# EU Community Metadata Infrastructure for Finding and Discovery

Daan Broeder – CLARIN

Presentation at Building Blocks of Global Research Commons, RDA P20
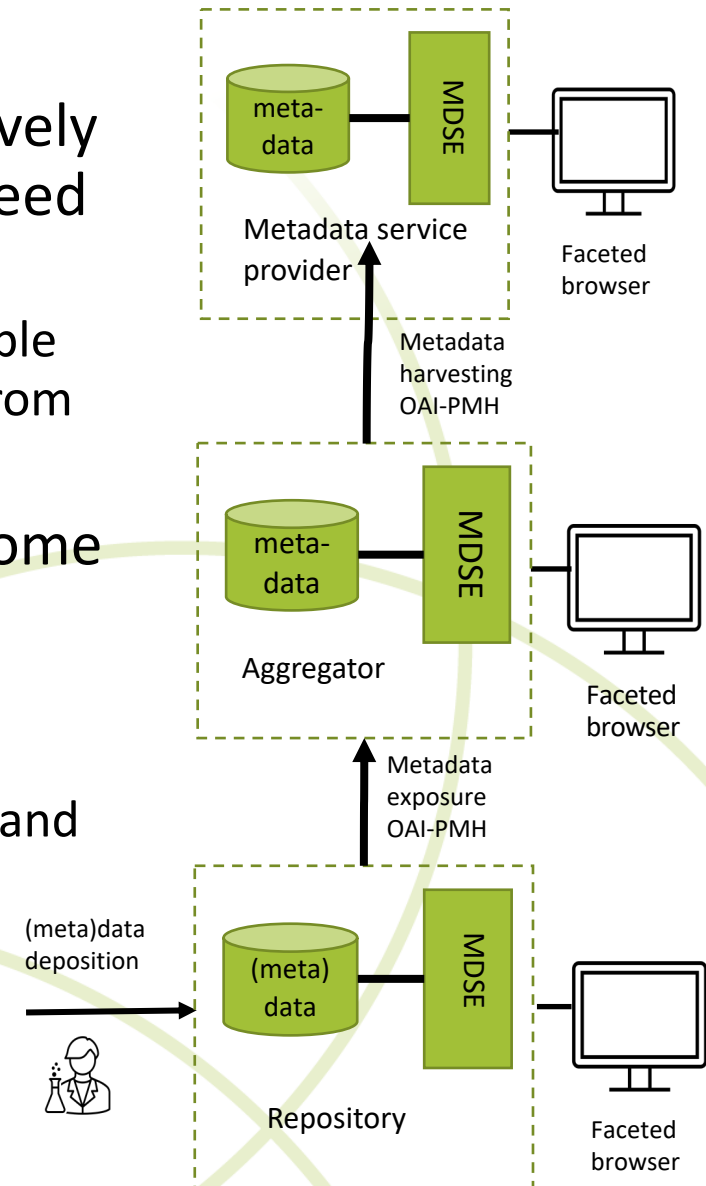
Goteborg March 20, 2023

**CLARIN**

# Finding and discovering digital assets

- Although finding and discovering can be seen as separate user experiences they are closely related, and rich metadata is required for both
    - To find a resource with sufficient precision
    - and determine if newly discovered resources are suitable
- Metadata catalogues used for finding are well positioned to support discovery next to finding
    - Explorative use eg. using facetted browser or keyword search
    - Pointing  to related resources eg. associated publications etc.
    - Providing suggestions for similar resources
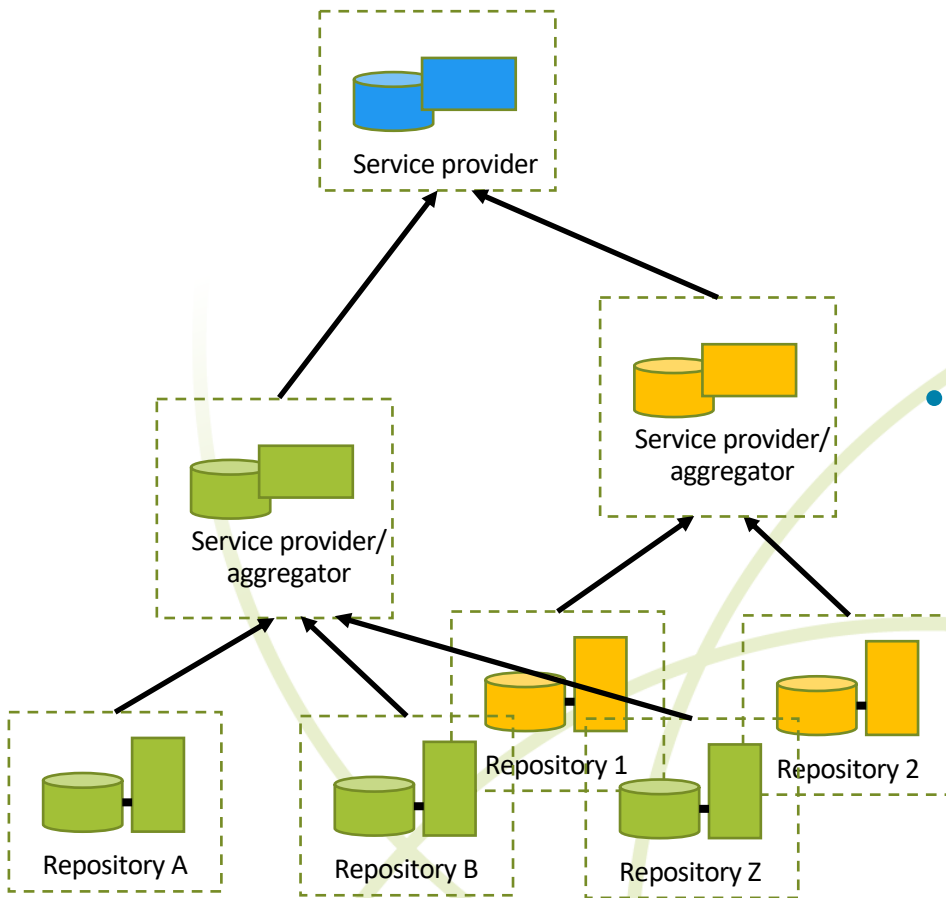
# Metadata infrastructure

- Offering a consolidated catalogue is relatively easy for a community sharing a single agreed metadata schema
  - For all search results is max metadata available and discoverability is high for all resources from the community

- If multiple metadata schema play a role some metadata mapping is needed.
  - Mapping to a common core metadata set, accepting the information loss
  - or to a rich superset, accepting empty slots and complex search GUIs
  - Note that information loss is detrimental for the discoverability aspect

meta-data | MDSE

Metadata service provider

Faceted browser

Metadata harvesting OAI-PMH

meta-data | MDSE

Aggregator

Faceted browser

Metadata exposure OAI-PMH

(meta)data deposition

(meta) data | MDSE

Repository

Faceted browser

# Harvesting the harvestable Harvesters

When does it make sense to harvest other organisations' metadata?
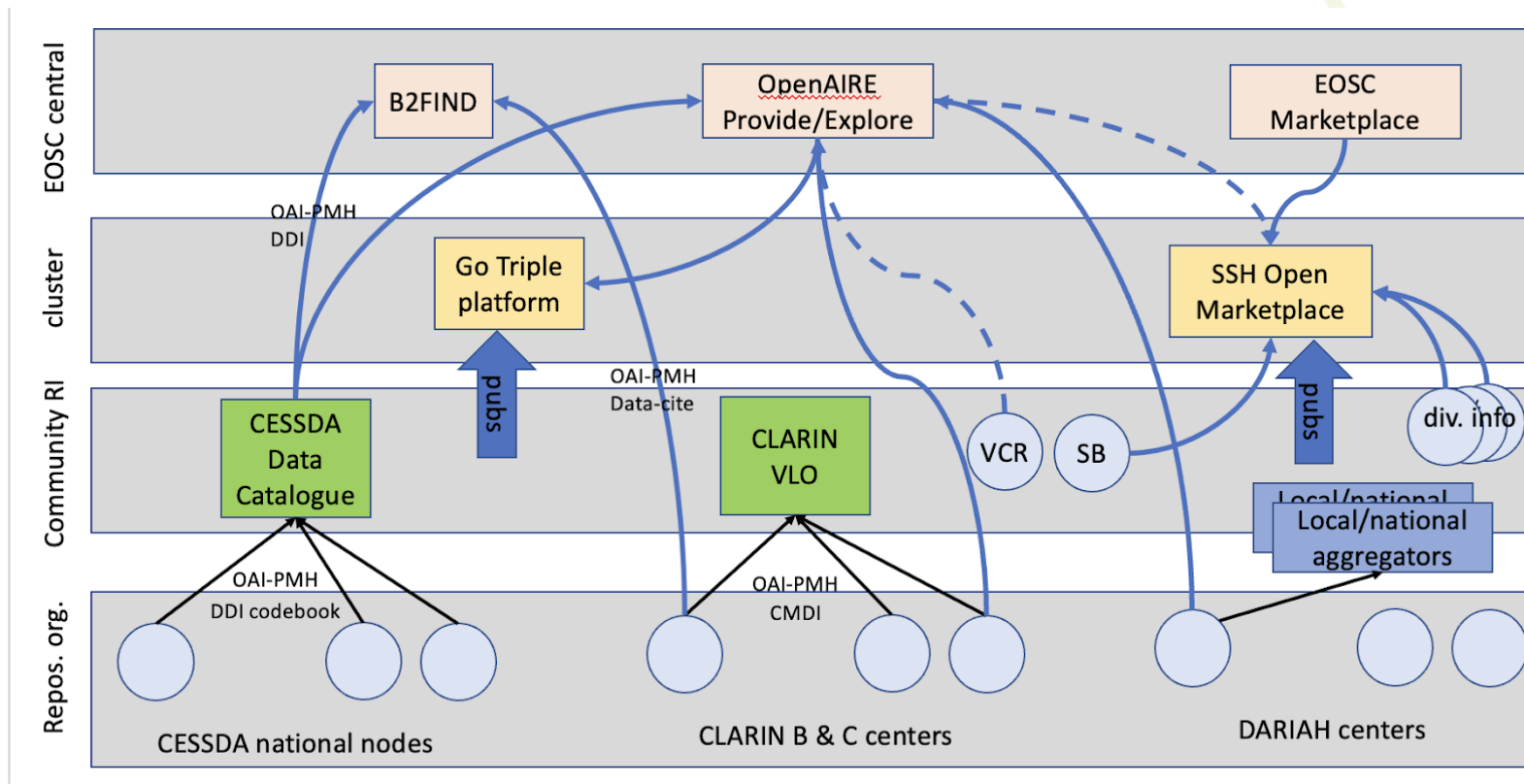
- What is the added value?
  - Community catalogue
  - Topical portal
  - Cross-disciplinary research
  - Additional discovery potential

- There is an effort involved and quality can be a problem
  - Metadata providers like to see their metadata in good company
  - Contact with responsible metadata providers is needed
  - Effort to keep content up-to-date

Service provider

Service provider/
aggregator

Service provider/
aggregator

Repository 1

Repository 2

Repository A

Repository B

Repository Z

# Who operates metadata catalogues and aggregators

- Discipline specific  RI organizations: eg. CESSDA, DARIAH, …

- Thematic Clusters,  clustering the RIs
  - however not automatically clustering also all RI activities such as a common metadata infrastructure

- General ICT service providers
  - OpenAIRE operating OpenAIRE Explore, integrated in the EOSC Portal
  - EUDAT operating B2FIND
  - Discovery options are high wrt. diversity, but more difficult for user to evaluate due to the information loss. Recommendations based on user-profiling can play a role (EOSC user panel)

- International topical catalogues eg. GEOSS

- Many national ICT providers and national thematic projects also operate catalogues and aggregators

# SSHOC - Social Sciences and Humanities Open Cluster

# Observations & recommendations

- The intended audience of a catalogue should match the metadata
  - Researchers need rich metadata for findability and discoverability
  - Other users/stakeholders (funders, meta-science, …) may need less
- Consider if a new consolidated catalogue can offer added value and for whom
  - Providing additional contextual information (discoverability)
  - Cross-disciplinary research
- For researchers familiar with the data it is often preferable to visit the data hosting repository or a community catalogue this limits discoverability in general
- This also suggests investigating options to exchange information between community catalogues, for instance to be able to recommend related resources