SVDE 2.0 Linked Data Management System and Entity Discovery Portal

Progress status of new developments

21st September 2021

Share-VDE Statement

In September 2021 the Share-VDE Advisory Council has approved and published an official statement that describes the role of the initiative in the broader context of Library Linked Open Data:

Share-VDE's Role in Library Linked Open Data

Share-VDE 2.0 is launched

The new SVDE 2.0 is now live at https://svde.org

- new back-end infrastructure for the Linked Data Management and the Cluster Knowledge Base;
- new Entity Discovery Interface (web portal).

This results have been achieved

- with the guidance of the Advisory Council and the active involvement of the Working Groups and parallel project Kubikat-LOD;
- with the support of the SVDE founding members, the full members, and the LD4P project.

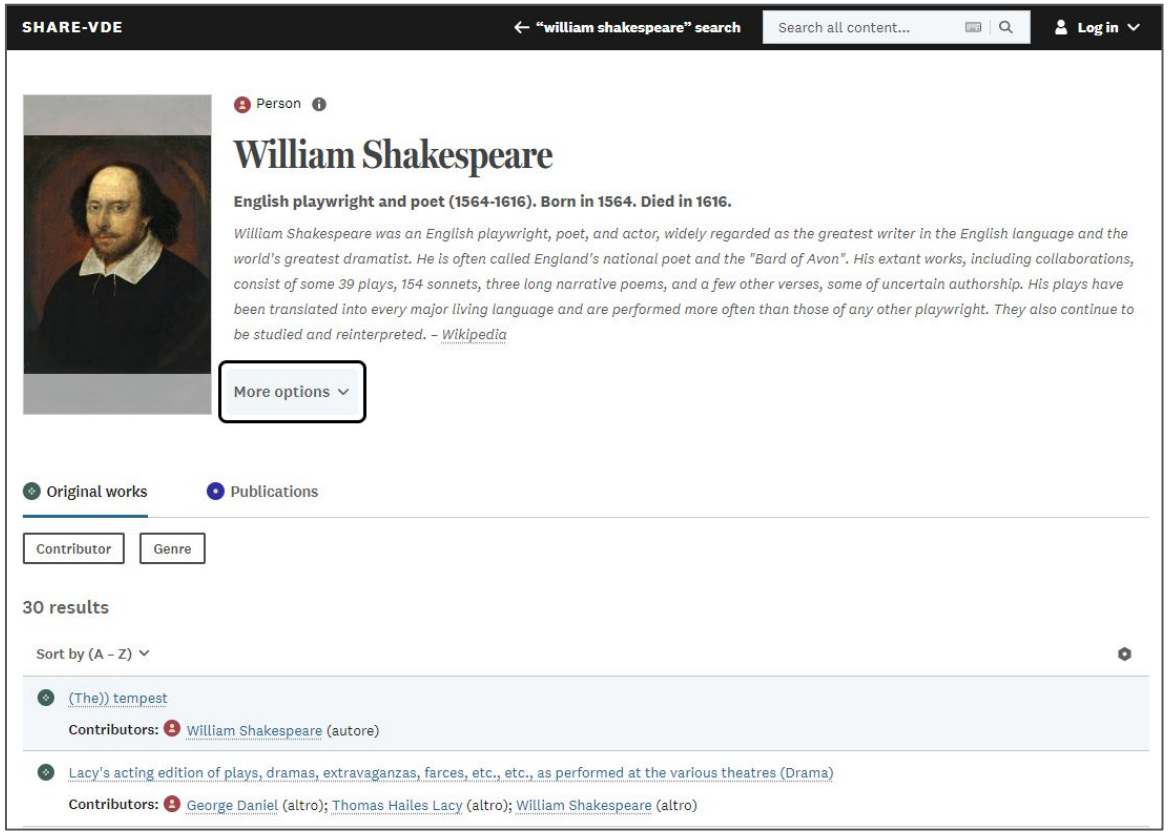
What was Share-VDE 1.0

A complex system designed from the perspective of expert users, where the user experience was quite complex

Back-end and front-end were not differentiated in the technological architecture

The system was not based on APIs

What is Share-VDE 2.0



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

What data is available

Progressive load of SVDE libraries data into the new system:

- Share-VDE 2.0 is available at https://svde.org
  - progressive upload of Stanford's bibliographic records + authority records from the Library of Congress → 14 millions of entity clusters
  - https://www.svde.org/about/about-share-vde
- Share-VDE 1.0 is available at https://share-vde.org

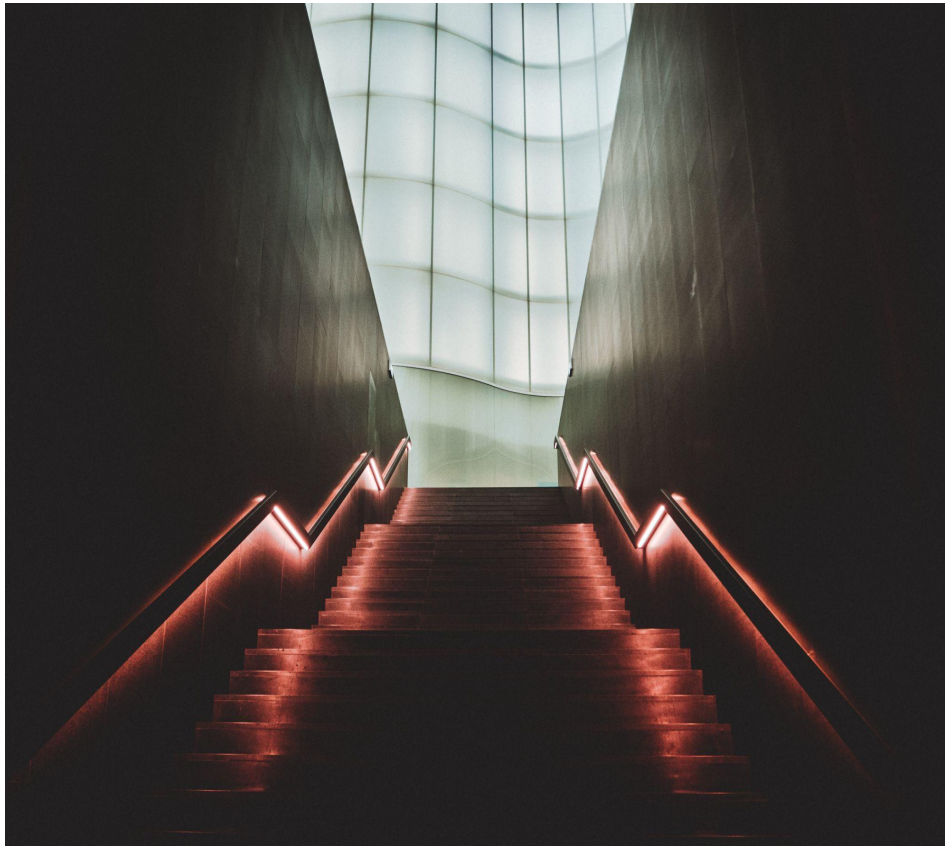
SVDE 1.0 and 2.0 will coexist until clustering iterations and data load on the new version will be completed.

The ultimate goal

The ultimate goal is to:

- create a linked data ecosystem where BIBFRAME entities benefit as much as possible from the wealth of data included in the original MARC catalogues
- act as a linked data node providing authoritative source of data through the CKB
- reconcile data from different libraries in a Union Catalogue and enrich with information from external sources (e.g. addition of URIs to entities from VIAF, ISNI, Wikidata etc.)
- provide a rich but simple user experience on the discovery portal
- expose the data on different layers that can serve many purposes (API layer, triple store, discovery portal)

How we got here



To meet this goal a much more powerful system is needed, and several steps to achieve it:

- complex search logic of the new discovery
- refactoring of the Cluster Knowledge Base including the data of many libraries (CKB 2.0)
- updates to the entity model and addition of many new attributes and properties

How we got here

- new API layer with different sets of APIs to support the search logics from the discovery portal and from external systems
- analysis from the SVDE team together with member libraries of the dedicated SEI - Sapientia Entity Identification WG for entity modeling and with members of Kubikat-LOD parallel project
- BIBFRAME extensions to support interoperability with other models, e.g. IFLA LRM (see Share-VDE entity model)
- updates to the entity model, updates to clustering specs, conversion specs in several iterations with the SEI WG
- new Cluster Knowledge Base, which means extended entity modeling and the whole refactoring of several storage mechanisms intertwined: RDBMS, Solr, triple store

Library-driven work

The work of the SVDE team is informed by member libraries. An example of some outcomes of the joint work around entity modeling with the Sapientia Entity Identification WG:

ENTITY	ATTRIBUTE (generic label)	PRESENT IN CKB 1.0	PRESENT IN TEST CKB 2.0 (formal name)	ADDED TO TEST CKB 2.0 ON DATE	ADDED TO PROD CKB 2.0 ON DATE (as of April 2021 this column does not apply)	VISIBLE ON UI	TEST / PROD
Agent (Person)	PREFERRED NAME	x	preferredHeading	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	SVDE URI	x	uri	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	BIRTH DATE		birthDate	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	DEATH DATE		deathDate	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	BIRTH PLACE		birthPlace	last update on April 20, 2021		x	T
	DEATH PLACE		deathPlace	last update on April 20, 2021		x	T
	VARIANT NAME/S	x	alternateHeadings	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	OCCUPATIONS		occupations	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	IDENTIFIER/S	x	identifiers	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	FULLER FORM OF NAME				last update on August 3, 2021		P
	PHOTO		photo	last update on April 20, 2021			T
	DESCRIPTION		description	last update on April 20, 2021		x	T
	SUMMARY		summary	last update on April 20, 2021			T
	OPUSES	x	opuses	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	PUBLICATIONS		publications	last update on April 20, 2021	last update on August 3, 2021		T/P
	GENDER				last update on August 3, 2021		P
	FIELD OF ACTIVITY				last update on August 3, 2021		P
Agent (Family)	PREFERRED NAME	x	preferredHeading	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	SVDE URI	x	uri	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	START DATE		startDate	last update on April 20, 2021		x	T
	END DATE		endDate	last update on April 20, 2021		x	T
	VARIANT NAME/S	x	alternateHeadings	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	IDENTIFIER/S	x	identifiers	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	PHOTO		photo	last update on April 20, 2021			T
	DESCRIPTION		description	last update on April 20, 2021		x	T
	SUMMARY		summary	last update on April 20, 2021			T
	OPUSES	x	opuses	last update on April 20, 2021	last update on August 3, 2021	x	T/P
	PUBLICATIONS		publications	last update on April 20, 2021	last update on August 3, 2021		T/P

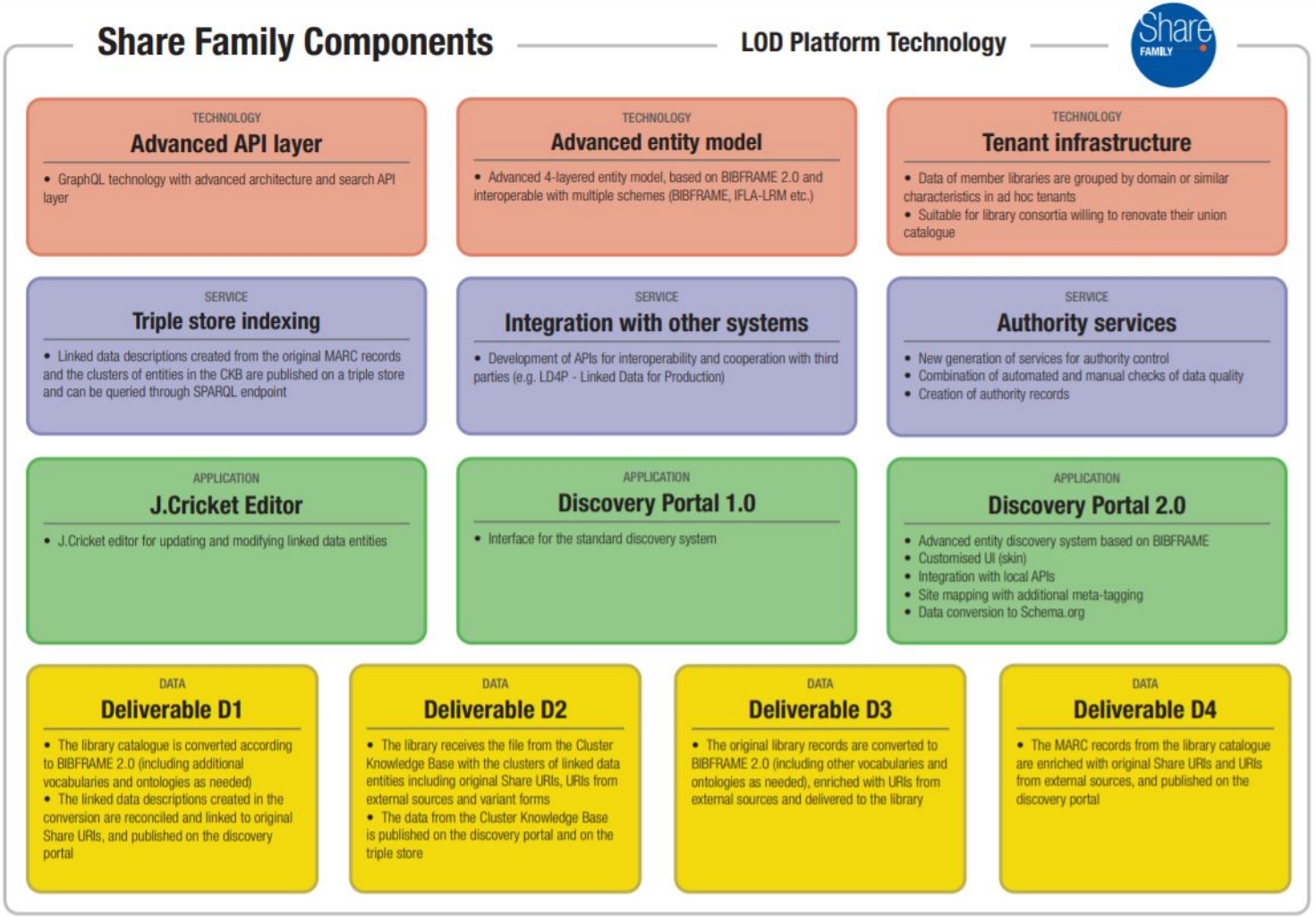
Overview of Share-VDE 2.0: front-end

- Simple search, including: exact match suggestions, explanations of search results; Wikidata descriptions
- Advanced search, including: search for any Agent type, for Original work, for Publication
- Entity pages (Agent, Original work, Publication, first version of Item), including wiki content (images from Wikimedia, summaries from Wikidata, descriptions from Wikipedia)
- Configuration of the system for the connection with local library services via API, for ad hoc customised skin portals
- Optimisation of the system for the J.Cricket editing features that will be developed over the next period
- Optimisation of accessibility features

Overview of Share-VDE 2.0: back-end

- Two API protocols: GraphQL API and REST API
- All Share-VDE entities are exposed through (read-only) API
- Search API provide several shapes / context behaviour (e.g. simple, advanced search, partial or full match, exact matches suggestions, terms modifiers, results explanation)
- Three query languages: TermsQL, SVDEQL, StructQL
- Search Quality Evaluation Tools
- Analysis and design of URI resolution and content negotiation mechanisms: dereference URIs and access to different formats of the entities
- Controlled vocabularies represented as entities (e.g. Roles, Places, Languages, Agent types, Forms, Genre etc.): this allows to dereference such vocabularies using URIs
- Authorization/Authentication infrastructure
- Continuous Integration

LOD Platform components



How to query Share-VDE and provide feedback

Share-VDE data can be queried through several methods:

- entity discovery portal (web user interface available at https://svde.org)
  - https://www.svde.org/about/about-share-vde
- via API through GraphQL and RESTful API endpoints
- via Stardog triple store (the Stardog db including the new CKB 2.0 will soon be available)

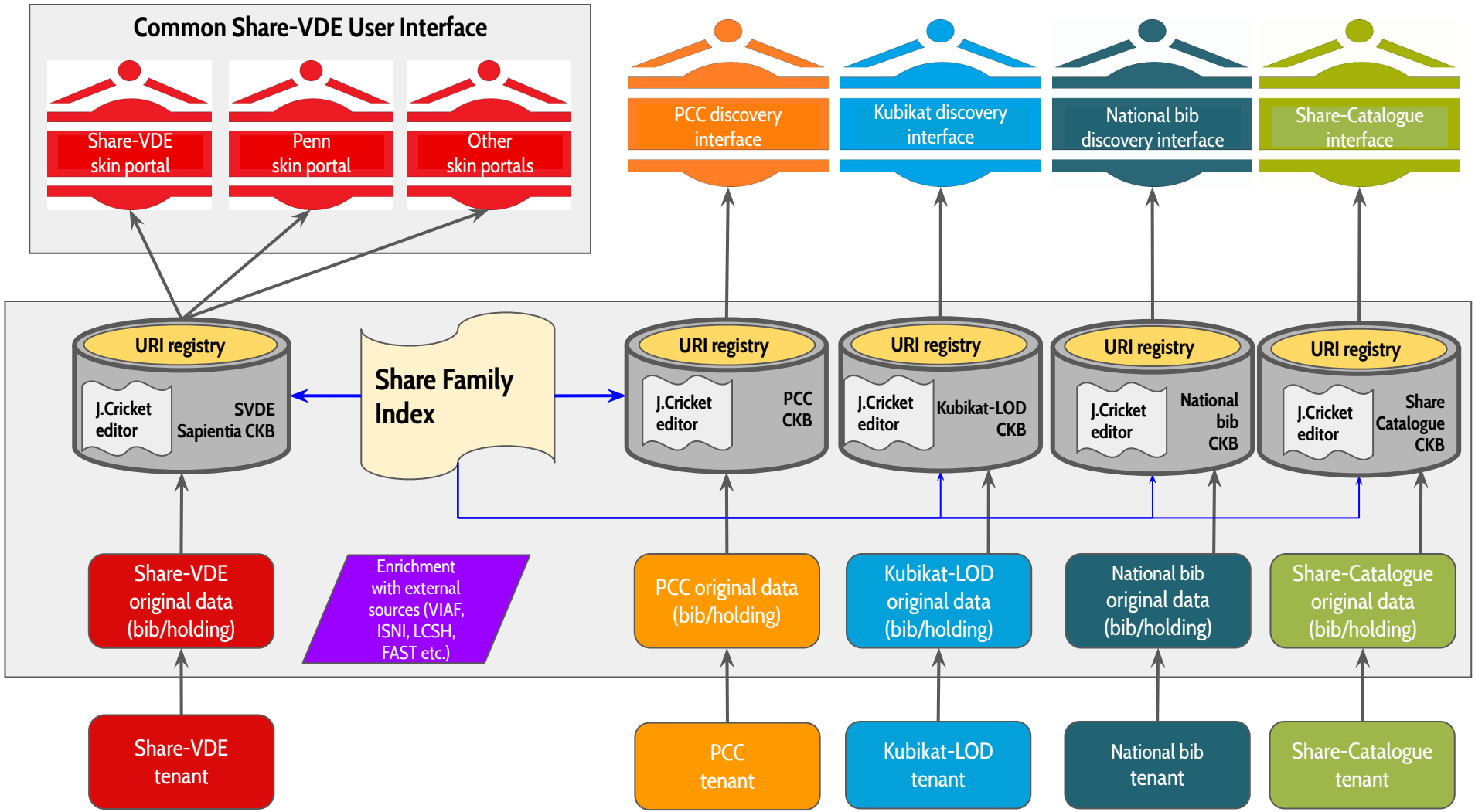
Report bugs and suggestions on the forum https://forum.svde.org/

Next steps

What comes next:

- progressive load of SVDE members' catalogues to populate the front-end portal
- complete connection with ad hoc skin portals
- continue the work on ad hoc features for Kubikat-LOD parallel project
- a huge work will be done to manage the Instances and related descriptions from the different libraries
- further enrichment of the CKB with new properties and refinements according to the joint work with the SEI WG
- developments of J.Cricket CKB editor
- Tenant architecture with Share Family Index (SFI) implementation

Further activities of the Share family


Common Share-VDE User Interface
Share-VDE skin portal
Penn skin portal
Other skin portals
PCC discovery interface
Kubikat discovery interface
National bib discovery interface
Share-Catalogue interface
URI registry
J.Cricket editor
SVDE Sapientia CKB
Share Family Index
URI registry
J.Cricket editor
PCC CKB
URI registry
J.Cricket editor
Kubikat-LOD CKB
URI registry
J.Cricket editor
National bib CKB
URI registry
J.Cricket editor
Share Catalogue CKB
Share-VDE original data (bib/holding)
Enrichment with external sources (VIAF, ISNI, LCSH, FAST etc.)
PCC original data (bib/holding)
Kubikat-LOD original data (bib/holding)
National bib original data (bib/holding)
Share-Catalogue original data (bib/holding)
Share-VDE tenant
PCC tenant
Kubikat-LOD tenant
National bib tenant
Share-Catalogue tenant

Further activities

- Kubikat-LOD: work is going on in parallel with the new SVDE infrastructure; important components emerged in the Kubikat group that serve the whole infrastructure are being tackled (e.g. serials) and work is progressing for the specific Kubikat tenant
- Parsifal, union catalogue of ecclesiastic libraries in Rome: work is going on in parallel to go live with the first version of the system
- Authority services: towards completion of the MARC-based services (tests ongoing at Stanford and among SVDE libraries); next steps: authority control based on linked data
- Continuous dialogue within the community and with other initiatives such as LD4P3, the PCC and the institutions involved

Further activities

- New working group 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
  - this could be a new tenant of the Share family

- Resources about the Share family
  - Share family presentation https://www.casalini.it/linked-data-for-libraries/
  - Share family resources https://wiki.share-vde.org/wiki/ShareFamily:Main_Page

Thank you!

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

https://wiki.svde.org/
https://svde.org
info@share-vde.org