Discovery: FAIR's next frontier

Peter Kraker (Open Knowledge Maps), Brigitte Mathiak (GESIS),
Alessia Bardi (OpenAIRE)

FAIR Festival 2021
21 June 2021

Agenda

1. Introduction to the topic and the Discovery IN
2. Results of the Discovery IN I: Use cases
3. Results of the Discovery IN II: Infrastructure
4. Summary and outlook

Motivation

Up to 85% of datasets are not reused (Peters et al. 2016)

→ Discoverability is a key challenge when it comes to research data

Lack of adequate user interfaces for data discovery

- Simple reuse of existing interface concepts for publications
- Design from the system‘s rather than the user‘s perspective

New market entrants following a closed/proprietary model

- Not suitable for the Internet of FAIR Data and Services
- Creates new (pay)walls and prevents innovation

GO FAIR

DISCOVERY

OPEN KNOWLEDGE MAPS

AfricArXiv · BASE · Institut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität zu Berlin · Bioschemas · cessda · CORE

DataCite · DARIAH-EU · elixir FRANCE · EUDAT · gesis Leibniz Institute for the Social Sciences · HZI Helmholtz Zentrum für Infektionsforschung

Alexander von Humboldt Institute for Internet and Society · hypothes.is · Impactstory · IGB · KNOW Center · net7 · NIOO

OpenAIRE · OPERAS open access in the european research area through scholarly communication · ORCID · RE figure · SCHOLIA · TIB Leibniz Information Centre for Science and Technology University Library · ZBW Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics · ZB MED Information Centre for Life Sciences

GO FAIR Implementation Network Discovery

Topic: Open User Interfaces for Increased Visibility of Research Results

Membership:

Personal members:

Julien Colomb - Humboldt-Universität zu Berlin
Francesca Di Donato - CNR
Tina Heger - University of Potsdam and Technical University of Munich
Aaron Tay - Singapore Management University

Organisational members:

Open Knowledge Maps (chair)
GESIS - Leibniz Institute for the Social Sciences (co-chair)
OpenAIRE (co-chair)
AfricArxiv
BASE
Berlin School of Library and Information Science, HU Berlin
Bioschemas
CESSDA ERIC
CORE
DataCite
DARIAH-EU
Elixir-fr
EUDAT
Helmholtz Centre for Infection Research (HZI)
HIIG
Hypothes.is
IGB - Leibniz-Institute of Freshwater Ecology and Inland Fisheries
Impactstory
Know-Center
Net7
NIOO-KNAW
OPERAS
ORCID
ReFigure
Scholia
TIB - Leibniz Information Centre for Science and Technology
ZB MED - Information Centre for Life Sciences
ZBW - Leibniz Information Centre for Economics

Findability vs. discoverability

Findability = attribute of the (meta-)data

Discoverability = attribute of the infrastructure

FAIR is a precondition to discoverability

Purpose

Provide user interfaces and other user-facing services for data discovery across disciplines

Explore new and innovative ways of enabling discovery (e.g. visualizations, recommender systems, semantics, content mining, annotation, responsible metrics)

Apply user involvement and participatory design, going beyond academia

Objectives

Improve visibility and discoverability of research data across disciplines

Increase reuse of FAIR data and therefore efficiency and effectiveness of research

Provide open alternatives to closed and proprietary infrastructures for data discovery

Workplan

Stocktaking of relevant use cases as well as indices, interfaces and services

Structuring: Defining the standards and structure of an open ecosystem for discovery that fulfils the use cases

Implementation: Working towards implementation of the ecosystem

Discussion

Use Cases

Background

Stocktaking started with the establishment of the group in 2019

In the first round, participants of the implementation network were asked to provide their use cases.

More from:
"Data Discovery Across Disciplines" workshop at the Open Science Fair 2019

supplemented with list from RDA Data Discovery IG had compiled

Use Case Solicitation & Processing

- Collect use cases in the form:

  "As a **x**, I want to **y**, where the benefit is **z**"

  - 100+
  - Clean up : common themes, organise, remove duplicates

- Classify 'role', 'object' and 'value/goal'
  - researcher, student, funder, …
  - paper, person, dataset, ..
  - value / benefit

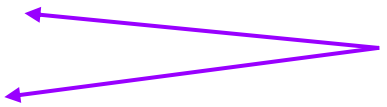
Use Case Solicitation & Processing

- Collect use cases in the form:

  "As a **x**, I want to **y**, where the benefit is **z**"

  - 100+
  - Clean up : common themes, organise, remove duplicates

- Classify 'role', 'object' and 'value/goal'
  - researcher, student, funder, …
  - paper, person, dataset, ..
  - value / benefit



functionality

Analysis I

data citations (14): datasets <-> papers (alternative media)

overview (14): discipline agnostic ‘quick query’

discovery metadata (12): find/rank by (general) size, format, license, (specific) expt’al design, instrument

documentation (9): reuse/interpretability; overlaps data citation to some extent

Analysis II

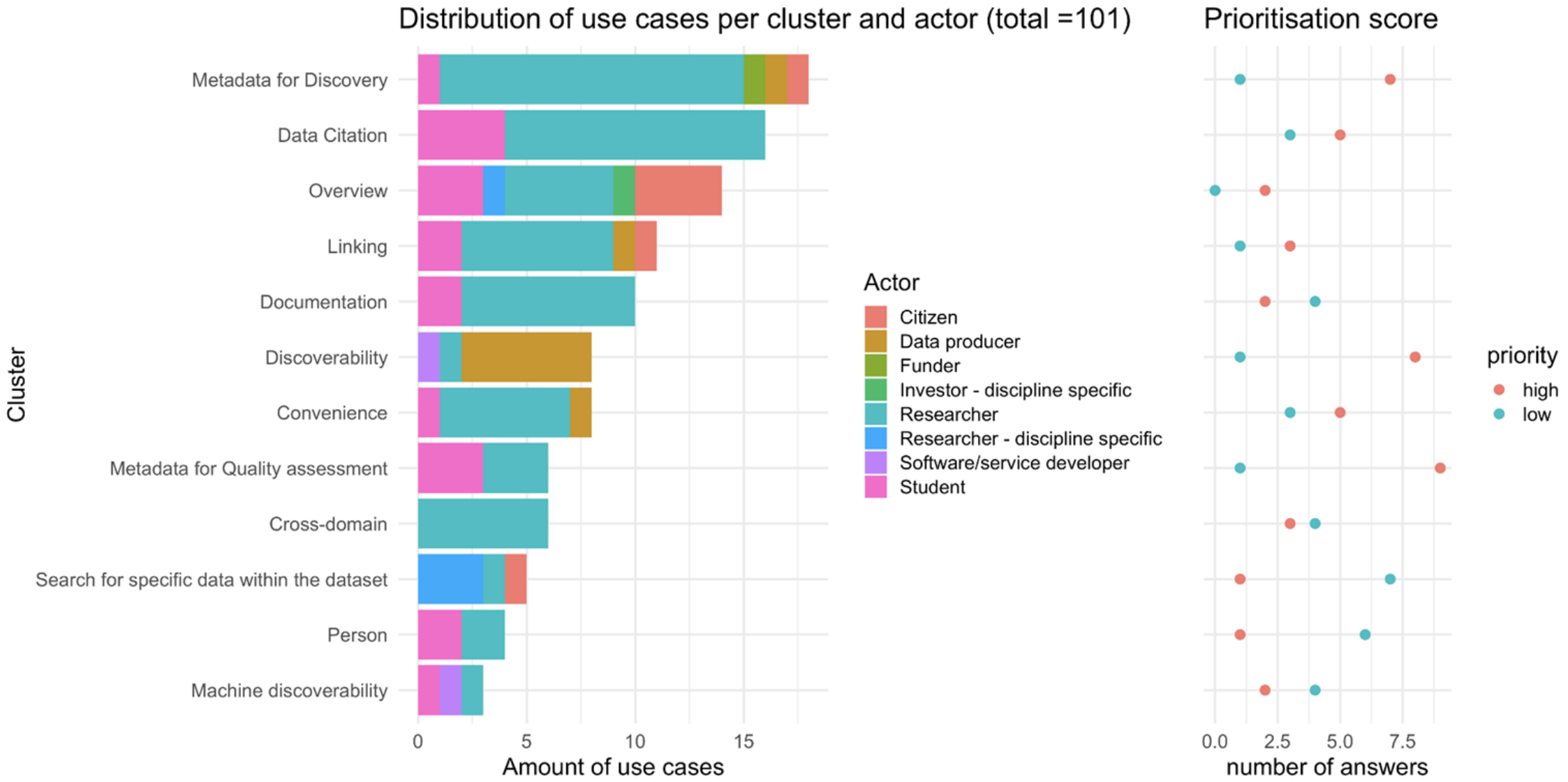
convenience (8): user centric; push notifications (RSS-feed type), autocomplete, thumbnail/previews

discoverability (8): promote own datasets; supplement metadata; feedback mechanism

‘quality’, ‘cross-domain’ & ‘linking’ (6 ea): trust/peer-review; discovery across domains; link datasets & related data

search within dataset (5): specifically within very large datasets

Prioritisation



Discussion

Open e-infrastructures for data discovery

Open e-infrastructures for data discovery

The e-infrastructures for data discovery with clear and established **open policies**, **open APIs**, and **open licenses** for data, metadata and source code, allowing for **community governance**.

Open e-infrastructures r**emove paywalls**, **avoid lock-in** effects and enable **community participation and outreach**.

GO FAIR

Status Quo : The Cosmos of Discovery services

Publishers

Researchers, organizations

Research infrastructures

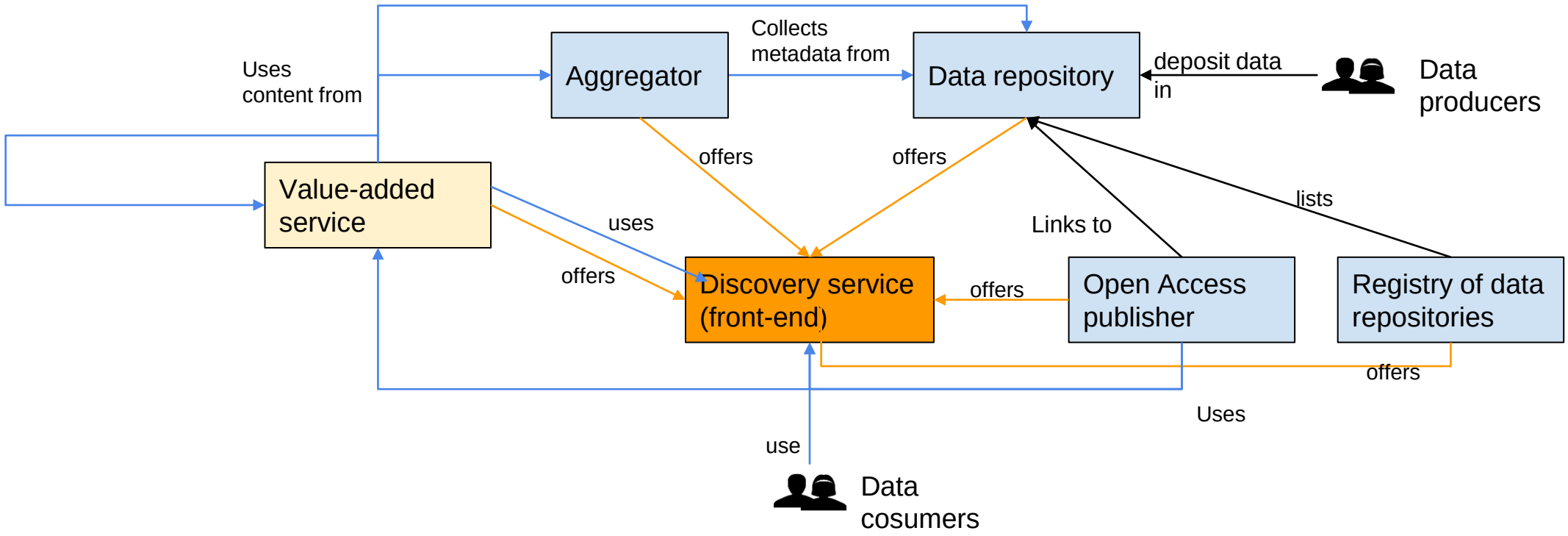
Thematic services and databases, data centres, repositories and archives

Aggregators

Registries

Value added services

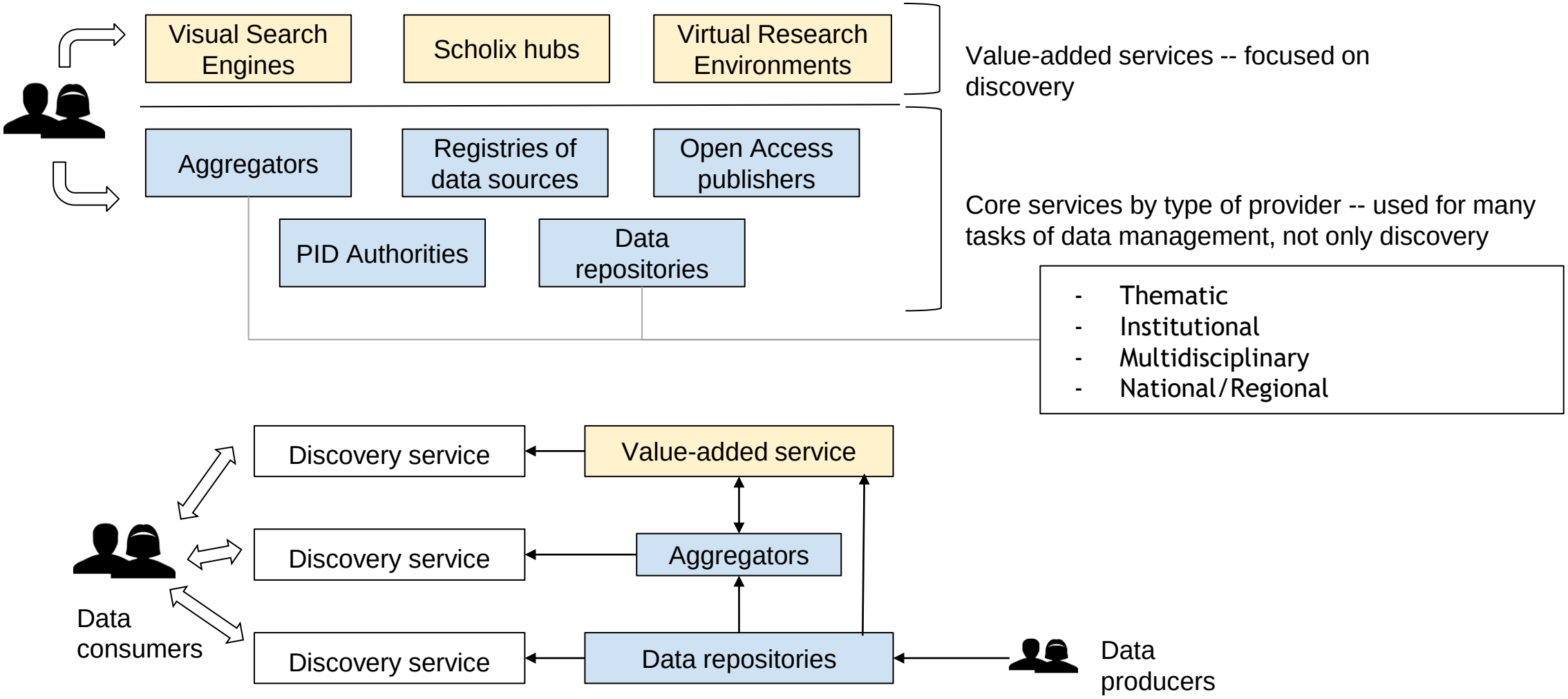
Categorize and analyse the ecosystem



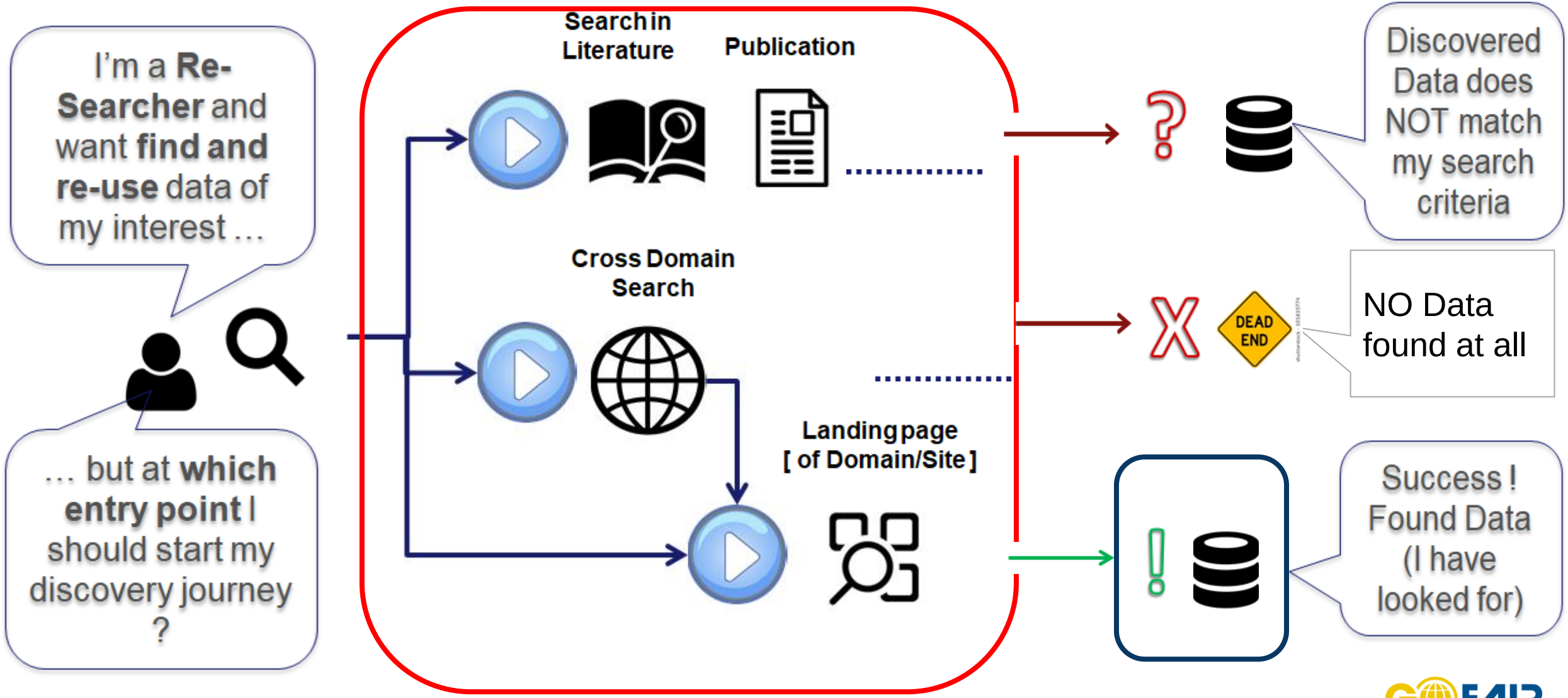
No one size fits-all

Evolution of requirements

Categorize and analyse the ecosystem



The long journey of discovery



Ecosystem of Data Discovery

Gaps and barriers to data discovery

- The ecosystem offers several pathways to data discovery
- Which path for which use-case? How can the ecosystem support the selection of the most proper path?
- Where can the implementers put effort to effectively increase the discovery of datasets and better serve specific use-cases?

Metadata Quality

Contextual information

Proper adoption of repository registries

Community standards

Community building

Link big data centres to the open ecosystem for data discovery

Do not leave the small institutions and the “long tail of science” behind

Sustainable business models for open services

Collaboration vs competitions

From the workshop at the International FAIR Convergence Symposium 03 December 2020

Discussion

Wrap-up

Next steps

Stocktaking of relevant use cases as well as indices, interfaces and services

Structuring: Defining the standards and structure of an open ecosystem for discovery that fulfils the use cases

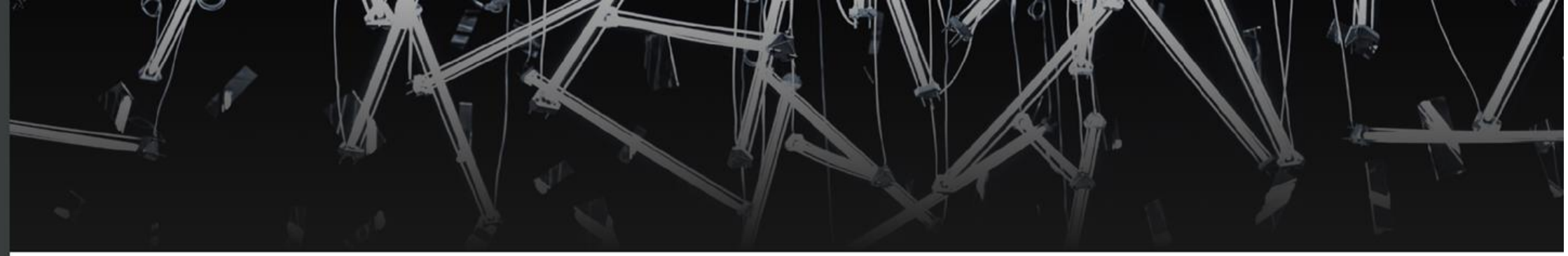
Implementation: Working towards implementation of the ecosystem

Wrap-up

Next steps

- Publication of the use case collection
- Soliciting further input for prioritizing use cases
- Putting together a first version of the infrastructure structuring
- Move on to analysis of coverage of use cases by the existing infrastructure
- Start to implement missing pieces/links between infrastructures to cover most important use cases

To join the Discovery IN, please apply at: https://www.go-fair.org/implementation-networks/overview/discovery/



Kaitlin Thaney

June 19th, 2020

The open scholarship ecosystem faces collapse; it's also our best hope for a more resilient future

Estimated reading time: 7 minutes

The COVID-19 pandemic is significantly impacting universities and higher education institutions, reducing budgets and presenting new design challenges that will fundamentally alter how research and scholarship operate. Economic volatility is also constraining support for key systems and services that the academy relies on, especially those that are community-led. **Kaitlin Thaney** *argues that there's a need to converge on community-controlled open scholarship projects, to both meet the demands of the moment, and build a more resilient system for scholarly communication for future crisis situations, and invites readers to participate in planning how such systems can be maintained.*

GO FAIR

… starts where you stop!

THANK YOU FOR YOUR ATTENTION!