

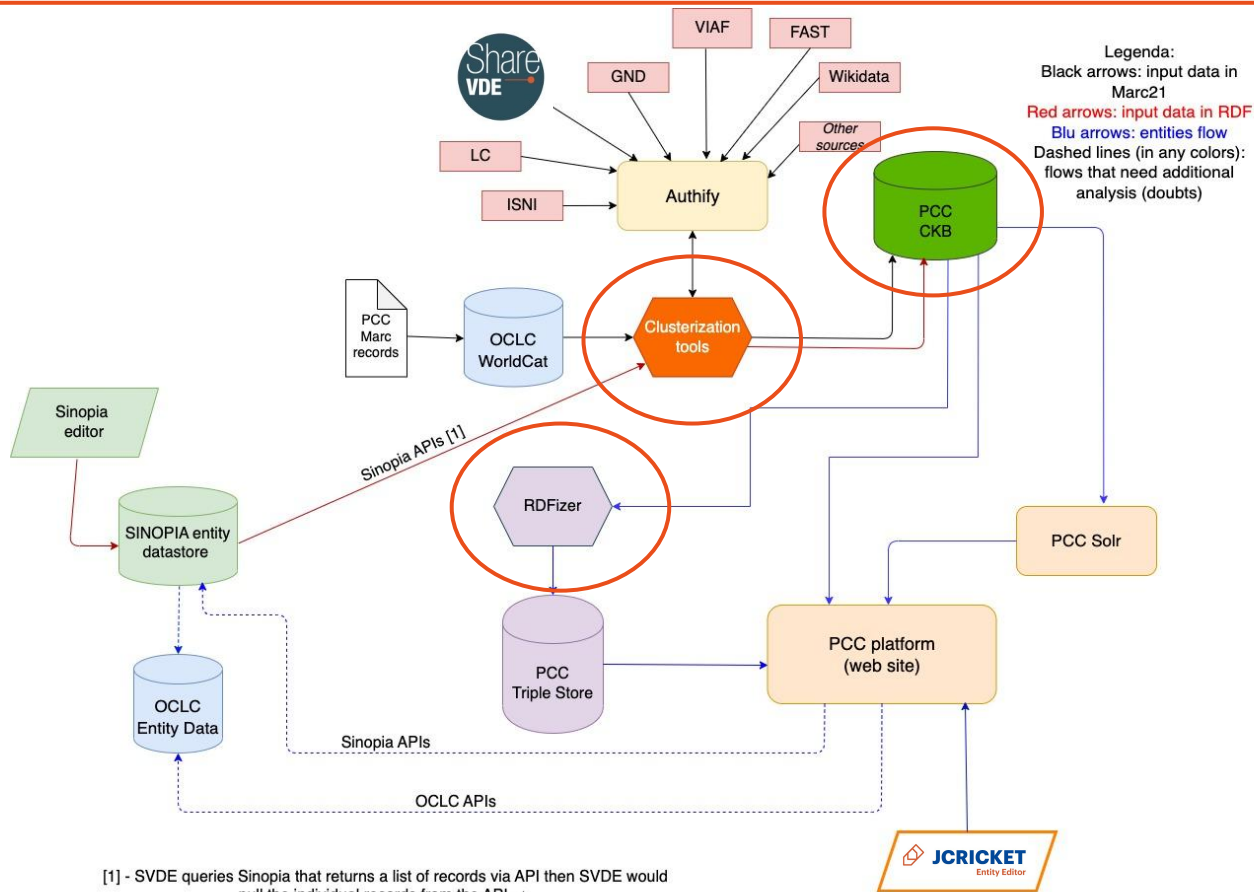


Sinopia, PCC data pool and Share-VDE interaction

PCC Sinopia Cataloging Affinity Group, 26th January 2023

<https://svde.org>
info@svde.org
<https://wiki.svde.org/>

PCC data flow - New version 2023

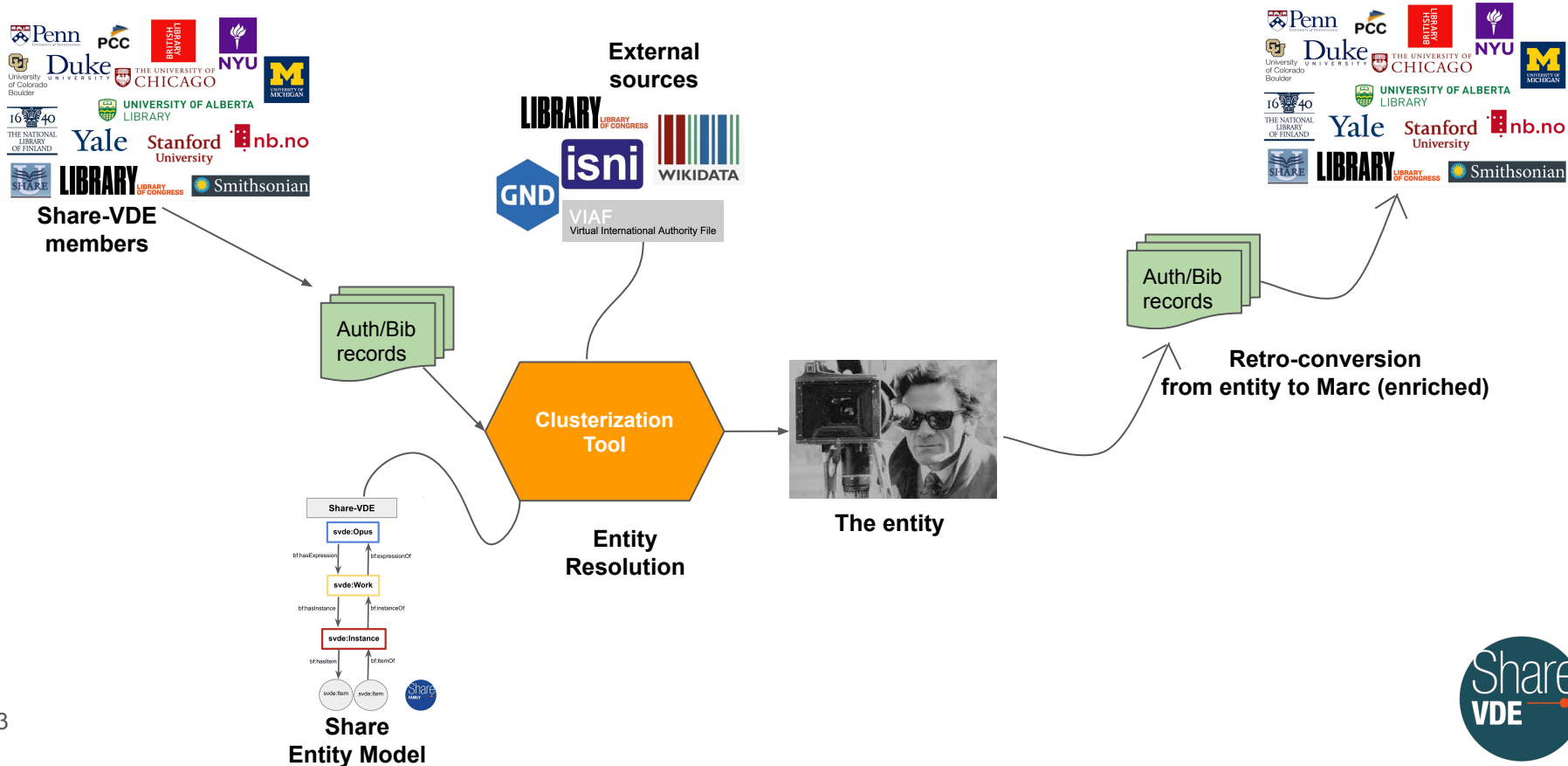


[1] - SVDE queries Sinopia that returns a list of records via API then SVDE would pull the individual records from the API →

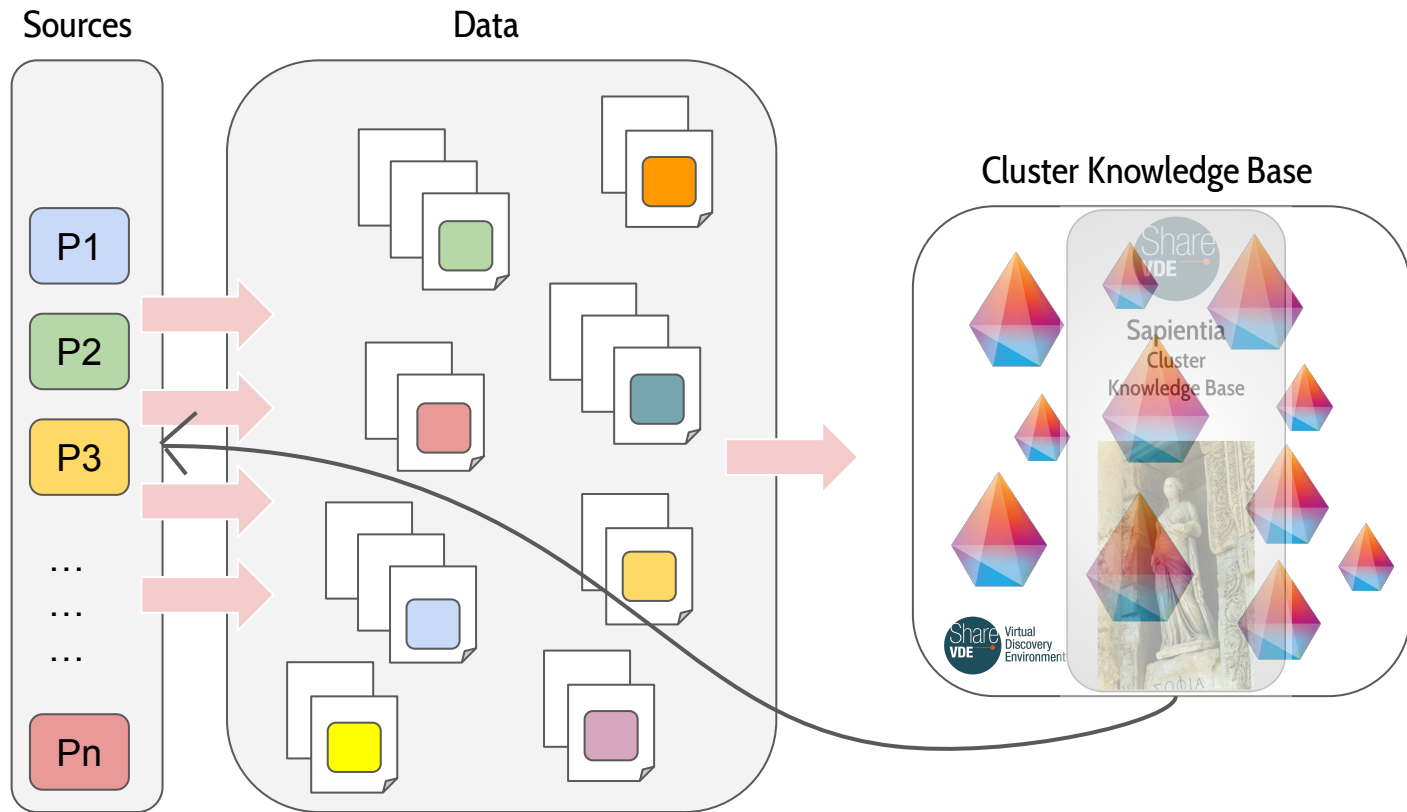
(https://docs.google.com/document/d/1xbS9X8oP7e8ztzKb6QfbsiU_aQAdDnq-82YuzrH_s/edit#)



The Clusterization tool: the Entity Resolution module

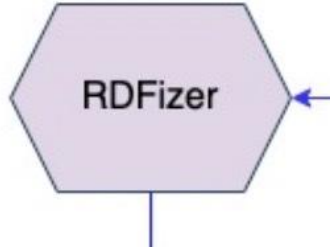


The Cluster Knowledge Base as the heart of the system



The RDFizer tool: how data is handled to become BIBFRAME data

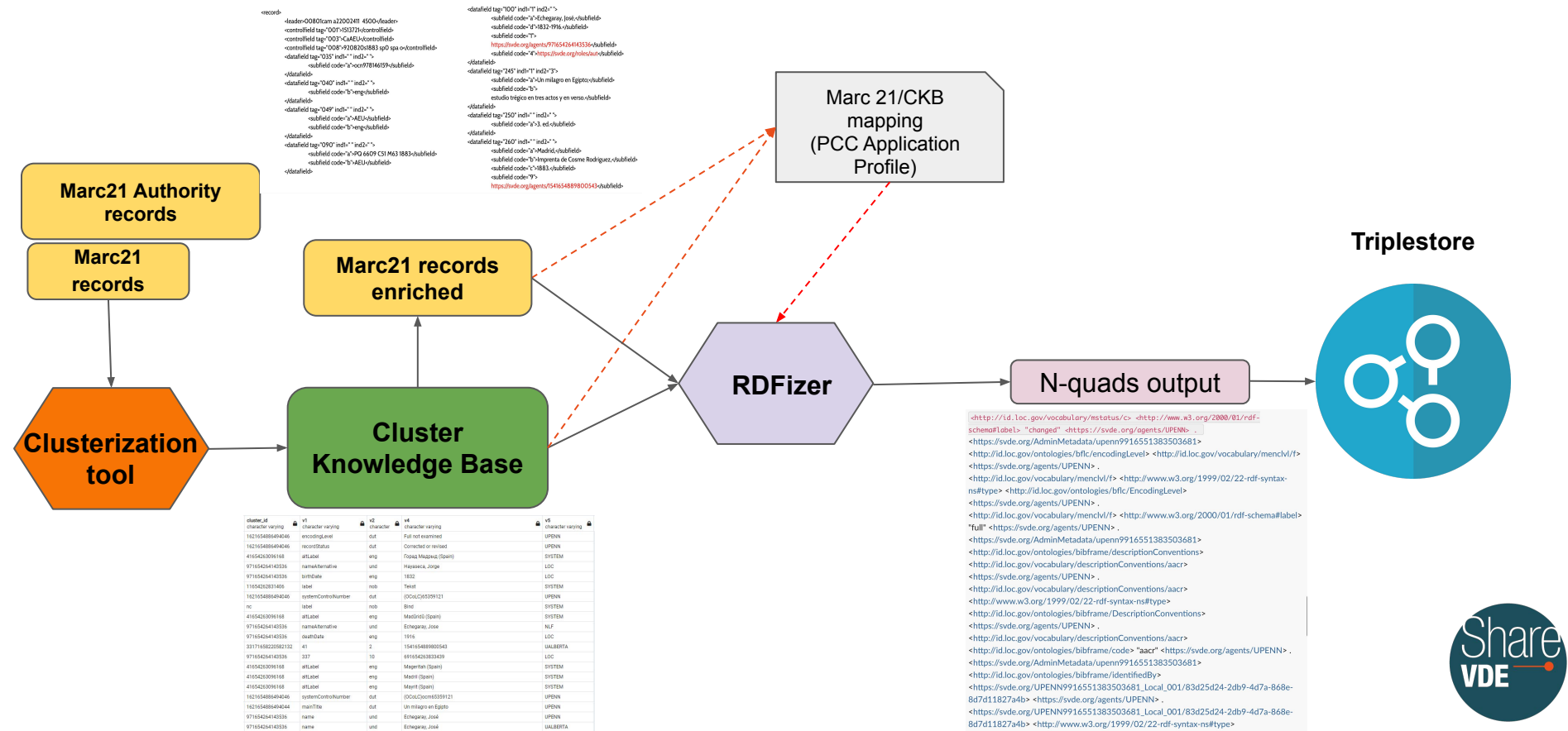
RDFizer tool is in charge of converting data in RDF, using ontologies such as Bibframe 2.0, MADSrdf, RDA, to produce N-Quad files output. It is based on Hadoop Map Reduce framework.



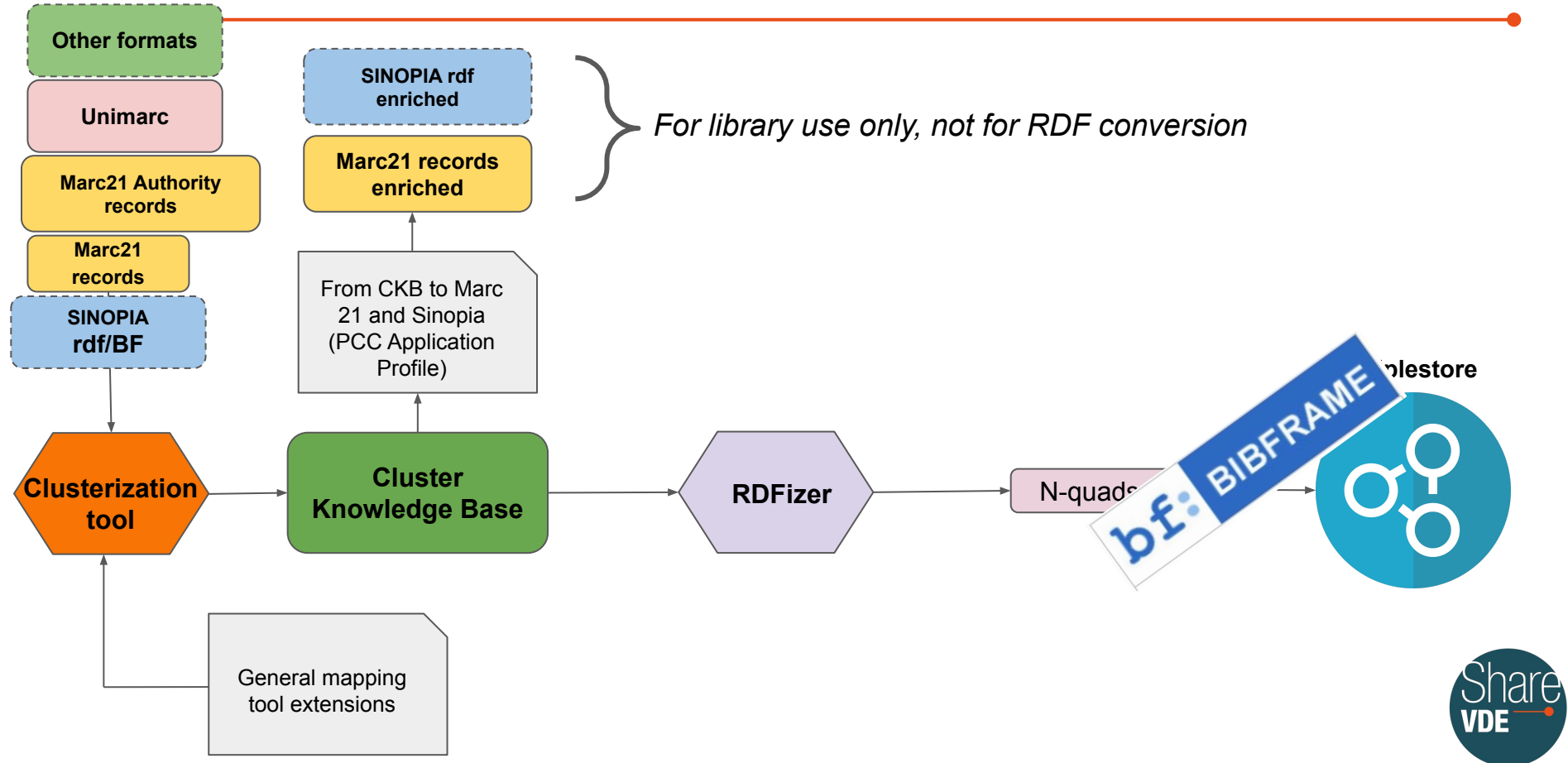
The conversion process is based on Library of Congress Conversion specifications (published on [Bibframe website](#)), adapted for Share-VDE model.

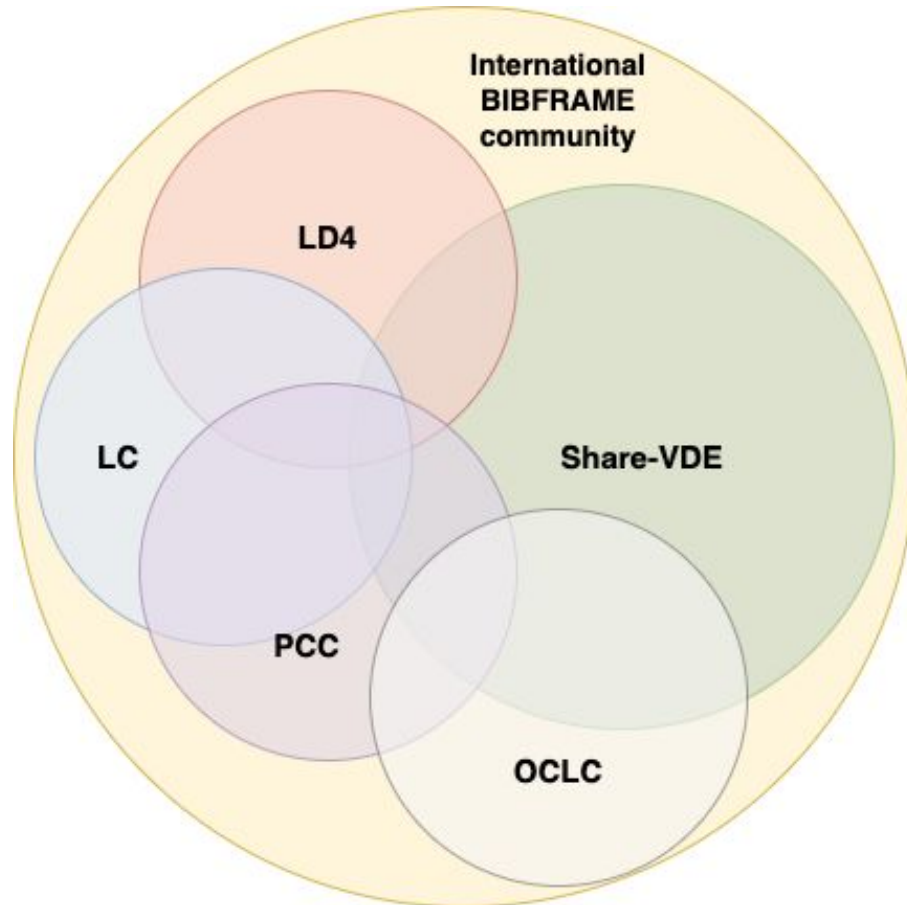
We have just finished reviewing the module to accept all the latest updates published by the LC (from version R2 to 1.7/2.0)

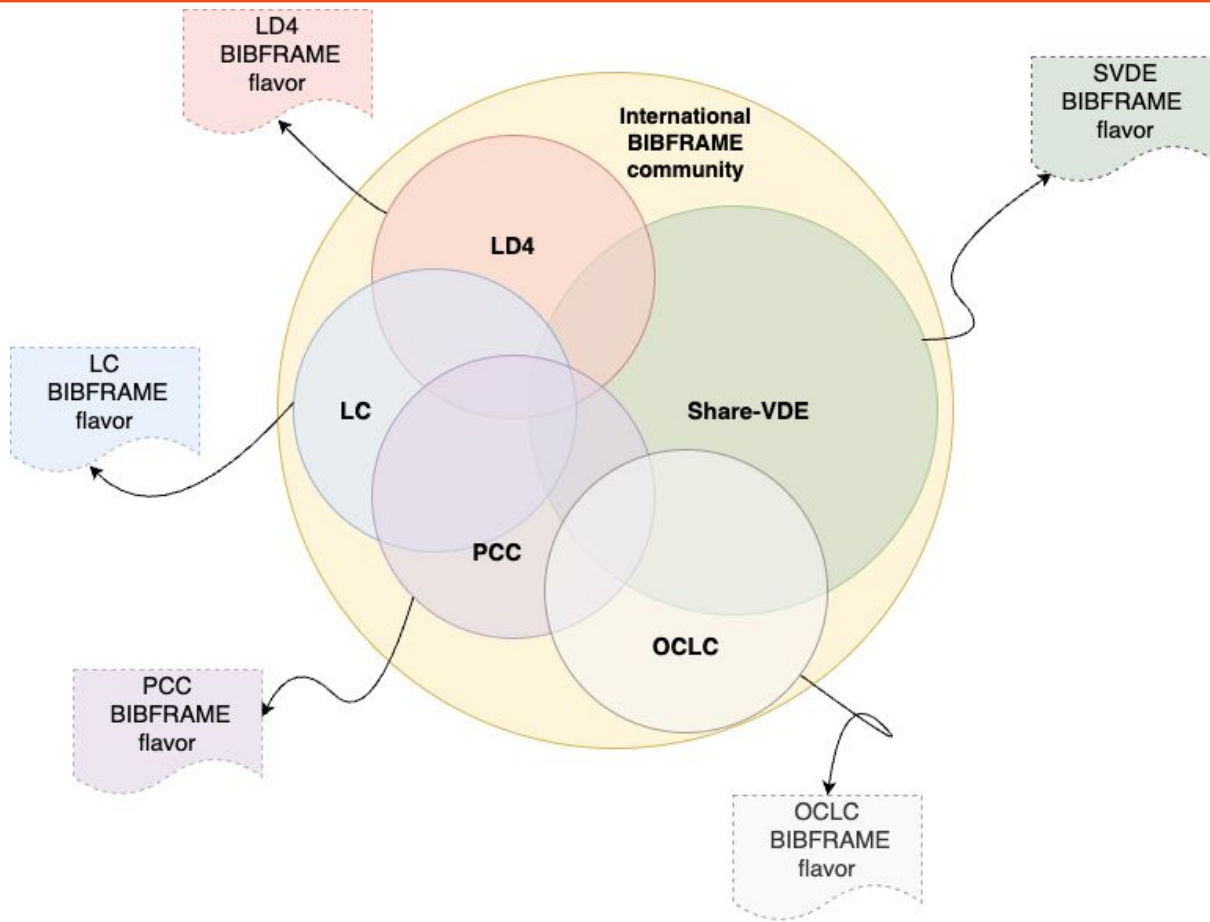
Conversion pipeline - Today



Conversion pipeline - Tomorrow







Who leads this multiplicity of voices?



Hector Berlioz, pioniere della direzione, dirige un coro. Disegno di Gustave Doré

Fonte:

https://it.wikipedia.org/wiki/Direttore_d%27orchestra

SVDE network and collaborations (1)

In order to guarantee interoperability between different BIBFRAME communities/nodes:

The Library of Congress - Network Development and MARC Standards Office (LS/NDMSO)

for exchange of ideas and principles bringing BIBFRAME into practice

Linked Data for Production (LD4P), 2016-2018

Linked Data for Production: Pathway to Implementation (LD4P2), 2018-2020

Linked Data for Production: Closing the Loop (LD4P3), 2020-2023

for LD4P's vision that intertwined SVDE activities for example interacting with

- Sinopia User Group
- Best Practices for Authoritative Data Working Group

SVDE network and collaborations (2)

PCC BIBFRAME Data Exchange Meeting

held online September 9-10, 2021

<https://www.loc.gov/aba/pcc/bibframe/PCC-BIBFRAME-Data-Exchange-Summary.pdf>

This meeting established two temporary working groups:

- Use Cases working group chaired by Philip Schreur (Stanford) that delivered its report in December 2021
[Use Case Working Group Final Report](#)
- Data Exchange working group chaired by Nathan Putnam (OCLC), in course

and one permanent group:

- PCC BIBFRAME Interoperability Group (BIG)
<https://www.loc.gov/aba/pcc/bibframe/TaskGroups/BIG/BIG-TOR.pdf>

OCLC & Casalini - Data Exchange Collaboration

for PCC's quality data collected and received in MARC21 from OCLC, processed and distributed in BIBFRAME by SVDE.

BIBFRAME Data Exchange Meeting

Meeting held virtually September 9-10, 2021 and represented groups were:

- **National Libraries:**

British Library; German National Library; Library and Archives Canada; Library of Congress (LC), LC Network Development and MARC Standards Office; LC Policy, Training and Cooperative Programs Division; National Library of Medicine; National Library of Sweden.

- **Program for Cooperative Cataloging (PCC):** Steering Committee; Policy Committee; Standing Committee on Standards; Standing Committee on Training; Standing Committee on Applications

- **LD4 Community:** Steering Committee; LD4P3; Questioning Authority (QA); Sinopia

- **Vendor Community:** EBSCO; ExLibris; Index Data; OCLC; Share-VDE (Casalini)

- **Others:** European BIBFRAME Group; ISSN International Centre; FOLIO community; Koha; RDA Steering Committee

BIBFRAME Data Exchange Meeting (cont.)

Purposes of the meeting:

- Discuss exchange of BIBFRAME data between systems and implementations
- Information sharing: What has been done so far, what are current plans and challenges encountered.

Major themes:

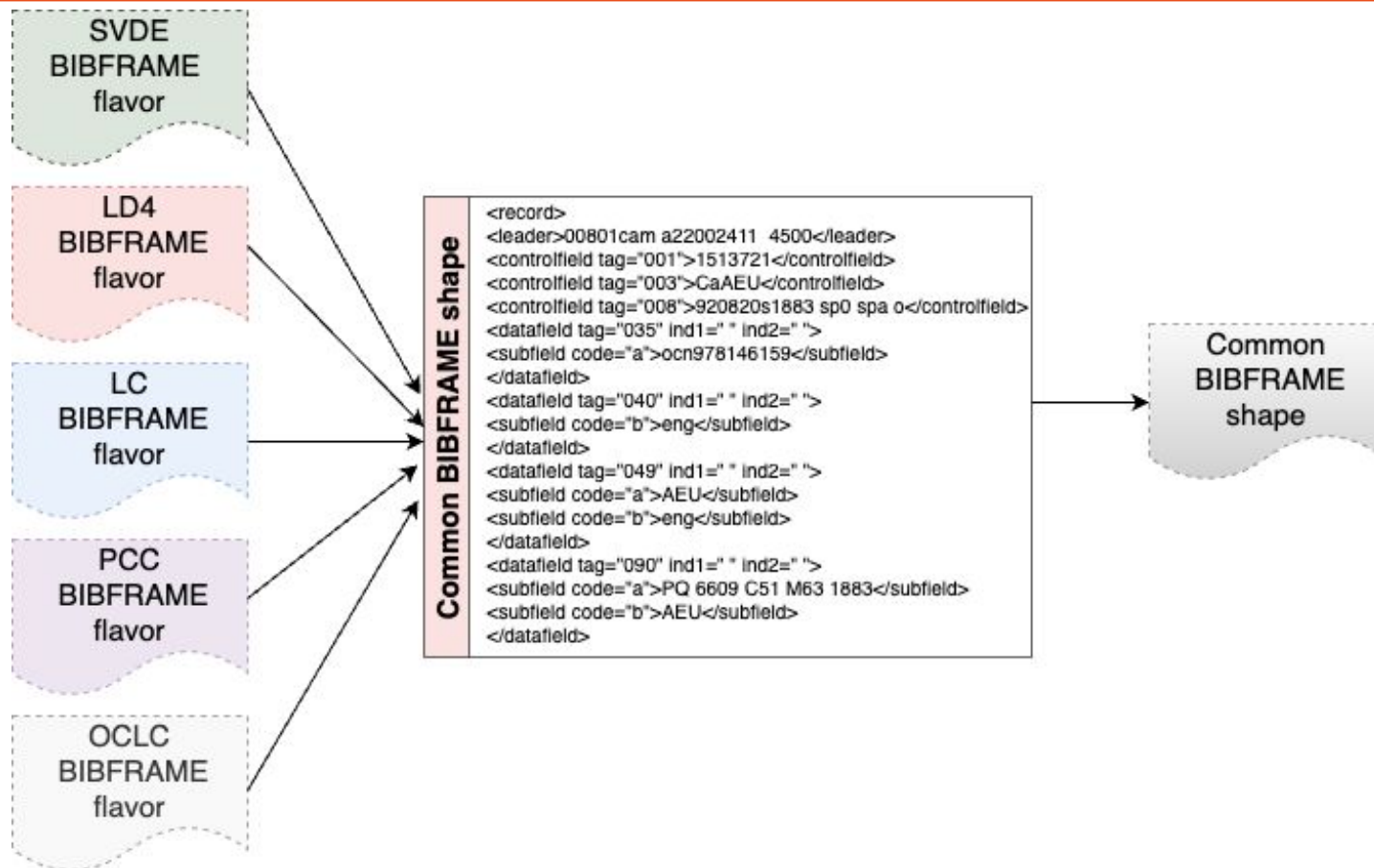
- Different choices in expressing the BIBFRAME ontology
- Challenges in interchange of BIBFRAME data
 - Choices of original data creation (e.g. LC BIBFRAME editor versus Sinopia)
 - Data conversion from MARC

BIBFRAME Interoperability Group - BIG

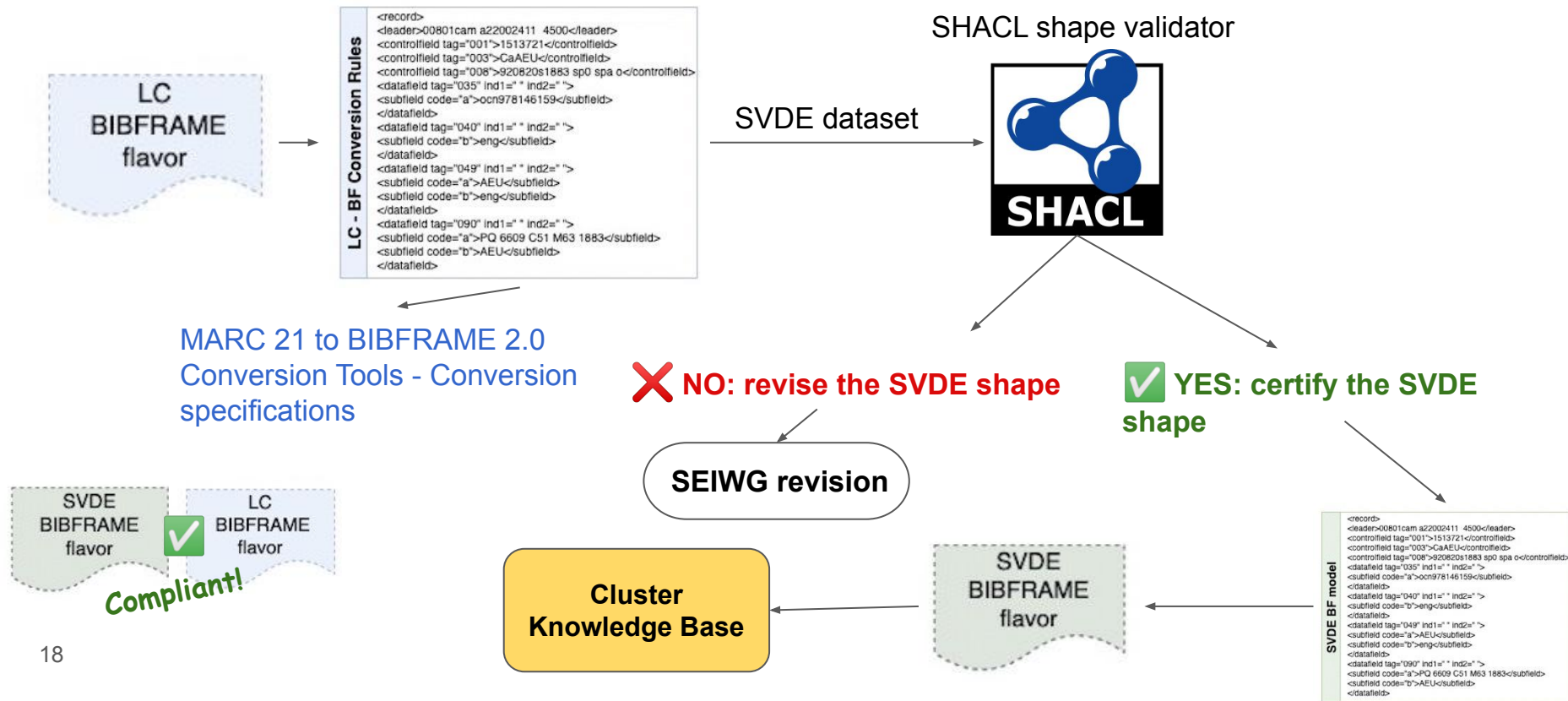
- Define a standard BIBFRAME “shape” to support data reuse including conversion to and from other formats.
- Explore defining BIBFRAME elements necessary for data exchange.
- Surface issues regarding the use of the Official RDA with BIBFRAME and propose strategies for their resolution.
- Collaborate and communicate with other groups working in the area of BIBFRAME interoperability to ensure the ability to reuse BIBFRAME created in one community in other BIBFRAME stores.

BIBFRAME Interoperability Group - BIG

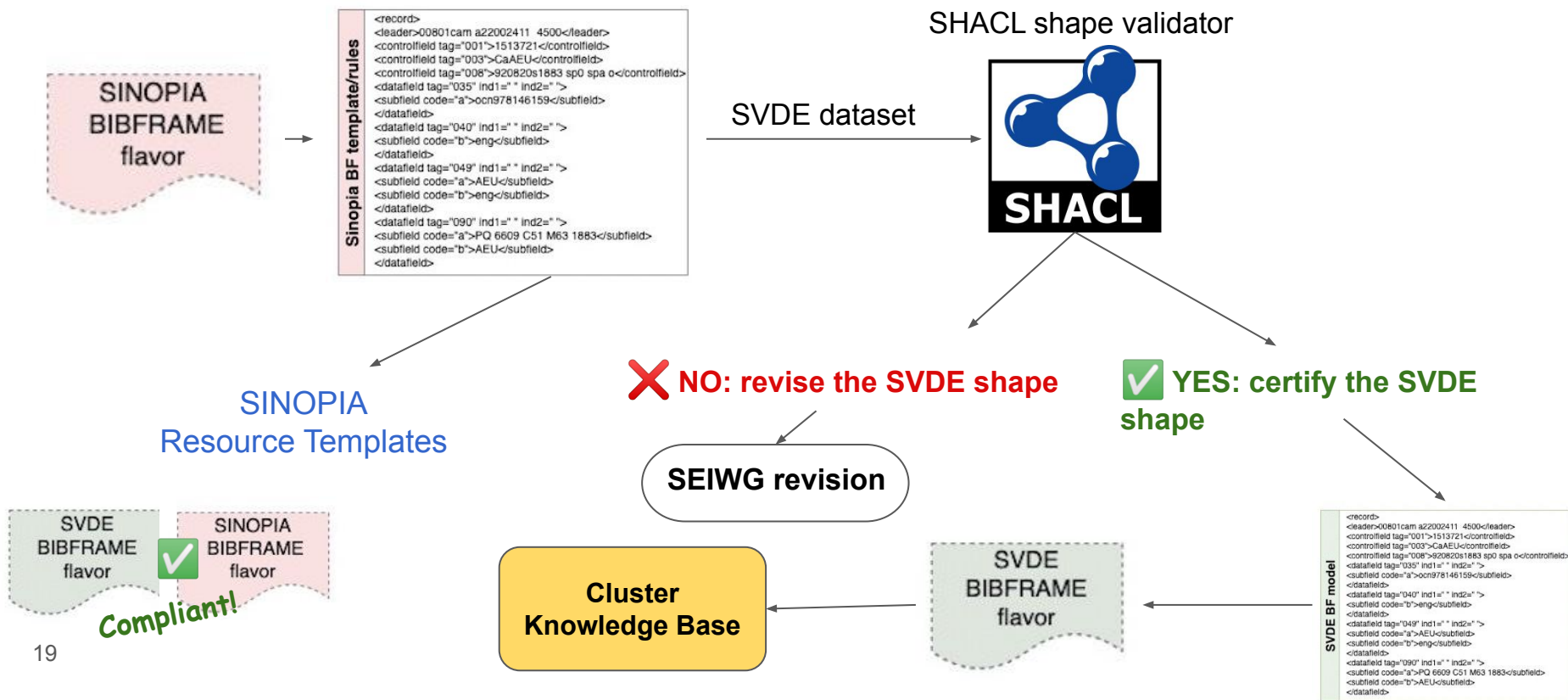
- Examine the work accomplished by the other BIBFRAME working groups and apply to this charge where appropriate.
- Gather use cases as necessary to inform decision making, expanding on the efforts of the Use Case Working Group and others.
- Provide an avenue for other interested parties to contact the BIBFRAME Interoperability Group and/or reach out to other stakeholders.



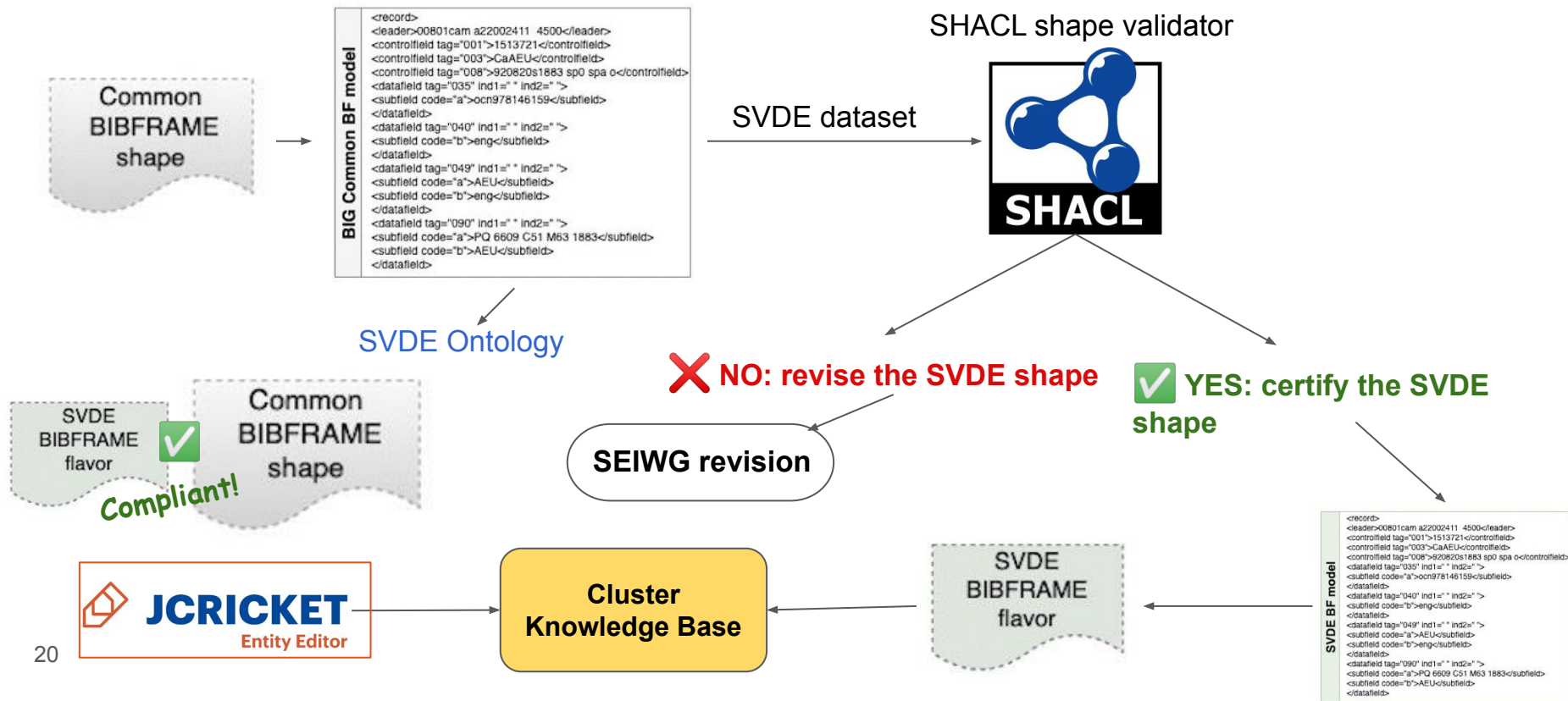
What the Share Family community is doing...



What the Share Family community is doing...



What the Share Family community is doing...



JCricket entity editor



What JCricket *is*



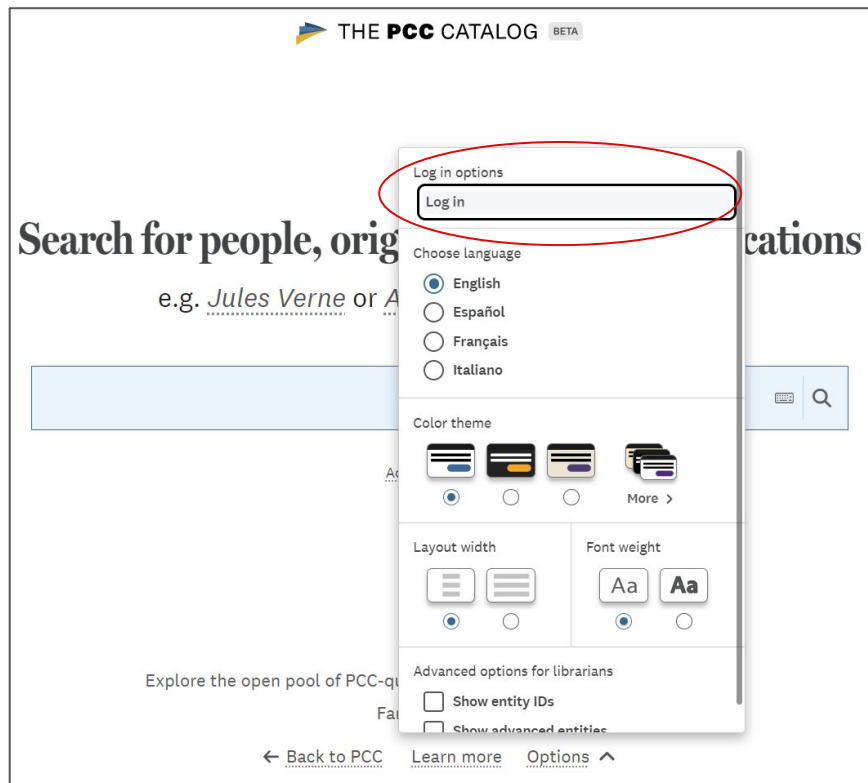
- ✓ it's a linked data entity / authority editor
- ✓ it applies to linked data entities created within all tenants of the Share Family (svde.org, pcc-lod.org, natbib-lod.org)
- ✓ it's a manual application that manages properties (attributes, relations and links) of entities in the CKB - Cluster Knowledge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

What it does



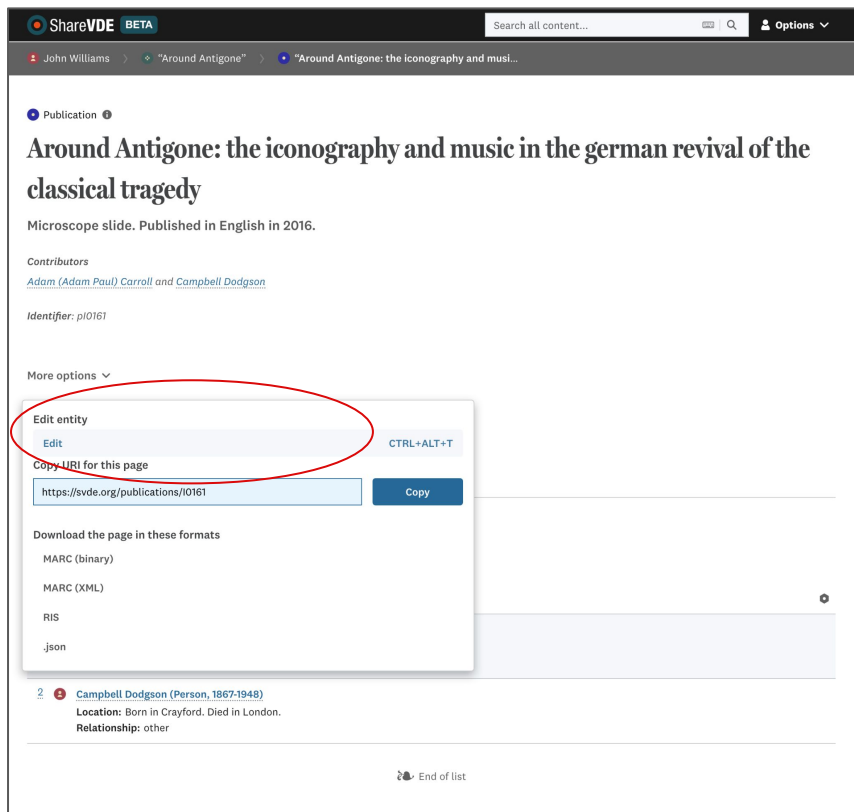
Integrated in the discovery portal web interface, **for authenticated users**

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

What it does



The screenshot shows the ShareVDE BETA web application. The top navigation bar includes the logo, a search bar, and user options. The main content area displays a publication titled "Around Antigone: the iconography and music in the german revival of the classical tragedy" by Adam (Adam Paul) Carroll and Campbell Dodgson. A "More options" dropdown menu is open, with the "Edit entity" option highlighted by a red circle. Below this menu, there is a section for downloading the page in various formats (MARC, RIS, JSON) and a list of entities, including Campbell Dodgson.



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

What it does

The screenshot shows the ShareVDE BETA web interface. At the top, there's a search bar and a user menu. The breadcrumb trail indicates the user is viewing a publication by John Williams, titled "Around Antigone", and is currently editing its metadata. The publication details include the title "Around Antigone : the iconography and music in the german revival of the classical tragedy", the publisher "Microscope slide", the language "English", and the year "2015". The contributors are listed as Adam (Adam Paul) Carroll and Campbell Dodgson. The identifier is p10161. Below the publication details, there's a section for "Edit metadata" with tabs for "Edit metadata" and "Log". The "Edit metadata" tab is active, showing a form with four rows of metadata fields: Title, Publication year, Print/online, and Auction/exhibition. Each row has fields for Title, Language (set to Undefined), Provenances (a dropdown menu), and Leader (a radio button). The "Add attribute" button is at the bottom left, and the "Save and close" button is at the bottom right.



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

What it does

The screenshot shows the ShareVDE BETA web interface. At the top, there's a search bar and navigation tabs. The main content area displays a publication titled "Around Antigone : the iconography and music in the german revival of the classical tragedy" by Adam (Adam Paul) Carroll and Campbell Dodgson. Below the title, it says "Microscope slide. Published in English in 2015." and lists the ISBN and Identifier. There are tabs for "Edit metadata" and "Log". Under "Log", there are buttons for "Operation type", "Provenance", and "Event date". Below these, it says "38 results" and "Sort by event date (9 - 0)". The log entries show property updates, such as "Property upsert end on 22/1/2023, 13:33:54" and "Property upsert begin on 22/1/2023, 13:33:20", with details on provenance, user, and changes.



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)

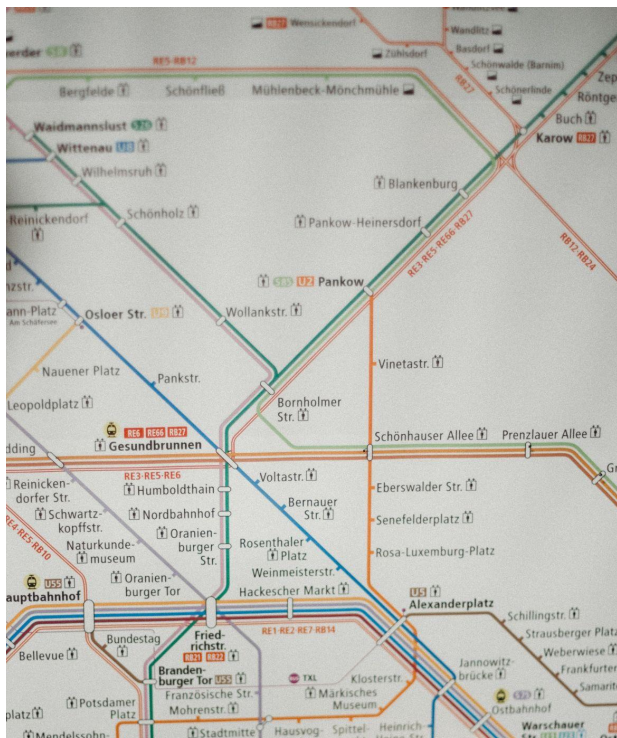


Why JCricket is valuable



- ★ 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 **SVDE community of libraries**, 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 🛠
 - early features from mid-February, for testing
 - progressive releases through 2023

Next generation cataloguing



The JCricket editor is an example of how Share-VDE technology, within the Share Family of initiatives, 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

References

Useful references:

- https://wiki.share-vde.org/w/images/c/c7/Share_Family_data_flow.pdf
- for more technical details on JCricket
https://wiki.share-vde.org/w/images/e/e8/JCricket_entity_editor_presentation.pdf
- on how JCricket has been conceived
https://wiki.share-vde.org/w/images/b/b3/Share-VDE_perspective_on_Cluster_Knowledge_Base_and_Provenance.pdf
-



Thank you!

tiziana.possemato@atcult.it
anna.lionetti@casalini.it

<https://svde.org>
info@svde.org
<https://wiki.svde.org/>