



# SSHOC

social sciences & humanities open cloud

**Advancing SSH Research  
with SSHOCingly good  
and sustainable resources**

6 & 7 April 2022  
Brussels & online







Advancing SSH Research with SSHOCingly  
good and sustainable resources

6-7 April 2022 - Brussels & Online  
**REGISTER NOW**

## Opportunities for Cross- Disciplinary Interoperability in SSH research data management



**Daan Broeder**  
CLARIN



**Willem Elbers**  
CLARIN



**Mari Kleemola**  
Tampere University  
CESSDA



**Matej Ďurčo**  
OEAW / DARIAH



**Claus Zinn**  
CLARIN

# Opportunities for cross-disciplinary interoperability in the SSH

## Programme:

- Introduction - Daan Broeder, CLARIN ERIC
- ConversionHub - Mari Kleemola, (Tampere University/FSD & CESSDA)
- Virtual Collection Registry and Switchboard use-cases - Claus Zinn, (Tuebingen University/ CLARIN) and Willem Elbers (CLARIN ERIC)
- The SSH Vocabulary Commons, Matej Durco (OEAW/ DARIAH)
- Discussion



# Introduction to cross-disciplinary interoperability

---

Daan Broeder - CLARIN ERIC





# Building Infrastructure for the SSH, the challenge of interoperability

*Interoperability*: the ability for different research infrastructure components to be easily connected and exchange information.

- Exchanging and using data and metadata eg. metadata catalogues
- Referencing and accessing data (Citation, PIDs, AAI, ...)
- Processing and analysing data
- ...

It is not trivial to maintain interoperability within a single organization, more so for research infrastructures consisting from many organisations

# Why Interoperability, and why work towards interdisciplinary interoperability?

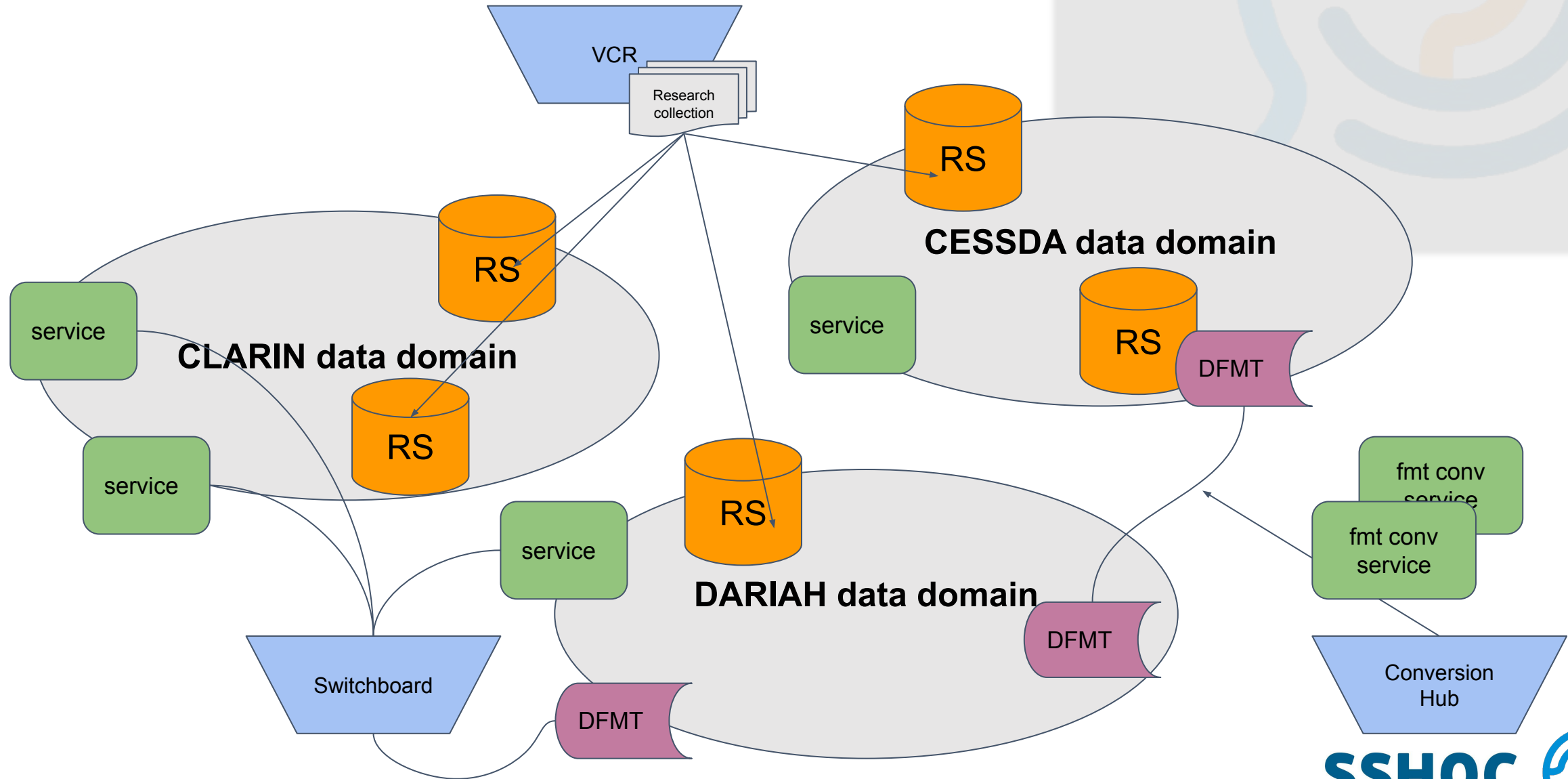
- Efficient data creation, processing & analysis by using combinations of interoperable tools and services, ie. not a single monolithic application
- Offers possibilities for sharing/reusing data and sharing services across communities
- More easy interdisciplinary research
- Scaling advantages



# Realising interdisciplinary interoperability

- Using standards, common recommendations, but note the
  - facing existing entrenched practices and existing investments
  - real problems & doubts about suitable common solutions eg. ONE metadata set, ONE ontology etc., in which case pragmatic conversions and mappings can be better solutions
- Common service development or selection of suitable existing services that are sufficiently flexible:
  - Support different required metadata sets
  - Support a sufficient number of data formats
- Common services that are found useful to researchers will also motivate further alignment of practices and stop fragmentation
- However it's not an automatic process and 'market research' and 'evangelization' is needed

# Between ussh







# Some final comments

Interdisciplinary Interoperability not only a matter of technical solutions but need also

- Cluster governance supporting such work, see the MoU formula
- Expert discussion & decision infrastructure
- Funding: either projects like SSHOC or in-kind

*Many interoperability related activities in SSHOC:*

- *T3.1 D3.1 interoperability challenges inventory and ConversionHub*
- *SSH Open Marketplace*
- *SSHOC Dataverse repository system*
- *CLARIN Switchboard and Virtual Collection Registry*
- *Common SSH vocabularies and their management*
- *SSH Data Citation*





# The SSHOC Conversion Hub

---

Mari Kleemola - Tampere University / FSD & CESSDA



# Task 3.5 Data and Metadata Interoperability Hub

- Metadata recommendations (D3.1)
- Conversion Hub
  - Content based on D3.1 and MS14. Reported in D3.6
- Format conversion solutions in development
  - eg. DDI-CMDI, DDI-DCAT-AP, DDI transformations,...
- Also still ongoing
  - Editorial work on Conversion Hub
  - Work on vocabulary commons

Task 3.5 Partners: CESSDA/FSD, CESSDA/SND, CESSDA/UKDS, CLARIN ERIC, CLARIN/EKUT, DARIAH/OEAW, ISTI/CNR, DARIAH.it/CNR, FORTH






# SSHOC recommended metadata standards by domain (D3.1)

All / common	Dublin Core, relaxed DataCite
Social Sciences / CESSDA, ESS, SHARE	DDI Codebook, DDI Lifecycle
Heritage Sciences / E-RIHS	CIDOC-CRM (and its extensions, especially PEM)
Language Sciences / CLARIN	CMDI
Arts and Humanities / DARIAH	CIDOC-CRM (and its extensions), EDM, TEI (teiHeader)

Each SSH community has its own established traditions and customs that may cause interoperability issues with other communities.

The existing diversity needs to be accepted since in most cases, the different formats and practices make good sense.

Conversion Hub can be used by researchers looking for solutions for metadata or data format conversions between what can be considered the major metadata standards and data formats used in the SSH.

 SSH OpenCluster  
Social Sciences & Humanities Resources

SSHOC Conversion Hub

Privacy policyLegal notice

Log in

Search for conversion solutions

Search

Expected Level of Knowledge

☐ advanced (18)  
☐ expert (13)  
☐ novice (11)  
☐ source and target schema knowledge  
-> XML schema expert (1)

Community

☐ CLARIN (20)  
☐ cross-community (15)  
☐ CESSDA (9)  
☐ DARIAH (5)  
☐ ERIHS (2)  
☐ LIBER (1)

Displaying 1 - 46 of 46

<https://conversion-hub.sshopencloud.eu>



# Kuha2



Solution

## Is a recipe

No

## Description

Kuha2 is a metadata server that provides descriptive social science research metadata for harvesting via multiple protocols and a growing variety of metadata standards. The software is a collection of applications and consists of three server applications, a client application and a database.

## Input format

[DDI Codebook 1.2.2](#)

[DDI Codebook 2.5](#)

[DDI Lifecycle 3.1](#)

## Input format comment

API available for format independent input

## Output format

[EAD3](#)

[Dublin Core](#)

[DDI Codebook 2.5](#)

## Invocation type

[source code/web application](#)

## Community

[CESSDA](#)

## Expected Level of Knowledge

[expert](#)

## Terms of Use / License

[EUPL-1.2](#)

## Link to landing page

[Documentation of Kuha2](#)

[Kuha2 repository](#)

## Status

maintained

## Source

[Documentation of Kuha2](#)

[FSD's Open Source Softwares](#)

## Contact

Toni Sissala / Finnish Social Science Data Archive

## Publication date

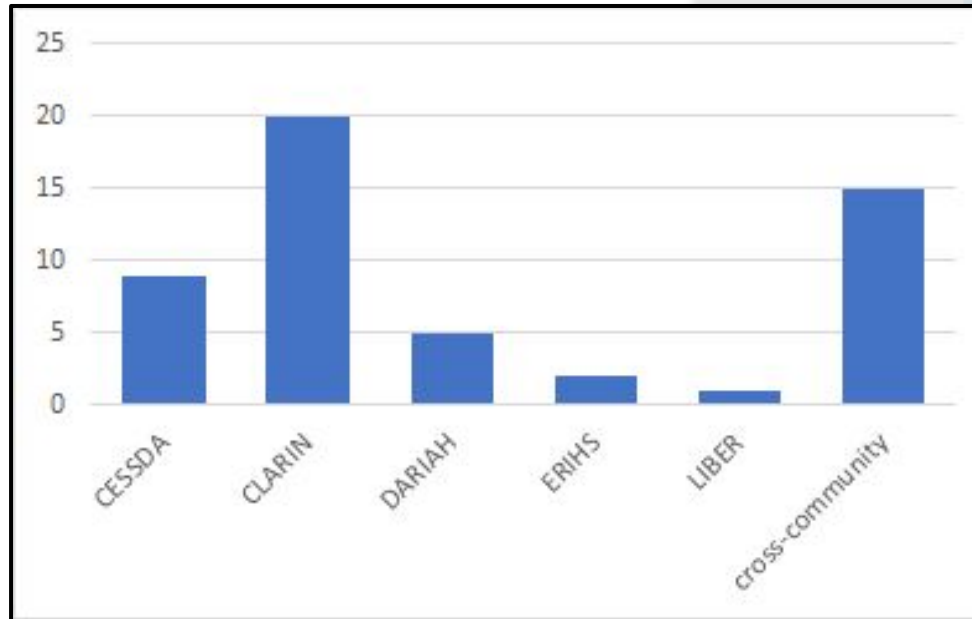
2017-10-25

## Last Modification date

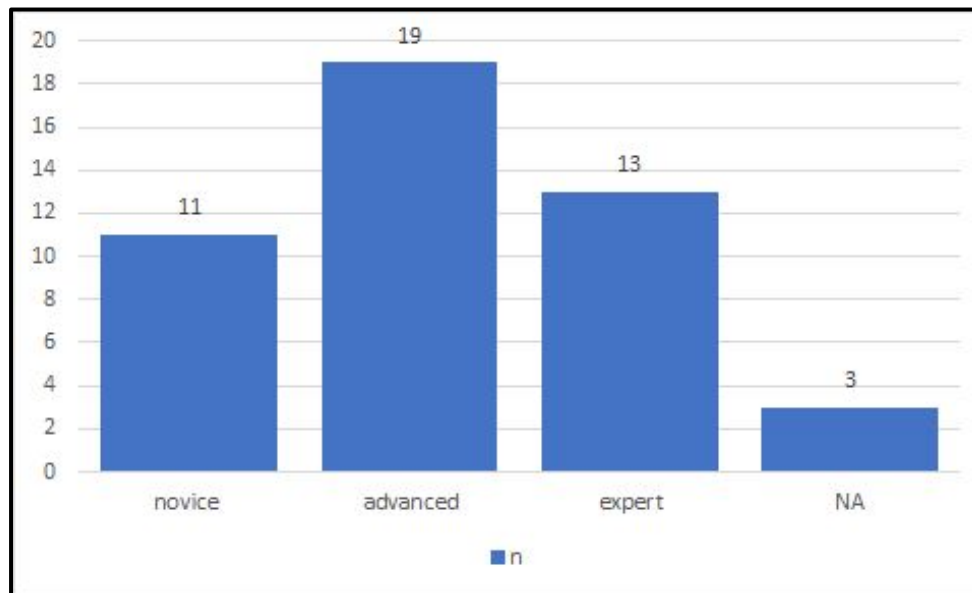
2021-05-05



- Number of solutions in Conversion Hub by community  
(the same solution can be listed under more than one community)



- Expected level of knowledge needed to use the solutions





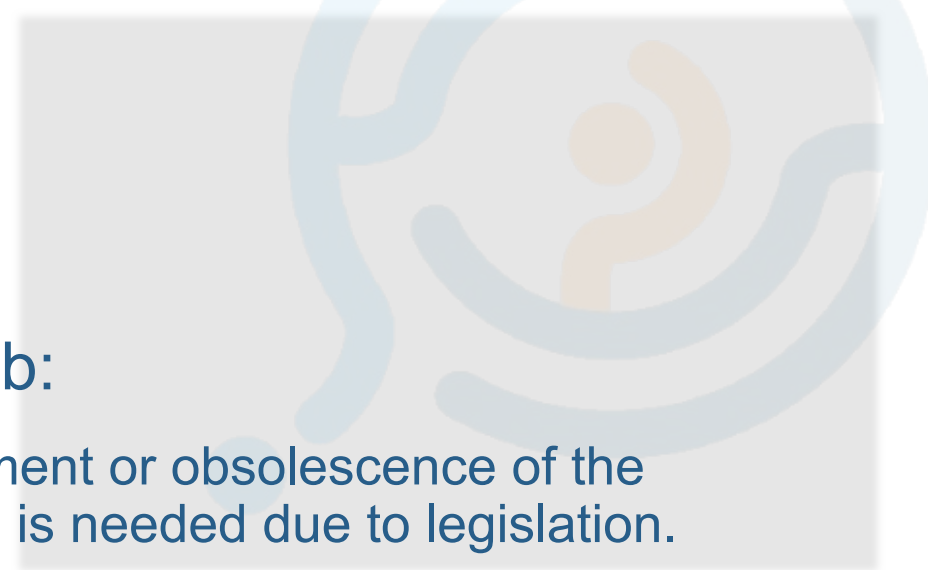
# Conversion Hub technology and functions

- Conversion Hub runs on Drupal 9
  - Data model specified in D3.6
  - Main functionalities that guided the implementation include: user management, content management, comments, API, structured data, faceted search, search options, and input form.
- => enable a good search and discovery experience for visitors, easy editing and creating of items by curators, and re-use of the data by external services - like the SSH Open Marketplace - following the FAIR data principles



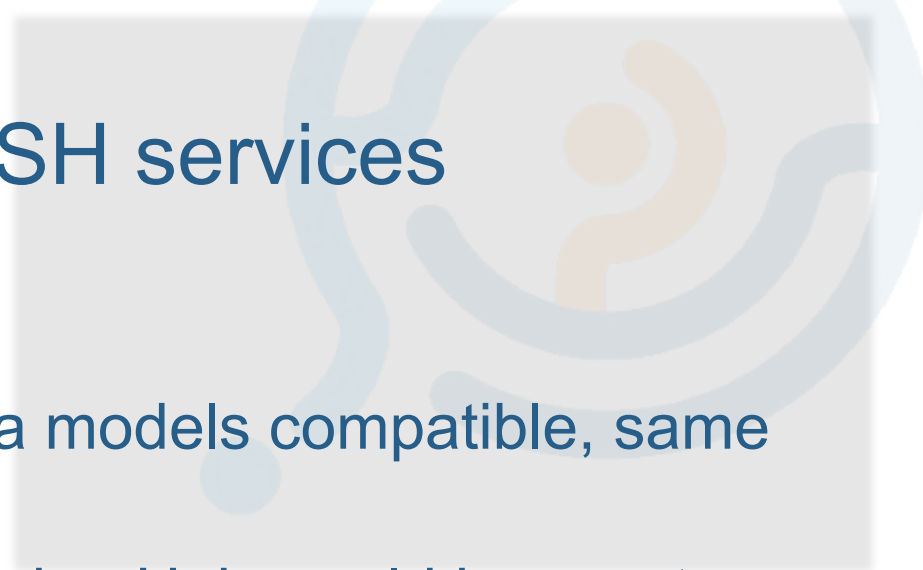
# Conversion Hub: sustainability

- D3.6 proposes scenarios for the Conversion Hub:
  - A. Basic hosting and maintenance through to replacement or obsolescence of the product, editorial actions only per request and if action is needed due to legislation.
  - B. A + ongoing editorial control and addition of solutions over time.
  - C. B + active research and development of content and technology, based on users' needs and technological advances.
- The costs will increase from A to C but overall, the costs of keeping Conversion Hub up and running post-SSHOC are moderate.
- Post-project (2 years): ACDH-CH / hosting technology; CESSDA / editorial work
- Individual conversion solutions created by T3.5: expectation is that the developing organisations will sustain them after the project



# Possible synergies with MP and other SSH services

- Conversion Hub is one of the sources for MP (data models compatible, same vocabularies used)
- Opportunities to maintain and develop the Conversion Hub would be greater, and costs lower, if it were part of a wider solution or a pool of interconnected services.
- It should be possible to find synergy by creating, for example, common curation, technical or governance teams across SSH services. The level to which costs could be defrayed by sharing roles across contractual services depends on the specific model and partnerships involved.





# Further information

Broeder, Daan, Trippel, Thorsten, Degl'Innocenti, Emiliano, Giacomini, Roberta, Sanesi, Maurizio, Kleemola, Mari, ... Ďurčo, Matej. (2019). **SSHOC D3.1 Report on SSHOC (meta)data interoperability problems** (Version v1.0). Zenodo. <https://doi.org/10.5281/zenodo.3569867>

Mari Kleemola, Daan Broeder, Thorsten Trippel, Iris Alfredsson, Matej Ďurčo, Klaus Illmayer, Eleni Tsoulouha, Maurizio Sanesi, & Emiliano Degl'Innocenti. (2020). **MS14 Inventory of existing (meta-)data format interoperability solutions** (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.4581290>

Mari Kleemola, Katja Moilanen, Daan Broeder, Matej Ďurčo, Klaus Illmayer, Maurizio Sanesi, Emiliano Degl'Innocenti, Hervé L'Hours, Benjamin Mathers, Johan Fihn Marberg, Eleni Tsoulouha, Athina Kritsotaki, & Cesare Concordia. (2021). **D3.6 Report on SSHOC format interoperability solution services, including new software**. <https://doi.org/10.5281/zenodo.5561604>



# Switchboard and Virtual Collection Registry in the SSH

---



# Final Conference – Interoperability demonstration

---

Building the VCR Service for the SSH

Willem Elbers, Researcher, CLARIN ERIC

**SSHOC Final Conference**

**April 6th, 2022**

Brussels & Virtual

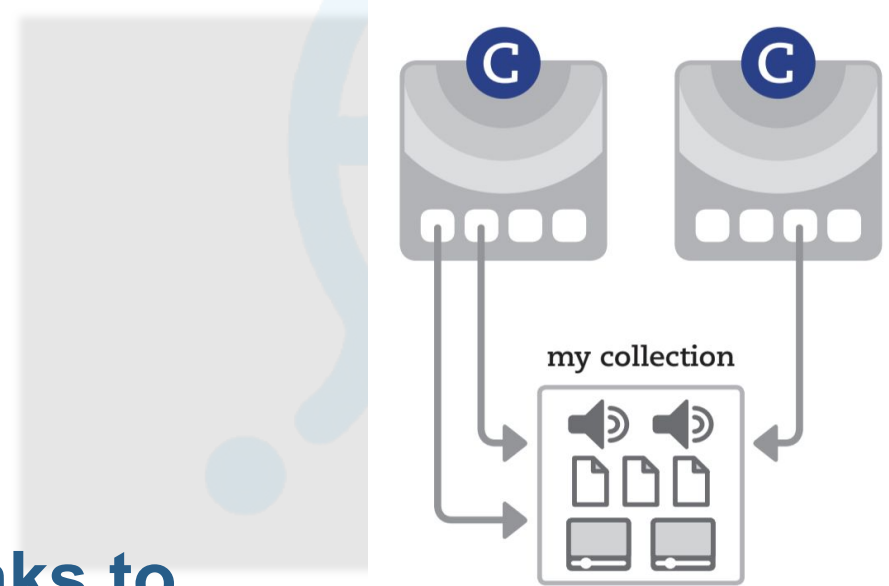


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782





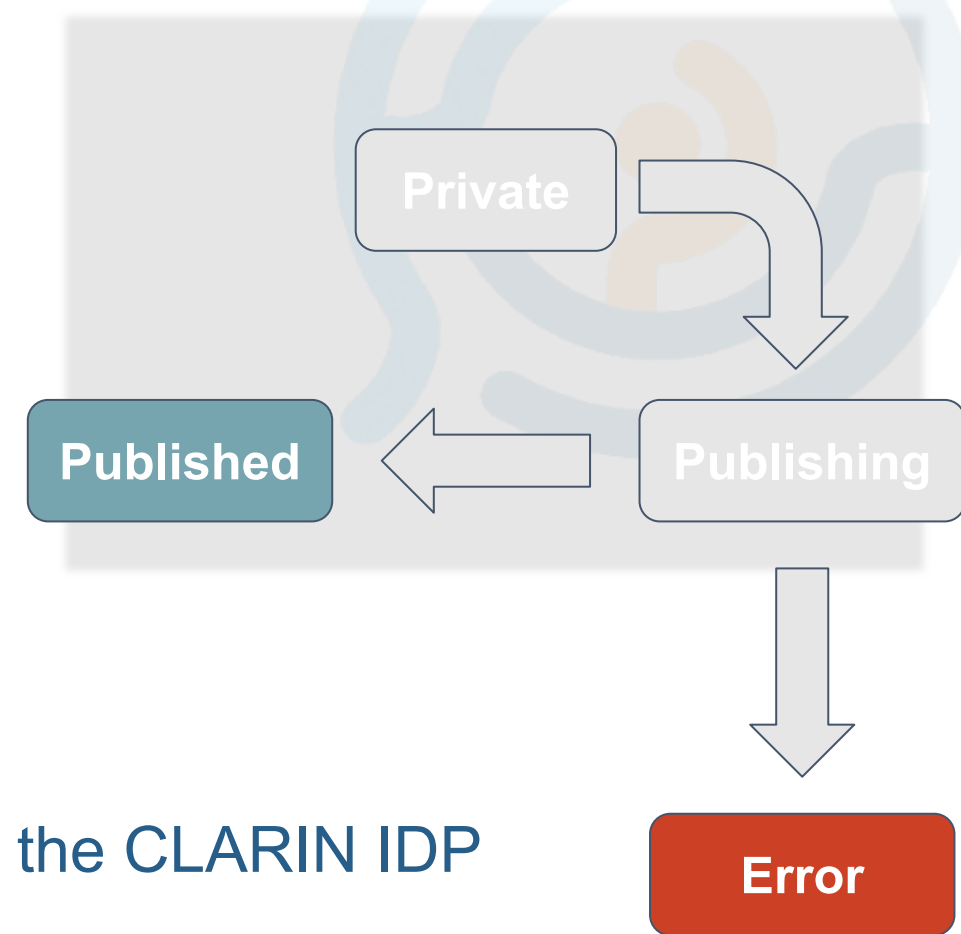
# Overview



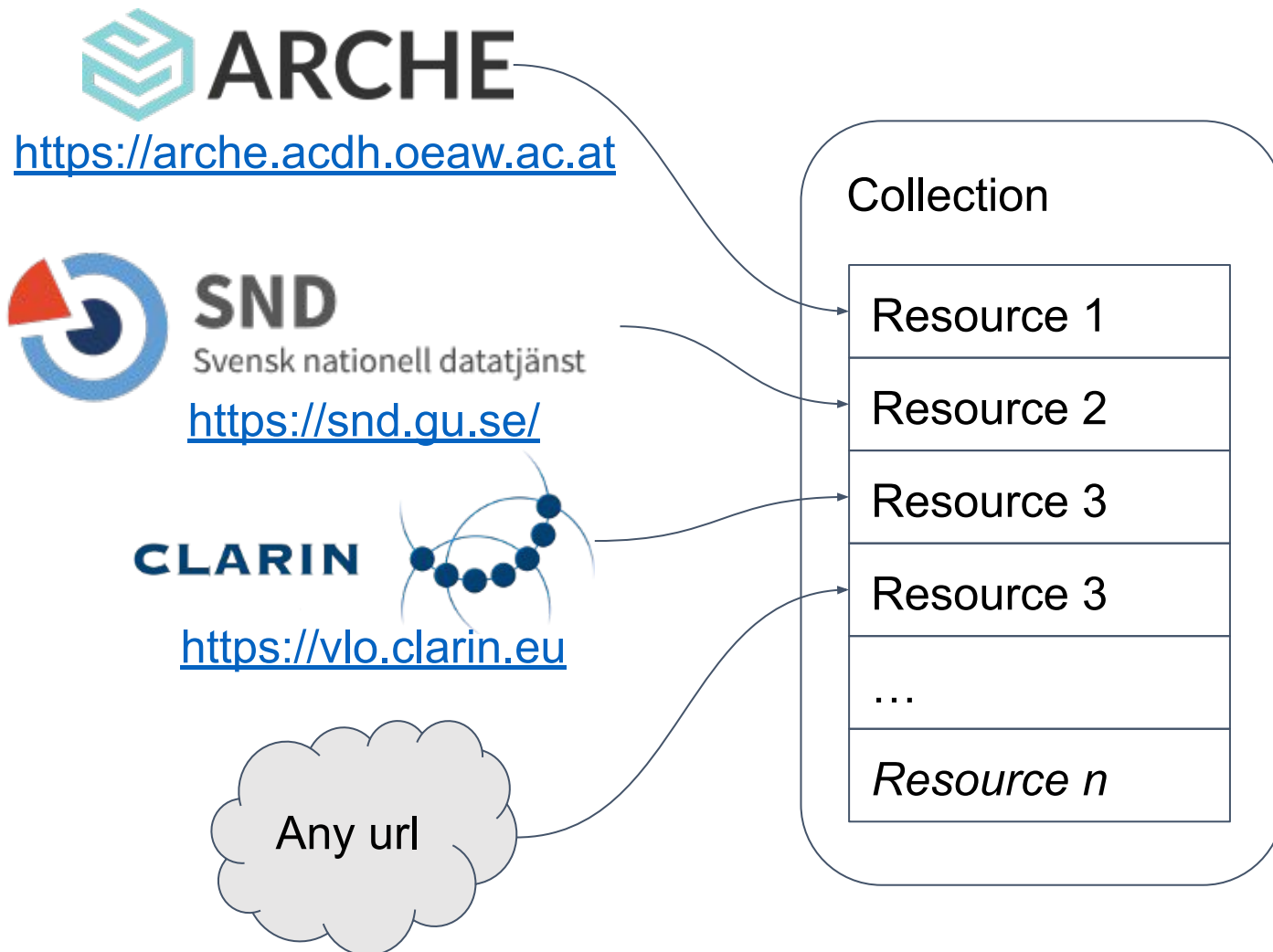
- A **virtual collection** is a coherent **set of links to objects** (e.g. annotated text, video) that can be
  - Easily created, accessed and cited
  - Originate from **different archives**, hence the term virtual
  - **Manually accessed** as well as **automatically processed**
- The Virtual Collection Registry is a **mature service** offered by CLARIN ERIC to **manage sets of links** to digital objects for **easy access** and **citation**

# Access

- All published collections are publicly accessible and citetable with a PID
- Login possible via most eduGAIN IDPs or the CLARIN IDP
- After login a user can create and edit their own collections.
  - While editing a collection is private and only accessible for the creator
  - When the collection is ready, it can be published after which it becomes publicly available (and gets assigned a PID, including a DOI)
  - Editing is discouraged after publication



# Integration with third party data catalogues



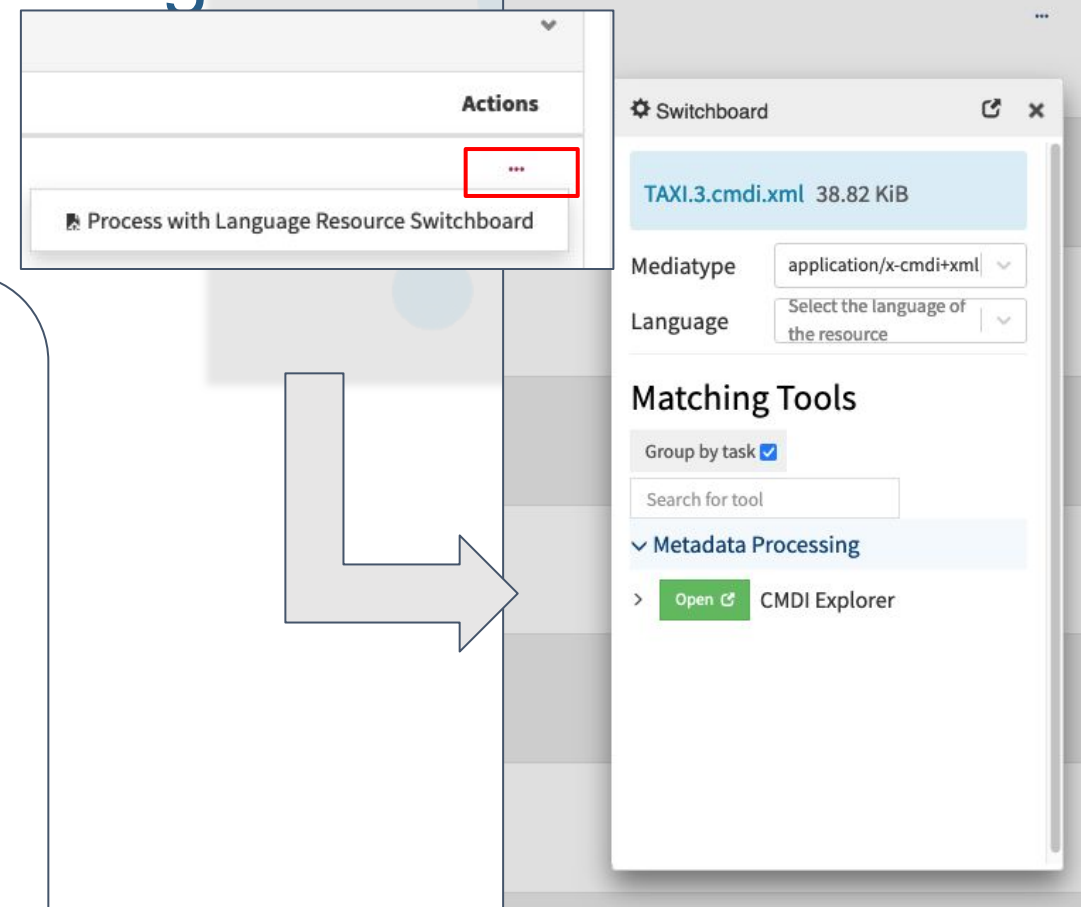
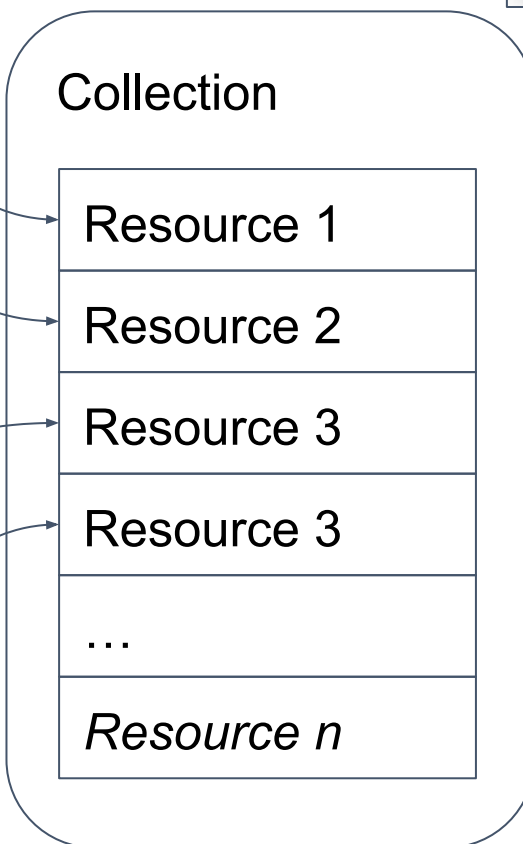


# Collection as input for further processing

 **ARCHE**  
<https://arche.acdh.oew.ac.at>

 **SND**  
Svensk nationell datatjänst  
<https://snd.gu.se/>

**CLARIN**  
  
<https://vlo.clarin.eu>



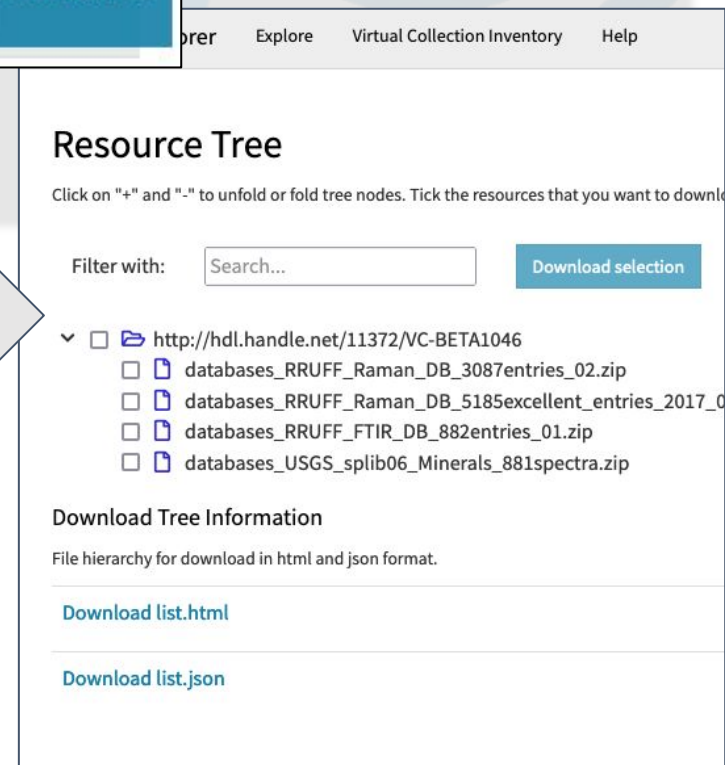
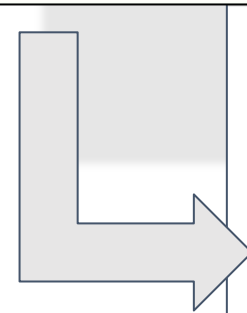
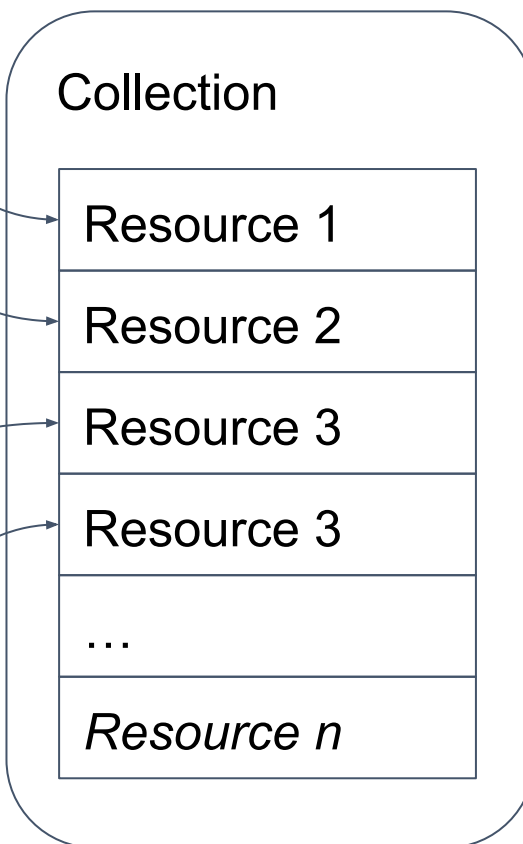
Processing collections with the  
Language Resource Switchboard

# Collection as input for further processing

 **ARCHE**  
<https://arche.acdh.oeaw.ac.at>

 **SND**  
Svensk nationell datatjänst  
<https://snd.gu.se/>

**CLARIN**  
  
<https://vlo.clarin.eu>



Download collections with the  
CMDI Explorer

# Demo

## Goal

A researcher is interested in germanic languages. By using various data catalogues an explorative study is conducted, resulting in a collection of resources related to germanic languages, preferably dialects or spoken by a minority. The results must be easy to share and cite via a DOI.



# Use cases



- Creation of collections on a specific topic
  - Publications
  - Training
  - ...
- Creating a neutral ground across repositories
  - Saving results from searches in different data repositories in a single collection
  - Adding results to an already existing collection
- Collections as input for workflows
  - Via switchboard integration
  - Downloading collections as an archive





# Thank you for your attention!

Virtual Collection Registry:  
<https://collections.clarin.eu/>

Source code:  
<https://github.com/clarin-eric/VirtualCollectionRegistry>

Integration instructions:  
<https://github.com/clarin-eric/VirtualCollectionRegistry/blob/master/doc/Integration.md>



# Final Conference – Interoperability demonstration

Building the VCR and Switchboard Services for the SSH

Claus Zinn, Researcher, University of Tuebingen

**SSHOC Final Conference**

**April 6th, 2022**

Brussels & Virtual

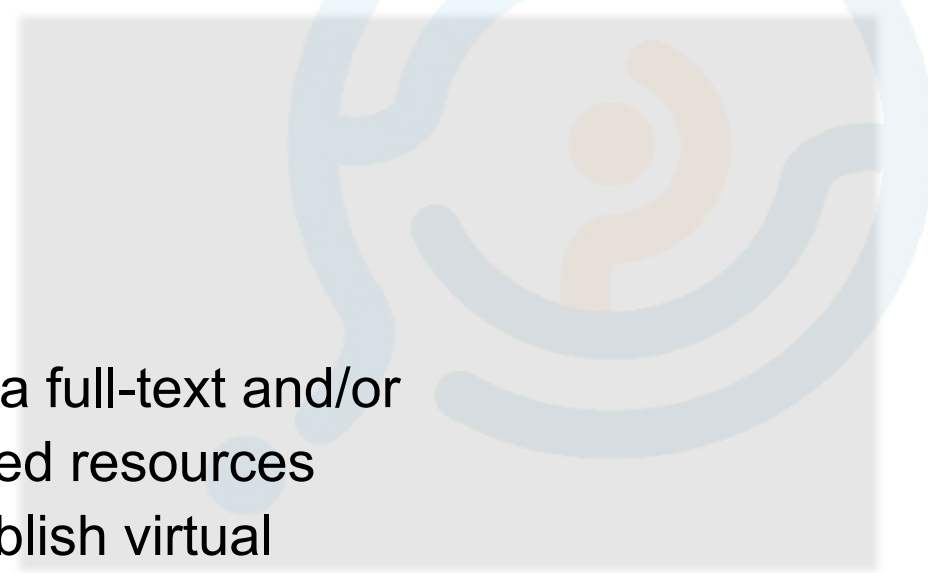


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

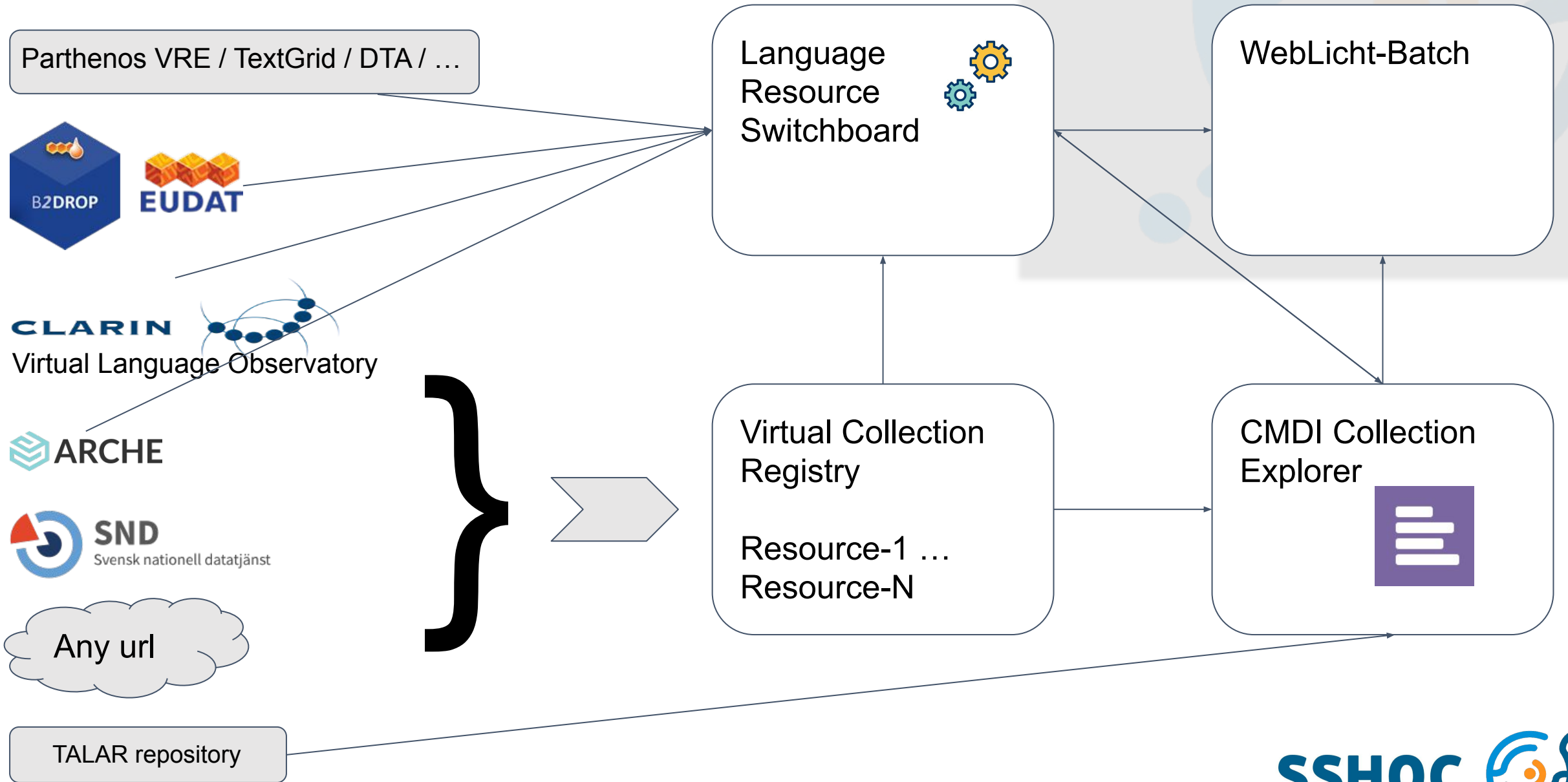


# Tool Overview and their interoperability

- [Virtual Language Observatory](#), where users can perform a full-text and/or faceted search to get access to 1 million+ language-related resources
- [Virtual Collection Registry](#), where users can build and publish virtual collections of resources
- [CMDI Explorer](#), where users can visualize and explore rich metadata descriptions and select resources they describe for further processing (e.g., download)
- [Language Resource Switchboard](#), where users can process a given resource with applicable tools
- [WebLicht-Batch](#), where users can batch-process sets of resources (or a large resource) with WebLicht, a workflow engine for linguistic analysis



# Interoperability (cont'd)



# Interoperability by design

- SSH repositories can
    - provide their metadata to the VLO (via OAI-PMH endpoint)
    - include the Switchboard plugin into the site (via JavaScript code)
    - make available the VCR from their site (via HTTP protocol)
  - SSH tool developers can integrate their tool with the Switchboard
- 
- Easy integration
  - Small investment, big return
  - Many efforts were spent in discussion with SSHOC communities





# Usage Scenario

## •Find resources

- from different venues / sites / repositories
  - e.g., in the Virtual Language Observatory (VLO), via their metadata
- manage resource set
  - e.g., in the Virtual Collection Registry
  - e.g., in CMDI Explorer via their metadata descriptions

## •Process resources

- with text analysis software
  - e.g., using the Language Resource Switchboard
  - invokable from the VLO, from CMDI Explorer, from repository sites

## •Manage processing results

- e.g., by storing them on cloud space
- and by adding references to their cloud location to the virtual collection



•

# DEMO – TIME

- <https://b2drop.eudat.eu/>
- <https://collections.clarin.eu/public?1>
- <https://weblicht.sfs.uni-tuebingen.de/CMDIExplorer/>
- <https://switchboard.clarin.eu/>
- <https://vlo.clarin.eu/?1>
- <https://weblicht.sfs.uni-tuebingen.de/weblicht-batch/>

The following slides are back-up only slides.



# Switchboard access via B2DROP

The screenshot shows a web browser window with the URL `b2drop.eudat.eu/apps/files/?dir=/&fileid=6034145`. The interface includes a sidebar with navigation options: All files, Recent, Favorites, Shares, Tags, and Shared to Circles. The main area displays a file list with columns for Name, Size, and Modified. A context menu is open over the file 'einstein2.txt', showing options: Add to favorites, Details, Rename, Move or copy, Download, B2SHARE, Switchboard, and Delete file. The file 'einstein2.txt' is 13 KB and was modified 6 months ago. The total size of the folder is 21.6 MB.

Name	Size	Modified
AP2-3-2-II	11.3 MB	5 months ago
Kooperation_Differenz_Projekt_II	< 1 KB	2 months ago
TRIAL sportbonden data	195 KB	9 days ago
DDR 205 Issue P02.zip	10.2 MB	4 months ago
einstein2.txt	13 KB	6 months ago
Readme.md	< 1 KB	4 days ago

# 1. Finding short stories in the VLO, & creating virtual collection

The screenshot shows the Virtual Language Observatory (VLO) search results page. The browser address bar displays the URL: `vlo.clarin.eu/search?0&fq=format:text/plain&fq=licenseType:PUB&fq=licenseType:RES&fqType=format:or&fqType=licenseType:or&q=short+story`. The page header includes the VLO logo, a search bar, and navigation links for Search, Contributors, and Help. The search results are displayed in a list format, showing 6 results within the selection for 'short story', 'text/plain', and 'Public or Restricted for individual'. The results are filtered by 'text/plain' and 'Public or Restricted for individual'. The results per page are set to 10.

Use the categories below to limit the search results to those matching the selected value(s).

- Language
- Collection
- Resource type
- Format
- text/plain

Temporal Coverage

Availability

Search options

Multiple value selection behaviour: OR

☐ Only include collection resources

☒ Create virtual collection from search results

**The squire's story**  
(Part of OTA Core Collection)  
No description  
English  
Landing page for this record

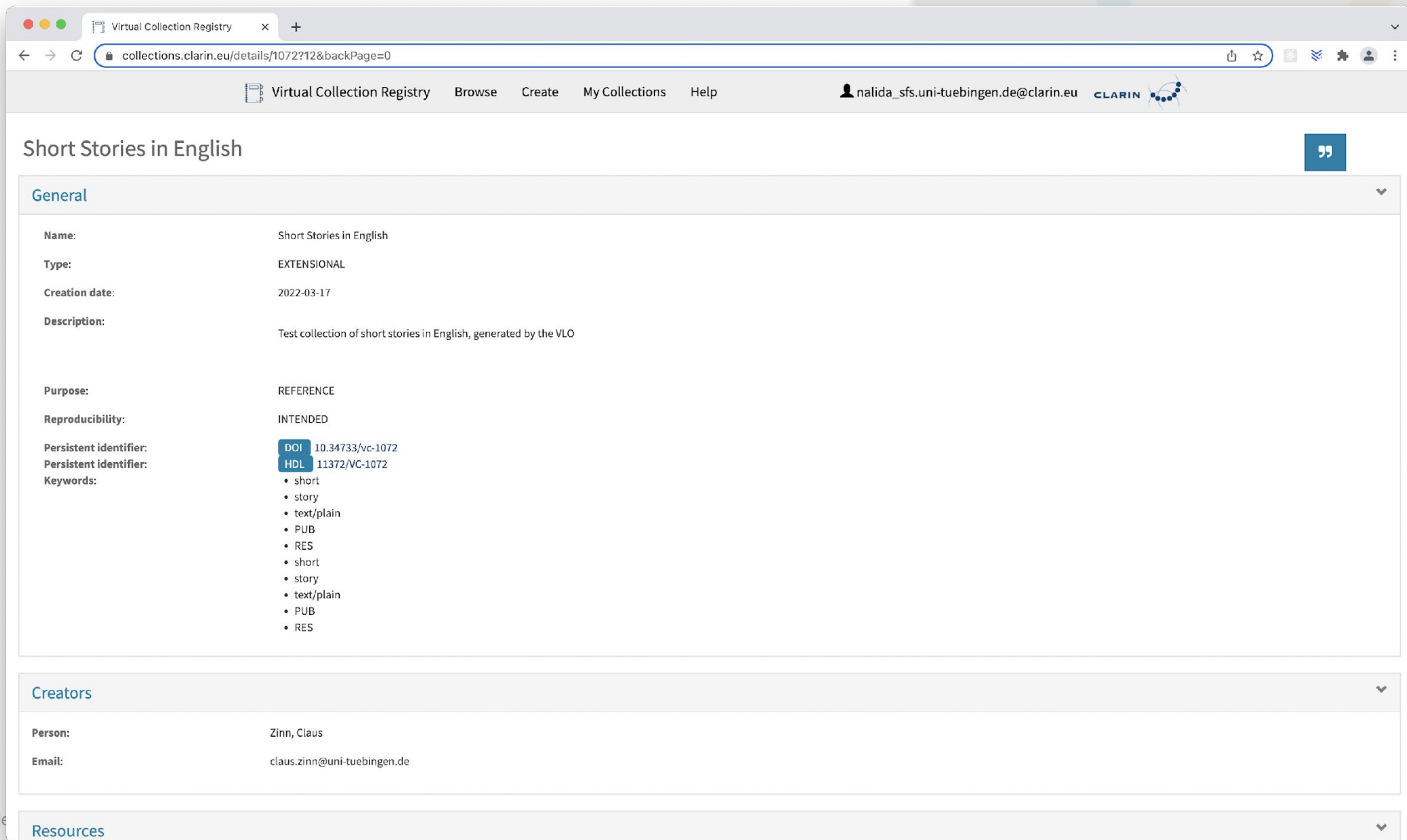
**The old nurse's story**  
(Part of OTA Core Collection)  
No description  
English  
Landing page for this record

**A ghost story / by Mark Twain**  
(Part of OTA Core Collection)  
No description  
English  
Landing page for this record

**Bartleby, the scrivener: a story of Wall-street**  
(Part of OTA Core Collection)  
No description  
English  
Landing page for this record



## 2. Viewing the newly created collection in the VCR



The screenshot displays the Virtual Collection Registry (VCR) web interface. The browser address bar shows the URL `collections.clarin.eu/details/1072?12&backPage=0`. The page title is "Short Stories in English". The interface includes a navigation bar with links for "Browse", "Create", "My Collections", and "Help", along with a user profile "nalida\_sfs.uni-tuebingen.de@clarin.eu" and the CLARIN logo.

The main content area is titled "Short Stories in English" and features a "General" tab. The details are as follows:

<b>Name:</b>	Short Stories in English
<b>Type:</b>	EXTENSIONAL
<b>Creation date:</b>	2022-03-17
<b>Description:</b>	Test collection of short stories in English, generated by the VLO
<b>Purpose:</b>	REFERENCE
<b>Reproducibility:</b>	INTENDED
<b>Persistent identifier:</b>	DOI 10.34733/vc-1072
<b>Persistent identifier:</b>	HDL 11372/VC-1072
<b>Keywords:</b>	<ul style="list-style-type: none"><li>• short</li><li>• story</li><li>• text/plain</li><li>• PUB</li><li>• RES</li><li>• short</li><li>• story</li><li>• text/plain</li><li>• PUB</li><li>• RES</li></ul>

Below the "General" tab is the "Creators" section, which lists:

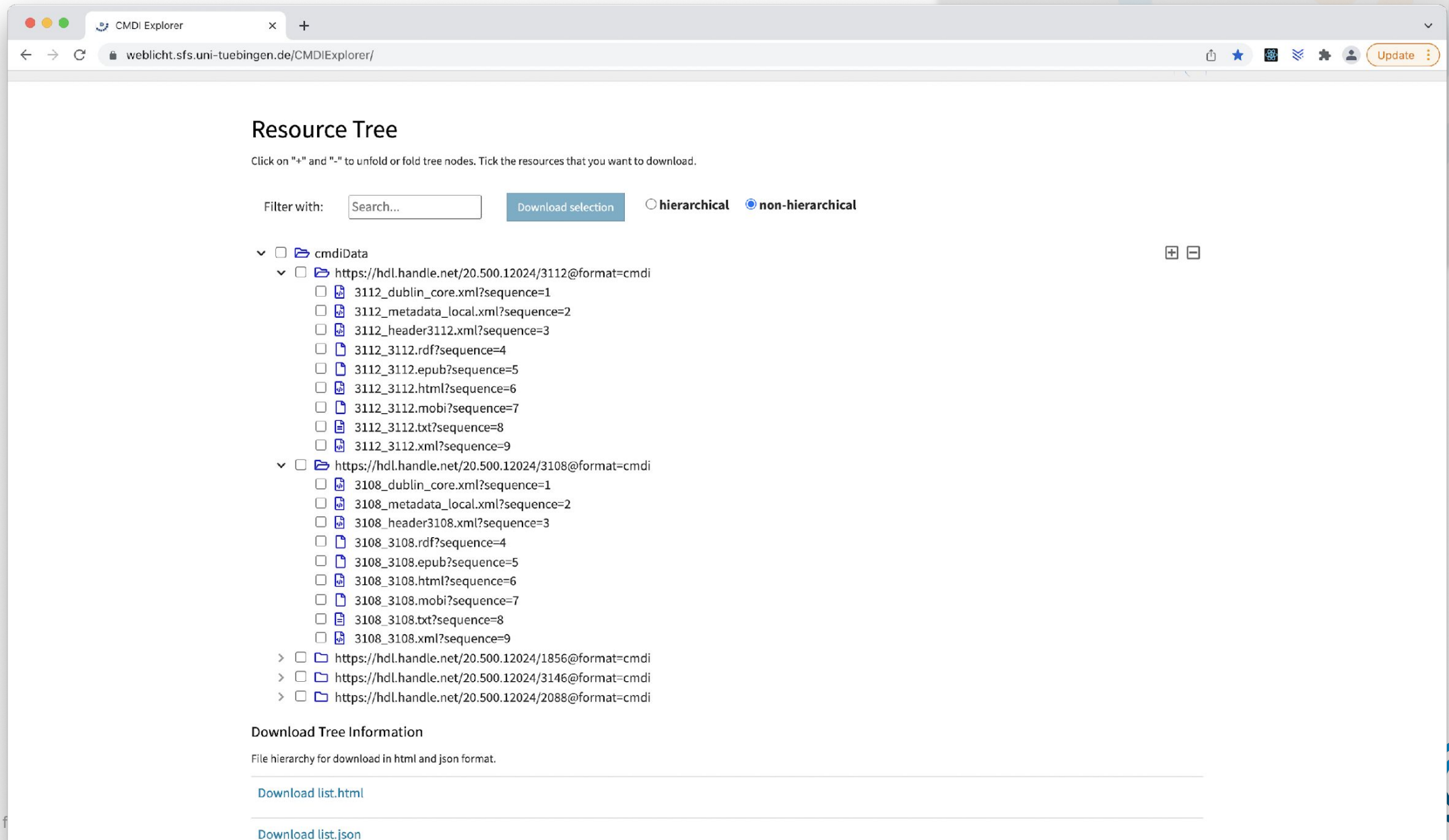
<b>Person:</b>	Zinn, Claus
<b>Email:</b>	claus.zinn@uni-tuebingen.de

At the bottom of the page is the "Resources" section, which is currently empty.





### 3. Sending collection to CMDI Explorer for further action



The screenshot shows the CMDI Explorer web application in a browser window. The address bar shows the URL `weblicht.sfs.uni-tuebingen.de/CMDIExplorer/`. The page title is "CMDI Explorer".

## Resource Tree

Click on "+" and "-" to unfold or fold tree nodes. Tick the resources that you want to download.

Filter with:  [Download selection](#) ☐ hierarchical ☒ non-hierarchical

- [-] ☐ cmdiData
  - [-] ☐ <https://hdl.handle.net/20.500.12024/3112@format=cmdi>
    - ☐ [3112\\_dublin\\_core.xml?sequence=1](#)
    - ☐ [3112\\_metadata\\_local.xml?sequence=2](#)
    - ☐ [3112\\_header3112.xml?sequence=3](#)
    - ☐ [3112\\_3112.rdf?sequence=4](#)
    - ☐ [3112\\_3112.epub?sequence=5](#)
    - ☐ [3112\\_3112.html?sequence=6](#)
    - ☐ [3112\\_3112.mobi?sequence=7](#)
    - ☐ [3112\\_3112.txt?sequence=8](#)
    - ☐ [3112\\_3112.xml?sequence=9](#)
  - [-] ☐ <https://hdl.handle.net/20.500.12024/3108@format=cmdi>
    - ☐ [3108\\_dublin\\_core.xml?sequence=1](#)
    - ☐ [3108\\_metadata\\_local.xml?sequence=2](#)
    - ☐ [3108\\_header3108.xml?sequence=3](#)
    - ☐ [3108\\_3108.rdf?sequence=4](#)
    - ☐ [3108\\_3108.epub?sequence=5](#)
    - ☐ [3108\\_3108.html?sequence=6](#)
    - ☐ [3108\\_3108.mobi?sequence=7](#)
    - ☐ [3108\\_3108.txt?sequence=8](#)
    - ☐ [3108\\_3108.xml?sequence=9](#)
  - > ☐ <https://hdl.handle.net/20.500.12024/1856@format=cmdi>
  - > ☐ <https://hdl.handle.net/20.500.12024/3146@format=cmdi>
  - > ☐ <https://hdl.handle.net/20.500.12024/2088@format=cmdi>

### Download Tree Information

File hierarchy for download in html and json format.

[Download list.html](#)

[Download list.json](#)



## 4. Selecting parts of collection with the Switchboard

The screenshot shows a web browser window with the URL `switchboard.clarin.eu/#/cmdiexplorer/https%3A%2F%2Fweblicht.sfs.uni-tuebingen.de%2FCMDIExplorer%2Fapi%2Fdeliverzip%2Ffea474ff6-2550-4984-be87-702718558fd4`. The page title is "Language Resource Switchboard". The navigation bar includes "Upload", "Tool Inventory", and "Help". The main content area is titled "Resources" and displays a list of files for selection. The files are listed in a table with columns for filename, size, and a "Hide content" button. The files are:

Filename	Size
downloaded/1856_ghost-1856.txt?sequence=3	13.07 KiB
downloaded/2088_bartleby-2088.txt?sequence=3	86.14 KiB
downloaded/3108_3108.txt?sequence=8	48.60 KiB
downloaded/3112_3112.txt?sequence=8	35.06 KiB
downloaded/3146_3146.txt?sequence=8	80.75 KiB
downloaded/list.html	27.76 KiB
downloaded/list.icon	7.80 KiB

Below the table, it states "No matching tools found." The footer includes the European Union flag, the text "This project is", the version "v2.4.0", the text "Service provided by CLARIN", and a "Contact" link.

## 5. Selecting parts of collection with WebLicht-Batch

The screenshot displays the WebLicht-Batch web interface. At the top, there's a navigation bar with tabs for 'WebLicht-Batch', 'Input Data', 'Inspect Task Status', and 'Help'. The 'WebLicht-Batch' tab is active. Below the navigation bar, there's a section for resource management showing a resource named 'unknown' (108.19 KiB) with a 'Delete resource' button. To the right, there are dropdown menus for 'Mediatype' (application/zip), 'Language' (English), and 'WebLicht Chain' (Named Entities). Below these, there are buttons for 'Cancel Processing' and 'Copy Userkey'.

**Task Progress**

A progress bar shows the task is approximately 27% complete. Below the bar, it says '(elapsed time: 27)'.

**Task Progress Info**

**3108\_3108.txt?sequence=8**  
(progress: 1, elapsed time: 10 seconds)

**Batch Progress Info**

batch item	current service	progress
1	Illinois Named Entity Recognizer	1

**1856\_ghost-1856.txt?sequence=3**  
(progress: 1, elapsed time: 5 seconds)

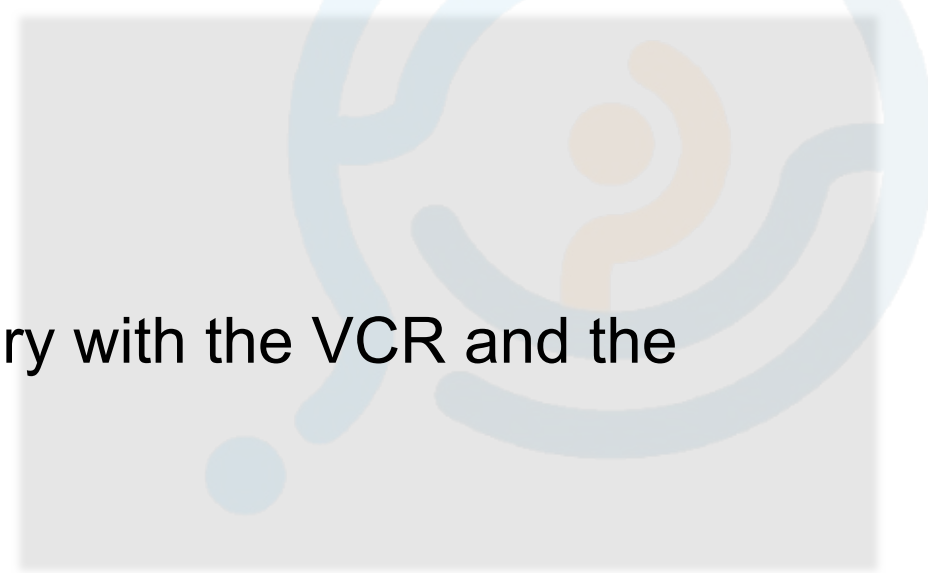
**Batch Progress Info**

batch item	current service	progress
1	Illinois Named Entity Recognizer	1

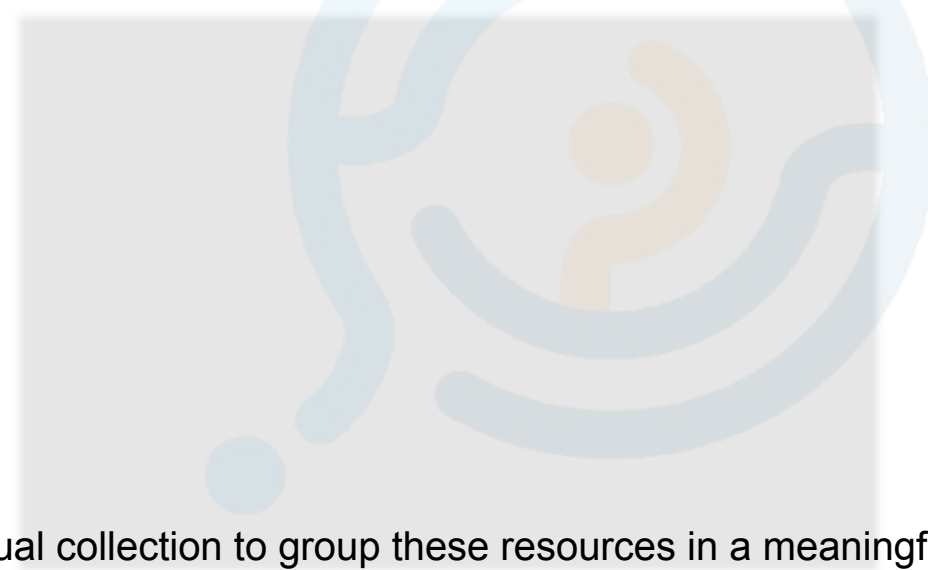


# Discussion

- Research Infrastructure Story
  - Easy, seamless integration of YOUR repository with the VCR and the Switchboard
  - Add your tools to the Switchboard !
  - Cross data domains with tools / services
- AAI still a big problem across data domains
  - While metadata is always available, the resources are often not easily accessible
  - Huge impact on RI success story
    - No one-click away from the next action, but detours as workaround
- Metadata quality & standards across data domains
  - E.g., metadata showing to landing pages rather than actual resources
  - CMDI not used in all data domains



# Demo - playbook



Use a variety of data catalogues to collect a number of relevant resources. Use a virtual collection to group these resources in a meaningful way.

## Collection of germanic language studies

- SND (<https://snd.gu.se/>): search for germanic
  - Documentation of Gutnish phonetics and phonology ([link](#))
  - Add to shopping basket
  - Submit collection
  - Login
  - Create new collection
  - Save collection
- ARCHE (<https://arche.acdh.oeaw.ac.at/>): search for DIAUMA
  - Add to collection
  - Merge with earlier created collection
  - Save collection



# Demo - playbook



- VLO: search for “germanic and swedish”
  - Nordic Dialect Corpus v. 4.0 ([link](#))
  - Submit to collection
  - Merge with earlier created collection
  - Save collection
- Edit collection
  - Add TextGrid link
  - (Optional) Add wikipedia link
    - [https://en.wikipedia.org/wiki/Germanic\\_languages](https://en.wikipedia.org/wiki/Germanic_languages)
  - Save collection
- Publish collection
- Show assigned PIDs
- Show downloading



# SSH Vocabulary Commons

---



# SSH Vocabulary Commons

---

Matej Durco (OEAW, DARIAH)



# Towards a SSH Vocabulary Commons

Common interest by RIs: CESSDA, CLARIN, DARIAH and E-RIHs

Collaborative use and management of vocabularies  
vocabularies as first-class citizens / FAIR data objects in their own right.

Mission: [vocabulary commons charter](#)

SSHOC activities wrt vocabularies:

- tasks 3.1 Multilingual terminologies, 3.5 Interoperability, WP7 SSH Open Marketplace usage of vocabularies etc.
- CLARIN Vocabulary Initiative inventorizing relevant SSH vocabularies and organizing info sessions on vocabulary management software and SSH requirements
- ICTeSSH workshop on vocabulary use in the SSH
- Wider scope for collaboration in the Humanities eg. TRIPLE project

# Priorities for the Vocabulary Commons

- Operating a Vocabulary repository
- SSH vocabulary overview
  - of all relevant SSH vocabularies
- Vocabulary federated content search
  - “Deep” search **also** in the terms (essential for recommendations)
- Harmonized versioning of vocabularies
  - yet no agreed way of how to manage update of vocabularies and individual vocabulary terms
- API standardisation:
  - to enable tools to interact smoothly with different vocabulary platforms
- Foster exchange between users and developers (eg SKOSMOS)





In many areas of scholarly work, controlled vocabularies (gazetteers, thesauri, etc.) play a crucial role as a stable reference for resources and for ensuring interoperability between disparate projects and their data. [Read more](#)

from all English

Search

## SSH Vocabulary Commons Categories

### SSHOC

Data Stewardship Terminology  
EOSC Geographical Availability List  
EOSC Life Cycle Status List  
EOSC Resource Category List  
EOSC Technology Readiness Level  
Expertise level for training resources  
Formats of training resources  
IANA Media Types  
Intended audience  
Invocation types  
Multilingual Metadata  
Software License  
SSK Standards List  
Status of training resources  
The Bibliographic Ontology (bibo) Concept Scheme (parts used in SSHOC)

SKOSMOS instance to  
publish the SSHOC  
vocabulary results and  
other SSH vocabularies

<https://vocabs.sshopencloud.eu/vocabularies/>

## CONTACT



ACDH-CH  
Austrian Centre for Digital  
Humanities and Cultural Heritage  
Austrian Academy of Sciences

Sonnenfelsgasse 19,  
1010 Vienna

T: +43 1 51581-2200  
E: [acdhelpdesk\(at\)oeaw.ac.at](mailto:acdhelpdesk(at)oeaw.ac.at)

## PARTNERS



## HELPDESK

ACDH-CH runs a helpdesk offering  
advice for questions related to various  
digital humanities topics.

[ASK US!](#)



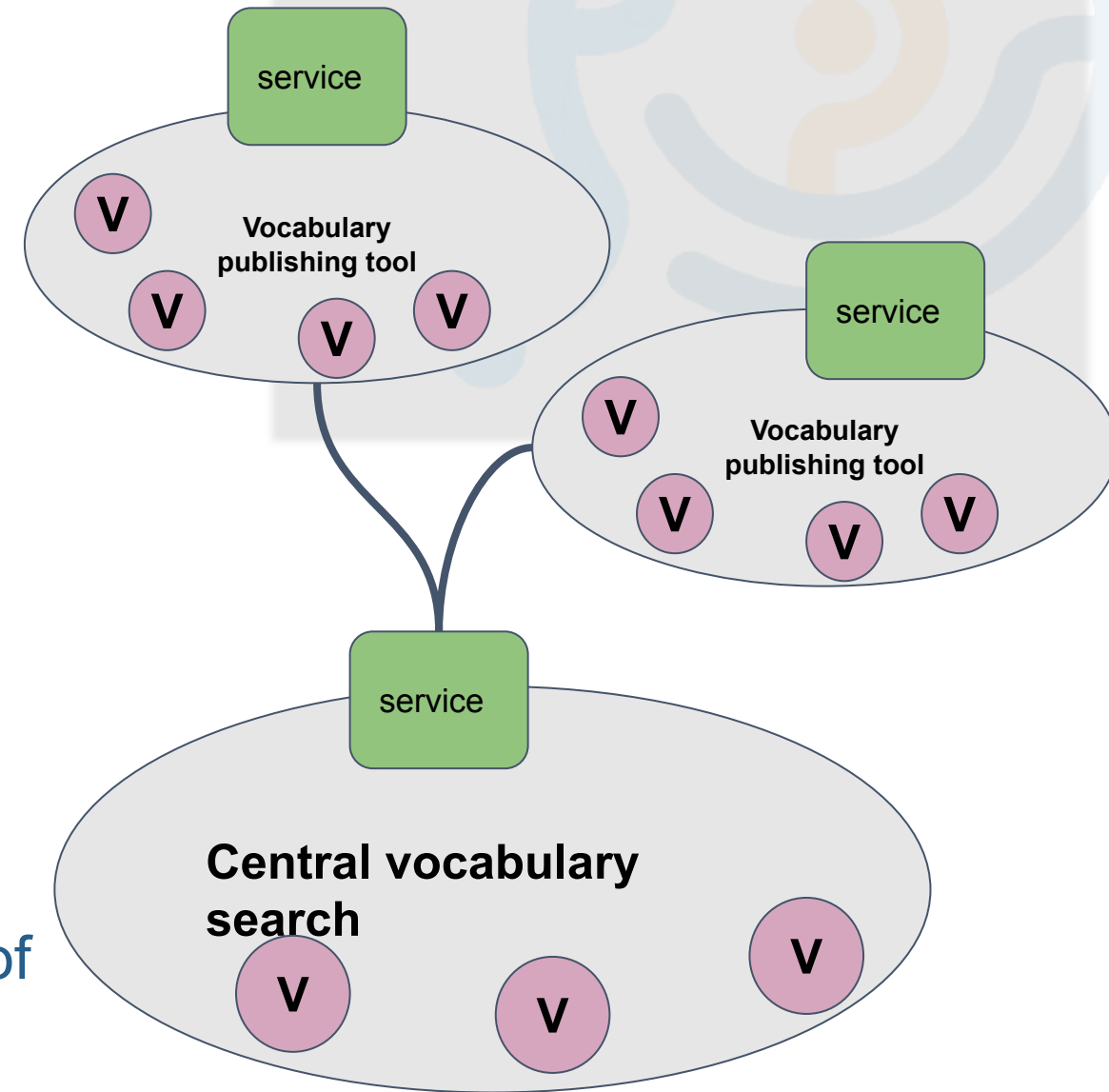
# Vocabulary Search

Effective vocabulary recommendations for researchers require:

- Browsing/searching via metadata AND
- Searching in the vocabulary terms

Considering two strategies:

- Central metadata and content harvesting and indexing  
(update/syncing problem)
- Federated search, possibly on the basis of SKOSMOS instances  
(performance problem)



# Collaborative Vocabulary Management

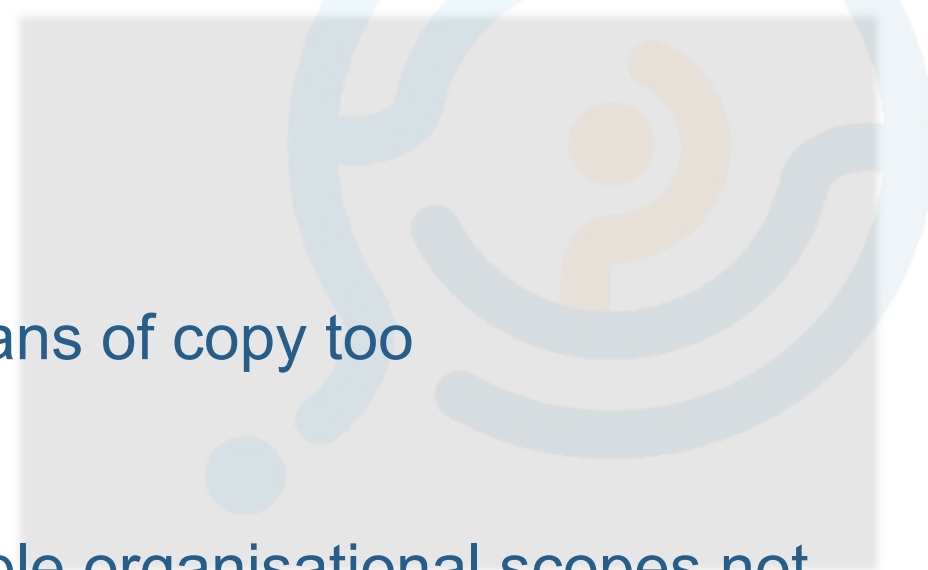
Vocabulary management is “solved” and reuse by means of copy too

BUT:

Reuse and management of a vocabulary across multiple organisational scopes not

- Ownership in a distributed (loosely coupled) setup => authority
- Procedures for evolution and agreement  
(new concepts, semantic shift)
- Synchronisation with source/target applications

Ideally, user wants to adjust vocabulary (add new concepts) at point of use



# Panel Discussion

## Questions:

1. Do you agree with rationale for interoperability as presented?
2. What are the practical limits of interoperability
3. How realistic is reusability of vocabularies across disciplines
4. What about collaborative management of vocabularies? (technical, governance, evolution procedure)
5. xxx

Advancing SSH Research with SSHOCingly  
good and sustainable resources

6-7 April 2022 - Brussels & Online  
**REGISTER NOW**



**Daan Broeder**  
CLARIN



**Willem Elbers**  
CLARIN



**Mari Kleemola**  
Tampere University  
CESSDA



**Matej Ďurčo**  
OEAW / DARIAH



**Claus Zinn**  
CLARIN





# SSHOC'n Tell Challenge

Try our tools in fun challenges!

Join us online on the **mornings of 6 & 7 April**

Your chance to win SSHOC gifts and vouchers  
of up to €1000 for the best user stories

# Thank you for your attention!

Join our community



<https://www.sshopencloud.eu>



[info@sshopencloud.eu](mailto:info@sshopencloud.eu)



[@SSHOpenCloud](https://twitter.com/SSHOpenCloud)



[/in/sshopencloud](https://in.sshopencloud)