The Share family, shared and integrated ecosystem for library linked data

Linked Data and International Standards for Cultural Heritage

Bruxelles, 13th September 2022 Tiziana Possemato

Share-VDE as a library-driven community



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

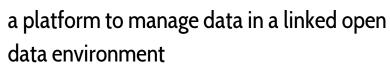
<u>https://svde.org</u> <u>https://wiki.svde.org/</u> <u>Casalini Lab</u> <u>Share - Linked Data Environment</u>

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 <u>www.svde.org</u>





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

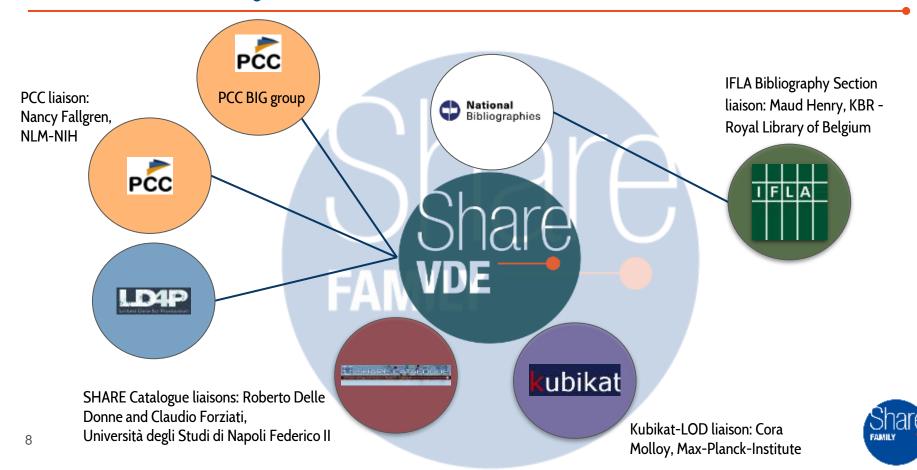
Active participation



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: crossdomain 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



The <u>Share-VDE AC</u> 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: <u>Share-VDE Statement</u>, 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.

13



Authority/Identifier Management Services WG

The <u>AIMS WG</u> 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;

14

- 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;

15

- reporting features providing complete details of the validation and corrections done to the records.
- \rightarrow 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 <u>CKBE WG</u> 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

Share-VDE	Q, Search all contant
Connect to external IDs	
Name	Wikidata Iben 0336030
OCLC VIAF Wikidata LC GND BNF Other Image: Construction of the state	Description Restrict States setting that an appropriate displayers on with kinds Description Restrict States on Appropriate States States on Approp



Sapientia Entity Identification WG

The <u>SEI WG</u> 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 <u>SVDE entity model</u>: svde:Opus | svde:Work | svde:Instance | svde:Item;
- svde:Opus and svde:Work are types of bf:Work \rightarrow this ensures interoperability;
- consolidating the definition of svde:Instance entity properties;
- review of clustering and conversion rules.

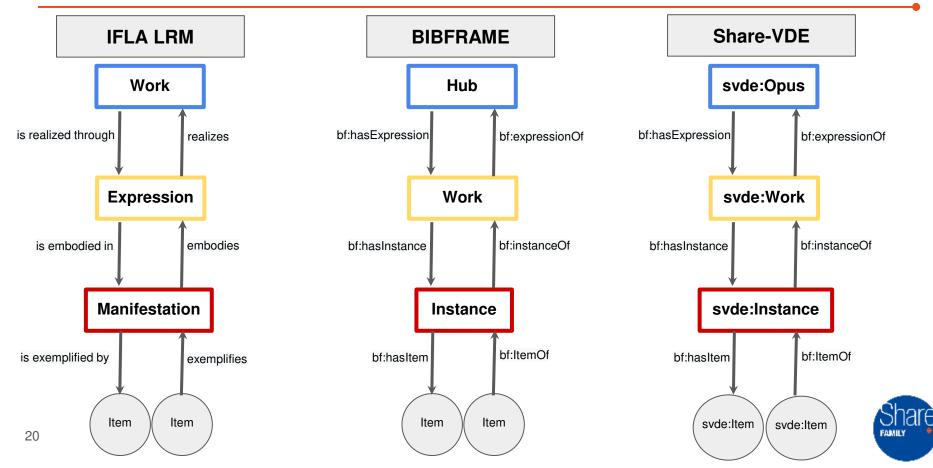


Share-VDE as a BIBFRAME node to put BIBFRAME into practice:

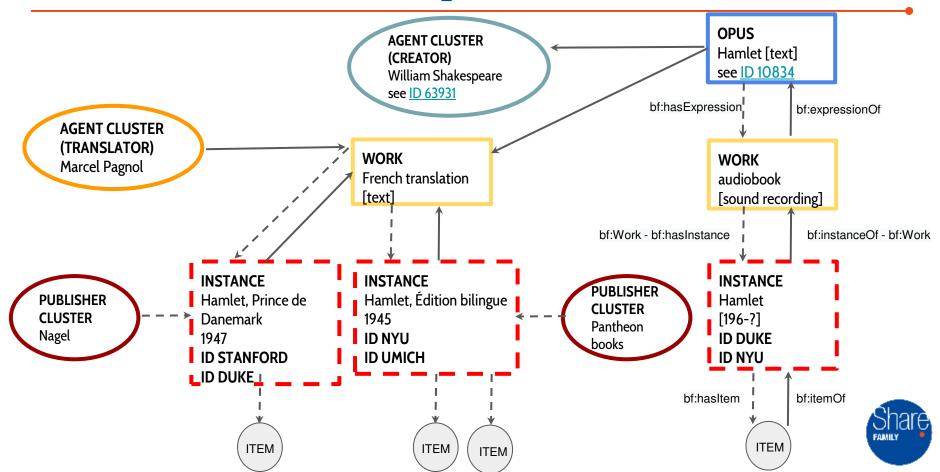
- Share-VDE provides enriched data that is <u>interoperable</u> with other BIBFRAME nodes and with other models;
- the Share-VDE working groups have reviewed algorithms and processed, and expanded the BIBFRAME model to meet real-world needs;
- focus on cooperation also in the IFLA context: the mapping 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 <u>UX-UI WG</u> 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 <u>Share Family discovery portals</u> supported by the same technology.



Focus on Share-VDE 2.0 Entity Discovery

SHARE-VDE	← "william shakaapaara" saarch	Search all content	🔲 Q 🕹 Login 🗸
-	🥹 Person 🚯		
100	William Shakespeare		
and a	English playwright and poet (1564-1616). Born in 1364. Died in 1616.		
	William Skakespeare was on English playwright, over, and actor, widely regards world's greatest dramatist. He is often called England's national poet and the " consist of some 30 plays, 164 sonnets, three long narrative poems, and a few of been translated into every major living language and are performed more often be studied and reinterpreted. – Wikipedia More optimus. ~	Bard of Avon". His extant is her verses, some of uncerta	orks, including collaborations, In authorship. His plays have
Original works	Publications		
Contributor Gen	re .		
30 results			
Sort by (A - 7) 🛩			0
(The)) tempest			
Contributors: 🔞	William Shakespeare (autore)		
Lacy's acting edition	tion of plays, dramas, extravaganzas, farces, etc., etc., as performed at the various theat	ras (Drama)	
Contributors: 🔞	George Daniel (altro); Thomas Halles Lacy (altro); William Shakespeare (altro)		

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

Search all content...

TE Kubikat

← leonardo da vinci search

C 2 Options V

A Lotter

Person O

Leonardo da Vinci

Italian Renaissance polymath (1452-1519).

Leonardo di ser Plera da Vinci was an italian polymoth of the High Renoissance who was octive as a pointer, droughtsman, engineer, scientist, theorist, sculptur, and anchitect. While his fame initially rested on his achievements as a pointer, he also became known for his notebooks, in which he made drawings and notes on a variety of subjects, including anatomy, astronomy, batany, cartography, pointing, and poleontology. Leonardo's genius epitamized the Renoissance humanist ideal, and his collective works campase a contribution to later generations of artists matched only by that of his younger contemporary, Michelangela. – Wikipedia

More options ~

Original works by
 Publications by
 Original works about

3,505 results

No,	Title A	Year	Genra	Ereptors	0
1	• 1478, a year in Leonardo da Vinci's career			Eduardo Villata (author)	
2	© 1499			Luisa Cogliati Arano (author)	
3	1515-2019 - La ville d'Amboise et son châtes au cœur d'un grand anniversaire	oa royal		Jean-Louis-Sureau (author) Christian Guyon (author)	
4	1519. La mort de Léonard			Leonardo da Vicci (artist) Gennaro Toscano (editor) Xavier Localite (editor) Schloss Ambeirs (host institution)	



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...
 Q
 Publication year
 Publication place

 49 results
 Title
 Language
 Publication year
 Format
 Contributors
 Image: Contributors

 • The inferno
 English
 2015
 • Stefano Albertini (editor)

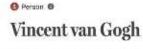
Entity Discovery: data enrichment from external sources

Search all content...

🏥 Kubikat

← van gogh, vincent search

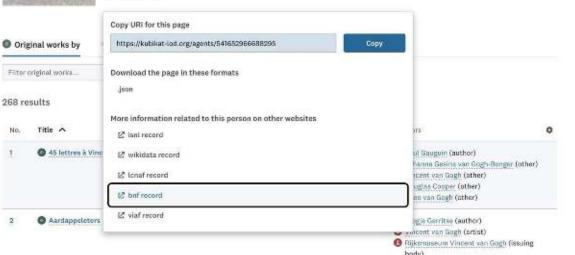
📼 Q, 🔒 Options 🗸



Dutch post-impressionist painter (1853-1890).

Vincent Willem van Goph was a Dutch Post Impressionist pointer who posthumously became one of the mest famous and influential figures In Western art history. In a decade, he created about 2,109 artworks, including around 860 all pointings, most of which date from the last two years of his life. They include landscapes, still lifes, partraits, and self-partraits, and are characterised by bold colours and dramatic, impulsive and expressive brushwork that contributed to the foundations of modern art. He was not commercially successful and, straggling with severe depression and poverty, committed suicide at the age of 37. – Wikipedia

More options ~





The <u>National bibliographies WG</u> 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: <u>https://natbib-lod.org/</u>
- 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 <u>https://natbib-lod.org</u>
 - with the skin for the British National Bibliography <u>https://bl.natbib-lod.org</u> (*Note: the skin for the British National Bibliography is <u>a preview of a beta site</u>)*
- 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

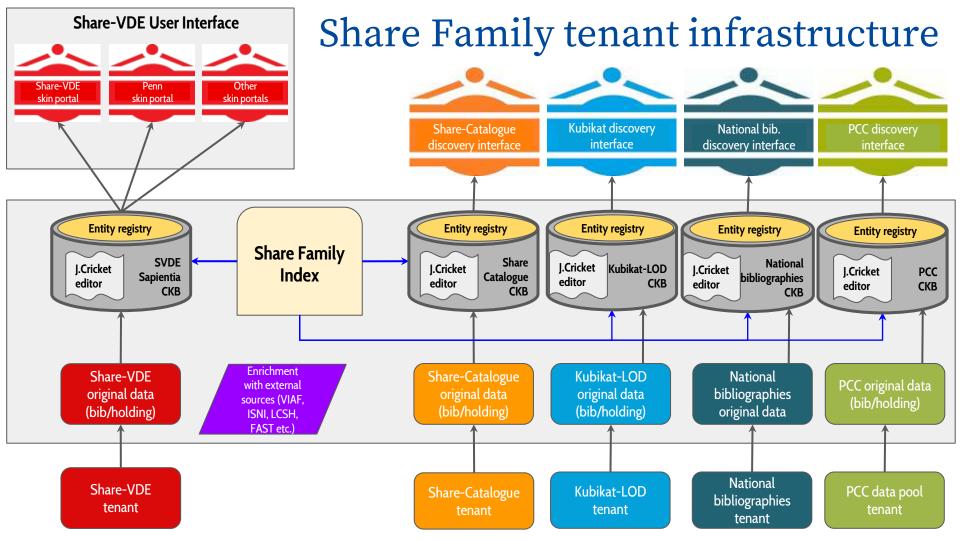


Creation of more branches in the Share Family, named tenants

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)







- 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** - <u>www.svde.org</u> => 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 <u>skins</u> for the following institutions containing their respective data are: <u>https://penn.svde.org</u> (integrations via APIs will follow) <u>https://stanford.svde.org</u>

- **PCC tenant** <u>https://pcc-lod.org</u> => with the PCC datapool.
- National Bibliographies tenant <u>https://natbib-lod.org</u>
 with the skin for the British National Bibliography <u>https://bl.natbib-lod.org</u> (*)
- Kubikat tenant <u>https://kubikat-lod.org</u> Kubikat art history libraries group

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



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

អ្វីវ Kubikat	💄 Options 🗸
vittorio sgarbi	🖂 🔍 Advanced search 🗸
Publications (1.9K) People and organizations (17) Original works	s (716)
Results that match both of the terms vittorio and sgarbi.	
Language Year Type Llbrary Electronic / print Format Sort by:	Relevance~
1,868 results	
 Publication Capolavori del Museo Sartorio scelti da Vittorio Sgarbi Volume. Published in Italian in 2017 in Santarcangelo di Romagna by Maggioli Musei. 	Facets and filters
Publication	
Intervista a Vittorio Sgarbi : gli antiquari visti da Published in <u>Giornale dell'arte</u> . Volume. Published in Italian in 1997.	
• Publication	
Le stanze segrete di Vittorio Sgarbi : Lotto, Artemisia, Guercino Volume. Published in Italian in 2016 in Cinisello Balsamo, Milano by <u>Silvana Editoriale</u> .	
Publication	

Default configuration: SVDE and PCC data pool

Simple search default configuration

ShareVDE BETA	THE PCC CATALOG
Scarch for people, original works and publications e.g. Jules Verne or Around the world in 80 days	Search for people, original works and publications e.g. Jules Verne or Around the world in 80 days
Q	٩
ácivancerá, teacth	Advanced search
Explore the library data of the 20+ Share-VDE member libraries	Explore the open pool of PCC-quality BIBFRAME data housed by the Share Family initiative
Learn more Options V	← Back to PCC Learn more Options ∨

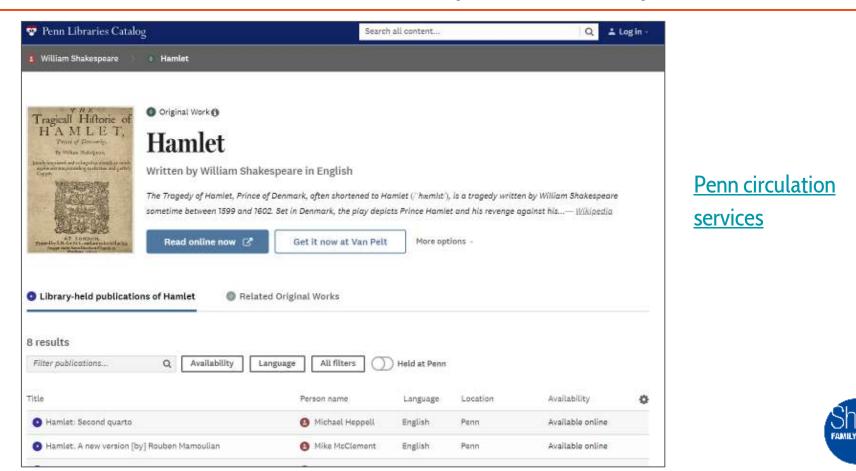


Default configuration: British National Bibliography

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

BRITISH NATIONAL BETA BIBLIOGRAPHY	• Published in English in 2006 in Leicester by Charrywood.
Search for people, original works and publications Publications Original works People and organizations	rdBMV 19440715209 RAME URAABBETT2 robest(Nev _ constructionsTheope Anders IV _ rot Abquildbillion Serrer: 2018 200 RMMARGyraphab (knew) - AnoppingAN Cathologing server: 500-000 Descriptive subsologing fairet: AACM domashioli, 24 cm. RMMAR Showeed: Lurge print ant domashioli, 24 cm.
e.g. the hitchhiker's guide to the galaxy adams	Retrieve with au (large print) Memory attacks inversion in encoding revel, from prepair Responsibility detrement: Artimus (Art) Memory attacks had org/publications/s00000000000000000000000000000000000
Language Year Type	Explore the ENIX at the Entitch Library (2) Green these formats gase
Advanced, search	Related agents Q Tabjects ABC (2001) HARC (2001) HARC (2001)
Explore linked open data for books and journals published or distributed in the United Kingdom and Ireland since 1950, and electronic publications since 2013	SI

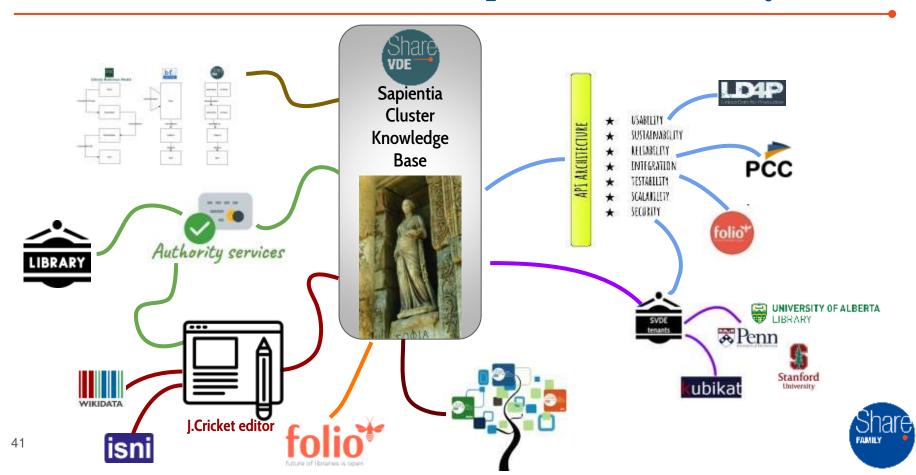
Local services: University of Pennsylvania



J.Cricket entity editor



Towards the Share-VDE Sapientia CKB ecosystem



From linked data publication to linked data editing

Ŵ	illiam (Gould	ing	
Brittis	h novelist, poet, play	wright and Nobe	Prize for Literature laureate	
X sewin	en Gandel Galetino CBF 119	Samerice 1911 - 1	Line 1993) was a British novellat, player	whit and post. Best known for
			o publich another eleven never's in his liter	
/ / // // //	ker Pripe for Rhas of Passag	pe, the first newel is wit	at become his sea trilogy. To the Ends of	- Witzenha
More op	diors -			
Original works by Orig	inal works about	Related people		
	arion Language	Related people		
			Veer	
Filter.	ation Language	Alleitan	Year 1954	
Sitian.	ation Language Cantributurs	All films	DOMEST 1	
Shar. Lac Publication title Dicord of the files	artion Language Castributors Withorn Galding	Att	1954	
Filier. Lee Publication Utile D. Londol the Ries D. Priches Martin, 1936 minut	ation Lenguage Contributions Withom Golding Withom Golding	Language English English	1854 1856 1899	Tick to add to merge list

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.

SHARE-VDE			Search all content_			<u>ৰ</u> ৩ >
2 ori 18F work affected	-	orks mer	ged			
Merge summa	ry Attachments	Title	Contributors	Language	Year	
Destination	<u>18</u>	Pinchar Martin	William Coulding	English	1956	
	0	Pincher Martin, 1956	Willham Goulding	English	1956	
		View destination o	riginal work			
		View graph on exter	malisite			
This agent has Name William Coulding	i been flagged for ma	inual review				
		Assign to	me			

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

45

- 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
 - \circ 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 🚥

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



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



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

022, 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	
022, Q3-Q4 🚽	Change management module	
	AKA "delta update" module, to enable regular imports of MARC	
	records for continuous update of SVDE data	
022, Q3-Q4 —	RDFizer (evolution of the LOD Platform RDF conversion tool)	
	It will support (among others) the import of Sinopia data to close this	
	Іоор	Continuous enhancement of the
	— Third parties' tools and protocols	
022, Q3-Q4	Third parties tools and protocols	Share Family Entity Discovery Portals
	Tools/Protocols for third parties' usage and data harvesting	
	Including APIs, UAI-PMH, Atom feeds and Activity stream	
2023, Q1 —	— Share Family Index	
	Set-up of interconnections among tenants	
2023 01 -	— Item as entity	
2020, QT	-	Share
	Creation of entity clusters of items	
	022, Q3-Q4	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 022, Q3-Q4 Change management module AKA "delta update" module, to enable regular imports of MARC records for continuous update of SVDE data 022, 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 022, 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



Thank you

tiziana.possemato@atcult.it tiziana.possemato@casalini.it

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