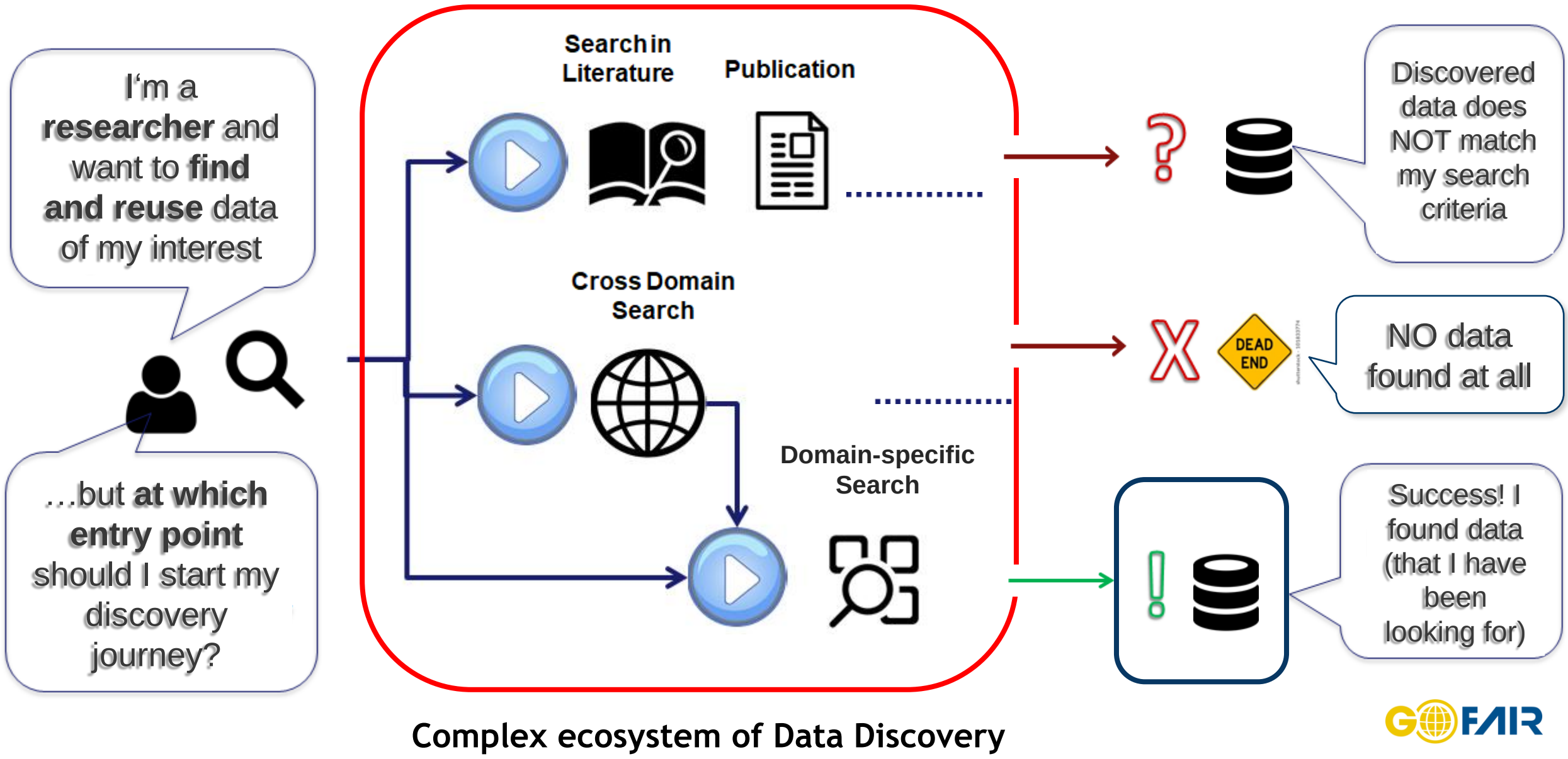
GO FAIR in Practice: The Discovery Implementation Network

Peter Kraker (Open Knowledge Maps)

Launch Event FAIR Office Austria
9 June 2021

Motivation



Complex ecosystem of Data Discovery

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

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

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

Use Cases

Large-scale Use Case Collection

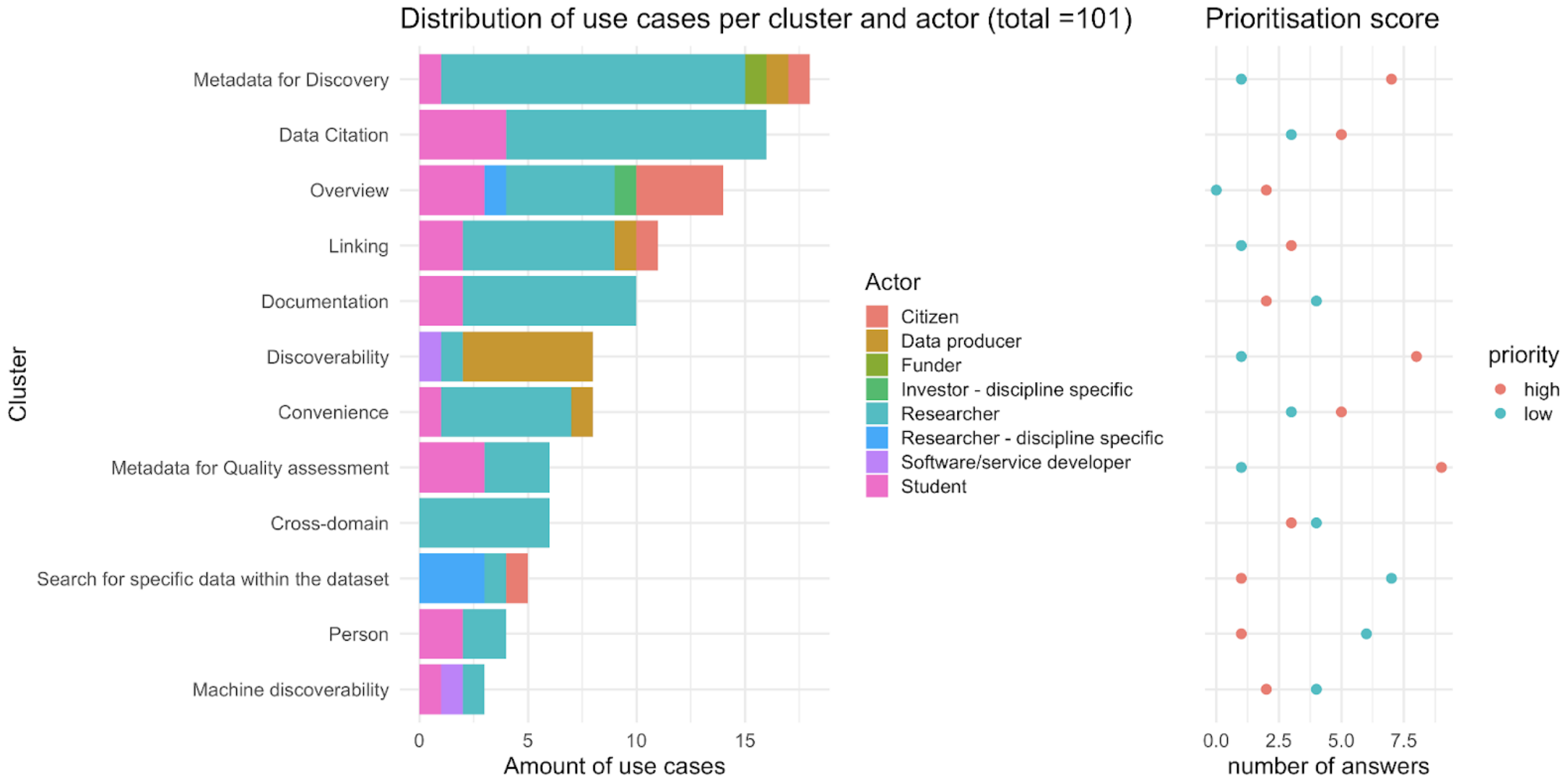
“As a [role] I want to [goal] so that [benefit]”
“As a researcher, I want to get an overview of datasets in my field so that I can determine which data I can reuse”

Over 100 such use cases were collected since 2019

Sources: own input, workshops, studies, interviews, existing use case collection from the RDA IG Data Discovery

Elaborate analysis: annotation with entities, thematic clustering, and initial community prioritization

Results


Distribution of use cases per cluster and actor (total =101)
Cluster
Metadata for Discovery
Data Citation
Overview
Linking
Documentation
Discoverability
Convenience
Metadata for Quality assessment
Cross-domain
Search for specific data within the dataset
Person
Machine discoverability
0
5
10
15
Amount of use cases
Actor
Citizen
Data producer
Funder
Investor - discipline specific
Researcher
Researcher - discipline specific
Software/service developer
Student
Prioritisation score
0.0
2.5
5.0
7.5
number of answers
priority
high
low

Summary/conclusions

Most use cases described scenarios that current infrastructure do not provide or provide badly

Example: “As a researcher, I want to find datasets that are similar to those that I used before, so that I can expand and compare my studies”

Reveals Crucial Requirements for the Data Discovery Infrastructure

Open dataset with accompanying paper will soon be published

Infrastructure

The cosmos of discovery services

Researchers, organizations

Research infrastructures

Thematic services and databases, data centres, repositories and archives

Publishers

Aggregators

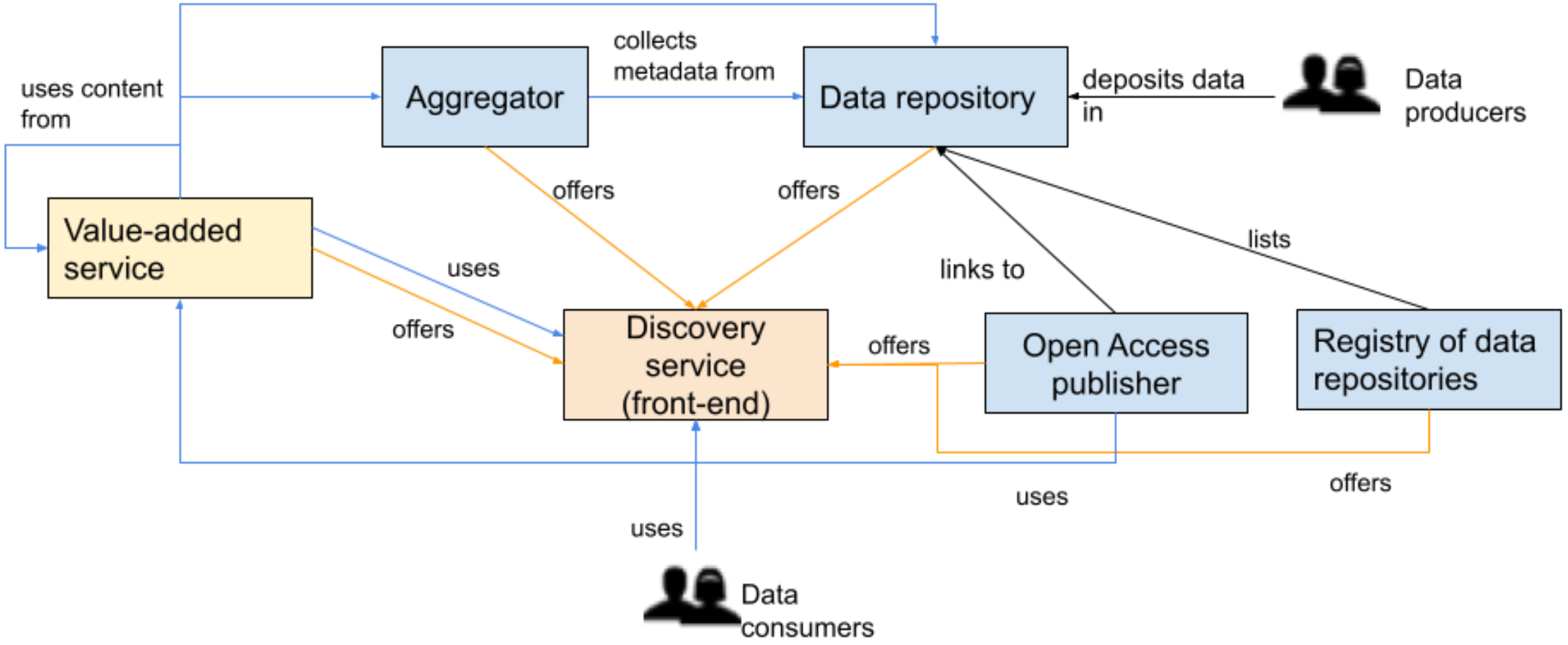
Registries

Value added services

Actions of the IN

01	**Analyse**	the open e-infrastructures for data discovery
02	**Categorize**	actors and their services, highlight their relationships
03	**Identify gaps**	in the data discovery workflow
04	**Suggest**	guidelines for users and implementers

Interactions among the actors in the system



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

How YOU can get involved

How YOU can get involved

Join an existing Implementation Network: https://www.go-fair.org/implementation-networks/overview/

Start your own Implementation Network

- Gather your collaborators and register your interest with GO FAIR
- Prepare a manifesto (brief, high-level description in 2 pages)
- Submit your manifesto to GO FAIR for approval

For more information, join us for the FAIR Festival, 21-23 June!

GO FAIR

… starts where you stop!

THANK YOU FOR YOUR ATTENTION!

Peter Kraker
pkraker@openknowledgemaps.org

OPEN KNOWLEDGE MAPS