Share Family and beyond
Cooperation and innovation
to bring Linked Open Data into practice

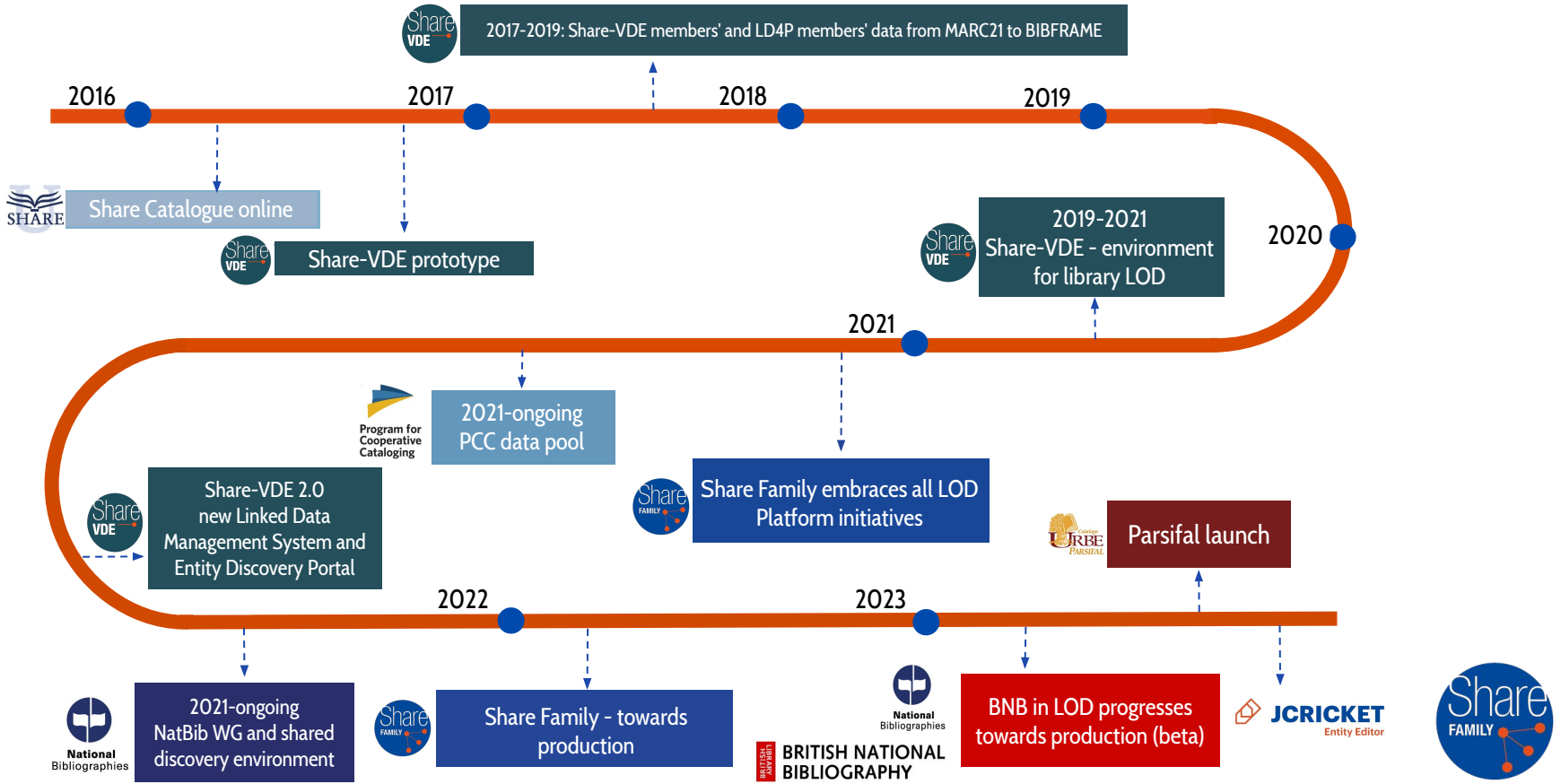
https://www.share-family.org/
https://wiki.svde.org
info@svde.org

Summary

- ✔ Share-VDE background and the Share Family
- ✔ Working Groups and cooperation
- ✔ Towards an operational environment
- ✔ JCricket - The Entity Management System

Share-VDE background and the Share Family

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 (LSPs) to join their knowledge in an ever-widening network of interconnected bibliographic data.

For further details please refer to https://www.share-family.org and the dedicated Share-VDE wiki section.

Share-VDE and 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.

INTEROPERABLE

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

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.

SUSTAINABLE

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.

OPEN

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.

AUTHORITATIVE

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-VDE and 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*

Outputs for consortia or single libraries

Linked Data Descriptions and Enriched MARC Records

- The Library catalogue is converted in linked data entities according to BIBFRAME 2.0. The entities are then enriched both with native and persistent SVDE URIs and URIs from external sources.
- MARC records from the original library catalogue are enriched both with native SVDE URIs and URIs from external sources.

Discovery Portal

- Advanced entity discovery system based on BIBFRAME
- Customised UI (skin)
- Integration with local APIs
- Site mapping with additional meta-tagging
- Data conversion to Schema.org

JCricket Entity Editor

- It's a manual collaborative tool designed to manage properties (attributes, relations, and links) of entities in the Cluster Knowledge Base, improving data quality through tasks like creation, merging, and splitting.
- Data can always be traced back to each Institution through the Provenance.
- It can potentially support other workflows and connections with systems external to the Share Family

Authoritative Services

- Innovative solutions that facilitate and improve authority control through automatic and manual procedures.
- Libraries to receive constantly updates on their bibliographic and authority records from authoritative sources.
- Authority Services currently available for MARC-based workflows offer automated URI enrichment, reconciliation and validation of library data.

Major benefits

Quality: enrichment with data from other authoritative sources and share responsibility for and contribute to sustainable maintenance of quality data.

Use: better exposition, data analysis opportunity, reuse in other projects, improve the visibility of hidden resources applying the entity data modeling structure.

Integration: intersection of different and multiple authoritative sources, increase circulation of data, ensure interoperability with local systems, enhance workflows.

Engagement: facilitate information exchange and collaboration with other communities, cross-reference between library resources.

Common priorities, challenges and concerns

- By adopting BIBFRAME as the main ontology in compatibility with IFLA-LRM, take advantage of the potentials of linked open data to facilitate interoperability among data pools, in coexistence with MARC.
- Transform library catalogs into research tools with structured access and visibility to original language research in all disciplines.
- Apply and support open metadata policies.
- Be independent of local practices and of ILS/LSP local choices.
- Foster international cooperation to maintain a growing wealth of information.
- By serving as an authoritative data source, contribute to a new bibliographic ecosystem where data modeling, enrichment and sharing are handled collectively.

Commitment to open data: triple store publication

Share-VDE data are open, and usable through an open endpoint to retrieve them in RDF format through SPARQL queries.

SERVICE

Triple Store Indexing

- Linked data descriptions are published on a triple store and can be queried through SPARQL endpoint

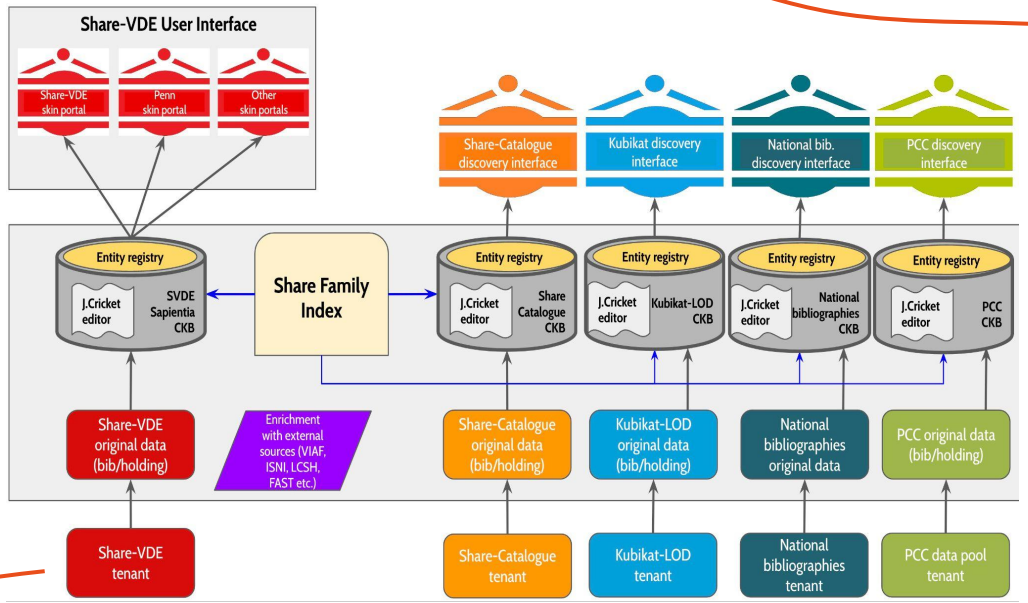
The core of Share-VDE integrated catalogue, ie. the Cluster Knowledge Base of linked data entities created from SVDE institutions' data, is published on a public query interface.

Data use policy: CC0 (unless expressly stated otherwise). Check out Share-VDE Open metadata policy for more information.

Versatile tenant infrastructure

According to the institution/consortia's needs and policies, it's possible to show their data ...

in their own tenant



in an institutional skin (customised sub-portal) in a pre-existing tenant

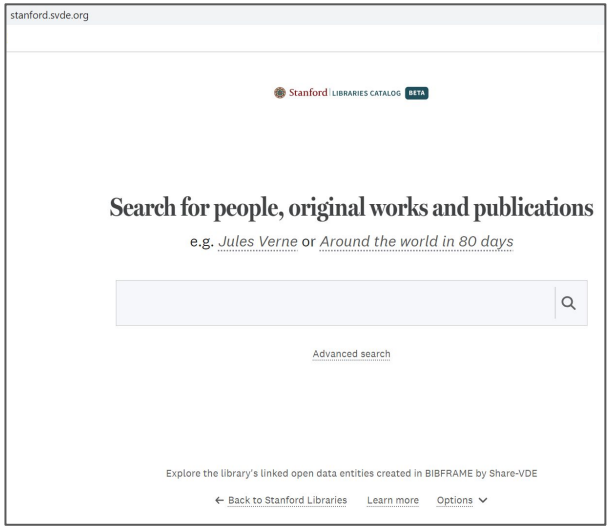
Institutional skin portals within a tenant

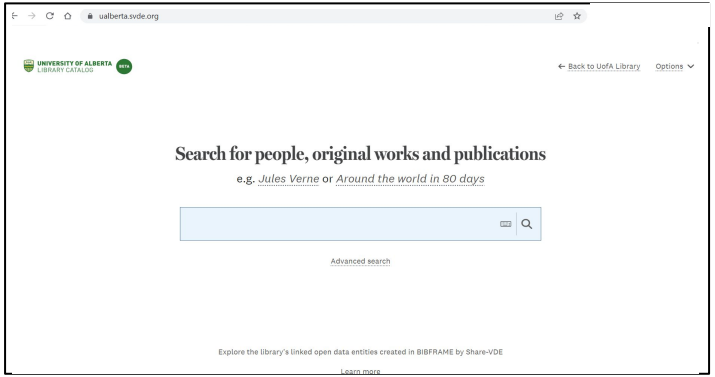
- While the main entity discovery portal of a tenant shows the data of all the institutions feeding the tenant's Cluster Knowledge Base, the skin portal gives the ability to filter only the data of the institution that the skin portal has been designed for.

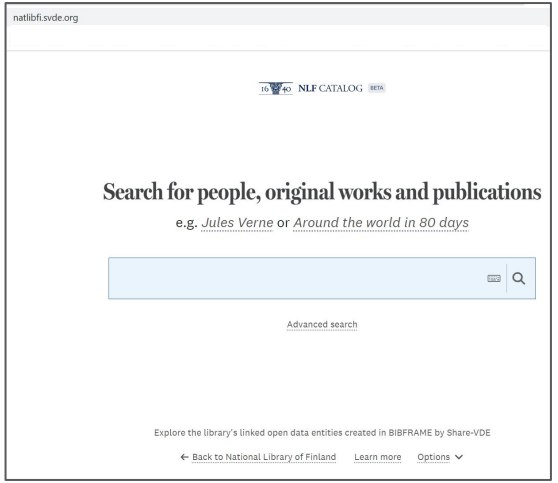
- To this aim, the “held at” filter was added, allowing to filter publications by what is available at the current library. It is enabled on skin portals at Publication (= Instance) level in these cases:
  - in advanced search, see e.g. NYU data pre-filtered here https://nyu.svde.org/advanced-search/publications?q=(title+does_not_contain+xyz)&heldAtLibrary=true (see the toggle on the right of the screen, you can turn it on / off)
  - in the Original work entity page that lists Publications, see e.g. https://nyu.svde.org/suite-de-la-mancha-flute-cello-piano-unknown-author-o781654264663247/library-publications (see the toggle on the right of the screen, you can turn it on / off)
  - in simple search results in cases where the simple search default on the home page is the Publication simple search (e.g. Natbib tenant)

Example: institutional skin portals in SVDE

- **SVDE tenant** - https://svde.org => with LC's authority data and the bibliographic data of member institutions
  - skin portals including: Penn, Smithsonian, Stanford, University of Alberta, New York University, National Library of Norway, National library of Finland (other skin portals will be set up following the load of libraries' catalogues to svde.org)







Future steps in other domains

The Share Family institutions and collaborative networks of libraries are engaging in discussions to establish specialized shared discovery environments, such as Share Art and Share Music.

Main steps to follow:

- Adapt the discovery interface to accommodate domain-specific objects and their respective metadata standards.
- Integrate relevant standards from the specific content domains into the Share Family technology and extend the Share Family ontologies to support diverse materials.
- Foster interconnections among collections from member institutions to enhance discovery options, facilitating cross-referencing between library resources and related materials within each domain.
- Uphold the core principles of the Share Family, emphasizing cooperation and member-driven participation across all initiatives.

Working Groups and cooperation

Share-VDE and Share Family Working Groups

Member institutions of Share-VDE and Share Family Working Groups and parallel projects are constantly contributing with their Subject Matter Experts to requirements gathering, functional analysis and feedback to developments.

Share Family Working Groups:

- National Bibliographies Working Group involving SVDE members and external institutions
- Italian group for the conversion UNIMARC - BIBFRAME
- discussions in the field of photo libraries and audio-visual collections

Share-VDE Advisory Council and Working Groups:

- Share-VDE Advisory Council
- Sapientia Entity Identification WG
- Authority/Identifier Management Services WG (currently on hold)
- Cluster Knowledge Base Editor WG (currently on hold)
- User experience/User Interface WG

Sapientia Entity Identification WG

The SEI WG reviews the use of entities, identifiers, and associated modelling in the Sapientia CKB; evaluates and refines processes for Sapientia entity clustering in Share-VDE and the creation of associated open and stable URI for use in Share-VDE and in the library community; reviews MARC to BIBFRAME and BIBFRAME to MARC conversion; engages with the library community to outline and/or develop best practices for use of Sapientia identifiers in BIBFRAME and MARC data.

Among the latest outcomes:

- definition of the SVDE Ontology; see also Jim Hahn's presentation at the BFWE 2023;
- svde:Work is subclass of bf:Work → this ensures interoperability;
- review of clustering and conversion rules;
- cooperation in the IFLA context: the mapping UNIMARC-BIBFRAME is being prepared and a formal liaison with SVDE has been approved by the IFLA Bibliography Section Standing Committee.

Share-VDE Ontology

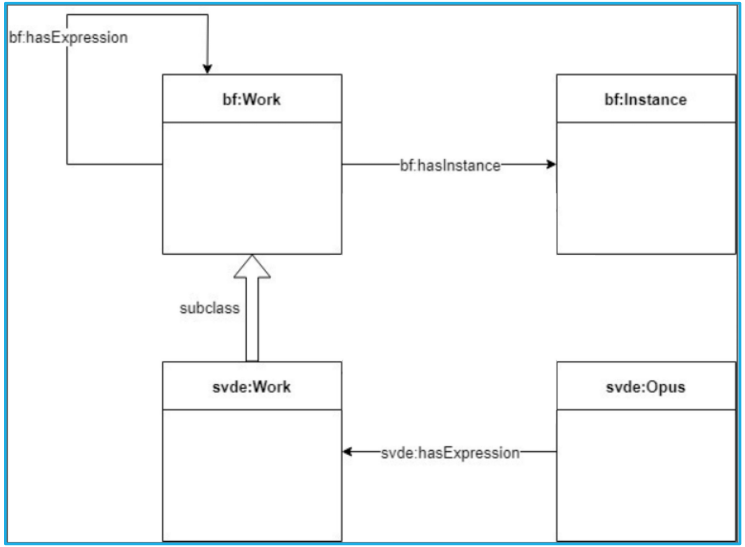
SVDE Ontology designed in SEI-WG as an extension for BIBFRAME.

Core model:

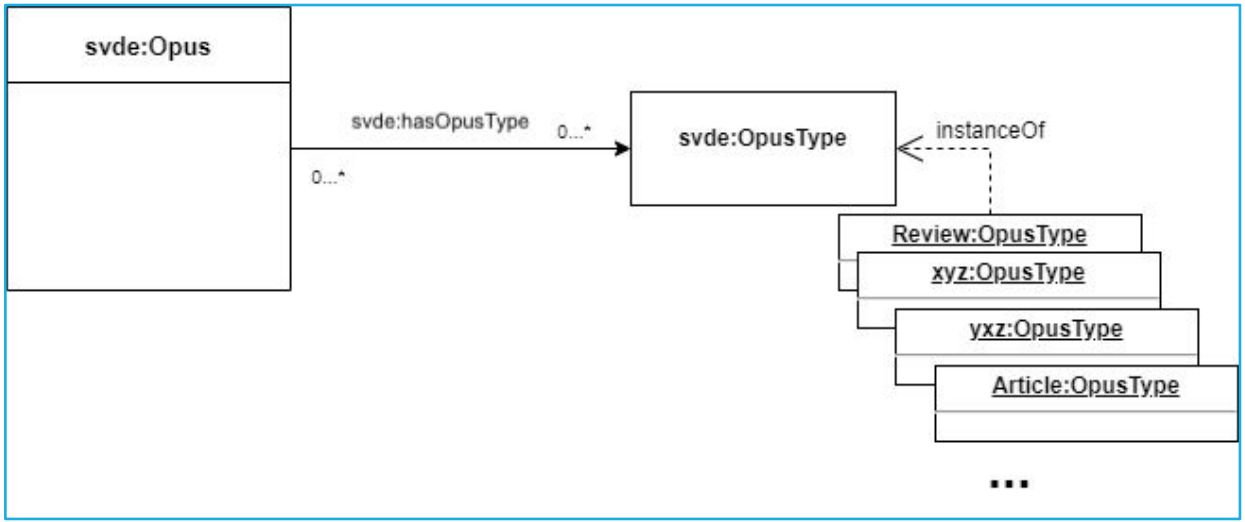
svde:Work,

svde:Opus,

svde:hasExpression



Share-VDE Ontology



Share-VDE Ontology re-use

While the ontology supports the discovery functionality of Share-VDE and the Share family search systems, it may be re-used in any system requiring a bridge among BIBFRAME, IFLA LRM and RDA.

Share-VDE Ontology Interoperability

The Share-VDE ontology achieves interoperability among the major bibliographic models by asserting that bibliographic entities are described by attribute sets.

The **attribute set modeling** approach is a departure from the conceptual modeling that has informed the development of nearly all modern linked data models.

Attribute set modeling

The **svde:Opus** is a parallel class to the IFLA LRM Work and the RDA Work. The set of attributes which comprise svde:Opus parallel those at- tributes in the IFLA LRM Work and the RDA Work.

The **svde:Work i**s a subclass of the BIBFRAME Work. The set of attributes which comprise the svde:Work parallel those attributes in the ILFA-LRM Expression and the RDA Expression.

Share-VDE Ontology takeaways

Direct entity mapping of the familiar and ubiquitous conceptual approach was not utilized to achieve Share-VDE ontology concordances – rather, minimal ontological commitment is made by observing the set attributes that define an entity.

Each linked data model, be it RDA, BIBFRAME or IFLA LRM, has a useful perspective, and each of these contribute to the task of bibliographic description.

Towards an operational environment

An integrated and hybrid environment

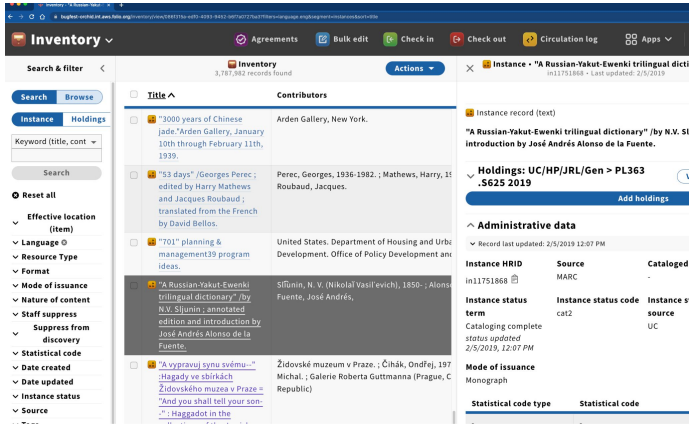
The mutual exchanges in the BIBFRAME / linked data community are bringing the Share Family towards:

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

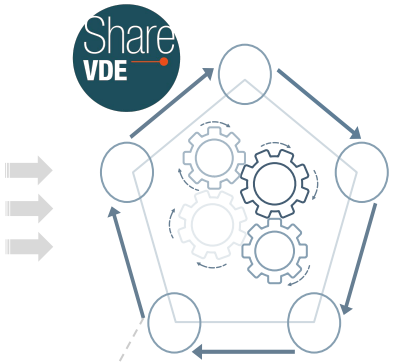
From Library Data to Share-VDE - From records to entities

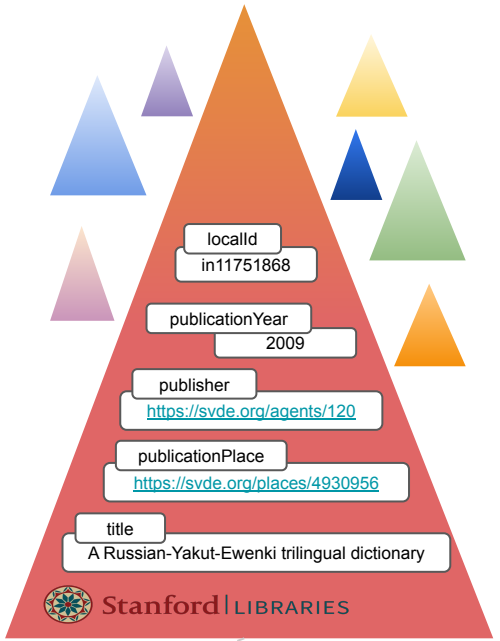
A Share-VDE member uses a local ILS/LSP for managing its data.

Stanford | LIBRARIES



The local record (or records in case of massive export) is sent to Share-VDE

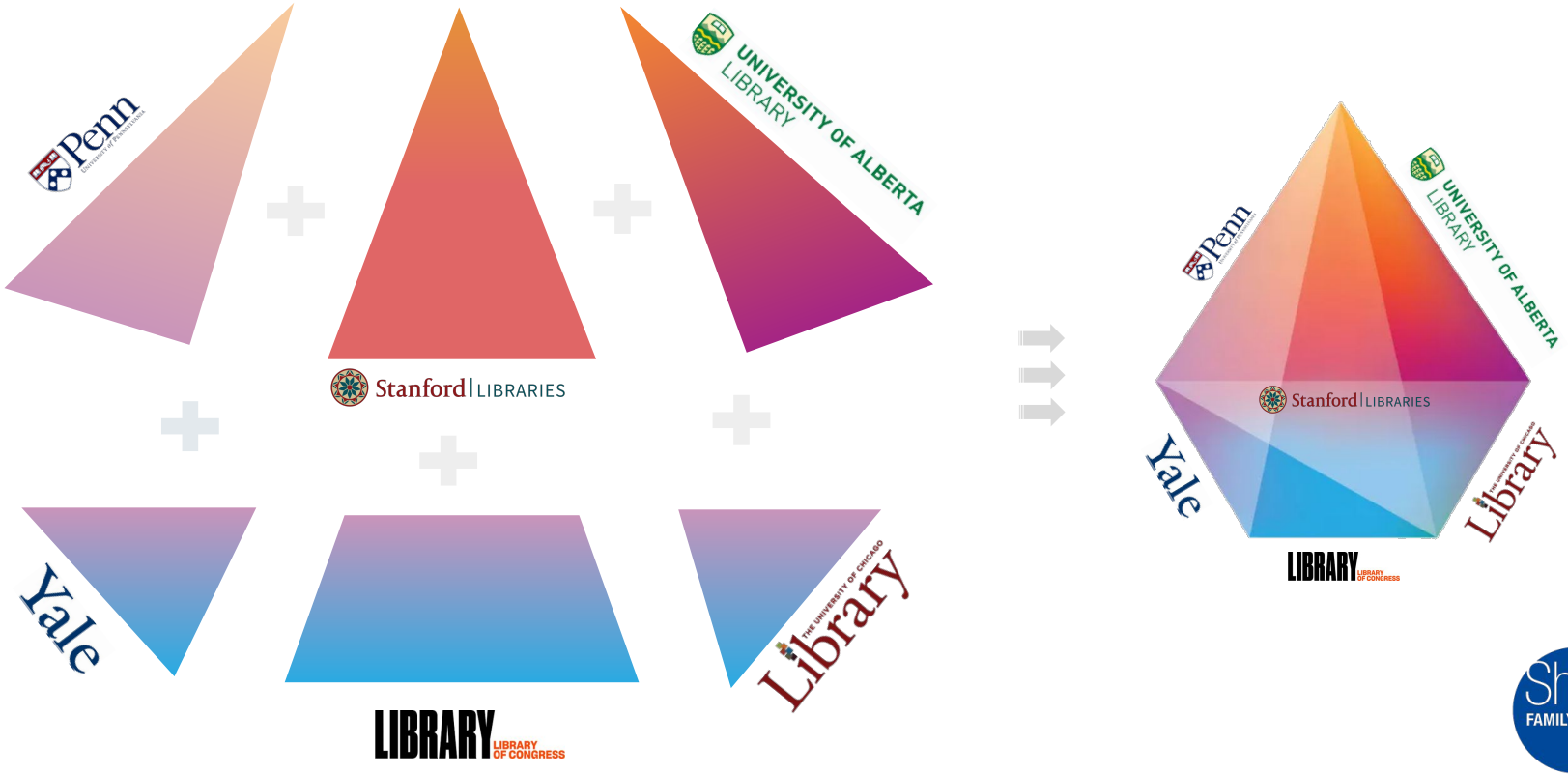




The original record (usually in Marc but also in other formats) is split across the entities that form the Share-VDE domain model. In this example we focus on the properties that are assigned to a Share-VDE instance (red triangle above)

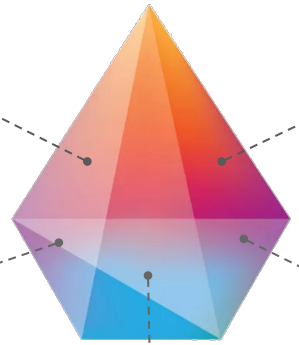
Share FAMILY

Prism, faces: the Share-VDE Entity



Share FAMILY

Faces (aka Contributions & Provenances)



UNIVERSITY OF ALBERTA LIBRARY

title	Alice in wonderland
titleAlternative	Alice's adventures under ground
author	https://svde.org/people/201

Stanford

title	Alice in wonderland
titleAlternative	Alice's adventures under ground
author	https://svde.org/people/201

LIBRARY OF CONGRESS

title	Alice in wonderland

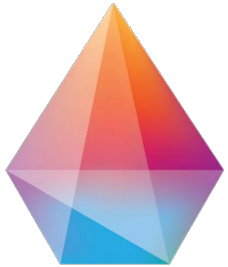
National Library of Norway

title	Alice's adventures under ground
titleAlternative	Journeys in Wonderland
author	https://svde.org/people/201

(links)

sameAs	http://dbpedia.org/resource/Alice's_Adventures_in_Wonderland	Dbpedia
sameAs	https://www.wikidata.org/wiki/Q189875	Wikidata
sameAs	https://data.bnf.fr/ark:/12148/cb358500385#about	bnf

Properties: Attributes, Relationships, Links



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

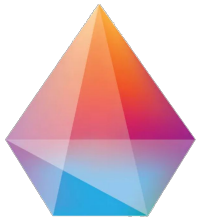
An attribute is a data property, having a literal as value

Share VDE		
sameAs	http://dbpedia.org/resource/Alice's_Adventures_in_Wonderland	Dbpedia
sameAs	https://www.wikidata.org/wiki/Q189875	Wikidata
sameAs	https://data.bnf.fr/ark:/12148/cb358500385#about	bnf

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

Name	Provenance
author	LIBRARY Stanford

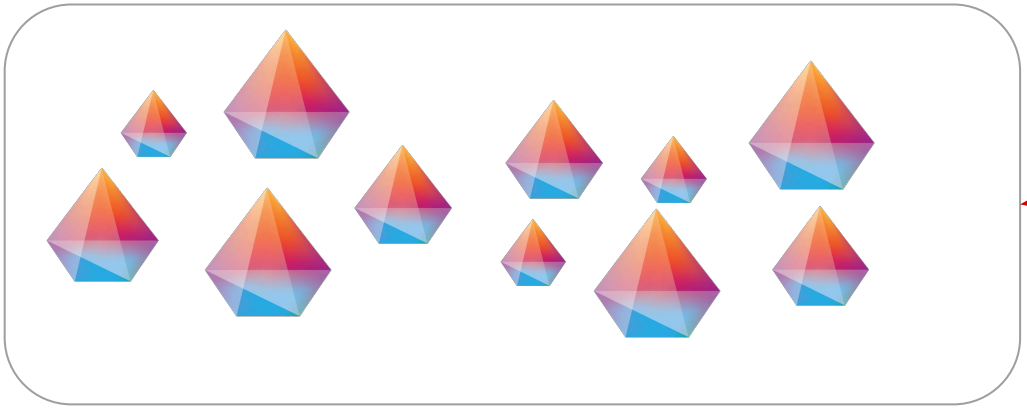
Name	Value	Provenance
name	Carroll, Lewis	LIBRARY Stanford
nameAlternative	Dodgson, Charles Lutwidge	LIBRARY Stanford
nameAlternative	Karol, Luis	National Library of Norway



A relationship is a connection between two Share-VDE Prisms

Sapientia - The Cluster (entities) Knowledge Base

The Share-VDE knowledge base (**Sapientia**) contains the integrated/clustered/enriched entities.



JCRICKET
Entity Editor

The Entity editor: how humans can improve the quality of machine data

Edit: a property is added/updated/deleted

~~Lewisss Carroll~~ → Lewiss Carroll

is author of ~~https://svde.org/opuses/1827349~~ → https://svde.org/opuses/920302

Invalidate

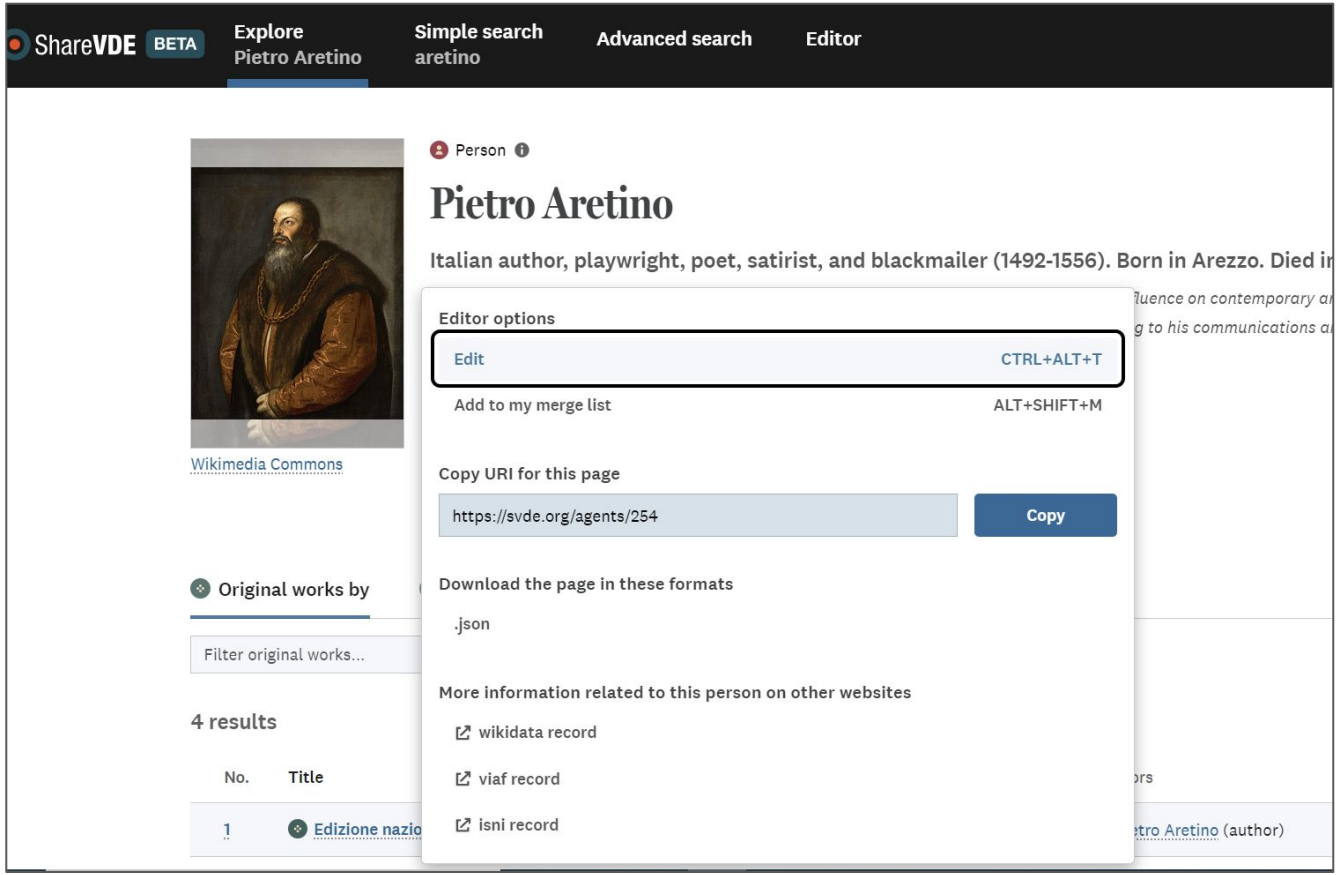
Merge: multiple prisms are merged into one

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

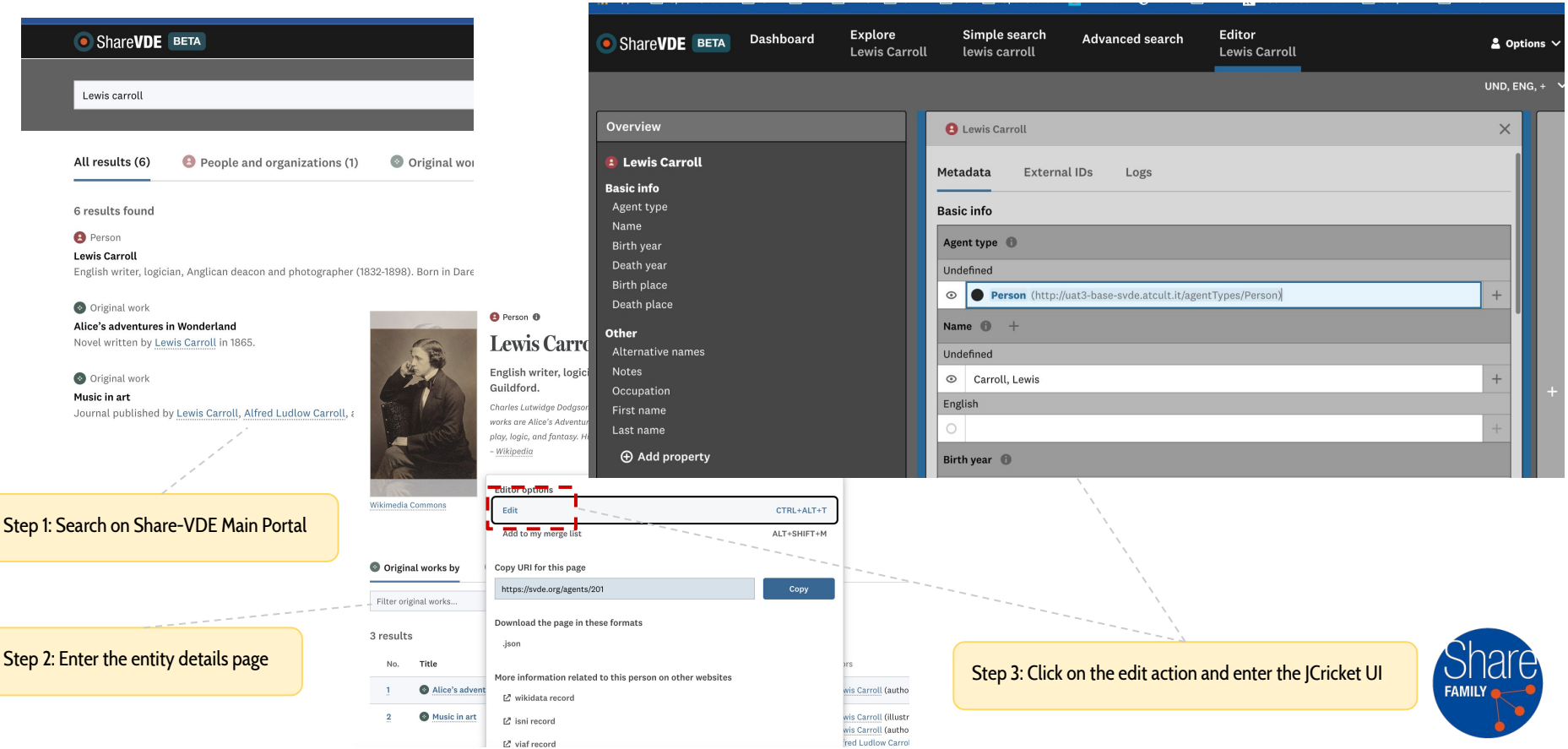
Split: a prism is split into multiple prisms

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

JCricket: enhance data quality and authority control



JCricket user interface



Step 1: Search on Share-VDE Main Portal

Step 2: Enter the entity details page

Step 3: Click on the edit action and enter the JCricket UI

Collaboration for enhancing library workflows and services

Member libraries send their records

both for their community of users and librarians

Sapientia Cluster Knowledge Base, a **collaborative source of high quality data**: the CKB includes the clusters of entities created in the reconciliation and conversion to linked data of the catalogues of all Share-VDE participating libraries

Data in the CKB is enhanced by JCricket Entity Editor

According to their local or network policies, libraries can implement their **services**

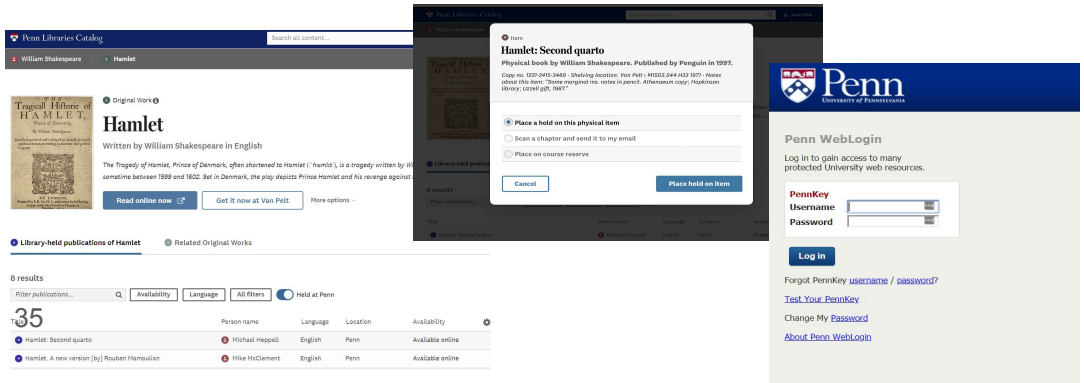
SERVICE

Integration with Other Systems

- Development of APIs for interoperability and cooperation with local LSPs and third parties (including FOLIO, Wikidata, LD4P - Linked Data for Production)

- discovery portal
- authoritative services
- shared cataloguing
- statistics and evaluation
- collection development
- interlibrary loan

and so on!



Focus on Authority services

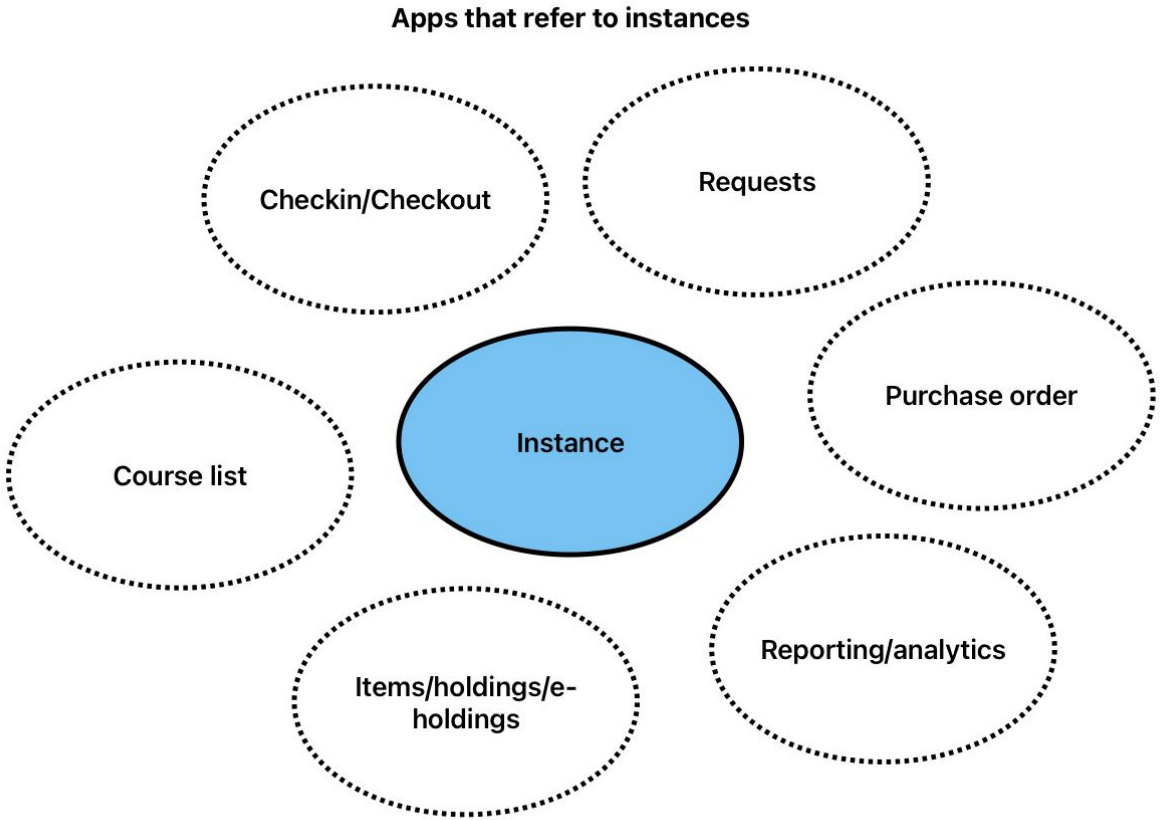
Automatic services for Share Family libraries, piloted by Stanford University:

- 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.

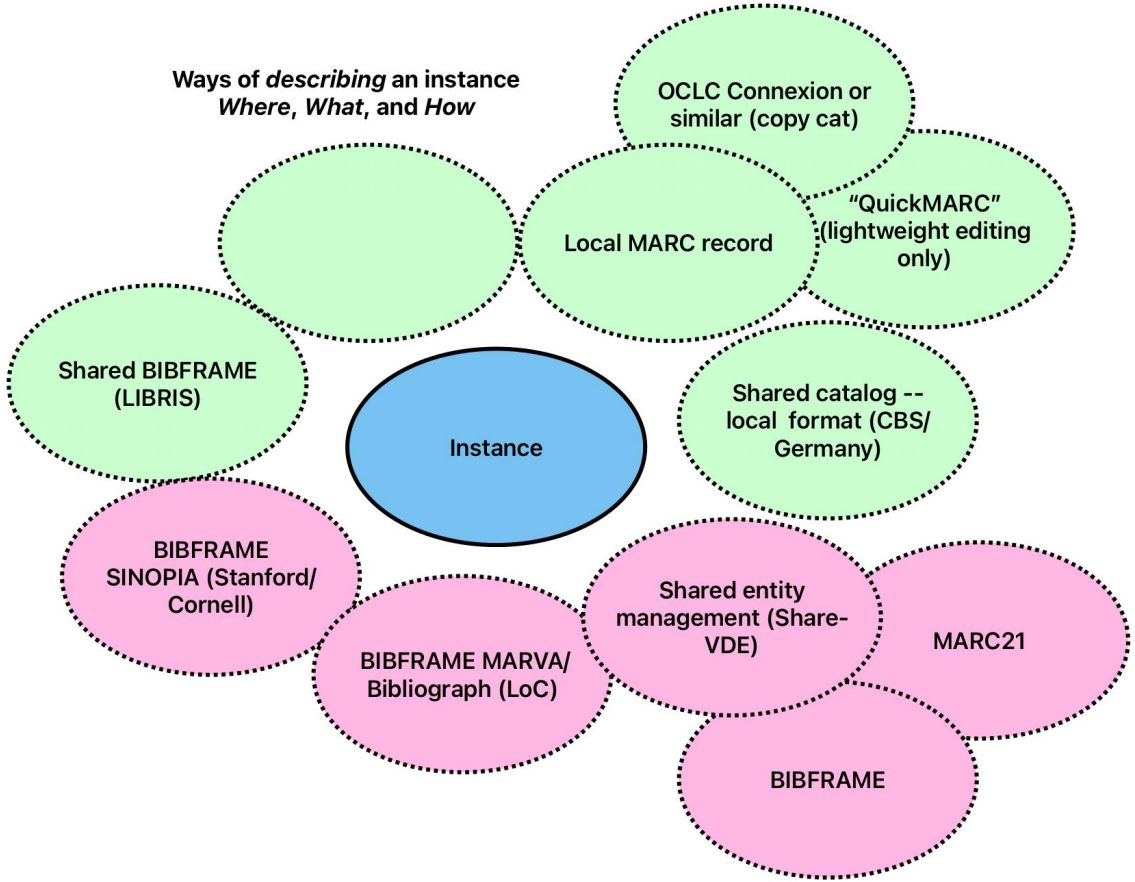
Next step developments: Authority Services fully integrated in the Linked Open Data environments.

Third party integration
FOLIO as a case study

FOLIO App ecosystem



Types of descriptive metadata ecosystems



SINOPIA integration: high-level milestones

✅ set-up the connector to fetch data from Sinopia

✅ ingested subset of Sinopia data from Stanford

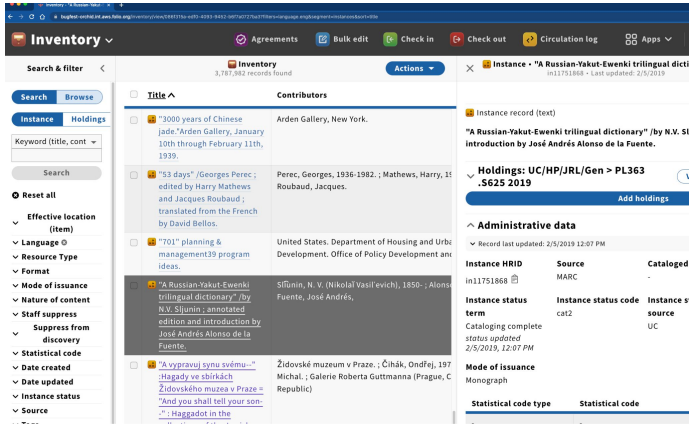
🚧 now creating the parser so that RDF data coming from Sinopia can be clustered by Share-VDE processes

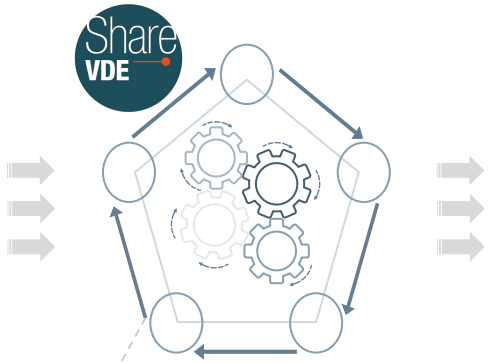
🎯 at the end of this process, Sinopia data will be included in the Share-VDE CKB - Cluster Knowledge Base

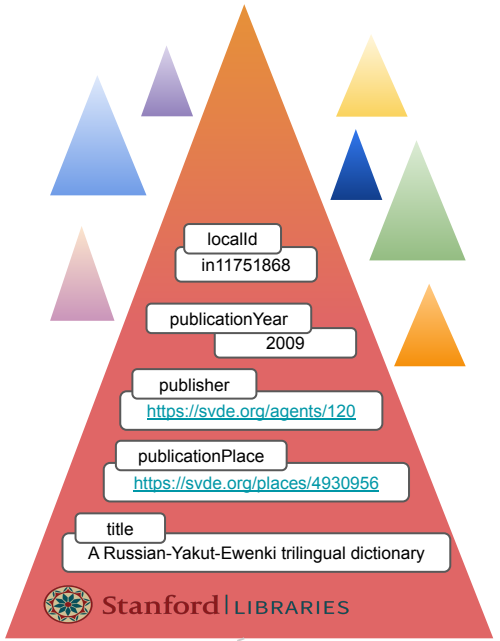
From Library Data to Share-VDE - Integration with folio

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

Stanford | LIBRARIES





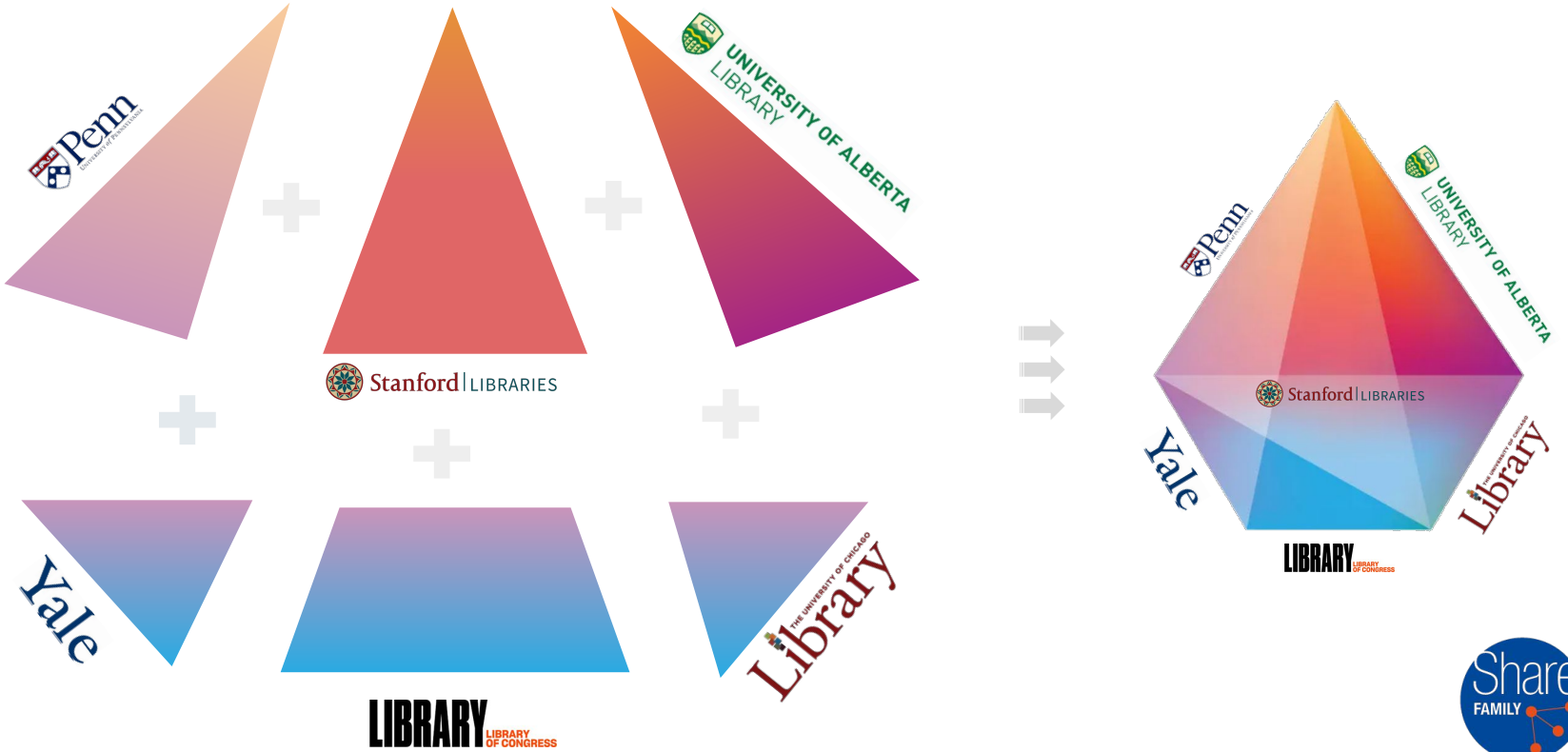


FOLIO instance (or instances in case of bulk export) is sent to Share-VDE

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)

Share FAMILY

Prism, faces: the Share-VDE Entity



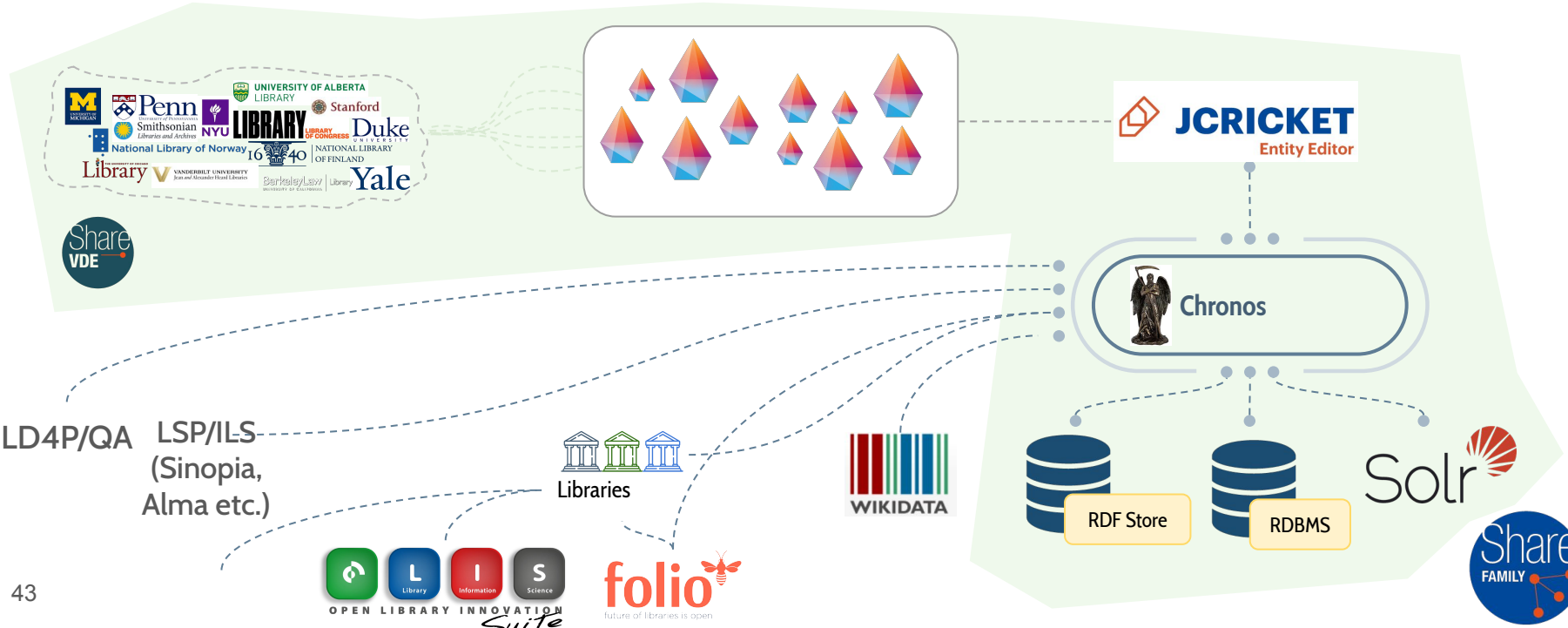
Share FAMILY

Third party integration - Outbound Connectors Architecture

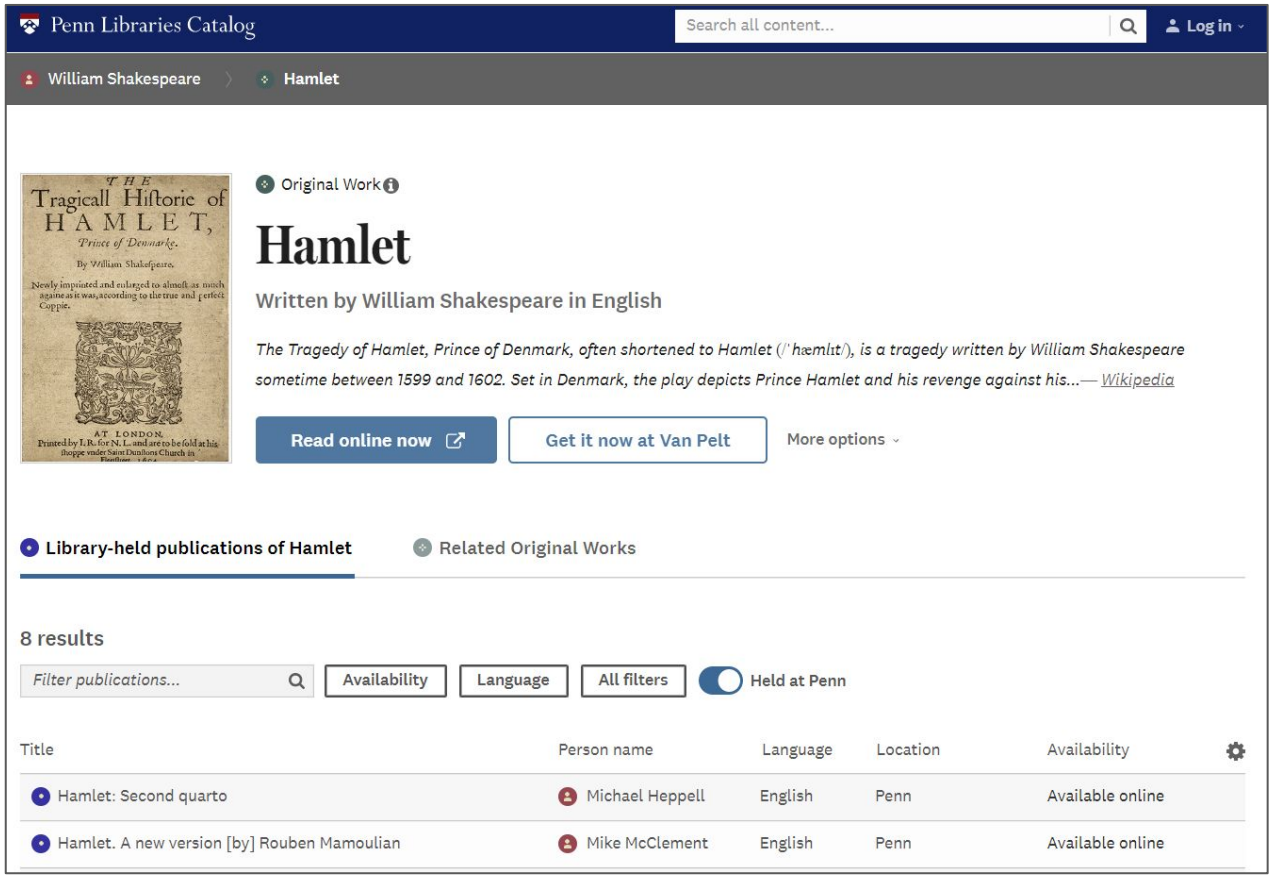
Data flows into Share-VDE from libraries, institutions and third-party sources (e.g. VIAF, ISNI, FAST)

The Share-VDE knowledge base (**Sapientia**) contains the integrated/clustered/enriched entities.

Data is mainly edited through JCricket, the Share-VDE entity editor.

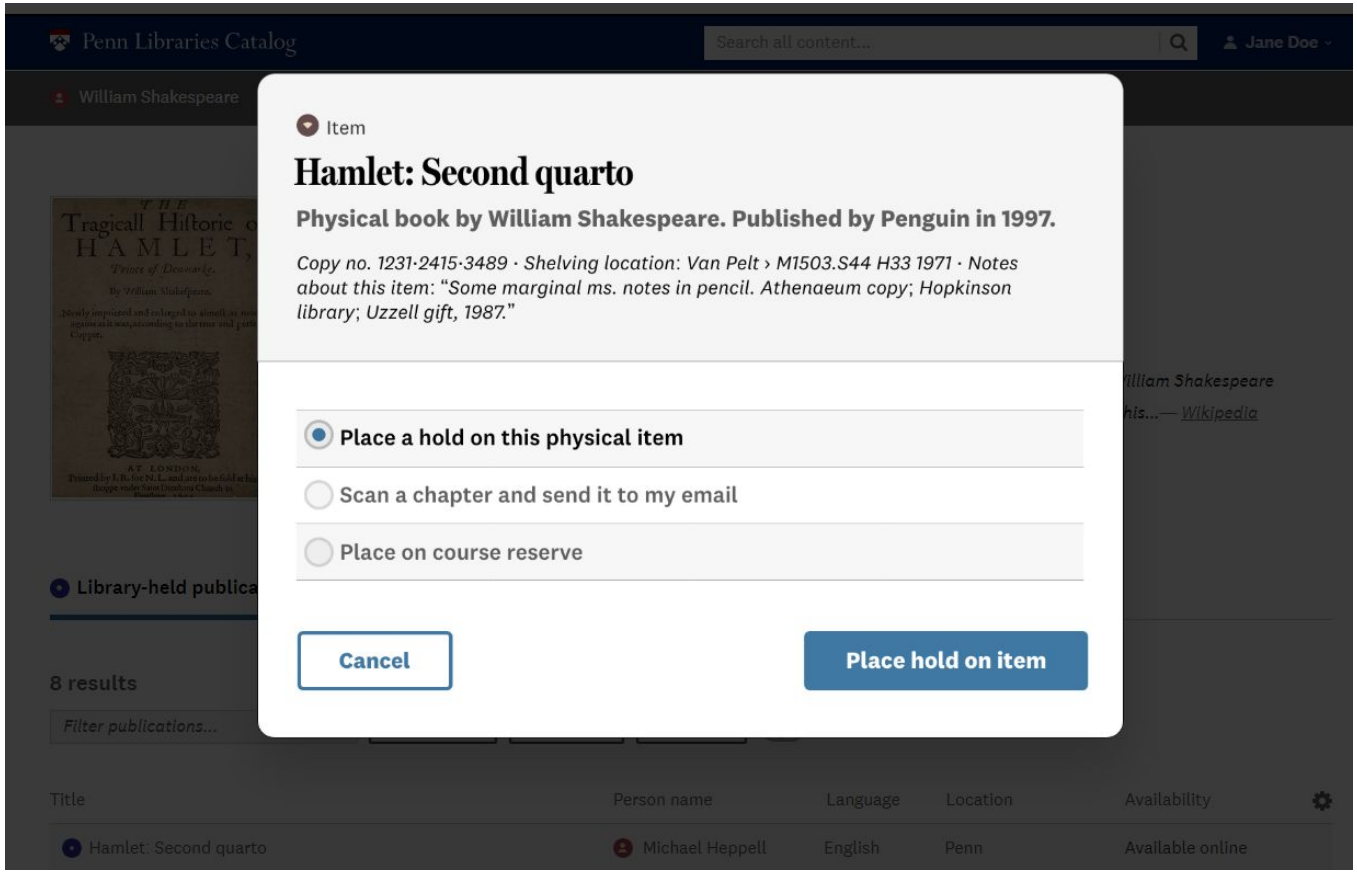


Integration with local services - circulation info



Integration with local services, e.g. connection to Alma APIs for [University of Pennsylvania circulation services]

Integration with local services - lending



folio integration: high-level milestones

[See a possible model for ILS/LSP interaction through FOLIO]

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

Beyond local workflows?

The need for bibliographic infrastructure is not limited to local library operations.

Consider resource sharing, collaborative collection lifecycle management, reporting.

Well-established, open bibliographic identities are critical for libraries to work together effectively.

Thank you!

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