level 2b: JCricket UI App in FOLIO

Using the FOLIO built-in "pluggable" nature, the FOLIO UI SDK and the Share-VDE (GraphQL) API

# Application to other domains - art and music





Discussions are arising among the Share Family institutions and collaborative networks of libraries (such as consortia) about setting up new shared discovery environments tailored to the **art and music domains**. Some of the applications:

- Share-Art and Share-Music will be Share Family branches dedicated to the linked data management and discovery for institutions in the art and music domains;
- in other Share Family tenants there could be the option of filtering data about art and music and make it available through specific discovery portals.

# Application to other domains - Share MIA



Exploration of related areas to build an enhanced linked data ecosystem: pilot projects for shared Linked Open Data environments in the domain of **Manuscripts**, **Incunabula and Ancient books**, an example:





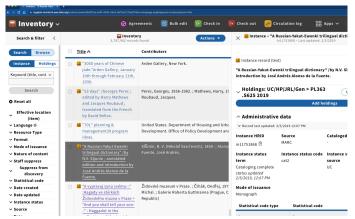
# JCricket Editor - The Entity Management System

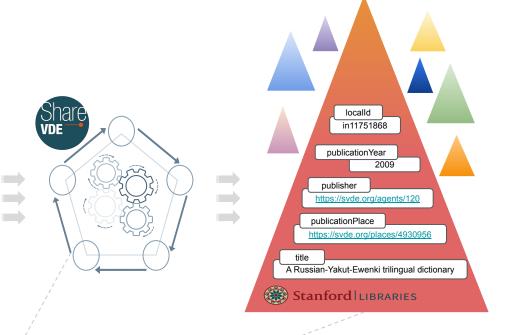


# From Library Data to Sapientia

A Share-VDE member (Stanford, in the example) uses FOLIO for managing its data.







FOLIO instance (or instances in case of massive export) is sent to Share-VDE

FOLIO instance data is split across the entities that form the Share-VDE domain model. In this example we focus on the properties that are assigned to a Share-VDE instance (red triangle above)

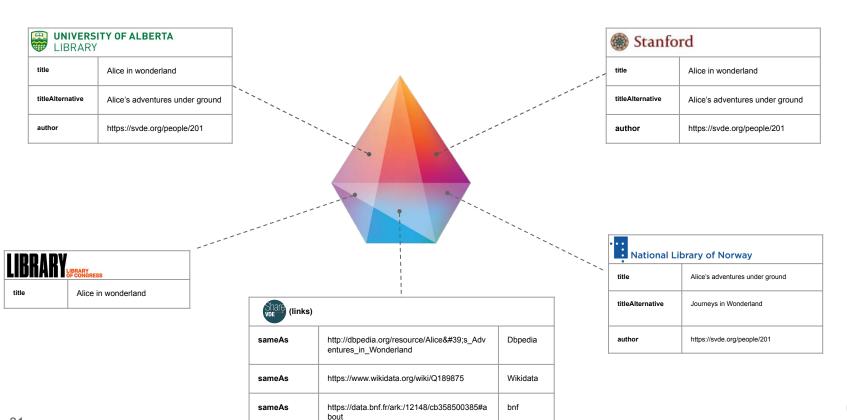


# Prism, faces: the Share-VDE Entity





# Faces (aka Contributions & Provenances)





# Properties: Attributes, Relationships, Links



Name	Value	Provenance
title	Alice in wonderland	LIBRARY Stanford
titleAlternative	Alice's adventures under ground	LIBRARY OF CONTINESS
titleAlternative	Journeys in Wonderland	National Library of Norway

An attribute is a data property, having a literal as value



A link is a connection between a Share-VDE Prism and an external reference

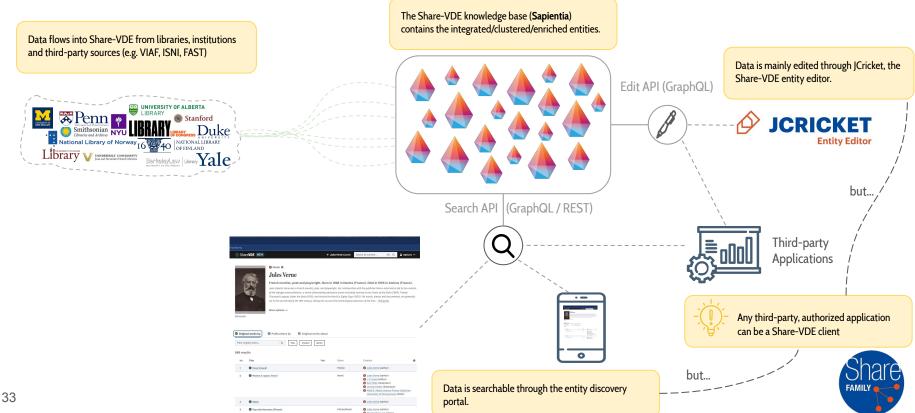
Name	Provenance	
author	LIBRARY LIBRARY ( Stanford	



Name	Value	Provenance
name	Carroll, Lewis	EIBRARY PERFECT
nameAlternative	Dodgson, Charles Lutwidge	LIBRARY MERCAL Stanford
nameAlternative	Karol, Luis	National Library of Norway



# The Big Picture: Genesis, Search, Edit



# JCricket: Available Operations

### Edit: a property is added/updated/deleted



Lowisse Carroll

Lewiss Carroll

is author of

nttps://syde.org/opuses/182734

https://svde.org/opuses/920302

### **Invalidate**







### Merge: multiple prisms are merged into one







For example, two prisms, "Mark Twain" and "Samuel Clemens", should be actually part of the same entity.

### **Split**: a prism is split into multiple prisms



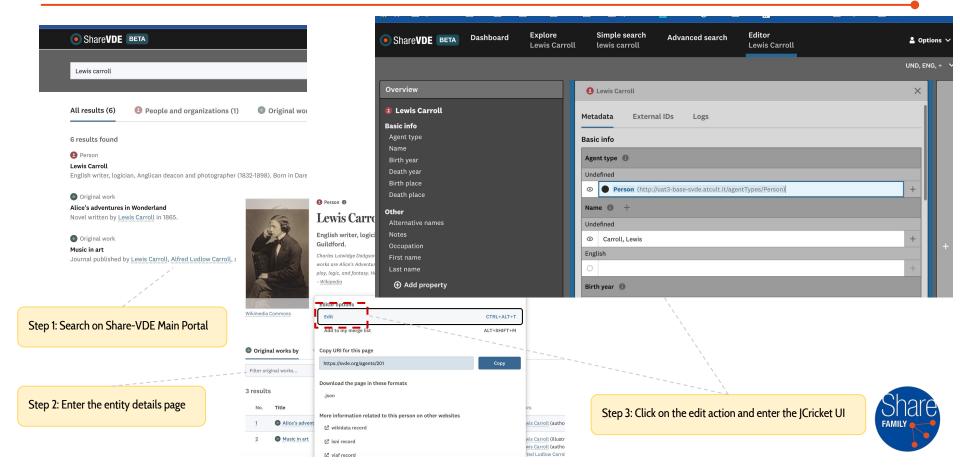




A prism (wrongly) contains information belonging to multiple entities (e.g., "Wallace David" and "David Wallace")



## JCricket user interface

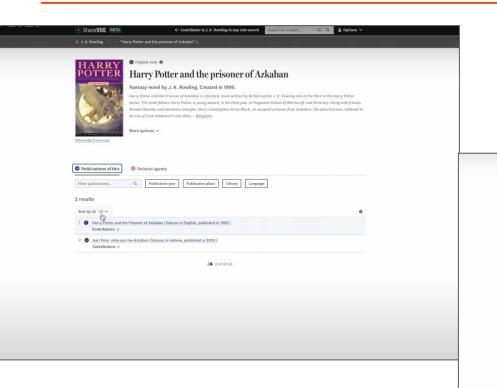


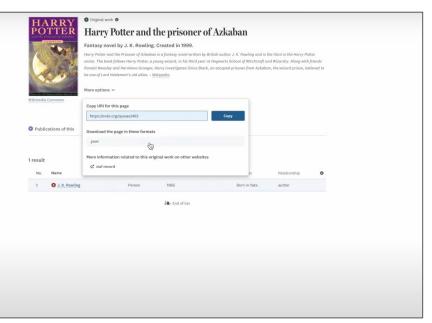


**ICricket demo** 

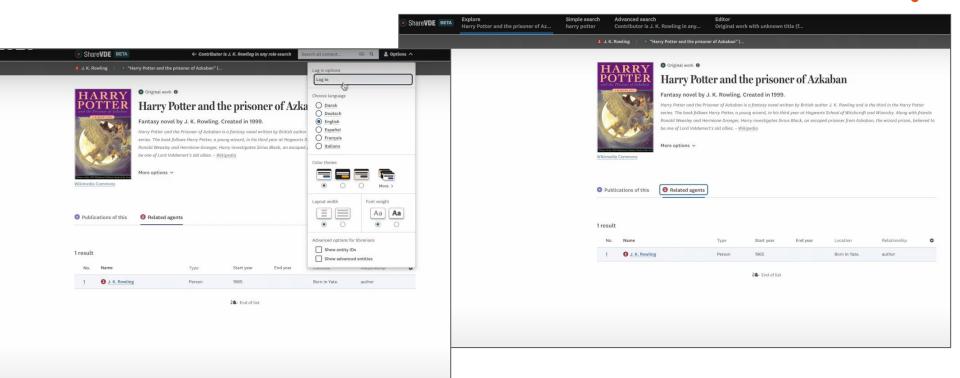


#### JCricket features - The user interface



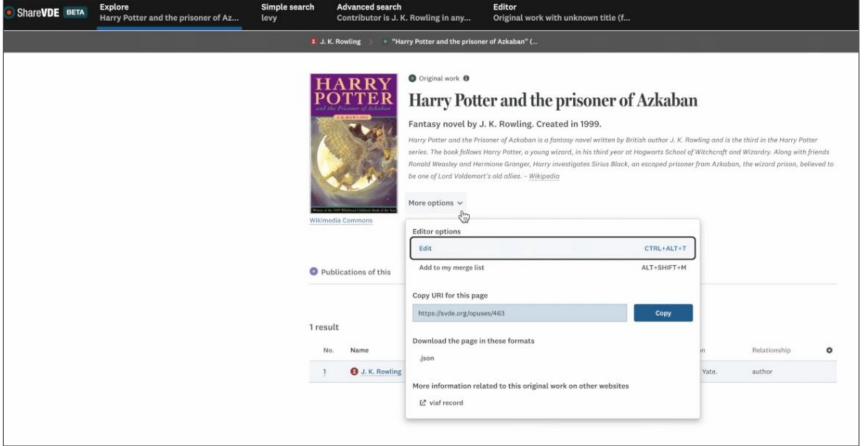


## JCricket log-in



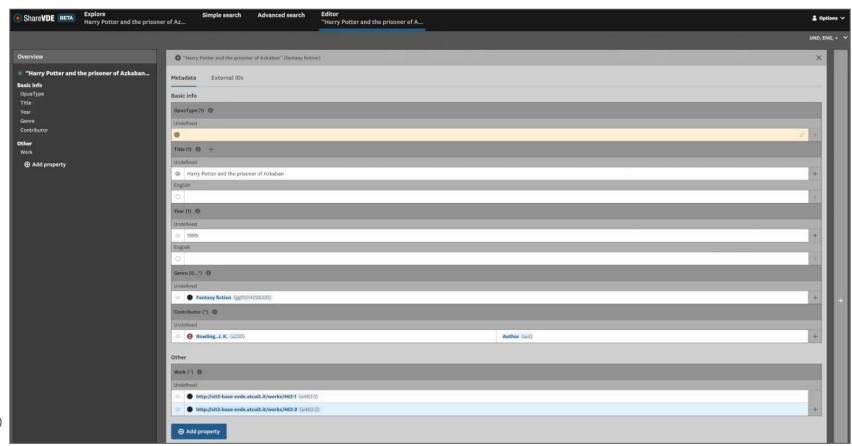


#### JCricket feature selection

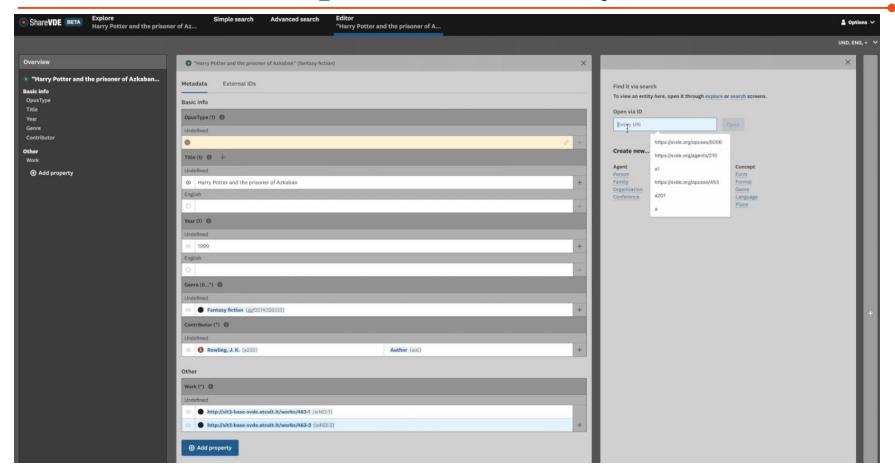




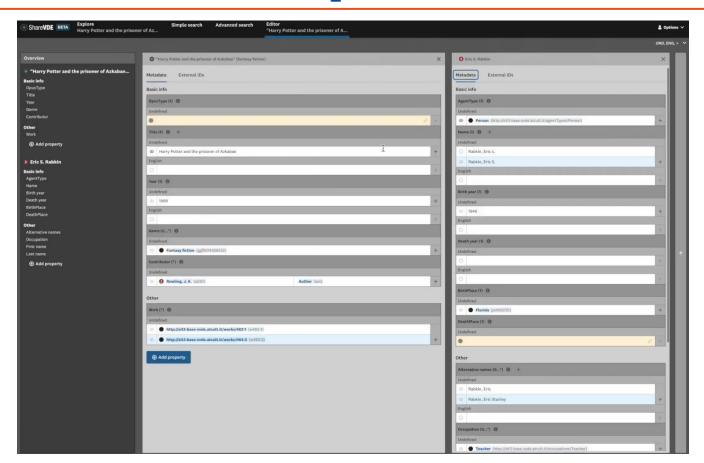
#### JCricket edit



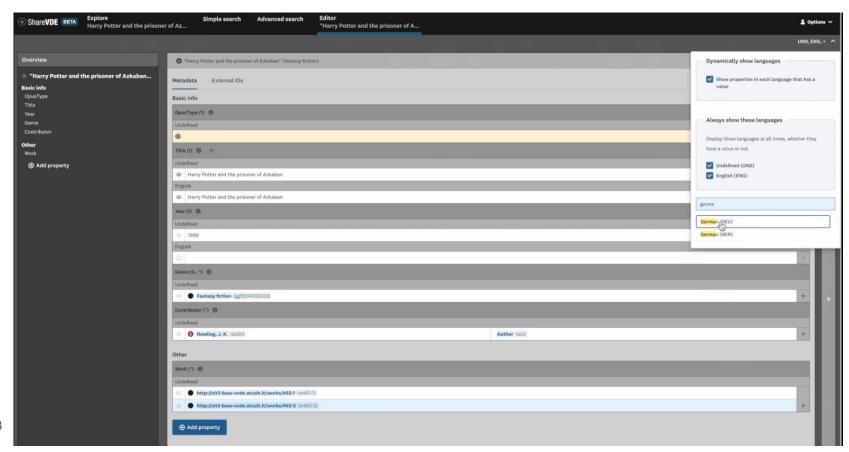
## JCricket edit - open a new entity via ID

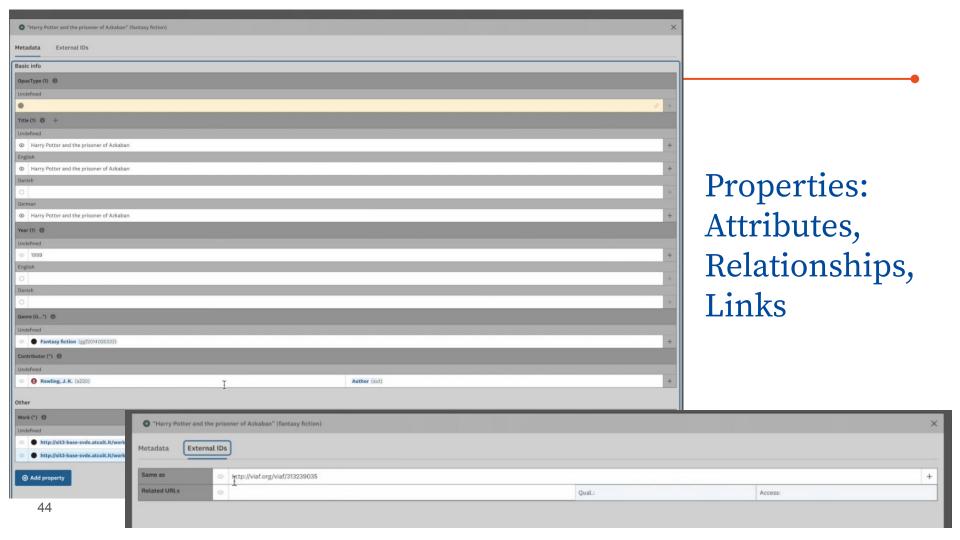


## JCricket edit - Multiple entities in one screen

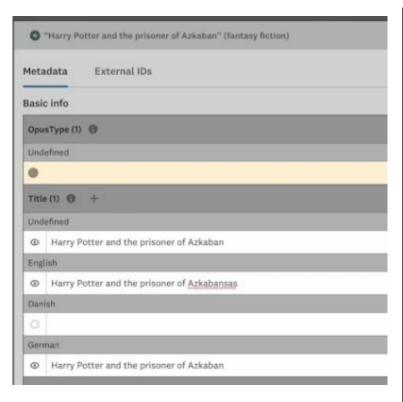


# JCricket edit - languages



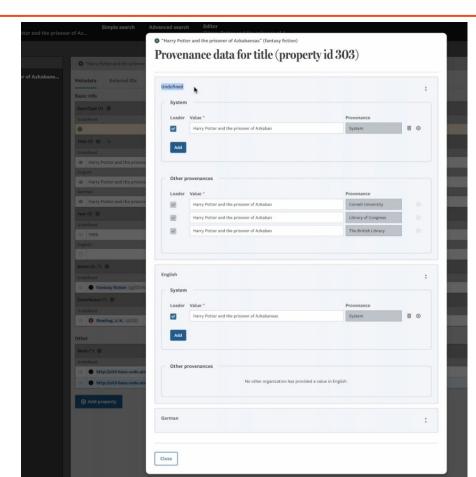


#### JCricket edit - Add variant forms

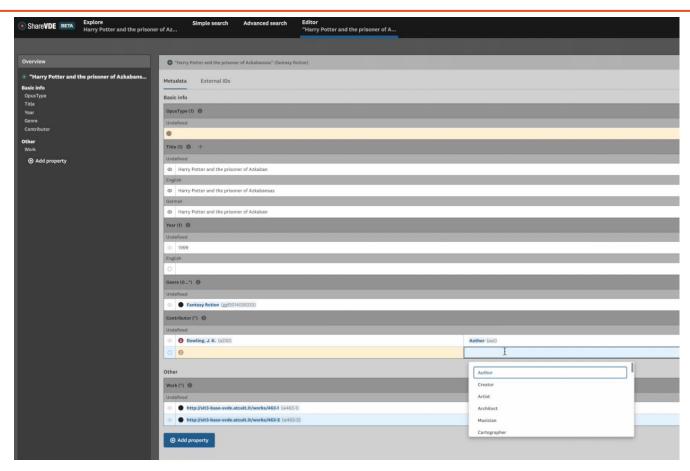




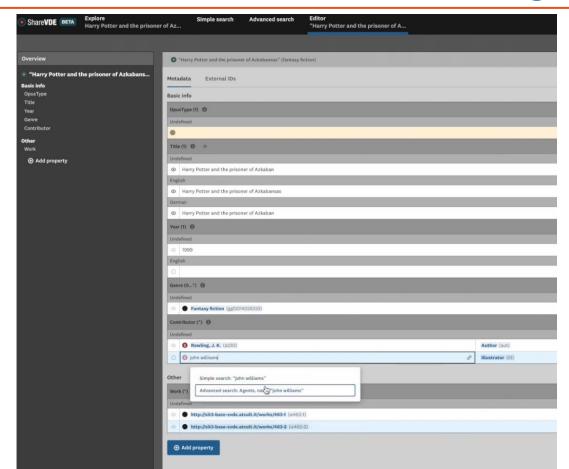
#### Provenances and leader value for end users



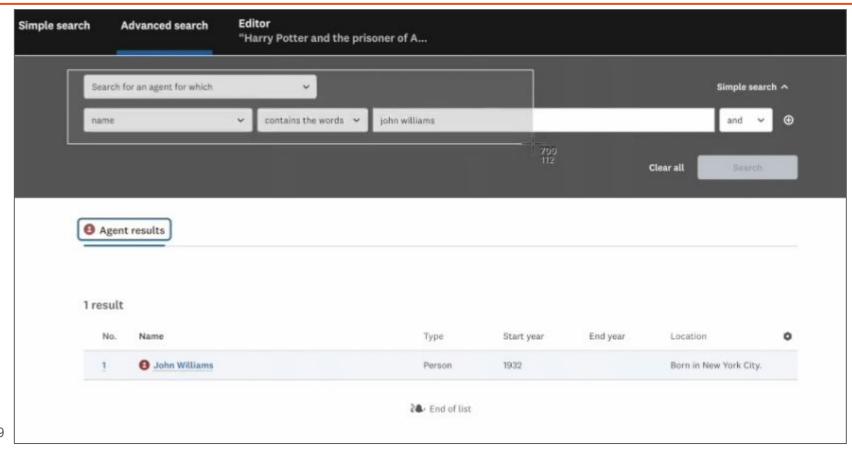
#### JCricket edit - controlled value list



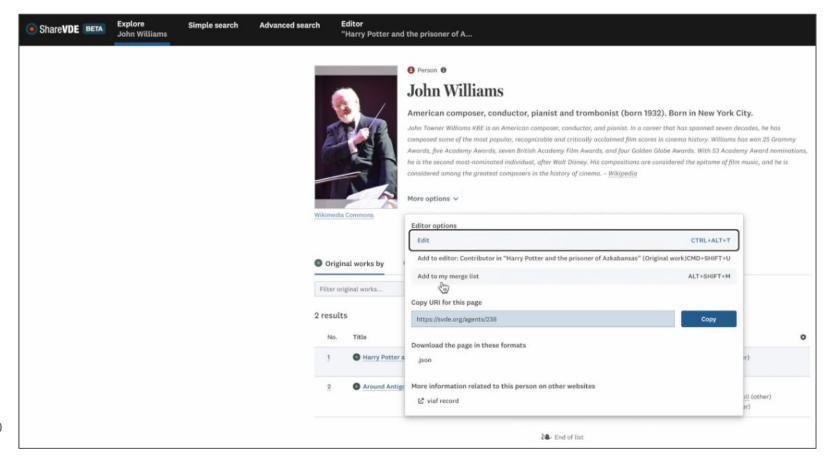
## JCricket edit - search while editing



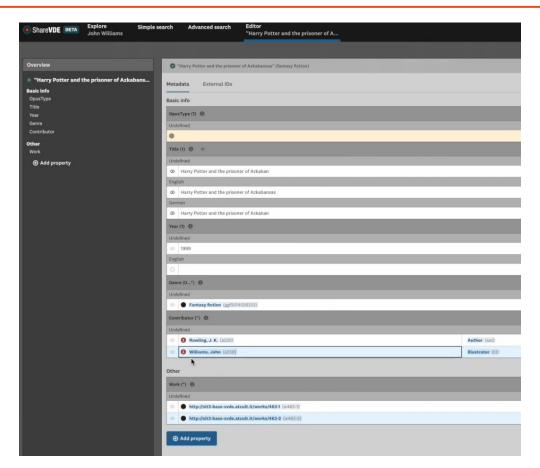
## JCricket edit - search while editing



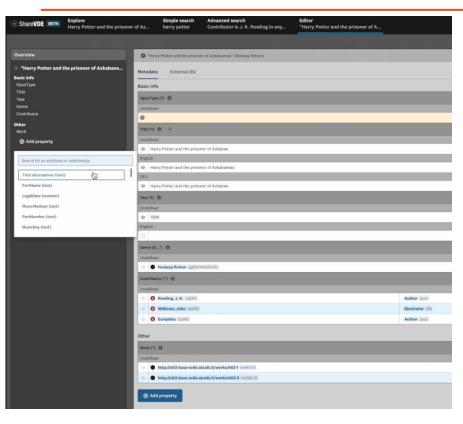
## JCricket edit - search while editing

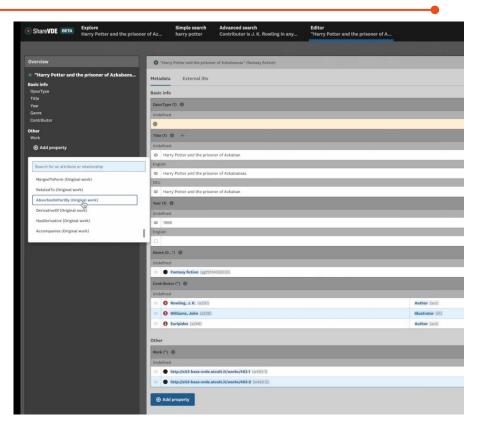


# JCricket edit - add agent

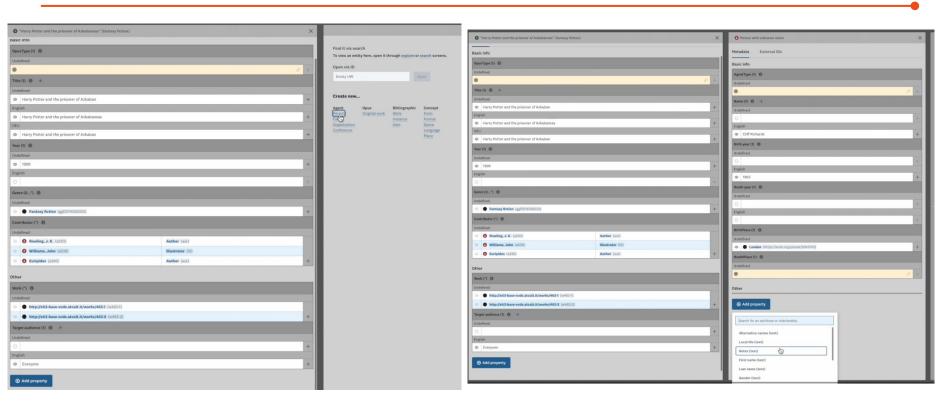


#### JCricket edit - add attribute



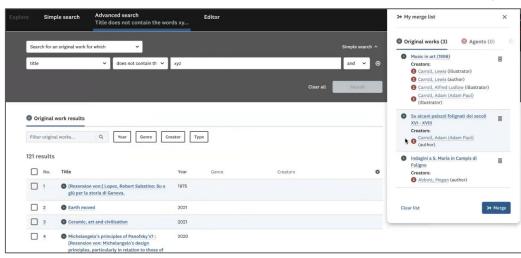


# JCricket - create an entity from scratch

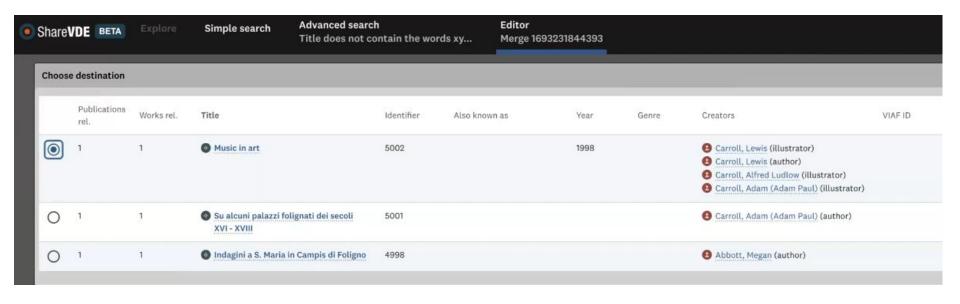


## JCricket merge - select merged entities

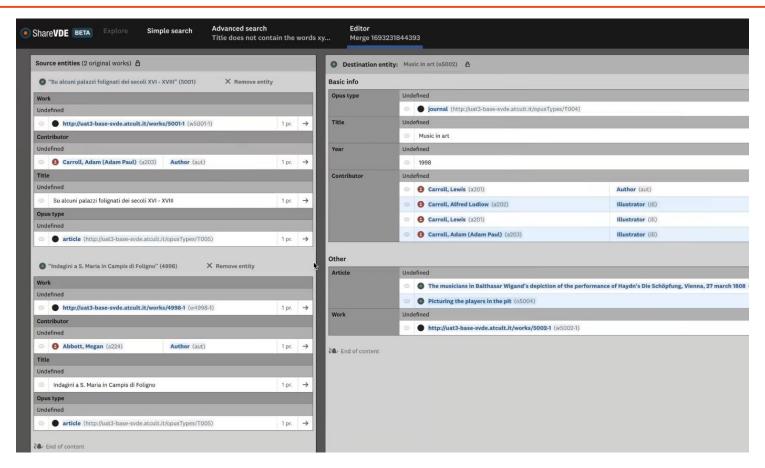




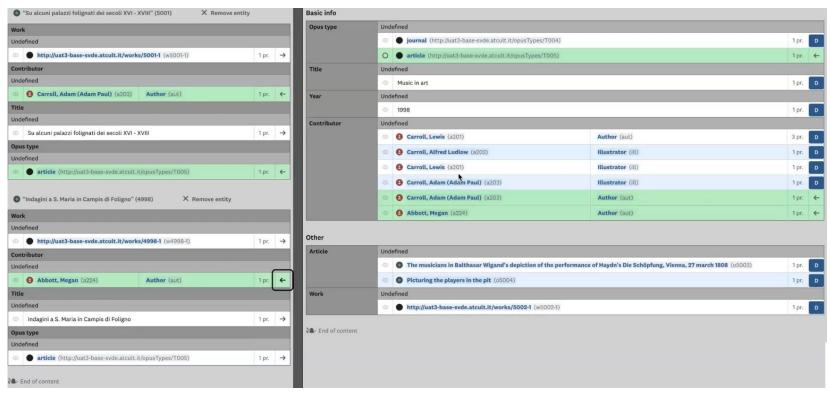
## JCricket merge - select destination entity



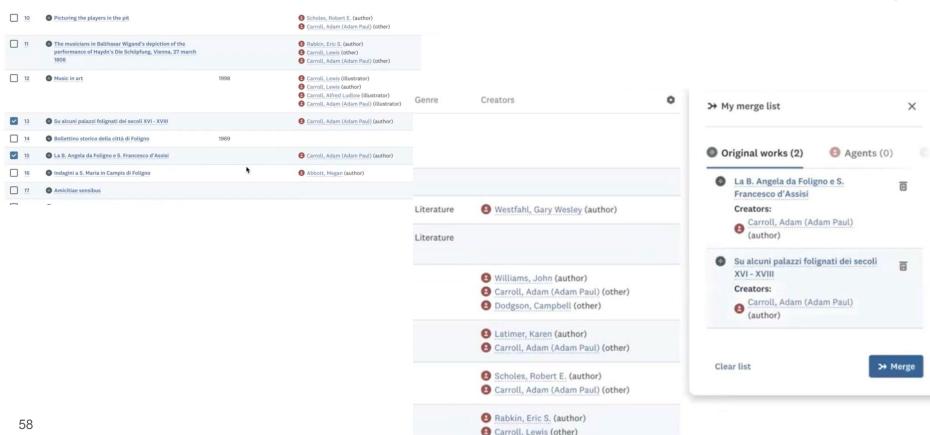
## JCricket merge - select properties



## JCricket merge - select properties

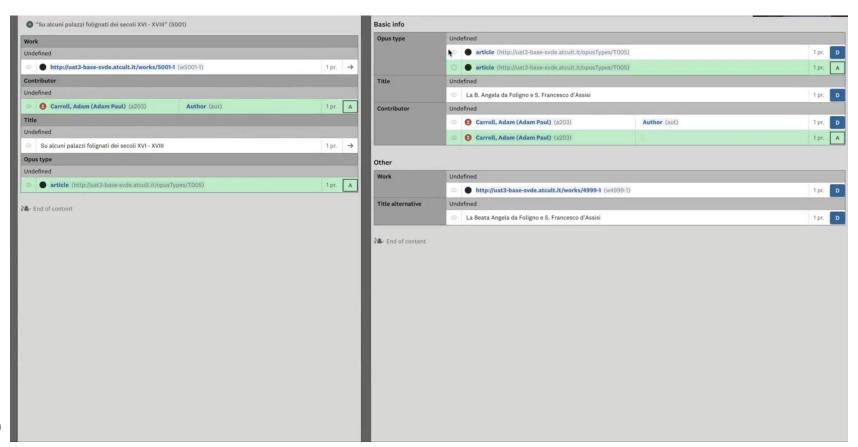


#### Merge entities where the role is the same

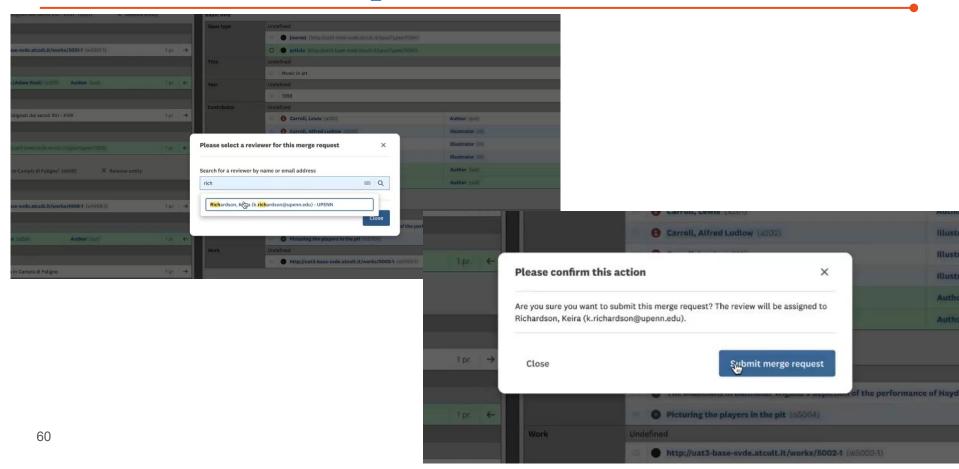


Carroll Adam (Adam Davil) (ather)

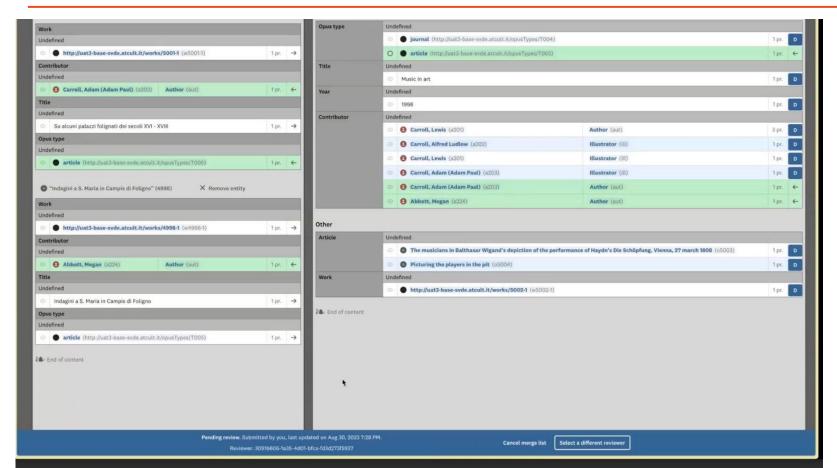
#### Properties <u>Automatically</u> added (full matching)



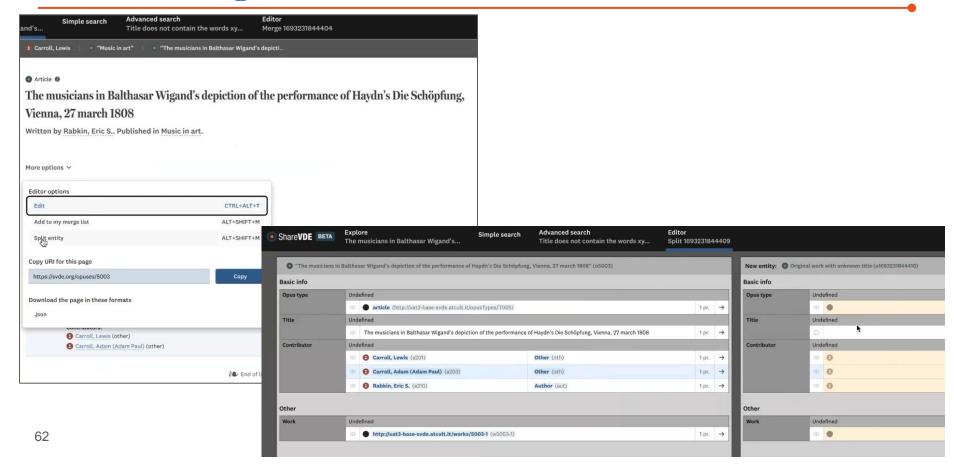
## JCricket - review process



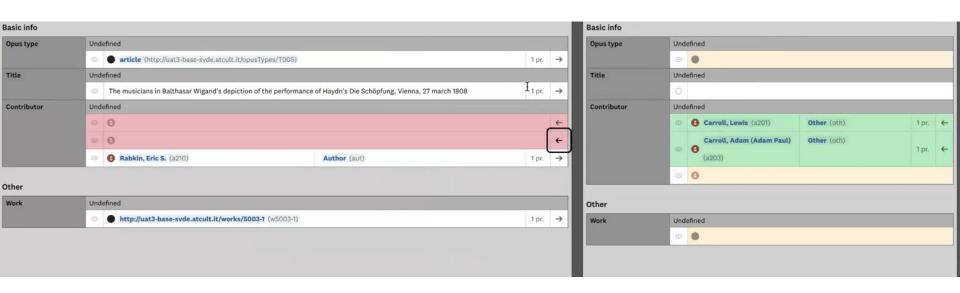
## Merge - pending review



## JCricket split



## JCricket split - select properties



#### What JCricket is



- it's a linked data entity / authority editor
- it applies to linked data entities created within all tenants of the Share Family (svde.org, pcc-lod.org, natbib-lod.org)
- it's a manual application that manages properties (attributes, relations and links) of entities in the CKB Cluster Knowledge Base
- it's a collaborative tool shared across member institutions
- it can be a new tool for entity sharing in LOD



#### What JCricket is not



- not a traditional bibliographic data editor
- not an original cataloguing tool
- not in contrast with Sinopia or Marva
- not impacting original data that reside in member libraries' systems (unless libraries want to use ad hoc APIs for entity updates both in SVDE and in their systems)



## Next generation cataloguing



The JCricket editor is an example of how the LOD Platform technology, within the Share Family Linked Data Ecosystem, is pursuing a new way of managing library cataloguing in a cooperative way:

- ★ aggregation of data from multiple sources
- ★ managed through standard protocols (linked data)
- ★ in a collaborative and integrated environment
- ★ that makes available open data and resources
- ★ to end users and professionals (researchers, scholars etc.)
- ★ for reuse in the library community and beyond

#### JCricket references

#### **Useful references:**

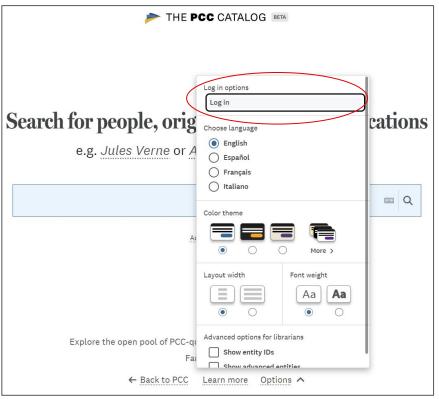
- JCricket overview <a href="https://wiki.share-vde.org/w/images/7/74/JCricket\_overview 2023-Jan-26.pdf">https://wiki.share-vde.org/w/images/7/74/JCricket\_overview 2023-Jan-26.pdf</a>
- for more technical details on JCricket <a href="https://wiki.share-vde.org/w/images/e/e8/|Cricket\_entity\_editor\_presentation.pdf">https://wiki.share-vde.org/w/images/e/e8/|Cricket\_entity\_editor\_presentation.pdf</a>
- on how JCricket has been conceived
  <a href="https://wiki.share-vde.org/w/images/b/b3/Share-VDE">https://wiki.share-vde.org/w/images/b/b3/Share-VDE</a> perspective on Cluster Knowledge Base and Provenance.pdf





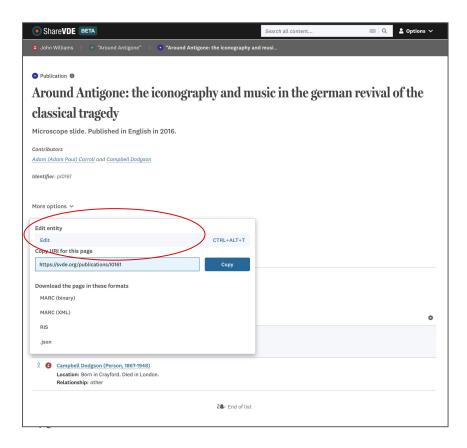
#### Thank you!

https://www.share-family.org/ https://wiki.svde.org info@svde.org



- Integrated in the discovery portal web interface, for authenticated users
- User types: basic and advanced
- Entities aggregate data from different contributing libraries (aka Provenances):
  - ★ an entity is also called Prism, where each face represents data coming from a given Provenance





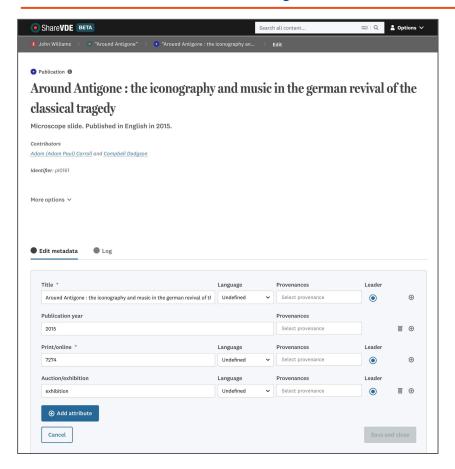


#### Edit function to change entities' properties

- ★ add, remove and amend attributes, relationships and links belonging to a single entity
- ★ real time notifications about cluster property changes

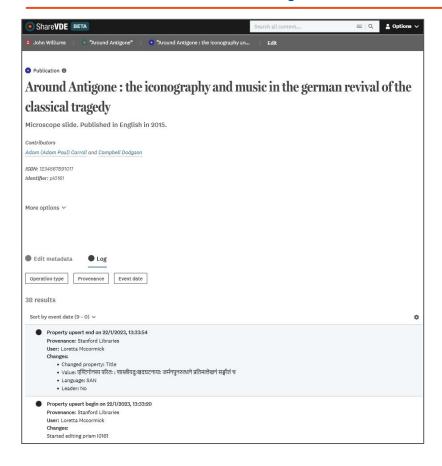


- Merge function: reconcile multiple entities into one (e.g. two authors who are in fact the same person)
- ★ choose the properties to copy to the merged entity
- multiple phases: create the merge list, edit the merge list, edit clusters, request for review, approve (or deny) the merge



- Split and Create functions: move one or more properties between two entities; optionally create a new one
  - ★ choose the properties to move from entities
  - ★ properties can be moved from an entity to create a brand new one
  - ★ multiple phases: create the split-set, edit the split-set, edit clusters, request for review, approve (or deny) the split





- Review workflow: edits are reviewed by advanced editors
  - ★ notifications to manage the review workflow
- Dictionary API: what are the available cluster types? Which attributes belong to a cluster type? What cardinality? Which relationships?
- Entity Event Log: tracks the history of changes
- Data changes synchronization across SVDE storages (e.g. RDF Store, Search Engine, RDBMS)



### Why JCricket is valuable



- ★ A collaborative tool, shared across the Share Family community and improving the data created within it
- ★ Linked data conversion in a shared discovery environment and direct entity management capabilities in one place
- ★ Increases data quality where massive automated processes are necessarily lacking
- ★ Collaboration ensures higher quality and authoritativeness of data
- ★ Conceived by <u>SVDE and Share Family community</u>, ie. real users
- ★ Potentially complementary to other tools, e.g. Sinopia as they cover different aspects of the entity management flow
- ★ Could potentially support other workflows and connections with systems external to the Share Family
- ★ You can always track back to your data through the Provenance
- ★ JCricket will extend authority capabilities through the integration with external data sources such as Wikidata



#### Where we are now



- ★ The back-end APIs that manage JCricket behind the scenes are ready
- ★ The respective front-end functions for the end users to actually use JCricket are under development
  - progressive releases through 2023

