The Share Family initiative: bringing Linked Open Data into practice

May 2024

info@svde.org
https://share-family.org
https://wiki.svde.org
https://svde.org
Share-VDE background and the Share Family
Share Catalogue online

Share-VDE prototype

2017-2019: Share-VDE members' and LD4P members' data from MARC21 to BIBFRAME

Share Family embraces all LOD Platform initiatives

PCC data pool

2017-2019: Share-VDE members' and LD4P members' data from MARC21 to BIBFRAME

2019-2021: Share-VDE - environment for library LOD

Share-VDE 2.0 new Linked Data Management System and Entity Discovery Portal

2021-ongoing PCC data pool

Share Family embraces all LOD Platform initiatives

BNB in LOD progresses towards production (beta)

2021-ongoing NatBib WG and shared discovery environment

Share Family - towards production

LILLIT launch

Parsifal launch

2022

2023

Stepping stones
Share-VDE and Share Family in a nutshell

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).
Main references

- Share Family brochure
- Share-VDE Executive Summary
- Share-VDE Open Metadata Policy
- Share Family Executive Summary for Consortia
- Executive summary of the National Bibliographies initiative
- The wiki website provide extensive resources to delve deeper into the community and technological aspects of the initiative: browse https://wiki.svde.org/ to explore news and updates, presentations and bibliographic resources, public documentation
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 that contribute to its development;
- 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.
Common priorities, challenges and concerns

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

For further details please refer to the following section of the wiki: About the Share Family.
Share-VDE - Virtual Discovery Environment

Berkeley Law Library
Duke University
Library of Congress
National Library of Finland
National Library of Norway
National Taiwan University Library
New York University
Smithsonian Institution
Stanford University
University of Alberta / NEOS Library Consortium
University of Chicago
University of Michigan Ann Arbor
University of Pennsylvania
Yale University

https://www.svde.org
Share Catalogue: Scholarly Heritage and Access to Research

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

Share Catalogue discovery portal
SHARE Catalogue - Participating Universities

Università degli Studi di Napoli Federico II (Naples)

Università degli Studi della Basilicata (Potenza)

Università degli Studi del Sannio (Benevento)

Università degli Studi di Cassino

Università degli Studi di Salerno (Salerno)

Università degli Studi di Napoli Parthenope (Naples)

Università degli Studi del Salento (Lecce)

Università degli Studi di Napoli L'Orientale (Naples)

Università degli studi della Campania Luigi Vanvitelli

Università degli Studi Suor Orsola Benincasa (Naples)
National Bibliographies in Linked Open Data

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

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

https://parsifal.urbe.it/parsifal/?l=en
Parsifal - Participating libraries (Unione Romana Biblioteche Ecclesiastiche)
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
Share Family - Linked Data Ecosystem: Principles

**CO-OPERATIVE**
Developed and driven by 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 resource description, with time, expertise and costs shared across the community for the benefit of all members.

**INTEROPERABLE**
By implementing the RDF-based BIBFRAME data model and facilitating interoperability with different data models and data pools, resource description can be 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.

**SUSTAINABLE**

**OPEN**

**AUTHORITATIVE**
Share Family processes and output

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.

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.

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

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.
Share Family tenant infrastructure

Share-VDE discovery portal and institutional skins
- Share-VDE portal
- Penn institutional skin portal
- other institutional skin portals

Share-Catalogue discovery interface
- Parsifal discovery interface
- NatBib discovery interface
- PCC data pool discovery interface

Entity registry
- SVDE Sapientia CKB
- Share-Catalogue CKB
- Parsifal CKB
- NatBib CKB
- PCC data pool CKB

Enrichment with external sources (VIAF, ISNI, LCSH, FAST etc.)

Share-VDE libraries
- original records

Share-Catalogue libraries
- original records

Parsifal libraries
- original records

National bibliographies
- original records

PCC libraries
- original records

Entity registry
- SVDE
- Parsifal
- NatBib
- PCC
Default configuration: SVDE and PCC data pool

Simple search default configuration on [SVDE](#) and [PCC data pool](#) portals
Default configuration: British National Bibliography

Simple search default configuration on Natbib tenant and the BNB - British National Bibliography skin*

* the British Library announced the launch of the beta version of the British National Bibliography Linked Open Data Portal
Default simple search configuration: the BNB

Simple search default configuration on Natbib tenant and the BNB - British National Bibliography skin* is set to Publications search, instead of the SVDE default.

This was done to comply with a different requirement whereby for the data stored in this tenant (ie. national bibliographies) it’s meaningful to direct users to publications.

Different communities or types of institutions might need customised features

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

● The Share Family of initiatives includes different branches and sister projects, supported by the same LOD Platform technology. Each branch or project is hosted in a specific tenant of the system architecture with a corresponding specific Cluster Knowledge Base and a dedicated web entity discovery portal.
   ○ For more details on the Share Family tenant infrastructure see the Summary of Share Family tenants.

● In some cases, within a single tenant a customised skin (ie. a sub-portal of the main entity discovery) can be created to address ad hoc needs of an institution, or group of institutions, willing to expose only their own data or to integrate local services in the Share environment.
   ○ For example, Share-VDE entity discovery portal at svde.org is one of such tenants, including a pool of data from a number of institutions, and the respective skin portals.
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](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)
Community work and outcomes
Among the active participant institutions

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.
The configurable components
Deliverables

Data, enriched with information (URIs and values) from external authoritative sources and converted following the BIBFRAME data model, are available for the publication on the Share portal and for other library projects, both in Marc 21 enriched and in RDF.

Data can be enriched with specific sources selected by each library, following their special and local needs.
Share Family and Share-VDE liaisons

- PCC liaison
- PCC BIG group
- SHARE Catalogue liaisons: Roberto Delle Donne and Claudio Forziati,Università degli Studi di Napoli Federico II
- IFLA Bibliography Section liaison: Tiziana Possemato, Casalini Libri and @Cult
- Kubikat-LOD pilot project in the art domain
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
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.
- 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.
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.

Among the latest outcomes:

- **Share Family Executive Summary for Consortia**, March 2024, which defines Share Family’s role in aiding consortia in adopting linked data methodologies for enhanced collaboration;
- **Share-VDE Executive Summary**, December 2022, summarising the scope of Share-VDE in the context of Linked Open Data for Libraries;
- Library and community events sub-group, dedicated to monitoring conferences/events/initiatives of interest for the Share community, and to submitting proposals for presentations as appropriate;
  - see the **SVDE wiki Resources page** for details about SVDE presentations at conferences and events.
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.

Among the latest outcomes:

- definition of use cases;
- functional analysis;
- analysis of interaction with Wikidata and ISNI (joint work with CKBE WG to design JCricket 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 released in 2022 Q2.

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 JCricket Editor for modifying (correcting / enriching), deleting, merging and separating clusters.

Among the latest outcomes:

- back-end developments for JCricket entity editor are completed, front-end features in progress;
- definition of use cases;
- design of manual editing features;
- analysis of interaction with Wikidata and ISNI to be incorporated into JCricket and authority dataflows that feed the Cluster Knowledge Base (joint work with AIMS WG to design JCricket functionalities).
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.

Among the latest outcomes:

- 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;
- definition of SVDE ontology;
- 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.
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.
The challenge of data models interoperability

Share-VDE Sapientia Entity Identification (SEI) Working Group decision of June 10th 2020

See the [SVDE entity model compared to BIBFRAME and IFLA-LRM](#) and an [example of application of the model](#).
SVDE Ontology designed in the dedicated Sapientia Entity Identification Working Group as an extension for BIBFRAME.

Core model:

svde:Work

svde:Opus

svde:hasExpression

SVDE ontology preliminary version

Share-VDE Ontology model
The **UX-UI WG** has re-designed Share-VDE user interface to respond to both patrons and library staff requirements and expectations. SVDE 2.0 entity discovery 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 skins, ie. customised sub-portals dedicated to individual institutions.

Among the latest outcomes:

- Share-VDE 2.0 Entity Discovery [https://svde.org](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;
- review and enhancements of portal features, in conjunction with the National Bibliographies Working Group.
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.

Among the latest outcomes:

- overview document [National_Bibliographies_Share_Family_initiative_2022-June.pdf](#)
- 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/](https://natbib-lod.org/); implementation of the skin portal for the BNB - British National Bibliography [https://bl.natbib-lod.org](https://bl.natbib-lod.org) (this is a preview of a beta site);
- joint work with the SVDE UX-UI working group to design end user services and user interface/discovery features.
National bibliographies WG - latest outcomes

- Study and address the needs of institutions that hold National Bibliographies linked data platforms;
- goal: build a shared discovery environment hosting LOD National bibliographies dataset;
- the British National Bibliography is the first national bibliography of this new tenant:
  - 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 Working Group is currently analysing use cases for ad hoc features of the shared National Bibliographies portal;
- review and enhancements of portal features, in conjunction with the SVDE UX-UI Working Group, are on going.
Technical overview
Share-VDE data flow - ER and Clusterization tool

Entity Resolution and Clusterization tool
Share-VDE Conversion pipeline: from records to entity
Lewis Carroll => *Alice’s Adventures in Wonderland*

Charles Lutwidge Dodgson => *The game of logic*
Profile: the information unit that expresses the identity of a particular and unique entity

Pasolini director

Pier Paolo Pasolini

Pasolini writer
Library cooperation: the cluster as the set of many faces

Sources

Data

Cluster Knowledge Base

P1
P2
P3
...

Pn
a cluster = an entity = a prism
The prism: attributes & relationships

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>Alice in wonderland</td>
<td>P1 P2 P3</td>
</tr>
<tr>
<td>titleAlternative</td>
<td>Alice’s adventures under ground</td>
<td>P1 P2</td>
</tr>
<tr>
<td>titleAlternative</td>
<td>Journeys in Wonderland</td>
<td>P4</td>
</tr>
</tbody>
</table>

sameAs
- http://dbpedia.org/resource/Alice’s_Adventures_in_Wonderland
  - Dbpedia
- https://www.wikidata.org/wiki/Q189875
  - Wikidata
- https://data.bnf.fr/ark:/12148/cb358500385#about
  - bnf
Towards the Share-VDE Sapientia CKB ecosystem
Share-VDE data flow - The entity management system

JCricket - The entity management system
JCricket is an entity editor that carries out the transition from Marc to a real Entity Management System.

JCricket acts on the entity database (CKB) created through Entity Resolution and clustering processes.

“The more the merrier” it’s a perfect vision in a collaborative community. But each library has also local needs: the architecture of JCricket allows to operate locally or centrally, creating a collaborative group that does not forget local specificities.
Next generation cataloguing

The JCricket 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
JCricket 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**: $C_1, C_2, C_3 \rightarrow C_1, C_2, C_3$
  - Multiple phases: create the merge list, edit the merge list, edit clusters, request for review, approve (or deny the merge)
- **Split (Cluster)**: $C_1 \rightarrow C_1, C_2$
  - $C_2$ 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
Share-VDE data flow - From record to entity: the RDFizer

RDFizer - Conversion tool
Conversion pipeline - Today

Marc21 Authority records

Marc21 records

Marc21 records enriched

Cluster Knowledge Base

Clusterization tool

RDFizer

Marc 21/CKB mapping (PCC Application Profile)

N-quads output

Marc21/CKB mapping

Tripstore

Marc21 Authority records

Marc21 records

Marc21 records enriched

Cluster Knowledge Base

Clusterization tool

RDFizer

Marc 21/CKB mapping (PCC Application Profile)

N-quads output

Marc21/CKB mapping

Tripstore
Conversion pipeline - Tomorrow

For library use only, not for RDF conversion

- Other formats
  - Unimarc
  - Marc21 Authority records
  - Marc21 records
- SINOPIA rdf/BF
  - Marc21 records enriched
    - From CKB to Marc 21 and Sinopia (PCC Application Profile)
- Clusterization tool
- Cluster Knowledge Base
- RDFizer
  - N-quads
  - BIBFRAME
- General mapping tool extensions
The Technology Stack (back-end)

Search API

GraphQL \{ REST \}
- Java
- spring

Curation API

GraphQL \{ REST \}
- Java
- spring

Stardog

PostgreSQL

Spark

Solr

Cassandra

Amazon SQS
The Technology Stack (front-end)

Component library

Front-end application

React

Sass

NEX.T.js

Vercel

TypeScript

GraphQl

Flagsmith

Share Family
Share-VDE: Infrastructure

Amazon EC2
Amazon Relational Database Service (Amazon RDS)
AWS Lambda
Amazon Simple Queue Service (Amazon SQS)
Amazon EMR
Amazon Keyspaces (for Apache Cassandra)
Enhanced Authority Services

The next step is to make Authority Services available also for linked data-based workflows - a truly new generation of features for the authority control.

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.

SERVICE
Authority services
- New generation of services for authority control
- Combination of automated and manual checks of data quality
- Creation of authority records
Cooperation & Interoperability

Cooperation and interoperability are key to Share technology: the use and the reuse of data, tools, ideas maximizes results and minimizes efforts.

Tools and protocols are being set-up for third parties' usage and data harvesting, including OAI-PMH, Atom feeds and Activity stream.

**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)
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

Hamlet: Second quarto


Place a hold on this physical item
Scan a chapter and send it to my email
Place on course reserve

Cancel
Place hold on item
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
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
Triple store publication - an open query endpoint

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

The core of SVDE integrated catalogue, i.e. the Cluster Knowledge Base of linked data entities created from SVDE institutions’ data, is published on a public query interface.
SVDE 2.0 back-end infrastructure leverages an advanced API layer orchestrating queries to SVDE data from the web discovery portal and from machine to machine applications.

The API layer is designed to respond to the increasingly complex search logic, the update to the entity model and the enhancement to the Cluster Knowledge Base.

- 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
The Share Family community includes different branches and sister projects, supported by the same technology. Each branch is hosted in a specific tenant of the system, ie. a group of institutions contributing to the same data pool.

This structure ensures autonomy of approach to data management for each tenant, but also cooperation, because all tenants are connected as part of the same “family”, and long term sustainability.
The approach to the SVDE entity model is to make it as much interoperable as possible, to facilitate data exchange with other systems.

It is based on BIBFRAME ontology, with ad hoc extensions to support interaction with IFLA LRM-based models.
Share Family entity discovery environments
SVDE 2.0 is supported by entity-based presentation layer reflecting BIBFRAME and the ad hoc SVDE extensions. The user experience is highly improved, hiding complexity to the end users. Each library can choose its own skin, to present a personalized image of its profile and its services, extending the functionality of the Portal from the shared environment to the local context.

**Application**

*Discovery Portal 2.0*

- 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
Share-VDE 2.0 Entity Discovery

Launched in September 2021 (beta version)

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

Improved user experience

Back-end infrastructure based on APIs

Initial list of the main features (we are going to see them live)

https://www.svde.org/about/release-log
Entity Discovery: data enrichment from external sources

Leonardo da Vinci

Italian Renaissance polymath (1452-1519).

Leonardo di ser Piero da Vinci was an Italian polymath of the High Renaissance who was active as a painter, draftsman, engineer, scientist, theorist, sculptor, and architect. While his fame initially rested on his achievements as a painter, he also became known for his notebooks, in which he made drawings and notes on a variety of subjects, including anatomy, astronomy, botany, cartography, painting, and paleontology. Leonardo’s genius epitomized the Renaissance humanist ideal, and his collective works comprise a contribution to later generations of artists matched only by that of his younger contemporary, Michelangelo. - Wikipedia
Entity Discovery: data enrichment from external sources

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

Filter publications... | Publication year | Publication place

49 results

<table>
<thead>
<tr>
<th>Title</th>
<th>Language</th>
<th>Publication year</th>
<th>Format</th>
<th>Contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferno</td>
<td>English</td>
<td>2013</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Entity Discovery: data enrichment from external sources
**Some of the main features**

**LOD Platform Entity Discovery portal overview**

- **Instance as entity** (aggregation of SVDE institutions’ data in a shared Instance, with link to resource in the local OPACs)

- **Default simple search configuration** (plain simple search vs. “OPAC like” search on publications)

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

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

- Present complex data aggregated from different sources and translated into linked data in a simple way (e.g. Google-like search);

- Design focus: provide intuitive access to complex data and make BIBFRAME easy to benefit from;

- The actual use of the discovery unveils issues that cannot be predicted in advance → (a lot of) practice should be combined with theoretical models;

- Mutual input among SVDE working groups: e.g. SEI group defining the treatment of data impacts the presentation layer of the data on the web discovery portal.
Some of the challenges

Ensure smooth user experience AND model interoperability
➢ end users should understand easily what content and data are represented by an Opus and the resources that embody it
➢ svde:Opus was chosen as highest level of abstraction

Strive to present data typically used by machines in a way that benefits also humans
➢ the SVDE UX group opted not to render the expression layer on the discovery interface
➢ so, this data are presented in the Publication layer of the discovery

Find a compromise to satisfy professional and generic users
➢ language and labels of the discovery portal are not pure “library language” (Opus = Original work; Instance ~ Publication; Agents = persons, families, organizations and conferences)
➢ this stems from the analysis by SVDE members and UX designers: we wanted to use labels that can be understood by non-expert users
Simple search #1

Simple search for Virginia Woolf:
- the system displays all results list
- user can select the view on People and organisations only or on Original works only
- headline under the Person name providing summary info
Simple search #2

- filters available on both People and organisations and Original works tabs

- sorting People and organisations tab by alphabetical order we see other results that apparently are not related with the search → explanation of results under the headline
Agent page for Virginia Woolf

- aggregation of all the info about the author: her works, publications derived from those works, works about the author

- the system pulls in author data from Wikipedia, localised in the language selected for the interface

- additional info can be toggled and record descriptions change dynamically

https://www.svde.org/virginia-woolf-a841654263859515/original-works-by
Entity page - Agent

More options button:

- SVDE URI for the entity
- Agent data representation in JSON
- link to IDs from external sources
Advanced search for all the entities fully implemented in the system:

- Agent, Original work, Publication

- search for Subjects will be added

- different search parameters can be combined
Advanced search for
https://www.svde.org/advanced-search/agents?q=(name+contains+foster+wallace)+and+(date_end+is_in_a_range+2007-2008)

- toggle extra info e.g. VIAF ID, Also known as
Entity page - Original work (= svde:Opus)

Filter for connected Publication data

Related agents

More options button:

- SVDE URI for the entity

- Original work data representation in JSON

- link to records from external sources

https://www.svde.org/infinite-jest-david-foster-wallace-o841654265592866/library-publications
Instance cluster:

- created with data contributed by all SVDE libraries. The description refers to the common Instance cluster

- breadcrumb retains relationship and shows snippet / preview of the content of connected entities on mouse over

- “Available at” function that enable connections with local library environments or connected services, including the link to the local OPACs and optionally interactive features like circulation request buttons
Entity page - Publication (~ bf:Instance)

More options button:

- SVDE entity URI

- data representation in different formats including JSON, MARC, MARCXML, RIS (other linked-data based representation formats are in progress, including JSON-LD, RDF XML, N-triples, N3, Turtle, N-Quads, TriX, TriG).

https://www.svde.org/infinite-jest-a-novel-p1401654885176149/subj ects
Subjects and Concepts

Subject entity - initial version:

- aggregates all subjects pertaining to the resource

- Concepts are the pieces that form a subject string and in SVDE they are entities of their own (initial version of Concepts that we will expand further)

- clicking on a Subject, the user is redirected to the resources attached to that Subject

Tracking of Provenance:

- the system can be queried via API to return the bibliographic records of a given Provenance (= institution) connected to an Instance

- On the interface it shows which institutions have contributed to a linked data cluster, or to subjects

- Provenance is key to support the editing of linked data entities with JCricket, to track updates and collaborative services

https://www.svde.org/infinite-jest-a-novel-p1401654885176149/subjects

<table>
<thead>
<tr>
<th>No.</th>
<th>Subject</th>
<th>Concepts</th>
<th>Subject provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FICTION (Topical subject)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>2</td>
<td>Family life + Fiction (Topical subject)</td>
<td>Family life (Topic) Fiction (Genre)</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>3</td>
<td>Motion pictures + Fiction (Topical subject)</td>
<td>Motion pictures (Topic) Fiction (Genre)</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>4</td>
<td>Compulsive behavior (Topical subject)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>17</td>
<td>Tennis + Fiction (Topical subject)</td>
<td>Tennis (Topic) Fiction (Genre)</td>
<td>Library of Congress</td>
</tr>
<tr>
<td>18</td>
<td>Compulsive behavior + Fiction (Topical subject)</td>
<td>Compulsive behavior (Topic) Fiction (Genre)</td>
<td>University of Alberta</td>
</tr>
<tr>
<td>19</td>
<td>Addicts + Fiction (Topical subject)</td>
<td>Addicts (Topic) Fiction (Genre)</td>
<td>University of Alberta</td>
</tr>
<tr>
<td>20</td>
<td>Compulsive behavior (Topical subject)</td>
<td></td>
<td>University of Alberta</td>
</tr>
<tr>
<td>21</td>
<td>Comportement compulsif + Romans, nouvelles, etc (Topical subject)</td>
<td>Comportement compulsif (Topic) Romans, nouvelles, etc (Genre)</td>
<td>University of Alberta</td>
</tr>
<tr>
<td>22</td>
<td>Addicts (Topical subject)</td>
<td></td>
<td>University of Alberta</td>
</tr>
<tr>
<td>23</td>
<td>englänin lieti (Topical subject)</td>
<td></td>
<td>National Library of Finland</td>
</tr>
<tr>
<td>24</td>
<td>koulukirjallisuus (Topical subject)</td>
<td></td>
<td>National Library of Finland</td>
</tr>
<tr>
<td>25</td>
<td>engläninkielinen kirjallisuus (Topical subject)</td>
<td></td>
<td>National Library of Finland</td>
</tr>
</tbody>
</table>
## Subjects and Concepts

- Initial version of Subject entity page, including subject strings linked in the Subject tab of the Publication page and the display of concepts.

- Concepts are the pieces that form a subject string and in SVDE they are entities of their own (initial version of Concepts that we will expand further).

[https://www.svde.org/family-life-fiction-s1031654873687244/original-works-about](https://www.svde.org/family-life-fiction-s1031654873687244/original-works-about)
Interesting Instance entity example


- many libraries contributed to this cluster
- subjects have different provenances
- initial rough display of item data
- additional resources extracted from the resource information, e.g. ToC, Publisher description etc.
- known issues: duplicated fields
- Provenance is retained in the system back-end along with original library records → key to support the editing of linked data entities with JCricket, to track updates and collaborative services
Default simple search configuration: the BNB

Simple search default configuration on Natbib tenant and the BNB - British National Bibliography skin* is set to Publications search, instead of the SVDE default.

This was done to comply with a different requirement whereby for the data stored in this tenant (ie. national bibliographies) it’s meaningful to direct users to publications.

Different communities or types of institutions might need customised features

(*) Note: the skin for the British National Bibliography is a preview of a beta site.
Support for ad hoc use cases

Implemented search for local ID (BNB number identifying all resources stored in this tenant)

https://bl.natbib-lod.org/advanced-search/publications?q=(nbn+contains+GBB9K2565)
### Support for ad hoc use cases

Display the information that the resource is a CiP record (Catalogue in Publication), ie. it has not been published yet (ad hoc data treatment was done to cover this case)

Other entities

Original work of type Series connected to all the Original works of the series volumes

https://svde.org/opus/961654264848228
Other discovery features and front-end design

Hover over linked information and the system loads info, snippet or even photo from Wikipedia extracted by matching the Wikipedia ID.

Localisation: discovery portal available in multiple languages.

Accessibility features (color themes, layout, font).

Layout can change dynamically depending on what data the user wants to see: from table list layout to card list layout.
Local services

Integration with local services, e.g. connection to Alma APIs for Penn circulation services
Integration with local services - lending
Default configuration: SVDE and PCC data pool

Simple search default configuration on [SVDE](#) and [PCC data pool](#) portals.
On/Off mechanism example

Facets and filters
Maximise efforts - Promote autonomy

The main purpose of this centralized architecture is to ensure long-term sustainability while promoting 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
The Entity Management System
A Share-VDE member (Stanford, in the example) uses FOLIO for managing its data.

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).

FOLIO instance (or instances in case of massive export) is sent to Share-VDE
Prism, faces: the Share-VDE Entity
Properties: Attributes, Relationships, Links

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>Alice in wonderland</td>
<td></td>
</tr>
<tr>
<td>titleAlternative</td>
<td>Alice's adventures under ground</td>
<td></td>
</tr>
<tr>
<td>titleAlternative</td>
<td>Journeys in Wonderland</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>sameAs</td>
<td><a href="http://dbpedia.org/resource/Alice's_Adventures_in_Wonderland">http://dbpedia.org/resource/Alice's_Adventures_in_Wonderland</a></td>
<td>Dbpedia</td>
</tr>
<tr>
<td>sameAs</td>
<td><a href="https://www.wikidata.org/wiki/Q189875">https://www.wikidata.org/wiki/Q189875</a></td>
<td>Wikidata</td>
</tr>
<tr>
<td>sameAs</td>
<td><a href="https://data.bnf.fr/ark:/12148/cb3585600385#about">https://data.bnf.fr/ark:/12148/cb3585600385#about</a></td>
<td>bnf</td>
</tr>
</tbody>
</table>

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

A relationship is a connection between two Share-VDE Prisms
The Big Picture: Genesis, Search, Edit

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 searchable through the entity discovery portal.

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

Any third-party, authorized application can be a Share-VDE client

Third-party Applications

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

but...

Penn
NYU
Duke
Library

but...

Third-party Applications

Any third-party, authorized application can be a Share-VDE client

but...

Any third-party, authorized application can be a Share-VDE client
JCricket: Available Operations

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

Lewiss Carroll → Lewiss Carroll

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

**Invalidate**

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

**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”)

116
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
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
Where we are now

★ As of May 2024, LOD Platform version 3.0.0 including JCricket has been released in test environment ✅
★ Major release that introduces many changes in the software and in data processing algorithms
★ A selected group of members will conduct the initial testing, initially focusing on JCricket Entity Editor
JCricket references

Useful references:

- JCricket overview
- general JCricket presentation
- 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
See the live demos!

JCricket premiered its demo at the ALA Conference 2023, during the Share-VDE Workshop:


JCricket demo at the LD4 Conference 2023:
https://www.youtube.com/watch?v=wbrqvWGuvl

Get involved!

Becoming part of the global Share Family means sharing data and co-operating with the greater international library community.

The family continues to expand as more and more libraries worldwide embrace the opportunity to be involved in an international network of information, creating dialogue, participation and partnership.

Get in touch with us to find out more about how the Share family can help your library:
- info@svde.org
- https://wiki.svde.org/
- https://share-family.org
- https://www.svde.org/about/about-share-vde
Thank you - please reach out for any question.