Share-VDE and the Share family

A shared and integrated ecosystem
for library linked data

Share-VDE as a library-driven community

Share-VDE in a nutshell

Since 2016, R&D work to facilitate libraries in the transition from MARC-based cataloguing to linked data

this expanded over time from the pilot project to Share-VDE and the Share Family of initiatives

https://svde.org
https://wiki.svde.org/
Casalini Lab
Share - Linked Data Environment

What Share-VDE does:
MARC data (or other traditional formats) are converted to linked data

data describing library resources are connected in a union catalogue, and can be queried as authoritative source

exposition for end users and professionals on the web platform www.svde.org

a platform to manage data in a linked open data environment

A cooperative and library-driven initiative

Share-VDE is a collaborative initiative based on the needs of libraries, developed and supported by:

the joint effort of the Share-VDE Advisory Council and of the Working Groups;

Casalini Libri, provider of bibliographic and authority data as member of the Program for Cooperative Cataloguing;

@Cult, provider of ILS, Discovery tools and Semantic web solutions for the cultural heritage sector;

the vision of Linked Data for Production initiative with special endorsement of Stanford;

with input and active participation from an international group of research libraries.

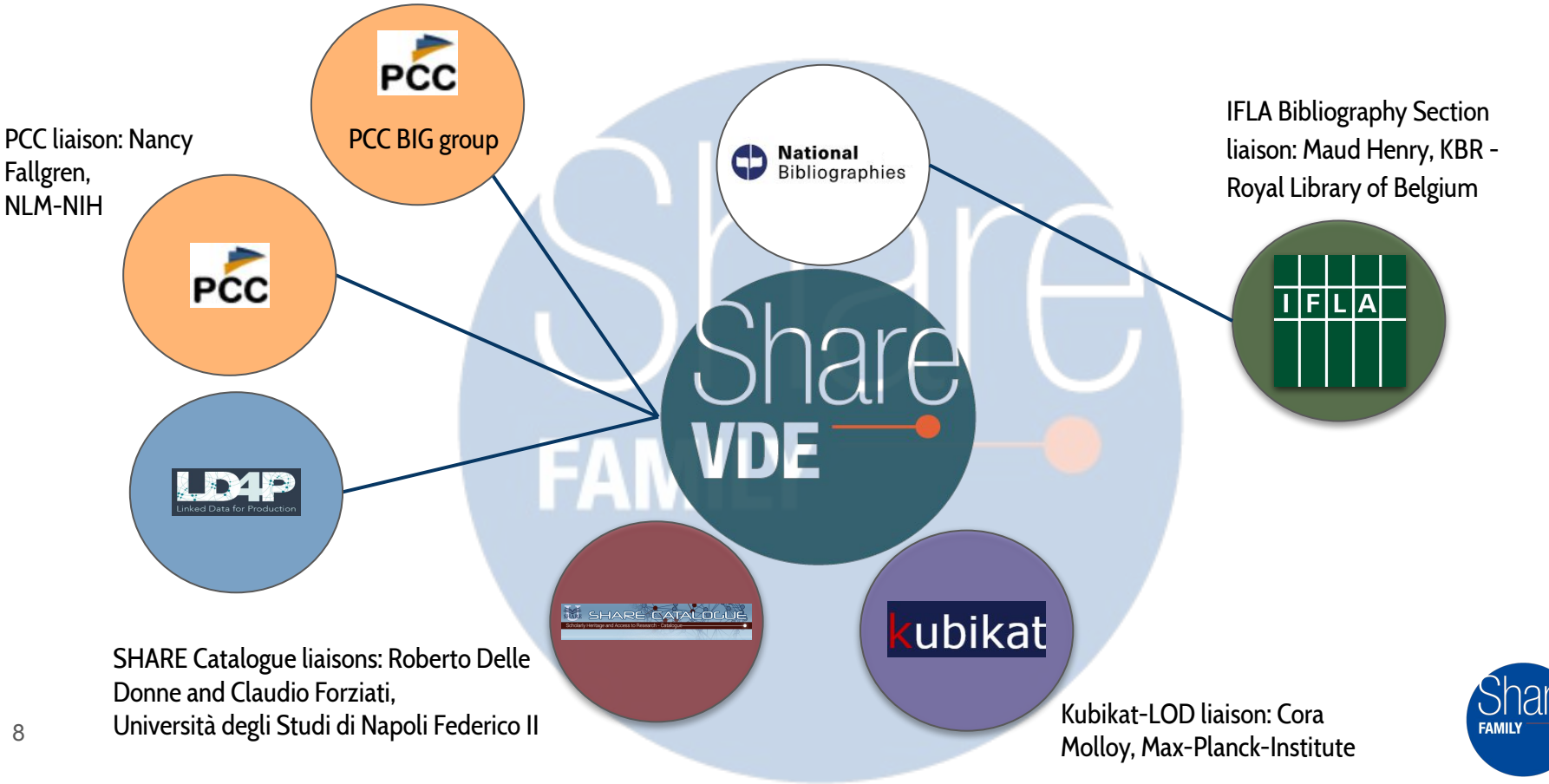

A collaborative community can produce a gravitational wave of energy that expands...


… and meets other
communities, with
other energy to Share

Active participation

Libraries members of SVDE 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 and Share-VDE liaisons



Community engagement: library community

Extended community: collaboration with heterogeneous initiatives and institutions in the library domain

Scientific value: sharing of data and services in different technological environments and diverse bibliographical and cultural context

Community engagement: World Wide Web

Mixed community: cross-domain cooperation across the Web community

Scientific value: same solutions serve scopes of different communities, data reuse

Share Family community work and outcomes

Active participation and concrete output

Libraries members 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-VDE Advisory Council and Working Groups:

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

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

The Share-VDE AC takes an active role in determining future uses and vision for the Share-VDE initiative; Develop future use cases for Share-VDE, and set development priorities as needed; Monitor and lead the work of the various Advisory Council Working Groups; Maintain communication among the Share Family member institutions.

Latest outcomes: Share-VDE Statement, September 2021:

- edited and approved by the Share-VDE Advisory Council;
- explanation of position in the broader context of Library Linked Open Data;
- Share-VDE has been a reference point in library linked open data since the initial R&D and prototype phase in 2016;
- cooperation: member libraries have contributed their data and are actively involved in the developments of the initiative.

Authority/Identifier Management Services WG

The AIMS WG defines guidelines and best practices for Authority/Identifier management; defines scope and data-flow for the creation and implementation of automated services based on preliminary documentation; proposes additional use cases identified as essential for effective knowledge base management.

Latest outcomes: new generation of services for the authority control

- definition of use cases;
- functional analysis;
- analysis of interaction with Wikidata and ISNI (joint work with CKBE WG to design J.Cricket functionalities);
- pilot of MARC-based authority services with Stanford University Libraries;
- initial analysis of services for authority control in linked data workflows.

Focus on Authority Services

Services for the authority control that combine automated and manual processes

For record environments:

- validation of MARC bibliographic records (correction of MARC fields and obsolete forms, update of tags and subfields etc.);
- enrichment of MARC fields with SVDE original URIs and URIs from external sources according to ad hoc profiling, including LCNAF, VIAF, ISNI;
    - Casalini Libri is ISNI registration agency creating and assigning ISNI to persons and organisations (e.g. publishers)
- matching processes on external authority files;
- import of authority records;
- reporting features providing complete details of the validation and corrections done to the records.

→ initial release of the authority control features for MARC records delivered to Stanford.

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

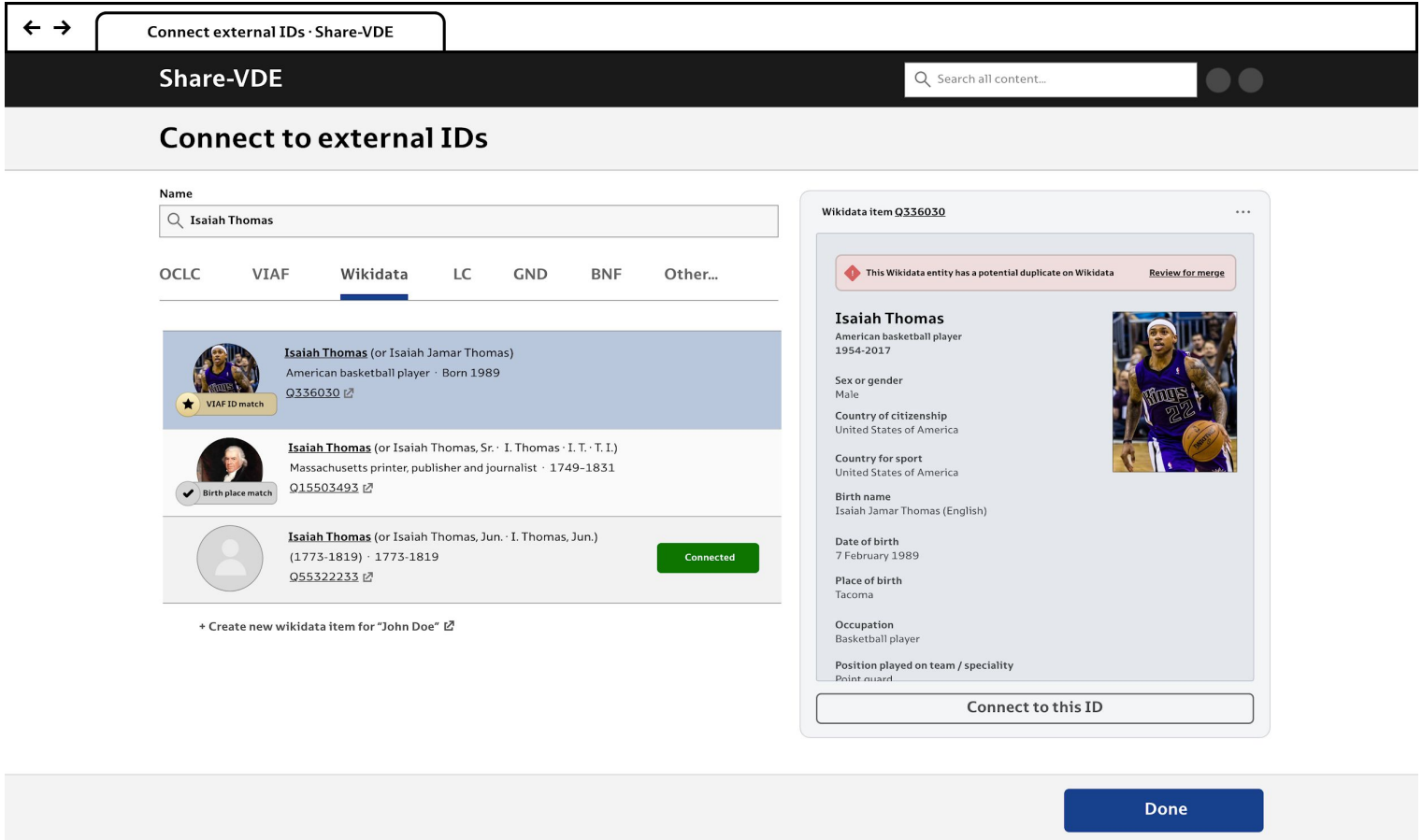
Cluster Knowledge Base Editor WG

The CKBE WG analyses how libraries interact with the *Sapientia* Cluster Knowledge Base (CKB) and their use of the J.Cricket Editor for modifying (correcting / enriching), deleting, merging and separating clusters.

Latest outcomes: back-end developments for J.Cricket entity editor started

- definition of use cases;
- design of manual editing features;
- analysis of interaction with Wikidata and ISNI to be incorporated into J.Cricket and authority dataflows that feed the Cluster Knowledge Base (joint work with AIMS WG to design J.Cricket functionalities);
- back-end developments started; respective front-end features will follow throughout 2022.

How J.Cricket interacts with Wikidata



Sapientia Entity Identification WG

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

Latest outcomes: svde:Instance as entity under definition

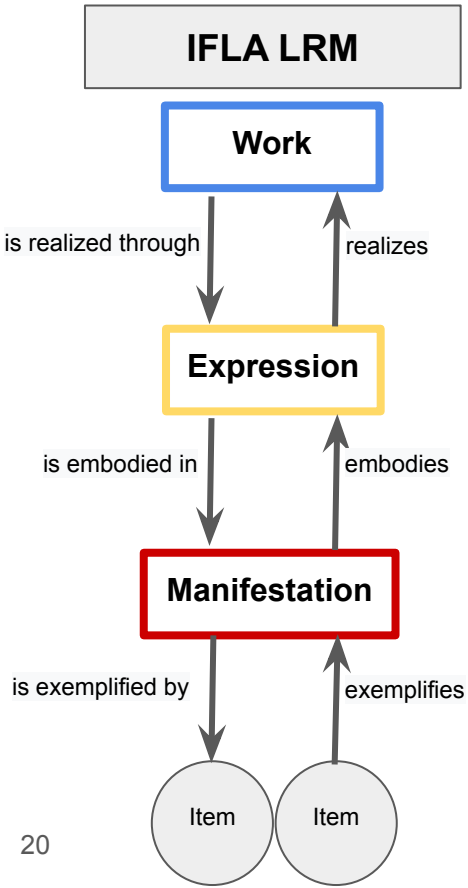
- 4 layers in SVDE entity model: svde:Opus | svde:Work | svde:Instance | svde:Item;
- svde:Opus and svde:Work are types of bf:Work → this ensures interoperability;
- consolidating the definition of svde:Instance entity properties;
- review of clustering and conversion rules.

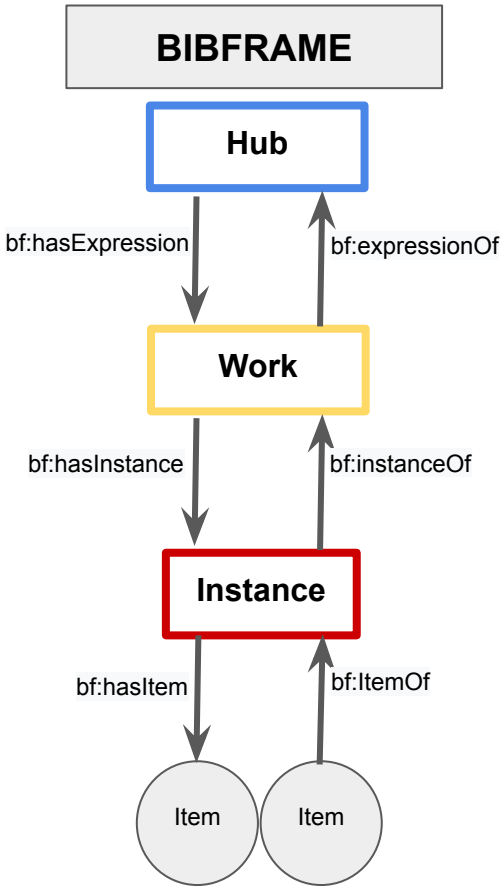
Focus on entity model

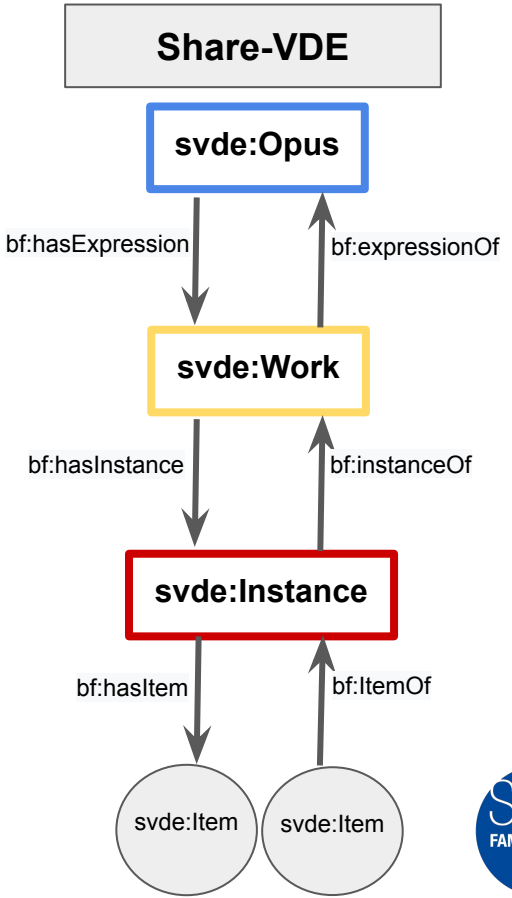
Share-VDE as a BIBFRAME node to put BIBFRAME into practice:

- Share-VDE provides enriched data that is interoperable with other BIBFRAME nodes and with other models;
- the Share-VDE working groups have reviewed algorithms and processed, and expanded the BIBFRAME model to meet real-world needs;
- focus on cooperation also 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.

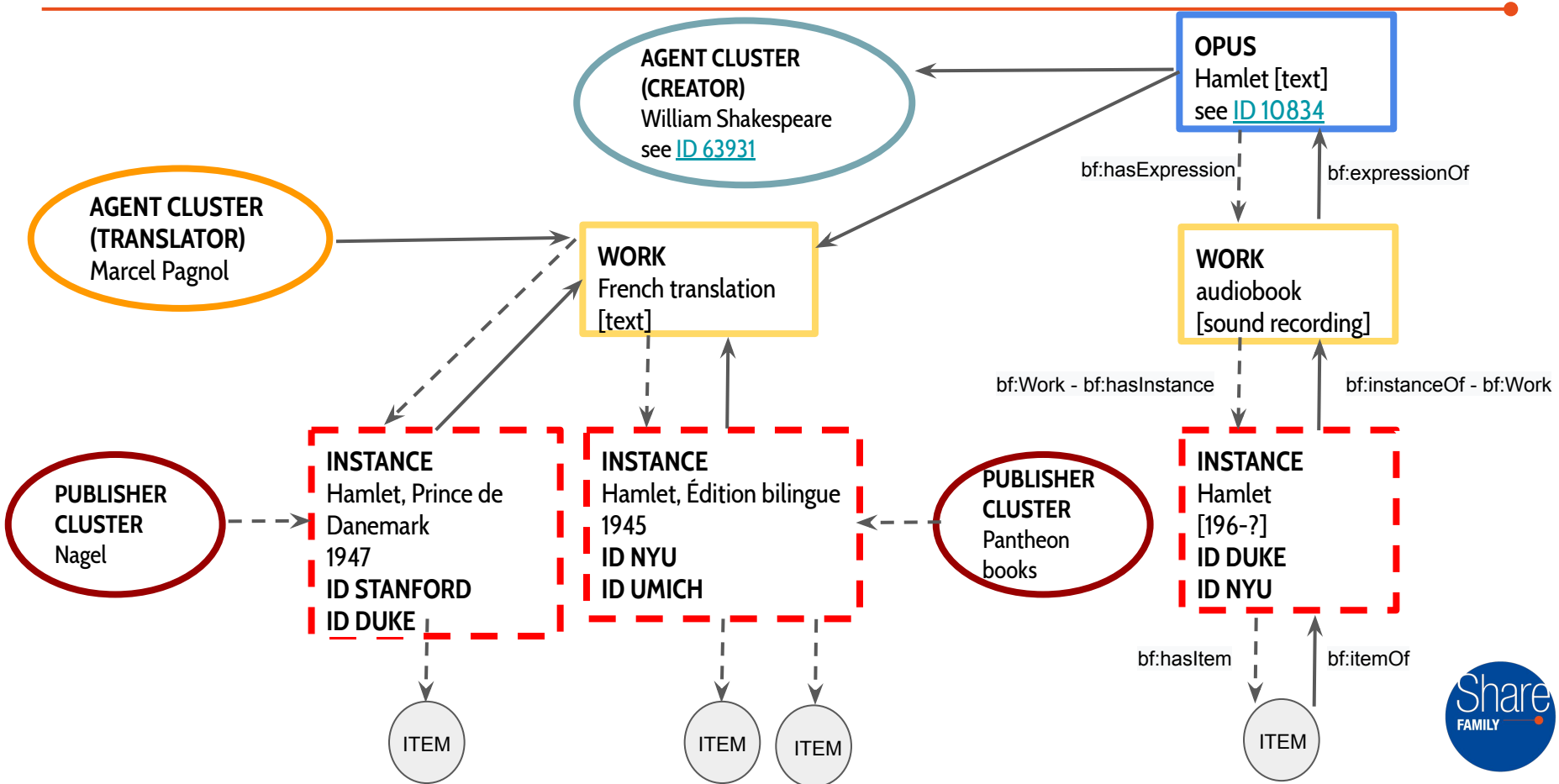
Comparison IFLA-LRM/BIBFRAME/Share-VDE







Share-VDE model (simplified version)



User Experience/User Interface WG

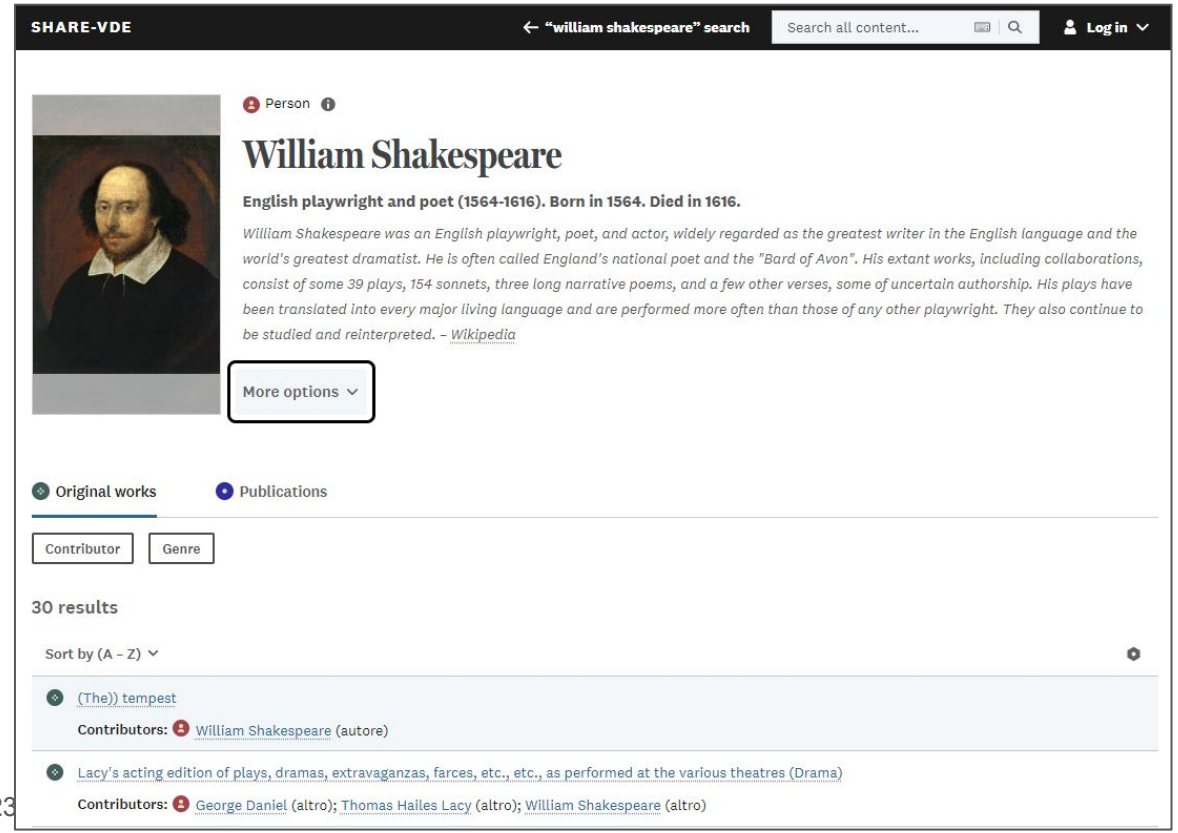
The UX-UI WG has re-designed Share-VDE user interface to respond to both patrons and library staff requirements and expectations. The new interface:

- reflects the components of the Share-VDE data model infrastructure;
- harnesses the potential of linked data and deliver wide-ranging and detailed search results;
- provides an intuitive user experience hiding the complexity of the underlying data model;
- embeds partner APIs for the interoperability with local library services (e.g. lending);
- allows dedicated skin portals (e.g. Penn the branding of the institution.

Latest outcomes: Share-VDE 2.0 Entity Discovery https://svde.org

- new Entity Discovery Portal and new back-end infrastructure for the Linked Data Management;
- other Share Family discovery portals supported by the same technology.

Focus on Share-VDE 2.0 Entity Discovery

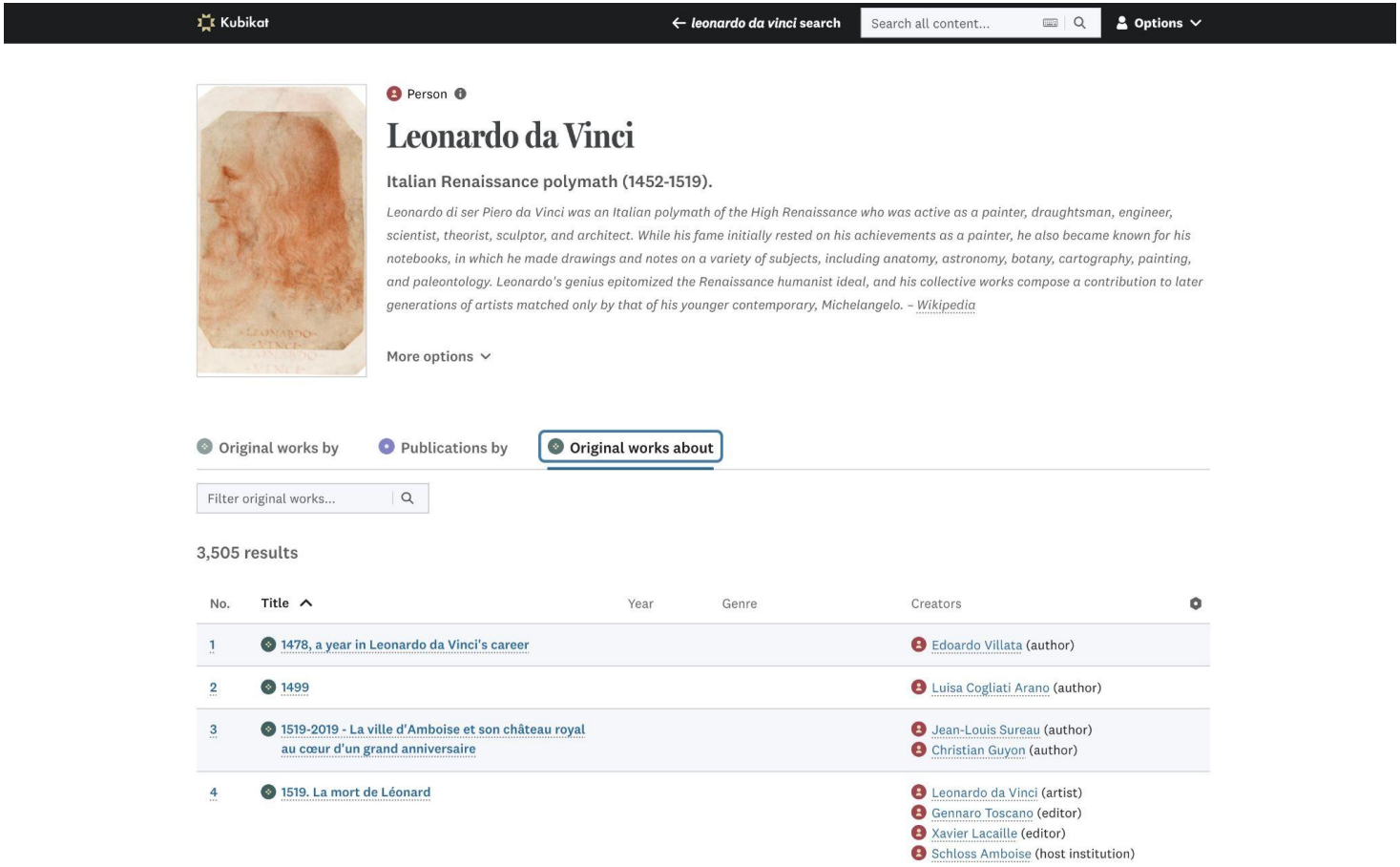


A (much more) complex system with entity-based presentation layer, reflecting BIBFRAME and the ad hoc SVDE extensions

Improved user experience

Back-end infrastructure based on APIs and enhanced with a new version of the [LOD Platform framework] and of the CKB

Entity Discovery: data enrichment from external sources



Entity Discovery: data enrichment from external sources

Volume

Inferno

First part of Dante's Divine Comedy. 1314 fiction.

Inferno is the first part of Italian writer Dante Alighieri's 14th-century epic poem Divine Comedy. It is followed by Purgatorio and Paradiso. The Inferno describes Dante's journey through Hell, guided by the ancient Roman poet Virgil. In the poem, Hell is depicted as nine concentric circles of torment located within the Earth; it is the "realm ... of those who have rejected spiritual values by yielding to bestial appetites or violence, or by perverting their human intellect to fraud or malice against their fellowmen". – Wikipedia

This is part of the series Penguin epics, Norton critical edition, Penguin classics, and Hesperus poetry.

More options

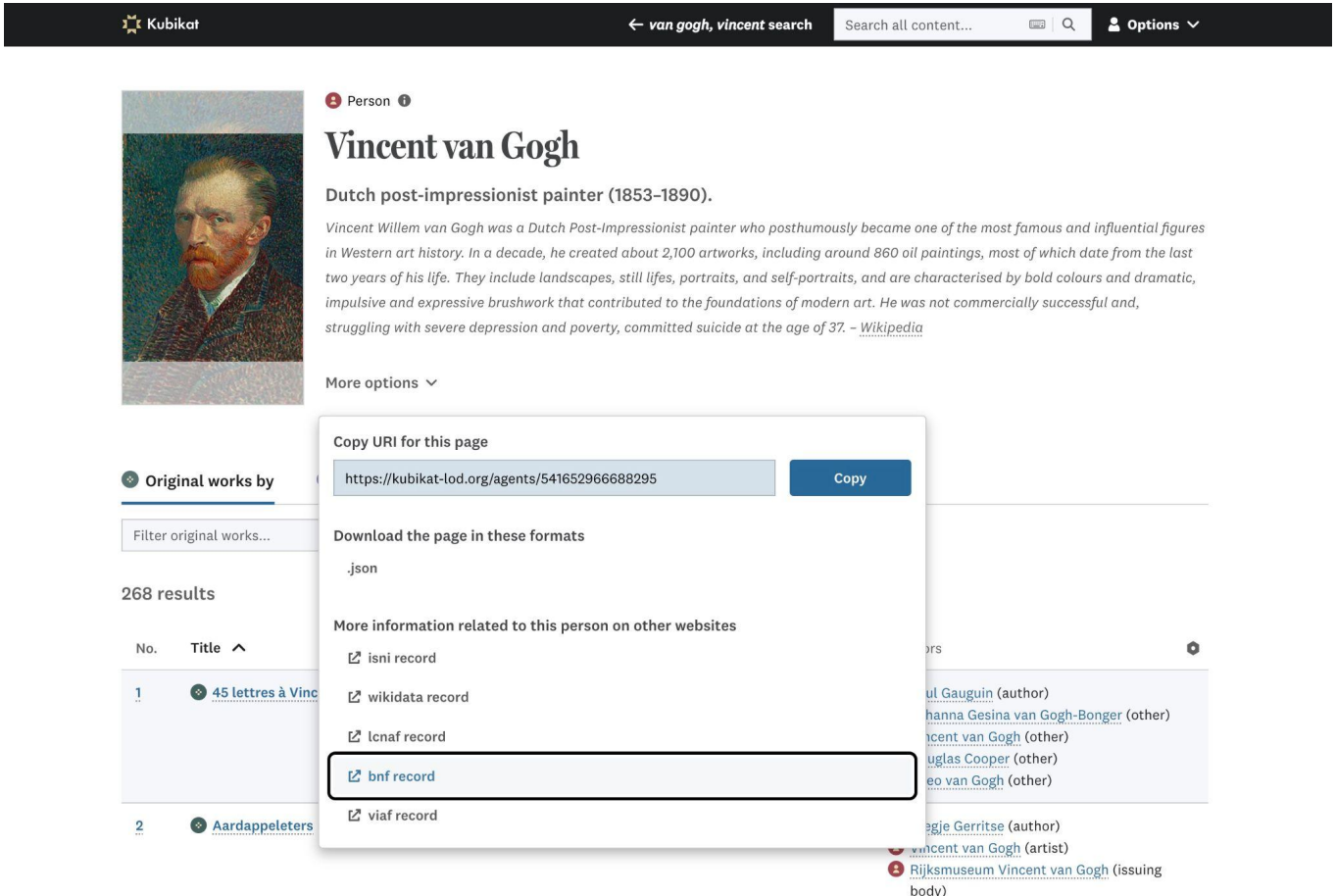
Library-held publications of this | Related agents

Filter publications... | Publication year | Publication place

49 results

Title	Language	Publication year	Format	Contributors
The inferno	English	2015		Stefano Albertini (editor)

Entity Discovery: data enrichment from external sources



National bibliographies WG

The National bibliographies WG is dedicated to the practical cooperation among the National Bibliographies, to address the needs of National Libraries and institutions that hold National Bibliographies in the framework of a shared entity discovery environment such as the Share Family of initiatives.

Main updates:

- involvement of SVDE / Share Family members and external institutions;
- IFLA Bibliographic Section liaison (Maud Henry from KBR - Royal Library of Belgium);
- discussion around topics of interest for an ad hoc tenant hosting national bibliographies;
- main tenant of the shared discovery environment for national bibliographies: https://natbib-lod.org/
- analysis of use cases for specific features.

National bibliographies WG - latest outcomes

- Study and address the needs of institutions that hold National Bibliographies WRT linked data platforms;
- goal: build a shared discovery environment hosting LOD National bibliographies dataset;
- British Library is early adopter: the British National Bibliography will be the first national bibliography tenant for the Share-VDE virtual discovery environment
  - **National Bibliographies tenant** - https://natbib-lod.org
  - with the skin for the British National Bibliography https://bl.natbib-lod.org (*Note: the skin for the British National Bibliography is a preview of a beta site)*
- the group is currently analysing use cases for ad hoc features of the shared National Bibliographies portal.

Share Family tenant infrastructure

The Share Family core principles

- Redundancy of data is complex to manage
- Linking entities is easier than duplicate data
- Cooperate and maintain autonomy at the same time
- Homogeneity of datasets and possible services to be shared
- Centralize core data through a lightweight method
- Distribute the technologic load to achieve long-term sustainability
- Profile levels of cooperation among systems and initiatives

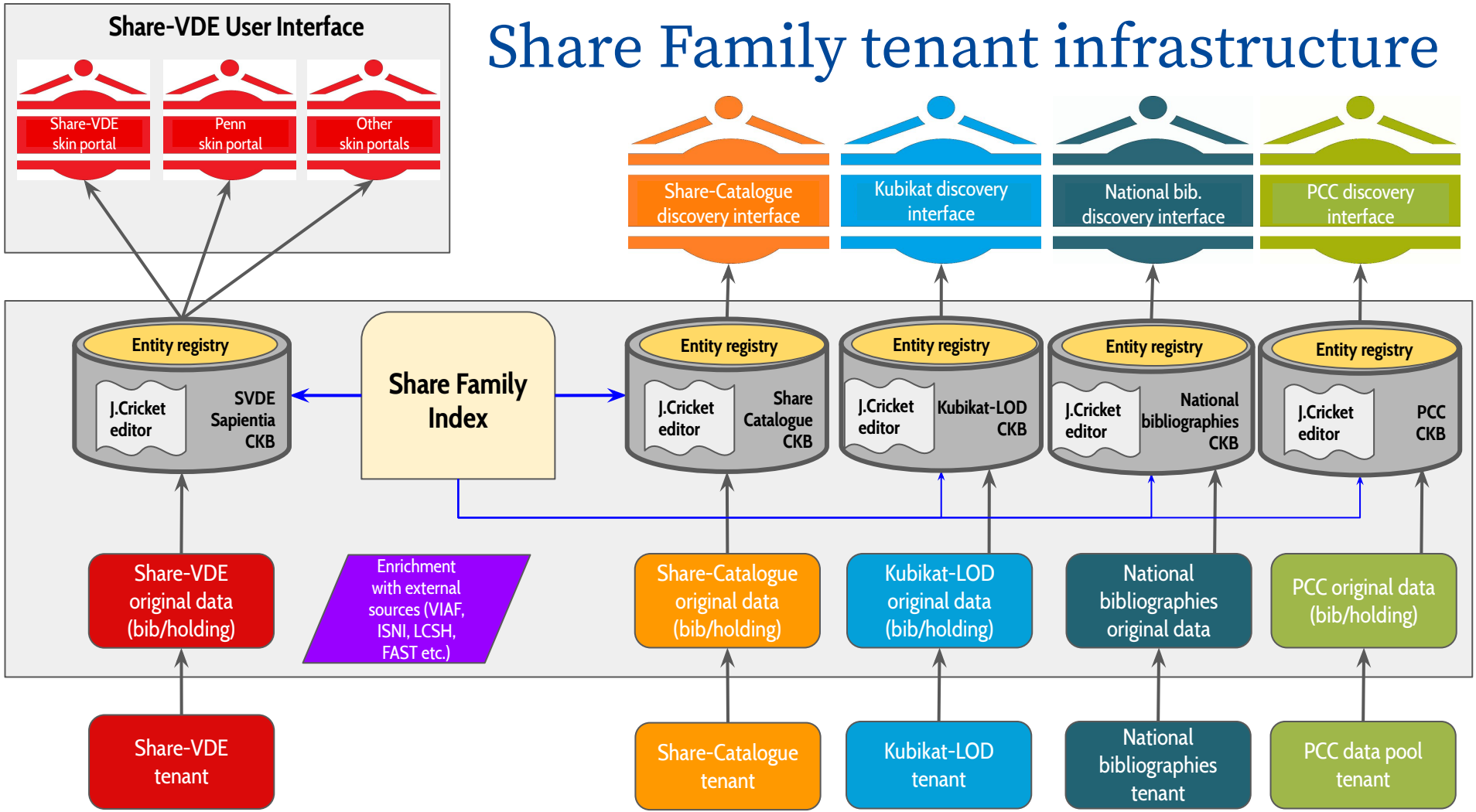
Solution in the Share Family architecture

Creation of more branches in the Share Family, named tenants

Consistent groups of institutions gathered by similar scope or from the same domain:

Share-VDE
Share-Catalogue
PCC data pool
National Bibliographies Group
Kubikat-LOD
Parsifal project (network of ecclesiastical university libraries in Rome)

Share Family tenant infrastructure



Benefits

- More efficient data management
- Technological sustainability
- Dedicated applications and services tailored to the institutions members of the various branches
- From the users perspective this enables richer and specialized sets of resources to be used.

Live tenants and skin portals

- **SVDE tenant** - www.svde.org => with LC's authority data and the bibliographic data of Stanford, UPenn and Smithsonian Institution. Further catalogues of participant libraries will follow in July.

  In addition the specific skins for the following institutions containing their respective data are:
  https://penn.svde.org (integrations via APIs will follow)
  https://stanford.svde.org

- **PCC tenant** - https://pcc-lod.org => with the PCC datapool.

- **National Bibliographies tenant** - https://natbib-lod.org
  with the skin for the British National Bibliography https://bl.natbib-lod.org (*)

- **Kubikat tenant** - https://kubikat-lod.org
  Kubikat art history libraries group

(*) *Note: the skin for the British National Bibliography is a preview of a beta site.*

Maximise efforts - Promote autonomy

The main purpose of this centralized architecture is to ensure **long-term sustainability** while favoring the **autonomy** of each tenant.

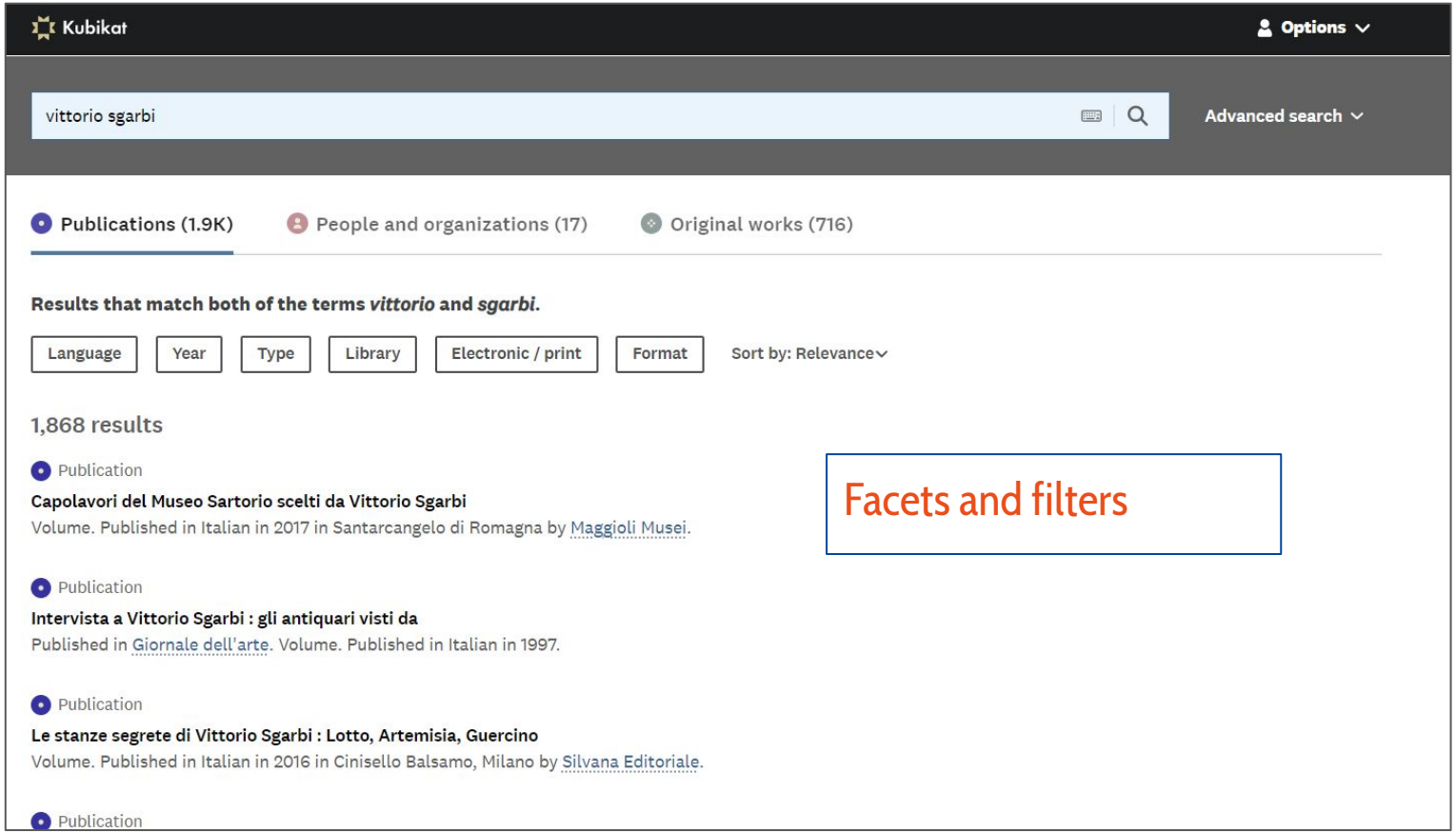
To foster this vision, it is essential to avoid ad hoc developments while ensuring the ability of local customizations. This flexibility is achieved through mechanisms that allow each tenant to selectively enable functions according to the purpose:

- on/off mechanism

- optional default configurations

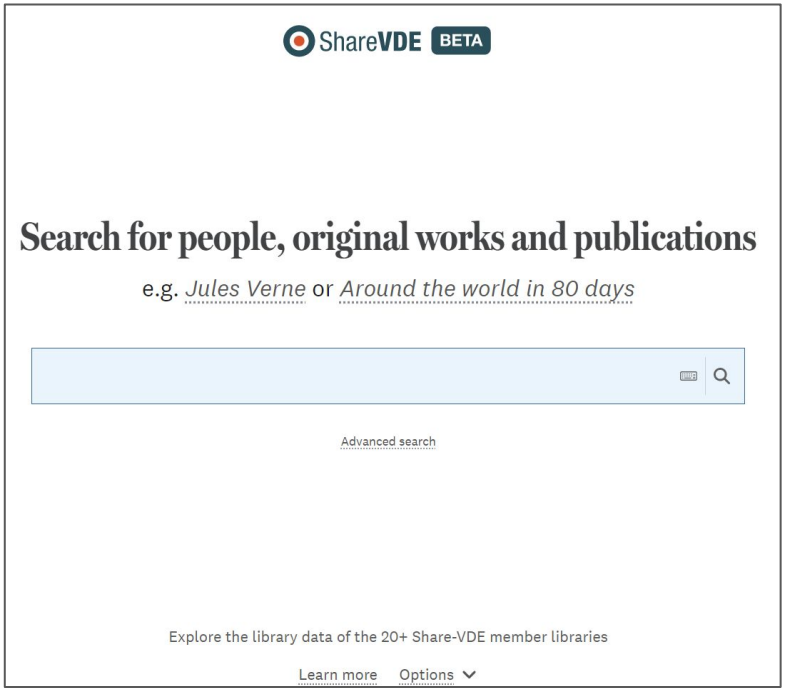
- local features/services

On/Off mechanism example



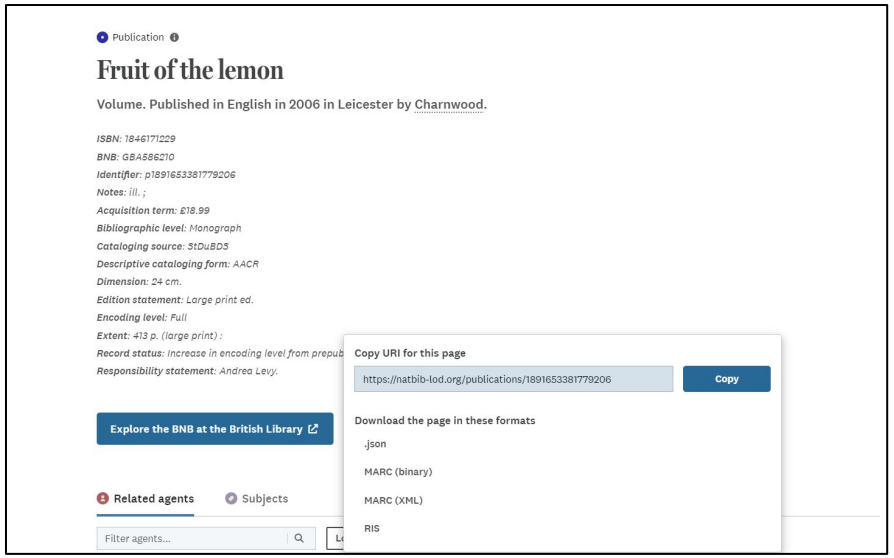
Default configuration: SVDE and PCC data pool

Simple search default configuration

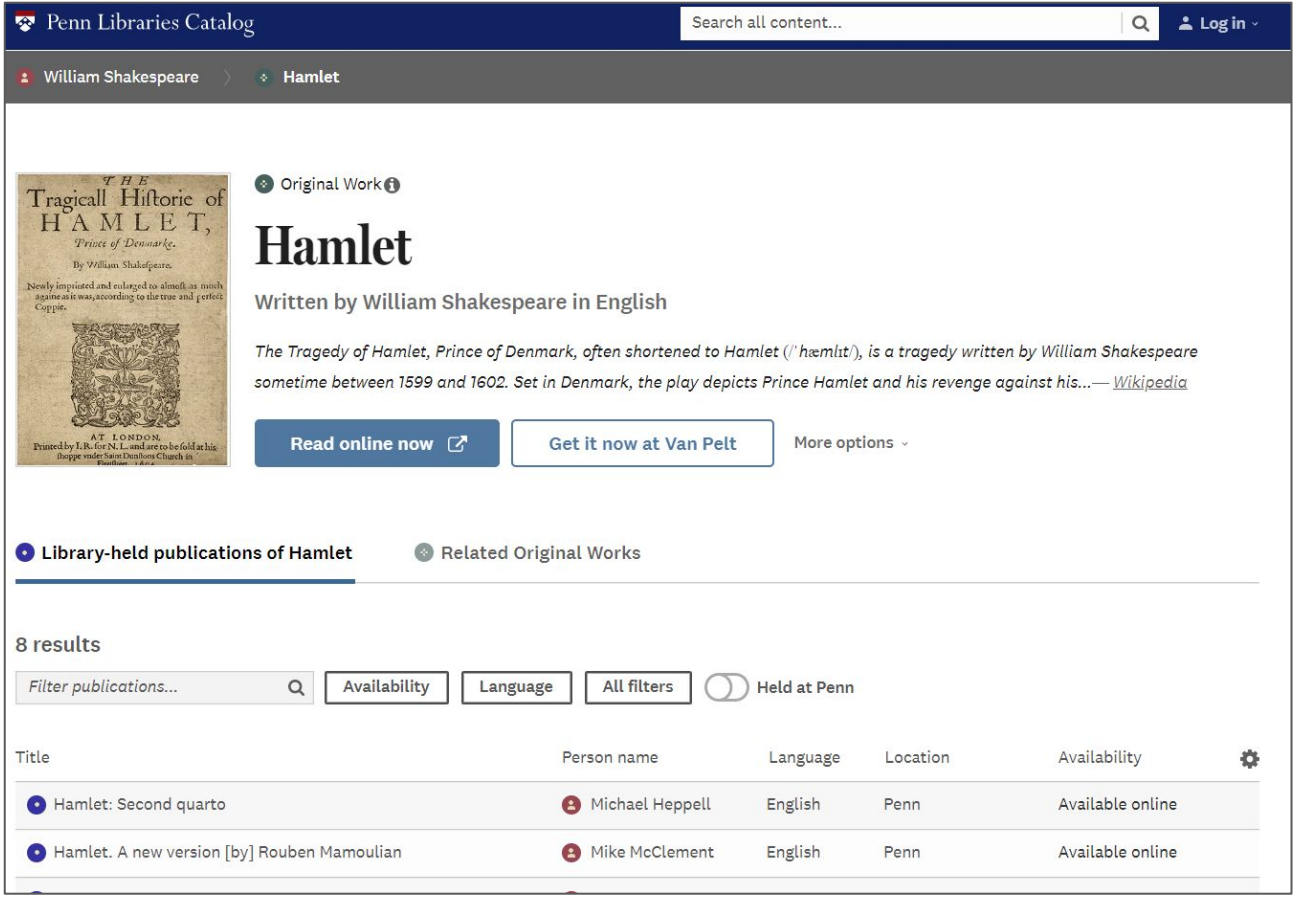


Default configuration: British National Bibliography

Simple search default configuration and the BNB - British National Bibliography skin



Local services: University of Pennsylvania

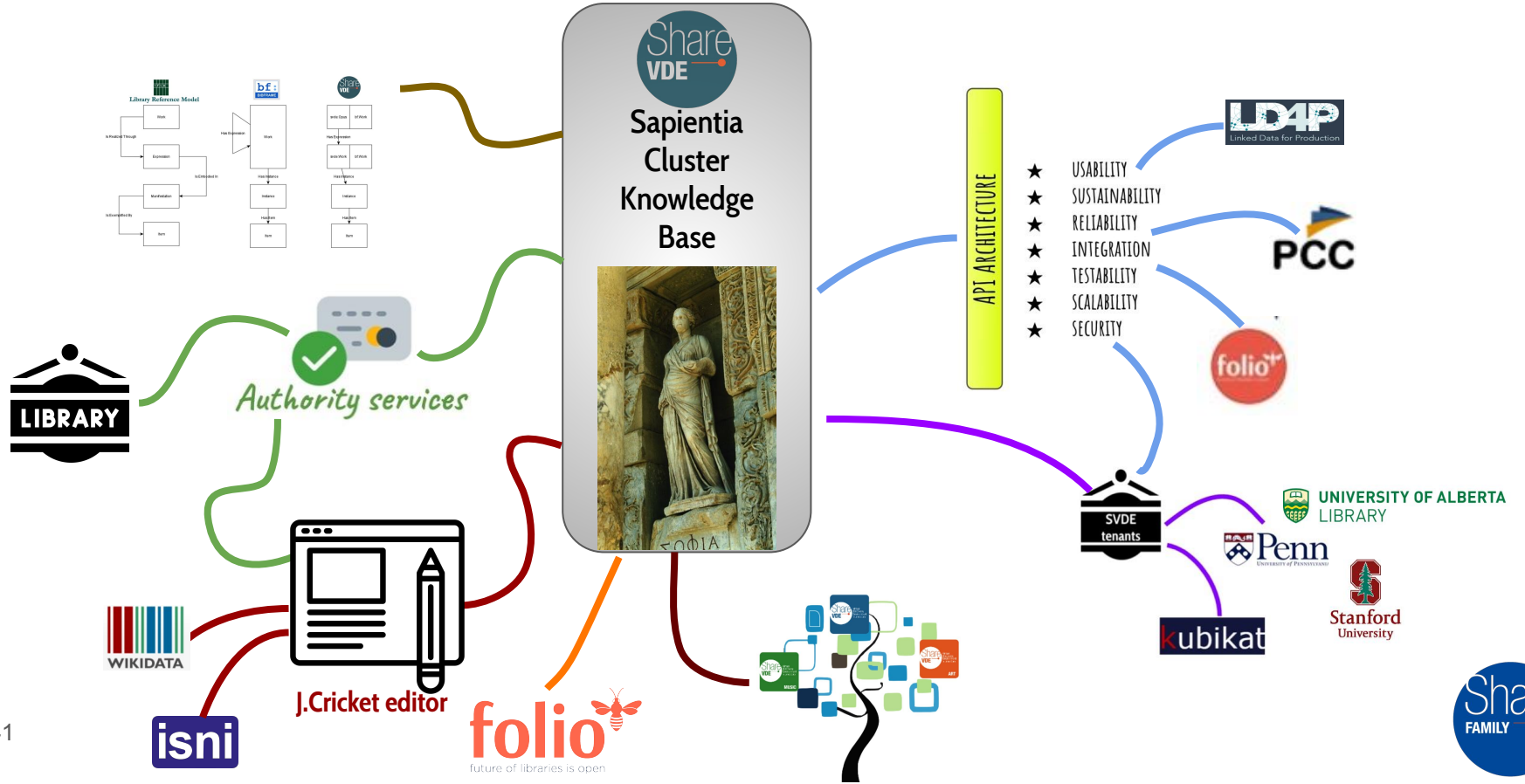


[Penn circulation services]

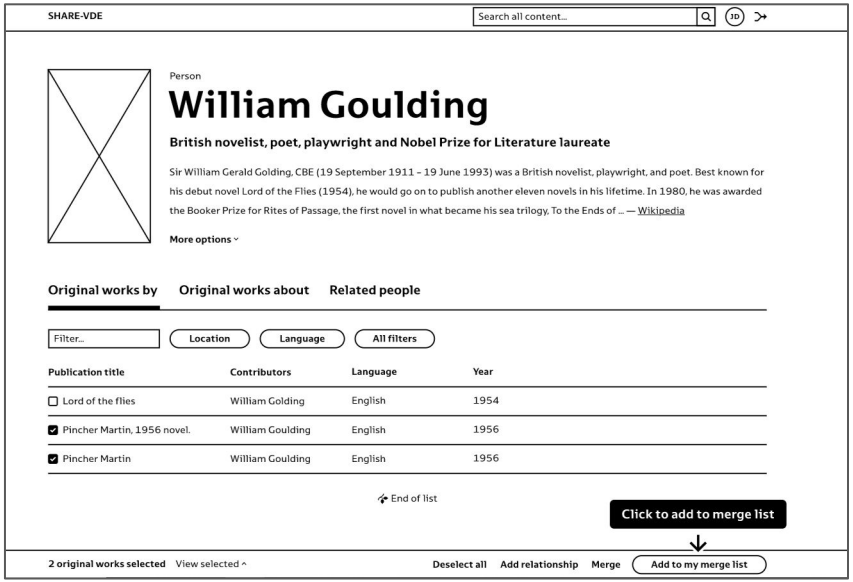
Share FAMILY

J.Cricket entity editor

Towards the Share-VDE Sapientia CKB ecosystem



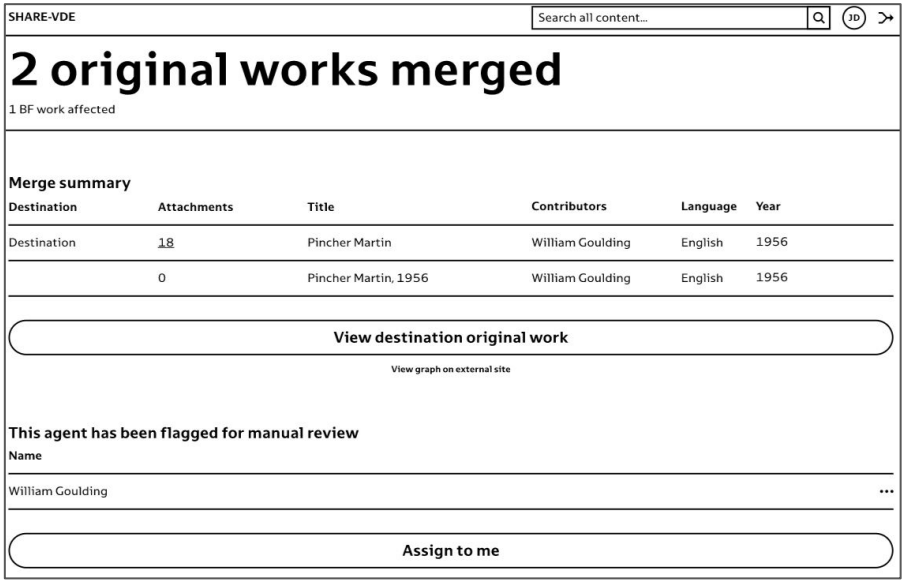
From linked data publication to linked data editing



The Share family platform is evolving from a discovery environment that converts traditional MARC data of libraries in Linked Open Data to an interactive authoritative source providing real services for libraries. This transition is happening through the editor named J.Cricket, that is the new application dedicated to the editing of the clusters of data in a collaborative and integrated environment.

From linked data publication to linked data editing

The editing tool J.Cricket will allow for editing the SVDE Cluster Knowledge Base, Sapientia, enabling several actions on the clusters (entities) saved in the SVDE database, including creation, modification, merge of clusters of works, of agents etc.
J.Cricket will extend authority capabilities through the integration with external data sources such as Wikidata and ISNI.



Next generation cataloguing

The J.Cricket editor is an example of how the Share family of initiatives is pursuing a new way of managing library cataloguing in a cooperative way:

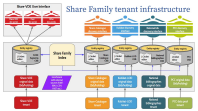
- aggregation of data from multiple sources
- managed through standard protocols (linked data)
- in a collaborative and integrated environment
- that makes available open data and resources
- to end users and professionals (researchers, scholars etc.)
- for reuse in the library community and beyond

J.Cricket 1.1.0: Features Recap

- **AAA:** Authentication + Authorization + Auditing
- **Cluster Status API**
- **Edit Cluster**
  - real time notifications (through GraphQL subscriptions) about cluster property changes
- **Merge:** C1, C2, C3 => ~~C1, C2,~~ C3
  - Multiple phases: create the merge list, edit the merge list, edit clusters, request for review, approve (or deny the merge)
- **Split (Cluster):** C1 => C1, C2
  - C2 could even be a new cluster
  - Multiple phases: create the split-set, edit the split-set, edit clusters, request for review, approve (or deny the merge)
- **Dictionary API**: What are the available cluster types? Which attributes belong to a cluster type? Which relationships? Given an attribute, which is its cardinality? Is it mandatory or not?
- **Data changes synchronization across Share-VDE storages (e.g. RDF Store, Search Engine, RDBMS)**
- **Entity Event Log (aka cluster changes)**: give me the history of changes of a given cluster
- **User notifications**: for managing the merge/split review lifecycle

Where we stand and a look ahead

Major achievements since September 2021



Release of **Share Family tenants** (beta version): Kubikat-LOD, Natbib-LOD, PCC-LOD

ShareVDE BETA

New release of the Entity Discovery Portal 2.0 supporting an enhanced CKB, including (among others):

Instance as entity

Default simple search configuration

Penn Libraries Catalog BETA

Skin portals supporting features specific to the individual skin (e.g. the British National Bibliography skin within the Natbib-LOD tenant)

Subjects

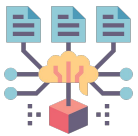
Initial version of **Subject management**, including various subject schemes (e.g. National Library of Finland subjects) linked subject strings and the display of concepts

Share FAMILY

Major achievements since September 2021

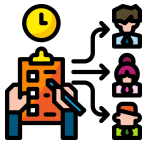
New system infrastructure migrated to **AWS - Amazon Web Services**

Start of developments of **J.Cricket** Entity Editor



Iterations of clustering process on svde.org: currently approx. 73 MARC authority and bibliographic records processed and approx. 139 millions entities created

now online: Duke University, Library of Congress, National Library of Finland, National Library of Norway, New York University, Smithsonian Institution, Stanford University Libraries, University of Alberta, University of Pennsylvania



In progress: refinement of **project coordination**
(e.g. aggregated release notes, issue tracking system)

Goals for the next future

2022, Q3-Q4

J.Cricket developments

Back-end: planned to be completed in October.
Front-end: end of back-end feature + ≥1 month. Features are being developed one by one, to enable testing iterations

2022, Q3-Q4

Change management module

AKA “delta update” module, to enable regular imports of MARC records for continuous update of SVDE data

2022, Q3-Q4

RDFizer (evolution of the LOD Platform RDF conversion tool)

It will support (among others) the import of Sinopia data to close this loop

2022, Q3-Q4

Third parties’ tools and protocols

Tools/Protocols for third parties’ usage and data harvesting
Including APIs, OAI-PMH, Atom feeds and Activity stream

2023, Q1

Share Family Index

Set-up of interconnections among tenants

2023, Q1

Item as entity

Creation of entity clusters of items

Continuous enhancement of the Share Family Entity Discovery Portals

Thank you

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