



DATA DISCOVERY ACROSS DISCIPLINES

Peter Kraker, Alessia Bardi, Johana Chylíková, Arnaud Gingold, Petr Knoth, Brigitte Mathiak, Christian Pietsch, Nancy Pontika, Jochen Schirrwagen, Nataliia Sokolovska & Heinrich Widmann

Open Science Fair 2019
Porto, 16 September 2019



Notice

Photographs and/or video will be taken at today's workshop

If you wish not to be recorded, please let the organizers or the film crew know: they will make sure not to include you.

Otherwise, you grant full rights to publish the images resulting from the photography/filming under a CC BY license for use by the FITT4RRI/OpenAIRE projects and the workshop organizers for documentation, dissemination, publicity, fundraising or other purposes to help achieve the objectives of these projects and the organizers.

Motivation

Discoverability is a key challenge when it comes to research data:
up to 85% of datasets are not reused (Peters et al. 2016)

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 FAIR data
- Creates new (pay)walls and prevents innovation

GO FAIR

DISCOVERY



OPEN
KNOWLEDGE MAPS



GO FAIR in a nutshell

Bottom-up, stakeholder-driven and self-governed initiative to implement the FAIR data principles

Open and inclusive ecosystem for individuals, institutions and organisations working together through Implementation Networks



GO FAIR Implementation Network Discovery

Topic: Open User Interfaces for Increased Visibility of Research Results

Membership:

Personal members:

Julien Colomb - Humboldt-Universität zu Berlin

Tina Heger - University of Potsdam and Technical University of Munich

Aaron Tay - Singapore Management University

Organisational members:

Open Knowledge Maps (lead)

AfricArxiv

BASE

Berlin School of Library and Information Science, HU Berlin

CESSDA ERIC

CORE

DataCite

DARIAH-EU

GESIS - Leibniz Institute for the Social Sciences

EUDAT

HIIG

Hypothes.is

IGB - Leibniz-Institute of Freshwater Ecology and Inland Fisheries

Impactstory

Know-Center

Net7

NIOO-KNAW

OpenAIRE

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 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

Findability vs. discoverability

Findability = attribute of the (meta-)data

Discoverability = attribute of the infrastructure

FAIR is a precondition to discoverability

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

Agenda

14:00 Welcome and introduction

14:05 Results of the Discovery IN I: Use cases
Including a quick-fire presentation by CESSDA

14:15 Discussion and audience input

14:30 Results of the Discovery IN II: Tools, services and interfaces
Including quick-fire presentations of BASE, CORE, DREAM, EUDAT & EOSC-Hub, OPERAS and Open Knowledge Maps

15:30 Discussion and audience input

15:50 Wrap-up and next steps



Results I: Use Cases



Results I: Use Cases

Roles

Researcher (data consumer and data producer)

Student (data consumer)

Categories

Overview

Recommendation

Visibility

Reputation

Results representation

Quality

Reuse

Use cases: overview

As a researcher I want to get an overview of datasets for my field of interest so that I can determine which datasets already exist that I can reuse

As a researcher I want to get an overview of datasets for my field of interest so that I can determine which TYPES OF datasets already exist that I can reuse so that I can decide whether to dive in deeper

As a researcher I want to know the people who are generating datasets in my field to have another way of searching and also create community

As a researcher, I want to find data (people and papers) outside of my field that can help me addressing research questions in my field, so that I can use these data from other disciplines to better tackle the research questions.

As a researcher I want to find datasets that are related to or originated from certain research domains I'm interested in to filter out records from irrelevant domains

As an undergraduate level student, I want to be able to find data around discrete research questions to do analysis and findings for short-term research projects

Use cases: results representation

As a researcher I want to filter search results by various fields, e.g. licences, amount of data points, format.

As a researcher or student, I want to receive information on available datasets in a structured and comprehensible way, so that it does not take me much time to get an overview

As a researcher I want to know the provenance and use licences of datasets in order to re-use (process, analyse, visualize etc.) foreign data resources by giving credit to, cite and share your results with the original data producer

As a researcher, I want to spend as little time as possible creating extra metadata, especially if the information was entered somewhere else (for example author information already available in the publication linked to the data)

Use cases: recommendation & quality

As a researcher I want to find datasets that are similar to those that I used before.

As a researcher, I want to have an overview of datasets, that can be linked to the dataset that I am currently using, so I can analyze them together.

As a researcher, I want to know what datasets have been cited along with the datasets that I am currently using, so I can be sure I am missing something relevant.

As a researcher, I want to have information on the quality and relevance of the dataset, I am currently looking at, so I can get idea on, if it has the quality I am looking for.

As a researcher, I want to see in which publications the datasets I'm considering to work with have been used.

As a student I want to know whether dataset have undergone peer review or whether they belong to papers that have been peer-reviewed.

Use cases: visibility, reuse and reputation

As a researcher, I want to know where I can put my data, insights and annotations for maximum discoverability

As a researcher I want to know whether my shared datasets have been reused.

As a researcher, I want to have metrics on the reuse of my datasets automatically sent to my CV (via Orcid?) so that I can show the importance of my work for the community (and get funding)

Group discussion 1

Get together in groups of 4-5

Write down **your own use cases** when it comes to data discovery. Please use the following format: “As a **[role]** I want to **[goal]** so that **[benefit]**”

You have **10 minutes** to discuss the use cases, and then each group will have **5 minutes** to present their top use case.

Facilitators will help you with any questions



Results II: Tools, services and interfaces



Results II: Tools, services and interfaces

Platforms and tools

BASE BibSonomy DOAB
EUDAT-B2FIND
GESIS search Hi Knowledge
OpenAIRE Explore Open Knowledge Maps Search
OpenEdition OpenAIRE Open Research Gateways for research communities ORCID
ploc
ReFigure ScholeXplorer TRIPLE Sourcedata
VIPER - The Visual Project Explorer Zenodo

APIs and services

B2FIND API based on CKAN API
DataCite REST API
Entity-fishing API
OpenAIRE Search API
ScholeXplorer API
Zenodo

Software and frameworks

Head Start
Scholix Framework

Group discussion 2

Get together in groups of 4-5

Write down **tools, services, and user interfaces** for data discovery. Please note only pieces of **open infrastructure**.

You have **10 minutes** to discuss, and then each group will have **5 minutes** to present their top piece of infrastructure.

Facilitators will help you with any questions

Wrap-up

Next steps

- Consolidation of workshop input
- Stocktaking of standards and funding opportunities
- Write-up of the outcomes of the stocktaking
- Move on to the structuring phase

To become a member of the GO FAIR Discovery IN, please apply at <https://www.go-fair.org/implementation-networks/overview/discovery/>



THANK YOU FOR YOUR ATTENTION!

Open Science Fair 2019
Porto, 16 September 2019

