

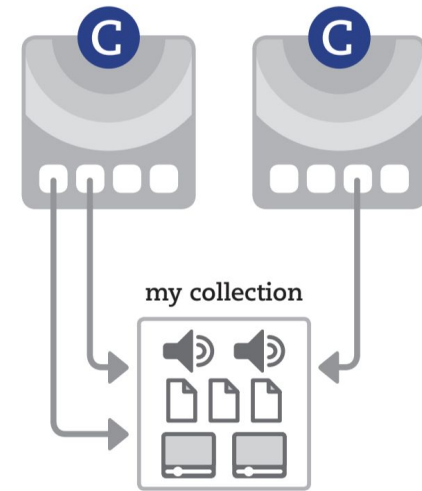
# Virtual Collections

CLARIN Centre Meeting  
9-10 June 2021

Willem Elbers  
[willem@clarin.eu](mailto:willem@clarin.eu)



# Overview



- A **virtual collection** is a coherent **set of links to digital objects** (e.g. annotated text, video) that can
  - Easily created
  - Easily accessed
  - Easily cited
- Links can originate from **different archives**, hence the term virtual
- A virtual collection is suitable for **manual access** as well as **automated processing**

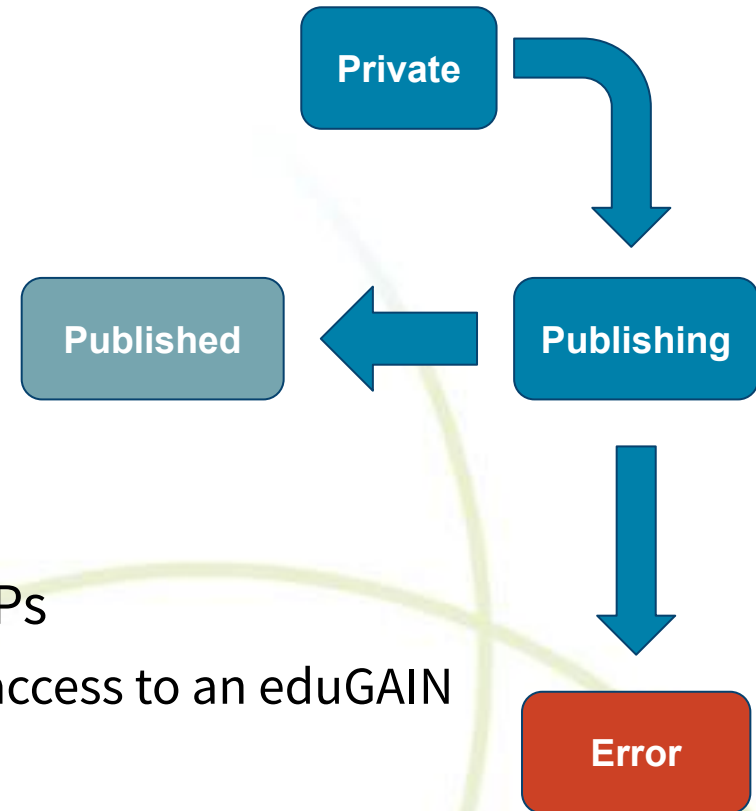
# Metadata

- Minimum amount of generic metadata
  - Not specific to any data domain
  - Working on DataCite DOI compatibility
- Descriptive metadata on the collection level
  - Name, type, author, keywords...
  - Type: extensional or intensional
- Descriptive metadata on the resource level
  - Label, description



# Access

- All published collections are publicly accessible
- Login possible via most eduGAIN IDPs
  - or the CLARIN IDP for users without access to an eduGAIN 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)
  - Editing is discouraged after publication

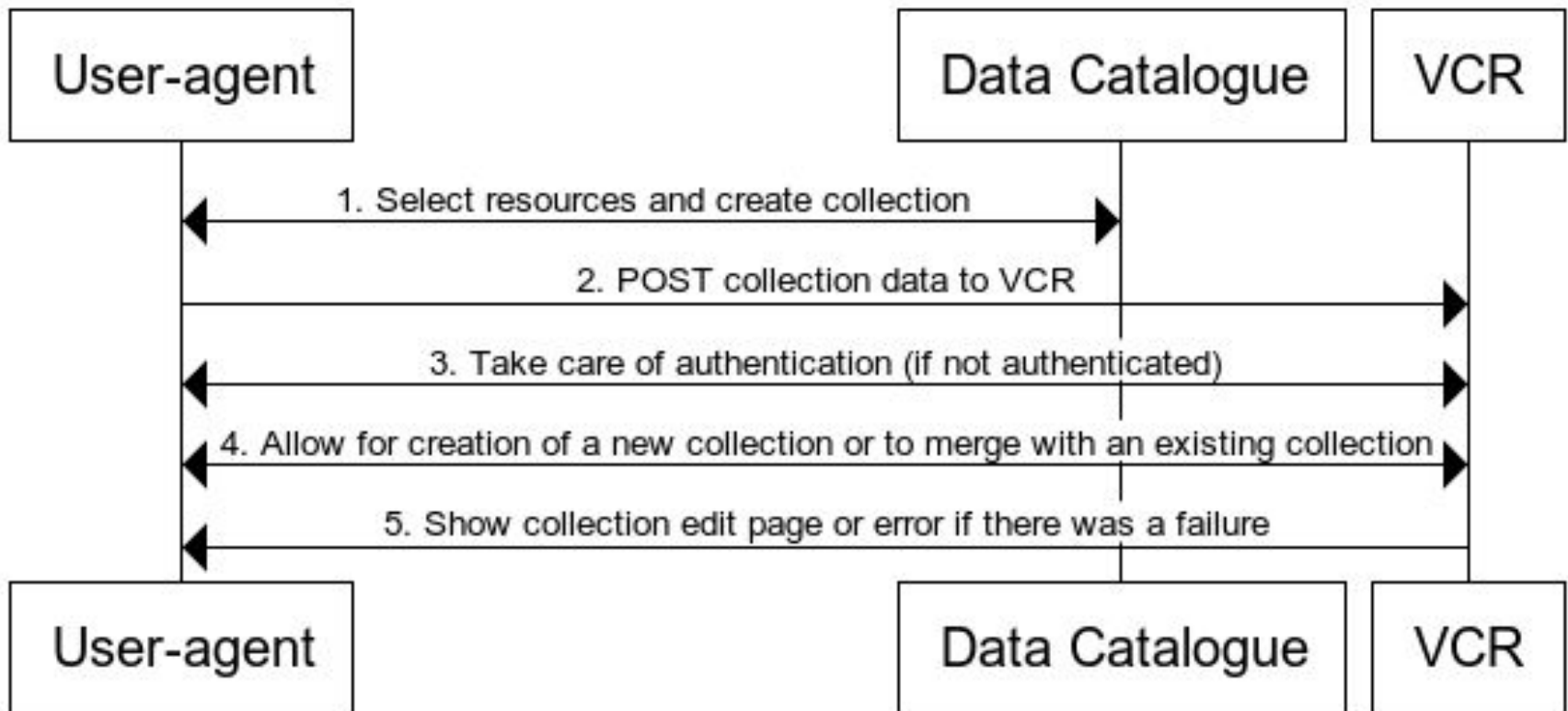


# External Data Catalogue Integration

- Easily create a virtual collection based on search results in an external data catalogue
- Goal: allow easy creation of a single collection over multiple data catalogues
- Integration effort in external data catalogue should be as minimal as possible
  - More complex integration possible via REST API
- Updated in latest release 1.6.x
  - Supports more metadata
  - Store origin and original query
  - Allow to merge into existing collections

# External Data Catalogue Integration

## VCR Integration workflow



# External Data Catalogue Integration

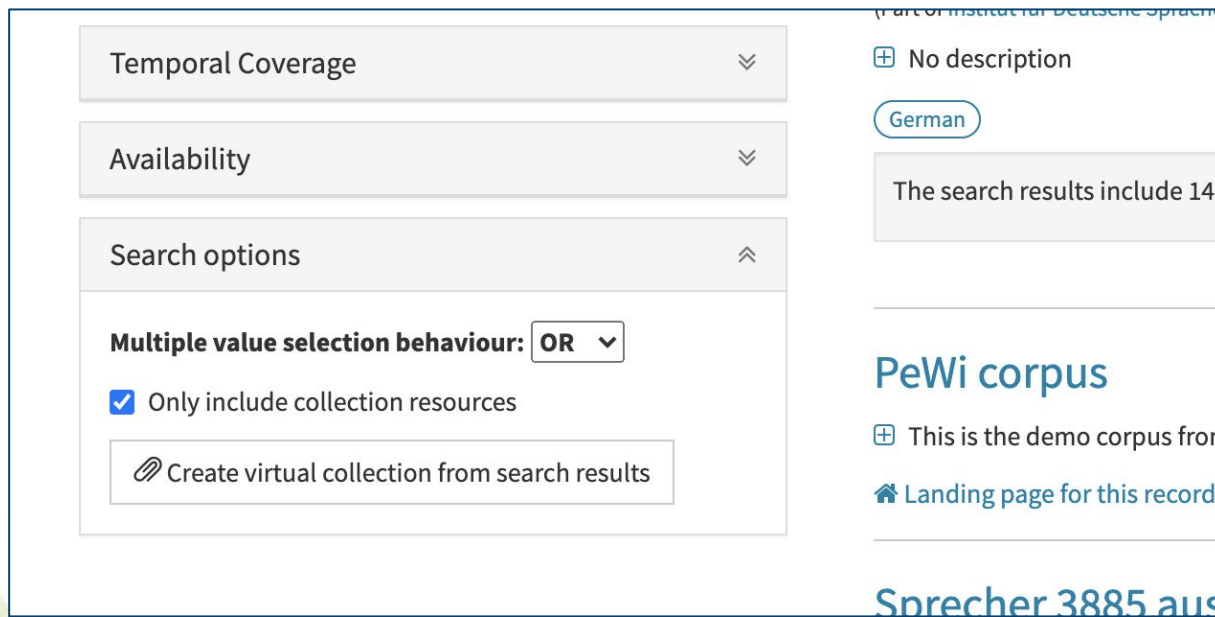
- Easiest to implement a form that submits the relevant data on the data catalogue side:

```
<form id="virtualCollectionForm" method="post" enctype="application/x-www-form-urlencoded"
name="vcrForm" action="https://collections.clarin.eu/submit/extensional">
  <input id="cName" type="text" name="name" value="Your collection name">
  <input id="cDescription" type="text" name="description" value="Your collection description">
  <input type="hidden" name="metadataUri" value="...">
  <input type="hidden" name="metadataUri" value="...">
  ...
  <input type="submit" value="Submit">
</form>
```

- Documentation available
  - via <https://collections.clarin.eu/help>
  - or directly on GitHub:  
<https://github.com/clarin-eric/VirtualCollectionRegistry/blob/master/doc/Integration.md>

# External Data Catalogue Integration

- Currently realized within the VLO
  - Perform a search
  - Open the “Search options” in the left column:



The screenshot displays the search interface of the VLO. On the left, there is a sidebar with three expandable sections: 'Temporal Coverage', 'Availability', and 'Search options'. The 'Search options' section is expanded, showing a 'Multiple value selection behaviour' dropdown set to 'OR', a checked checkbox for 'Only include collection resources', and a button labeled 'Create virtual collection from search results'. On the right, the search results are displayed, including a section for 'PeWi corpus' with a 'German' filter and a button to 'Create virtual collection from search results'.

- Follow the instructions after pressing the “Create virtual collection from search results” button



# SSHOC Use-Cases



- Discussions with various communities to collect input on use-cases
  - EHRIS, CESSDA, DARIAH, ...
  
- Generic use-cases:
  - VC creation for bookmarking, sharing and publication
    - The classical scenario
    - Versioning (vs forking)
      - New version if collection is updated by author
      - Forking if a new collection is created based on a different existing collection
    - Collaborative editing of collections
      - Invite based on username or group / organisation
      - Invite based on unique link

# SSHOC Use-Cases



- Several generic feature requests (continued):
  - Making virtual collections interoperable with other services
    - E.g. switchboard
    - Investigate how the metadata model can be extended
  - Saving query results in virtual collections
    - Enables support for incremental aggregation of result-sets potentially spanning different portals or data-catalogues
    - Add new results (from third party portal or data catalogue) to an existing collection
    - Add support for token based authentication to API to allow third party

# Future plans

- Promote external data catalogue integration
- Next milestone 1.7 release
  - Versioning and forking
  - Collaborative editing
- Milestone 2.0 release (to be decided)
  - Possible new data model
  - Possibly new UI framework
  - Updated API

# 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>