# Share-VDE entity discovery interface

## Design process and demo of the system

Recorded live demo: https://bit.ly/SVDE-discovery-live-demo
This slide deck: https://bit.ly/SVDE-discovery-demo-slides

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

# Overview

➢  Background of SVDE work on discovery
➢  Outcome: SVDE and Share Family Entity Discovery 2.0
➢  Live demo (and supporting slides with examples)
➢  Going forward with discovery...

Main content supporting the recorded live demo

➢  J.Cricket – collaborative entity editing use case
➢  Future plans
➢  Technical overview

Additional materials for reference

# Background

# SVDE UX design process

**2018-2019**: study and analysis for the enhanced SVDE discovery portal

➢ surveys with users run by SVDE partner designer and input gathered from real users (e.g. library patrons, scholars)

➢ dedicated UX-UI working group run by SVDE institutions: iterative feedback between stakeholders and UX designers

**2020**: user interface prototype

➢ not merely decorative but shaped on the SVDE BIBFRAME-based entity model

➢ further design input from SVDE institutions (Penn) for connecting SVDE with local circulation services

**2021**: SVDE 2.0 entity discovery portal development

➢ https://svde.org launched in September 2021 with other Share Family discovery portals supported by the same technology

**2022**: review and enhancements of SVDE 2.0

➢ build on SVDE users' feedback and researches (e.g. Penn study on UX testing: Slides, Video)

➢ SVDE UX-UI group re-convened to collect new input for refinements and enhancements

# User Experience/User Interface Working Group

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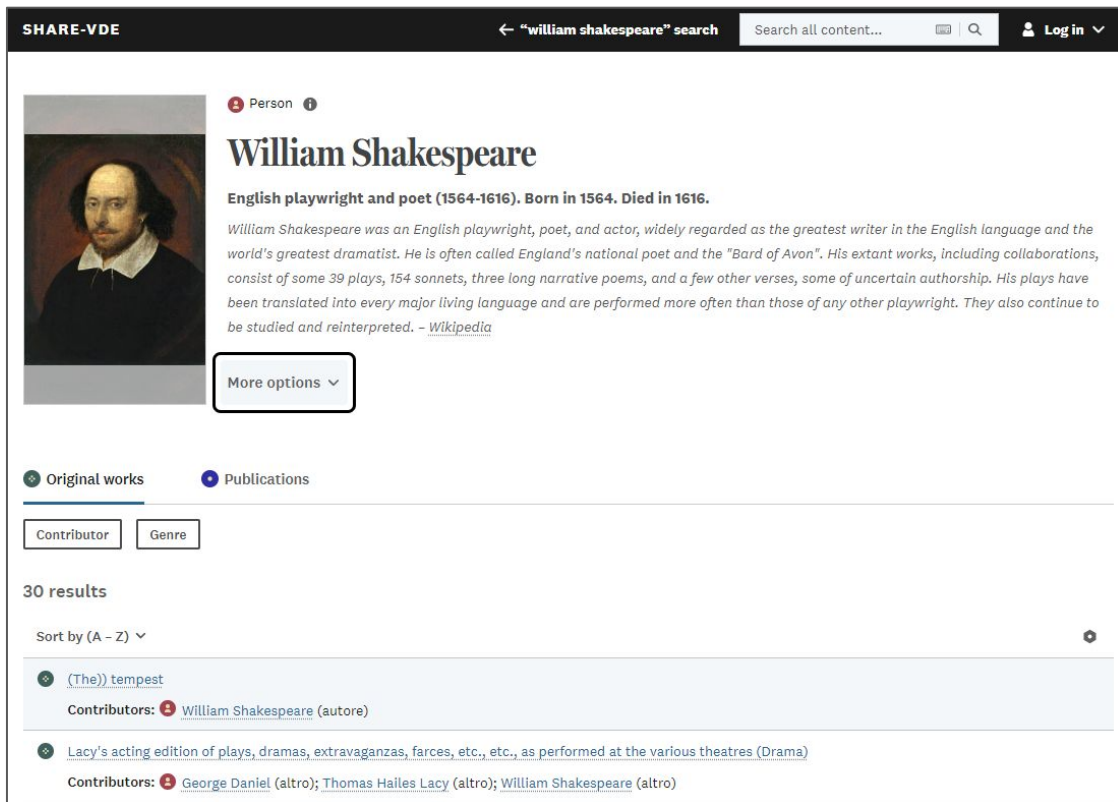
Main outcomes: Share-VDE 2.0 Entity Discovery https://svde.org

- new Entity Discovery Portal and new back-end infrastructure for Linked Data Management;
- other Share Family discovery portals supported by the same technology;
- iterative review and enhancements of portal features, in conjunction with the National Bibliographies Working Group.

# Outcome: SVDE and Share Family Entity Discoveries 2.0

# 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
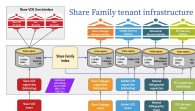
Improved user experience

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

# Some of the main features

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

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

**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 with discovery

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": Original Work = svde:Opus which is equivalent to the RDA Work (mostly). Publication is equivalent to the bf:Instance (RDA manifestation) and may list multiple expressions of the work. You can filter by various expression aspects in the filters

➢ this stems from the analysis by SVDE members and UX designers: we wanted to use labels that can be understood by non-expert users

# Live demo

# Supporting slides with examples from the SVDE and Share Family entity discoveries

# 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

# Entity page - Agent



Agent page for Virginia Woolf
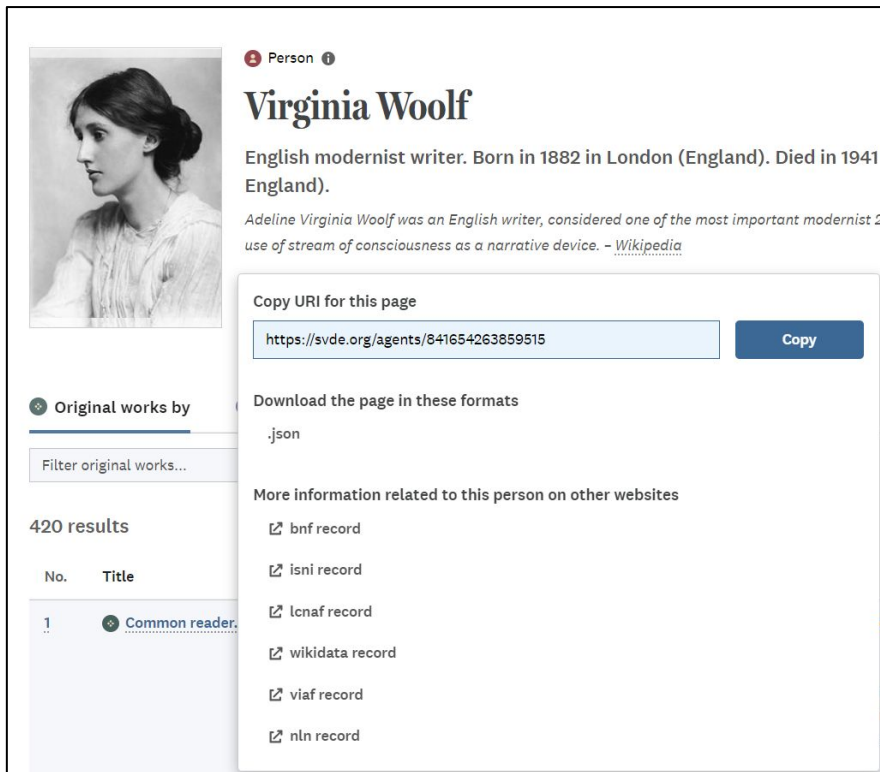
- aggregation of all the info about the author: her works, publications derived from those works, works about the author

- 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 #1



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 #2

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

# Advanced search #3



All tenants of the Share Family can now search for classification numbers in advanced search:

050 - LOC Call Number

051 - LOC Copy statement

052 - Geographic Classification

055 – LAC Classification

060 - NLM Call Number

061 – NLM Copy Statement

070 – NAL Call Number

071 - NAL Copy Statement

080 – UDC number

082 – DDC Number

083 - Additional DDC Number

084 - Other Classification Number

086- Government Document Classification Number

# Entity page - Original work (= svde:Opus)

# 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

# Entity page - Publication (~ bf:Instance)
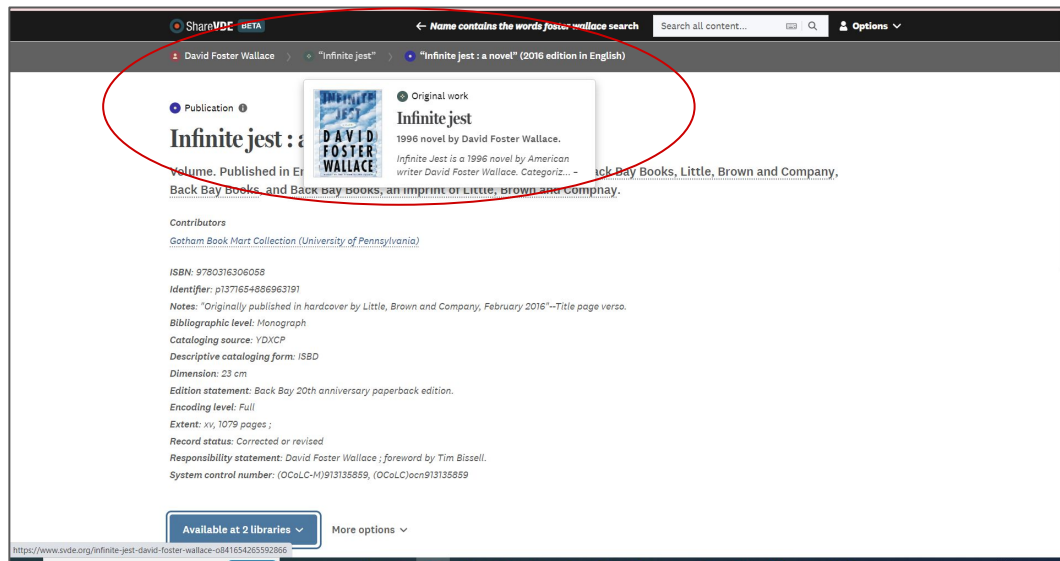


**Instance cluster:**

- 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/subjects

# Subjects and Concepts

| No. | Subject | Concepts | Subject provenance |
|---|---|---|---|
| | **45 results** | | |
| 1 | FICTION (Topical subject) | | University of Pennsylvania |
| 2 | Family life › Fiction (Topical subject) | Family life (Topic) Fiction (Genre) | University of Pennsylvania |
| 3 | Motion pictures › Fiction (Topical subject) | Motion pictures (Topic) Fiction (Genre) | University of Pennsylvania |
| 4 | Compulsive behavior (Topical subject) | | University of Pennsylvania |
| 17 | Tennis › Fiction (Topical subject) | Tennis (Topic) Fiction (Genre) | Library of Congress |
| 18 | Compulsive behavior › Fiction (Topical subject) | Compulsive behavior (Topic) Fiction (Genre) | University of Alberta |
| 19 | Addicts › Fiction (Topical subject) | Addicts (Topic) Fiction (Genre) | University of Alberta |
| 20 | Compulsive behavior (Topical subject) | | University of Alberta |
| 21 | Comportement compulsif › Romans, nouvelles, etc (Topical subject) | Comportement compulsif (Topic) Romans, nouvelles, etc (Genre) | University of Alberta |
| 22 | Addicts (Topical subject) | | University of Alberta |
| 23 | englannin kieli (Topical subject) | | National Library of Finland |
| 24 | kaunokirjallisuus (Topical subject) | | National Library of Finland |
| 25 | englanninkielinen kirjallisuus (Topical subject) | | National Library of Finland |

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 J.Cricket, to track updates and collaborative services

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

# 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-s10 31654873687244/original-works-about

# Interesting Instance entity example

https://www.svde.org/your-statistical-consultant-answers-to-your-data-analysis-questions-p26341656929688702/subjects

- 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 J.Cricket, to track updates and collaborative services

# Other entities



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

https://svde.org/opuses/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
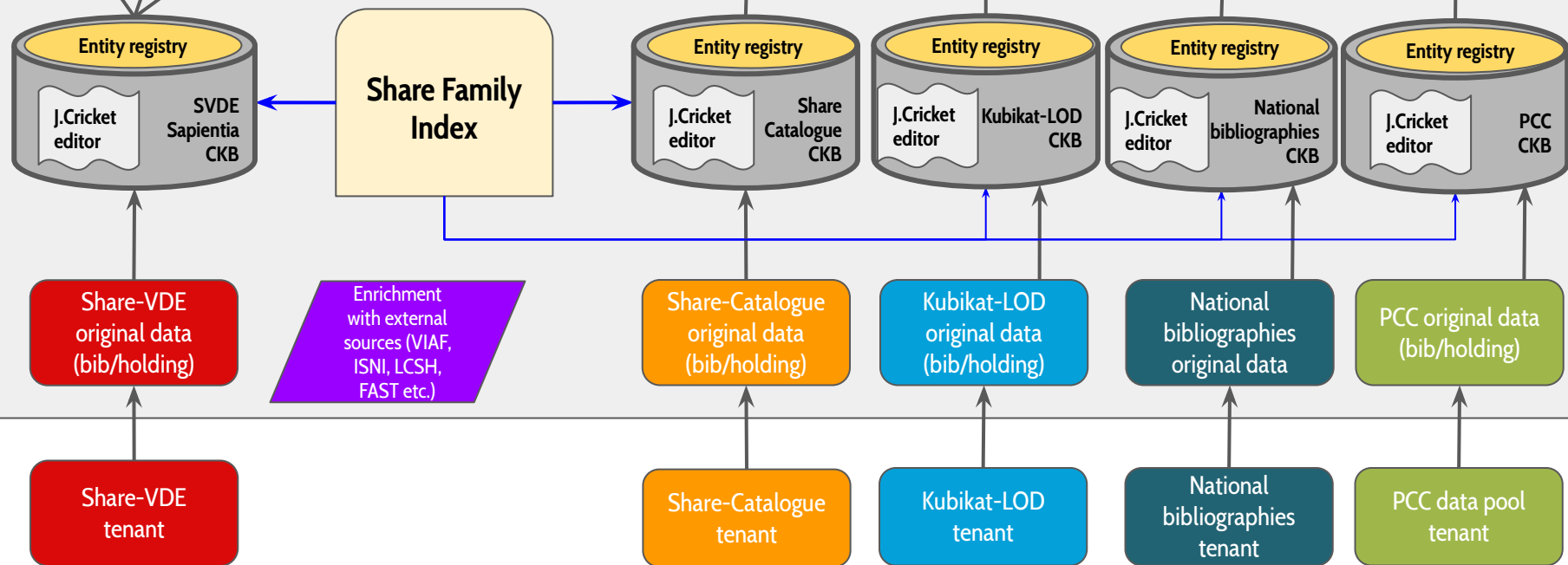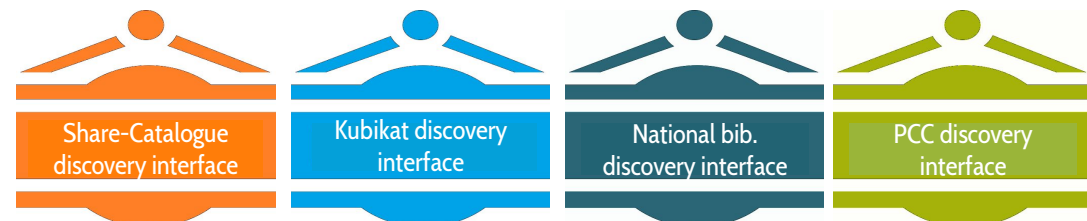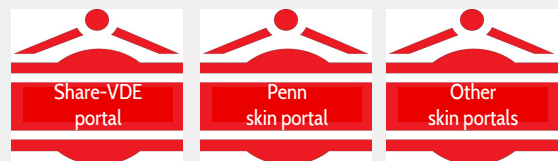
Localisation: discovery portal available in multiple languages

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

Share Family tenant infrastructure

# 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.

# Skin portals

- 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)

# Live tenants and skin portals

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

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

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

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

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

# Maximise efforts - Promote autonomy

The main purpose of this centralized architecture is to ensure **long-term sustainability** while 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

# On/Off mechanism example



Facets and filters

# 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*



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

# 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)

https://bl.natbib-lod.org/return-to-stringybark-creek-p19016533786O3115/related-agents

# Going forward with discovery...

# Identify areas for improvement

The elaboration of data from many sources and the creation of millions (and millions) of entities needs a consistent presentation layer for all cases

➢ e.g. in bibliographic models *series* are typically at abstract work level (ie. svde:Opus) → this creates Instances tied to more than one svde:Opus

➢ this is material for the SVDE SEI working group to define and give input to UX on how to render multiple Opuses

Continuous optimisation of:

➢ system performance

➢ ranking and relevance of search results based on users' feedback

Other known issues are being collected by the SVDE UX-UI group that will give input for enhancements

# Local services: University of Pennsylvania



Integration with local services, e.g. connection to Alma APIs for Penn circulation services

# Penn integration with local services - lending

# Future updates

The recorded live demo will be updated as soon as new functions will be available for the Share-VDE and Share Family entity discovery portals

Share the live demo: https://bit.ly/SVDE-discovery-live-demo

# Additional materials for reference

# J.Cricket - collaborative entity editing use case

# From linked data publication to linked data editing



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

# From linked data publication to linked data editing

The editing tool J.Cricket will allow for editing the Cluster Knowledge Base of each tenant where it is installed, enabling several actions on the clusters (entities) saved in the 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...

## 2 original works merged

1 BF work affected

**Merge summary**

| Destination | Attachments | Title | Contributors | Language | Year |
|---|---|---|---|---|---|
| Destination | 18 | Pincher Martin | William Goulding | English | 1956 |
| | 0 | Pincher Martin, 1956 | William Goulding | English | 1956 |

**View destination original work**

View graph on external site

**This agent has been flagged for manual review**
Name

William Goulding    ...

**Assign to me**

# J.Cricket 1.1.0: Features Recap

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

# Share-VDE Cluster: the Prism

**P1**

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

**P2**

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

**P3**

| title | Alice in wonderland |
|---|---|

**P4**

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

| sameAs | http://dbpedia.org/resource/Alice&#39;s_Adventures_in_Wonderland | Dbpedia |
|---|---|---|
| sameAs | https://www.wikidata.org/wiki/Q189875 | Wikidata |
| sameAs | https://data.bnf.fr/ark:/12148/cb358500385#about | bnf |

# How J.Cricket will interact with Wikidata

# J.Cricket - Postponed Features

- Create new Cluster
- Split cluster outputs n clusters (n >= 2)
- Unauthorized users should be able to request changes to entities
- Ad-hoc alert system for misaligned clusters with respect to bibliographic records

# 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

# Future plans

# Contribute to BIBFRAME interoperability
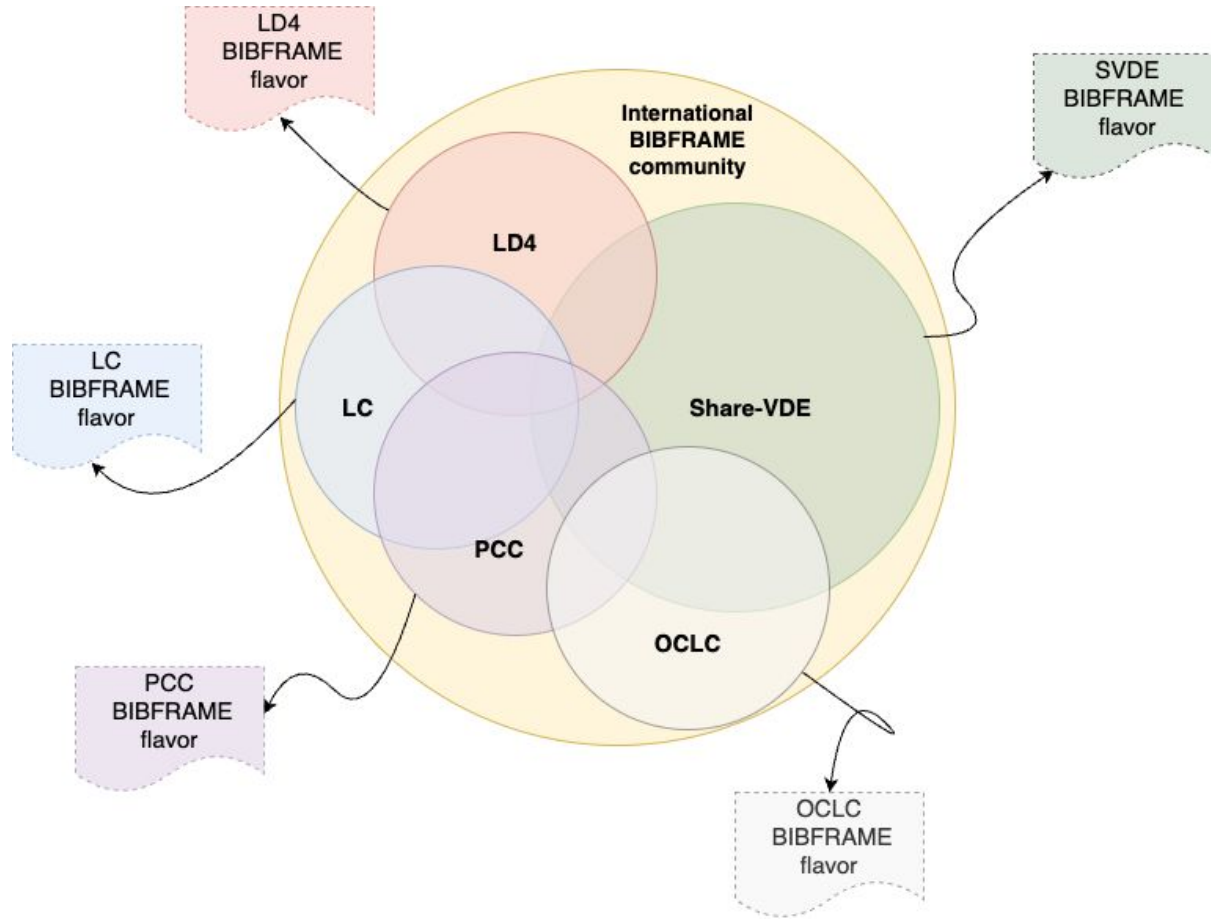
# Contribute to BIBFRAME interoperability
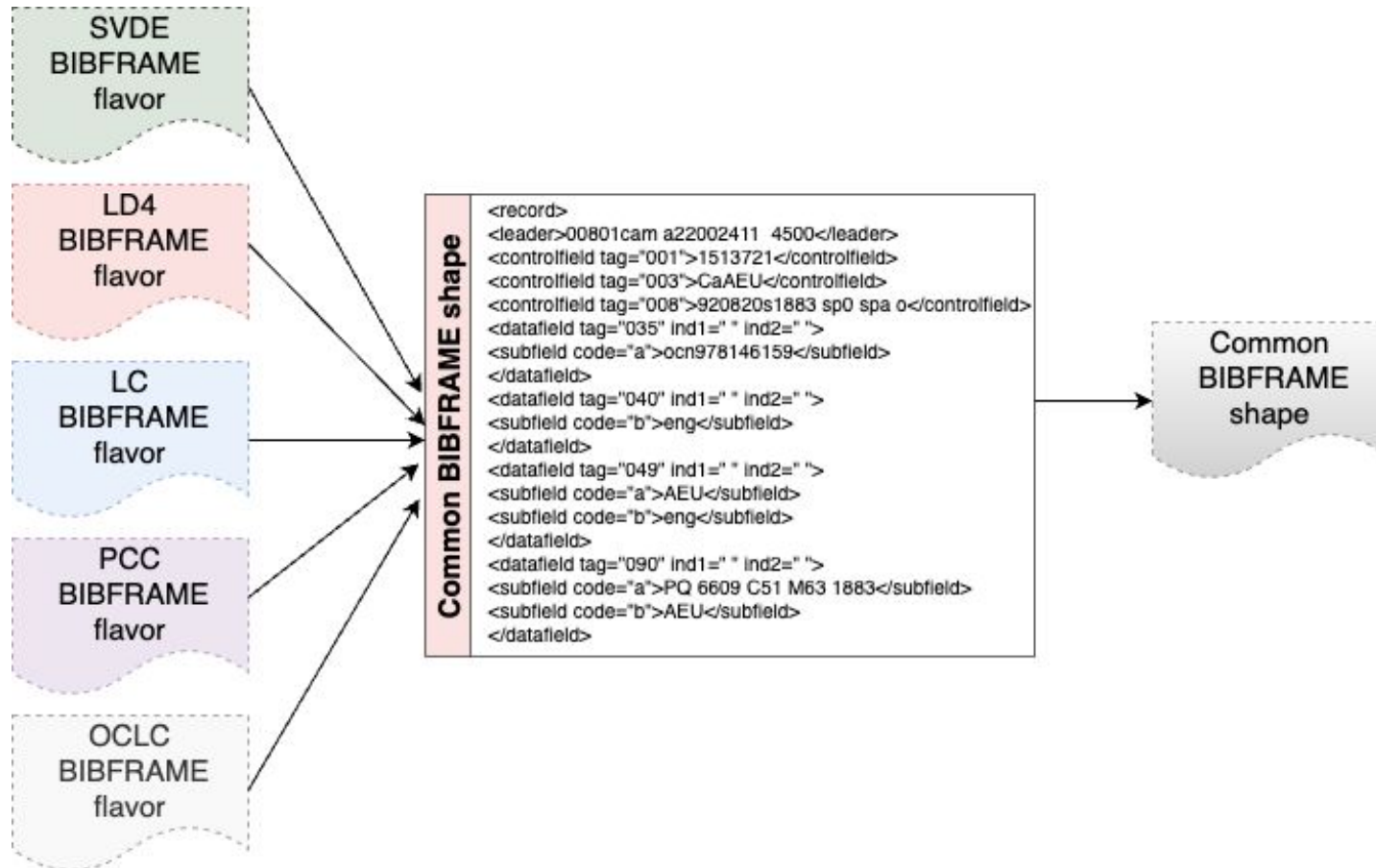
# Overall SVDE goals for the next future

**2022 Q4**
**2023 Q1**

## J.Cricket developments

Back-end: completed in November 2022
Front-end: developments in course

**2022, Q4**

## RDFizer (evolution of the LOD Platform RDF conversion tool)

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

**2023, Q1**

## Change management module

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

**2023, Q1-Q2**

## Third parties' tools and protocols

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

**2023, Q2**

## Item as entity

Creation of entity clusters of items

**2023, Q3**

## Share Family Index

Set-up of interconnections among tenants

**Continuous enhancement of the
Share Family Entity Discovery Portals**

Share
VDE

# Technical overview

# Share-VDE's engines: the API Layers

## Search API



A powerful and highly standardized **API** exposing Share-VDE' search capabilities over the **HTTP** protocol.

It comes in two flavours:

- **GraphQL**, with a rich data dictionary and the flexibility to query and extract only the fields of interest.

- **RESTful**, returning data represented according to the **HATEOAS** standard.

## Knowledge Base



## Curation API
### J.Cricket



An advanced API offering **Prism curation** capabilities.

It takes advantage of the most modern features of **GraphQL** to establish a continuous informational dialogue between the server and the connected clients, so to offer them the most reactive experience possible.
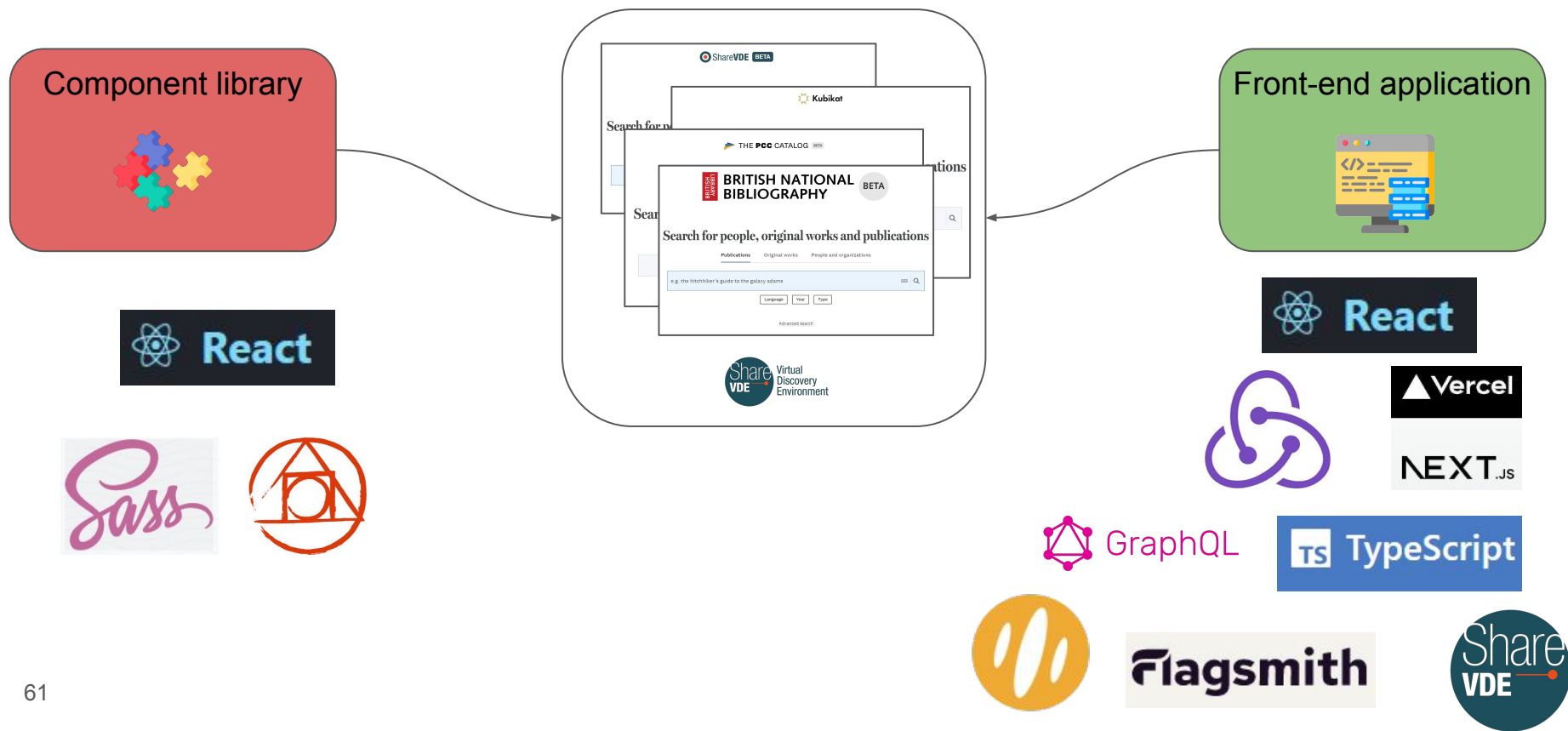
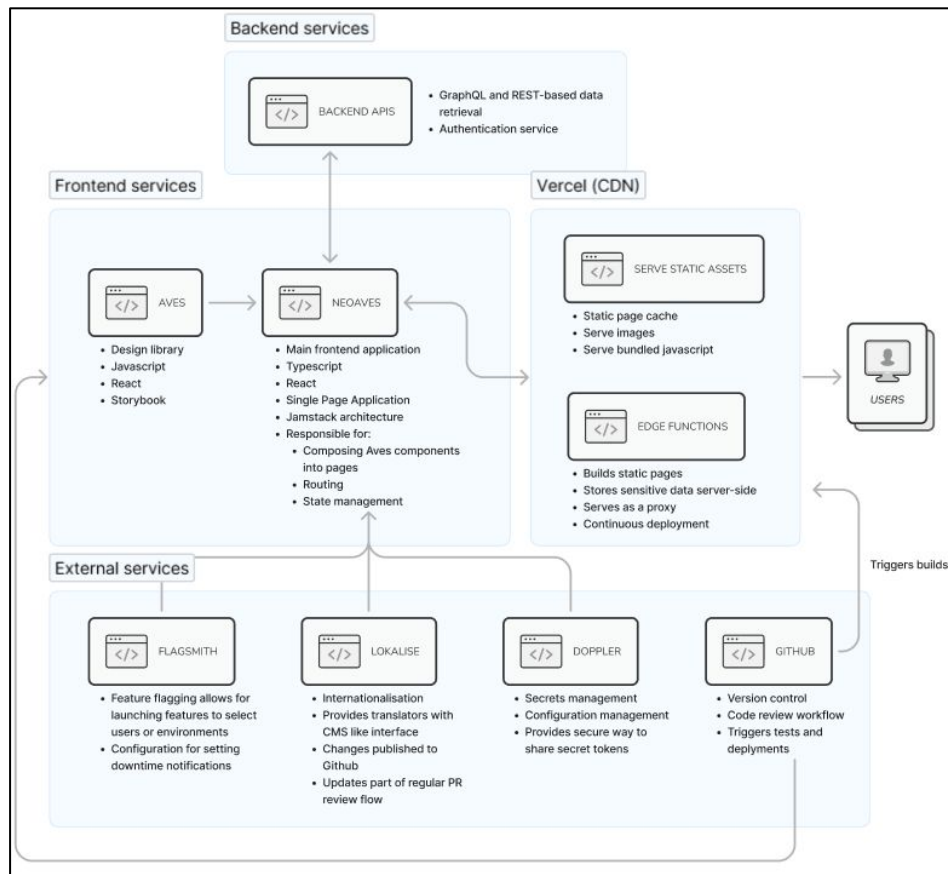# The Technology Stack (back-end)

# The Technology Stack (front-end)

# The Technology Stack (front-end architecture)

# Share-VDE: Infrastructure
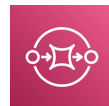


powered by aws

Amazon EC2

Amazon Relational Database Service (Amazon RDS)

AWS Lambda

Amazon Simple Queue Service (Amazon SQS)

Amazon EMR

Amazon Keyspaces (for Apache Cassandra)

# Thank you!

December 2022

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