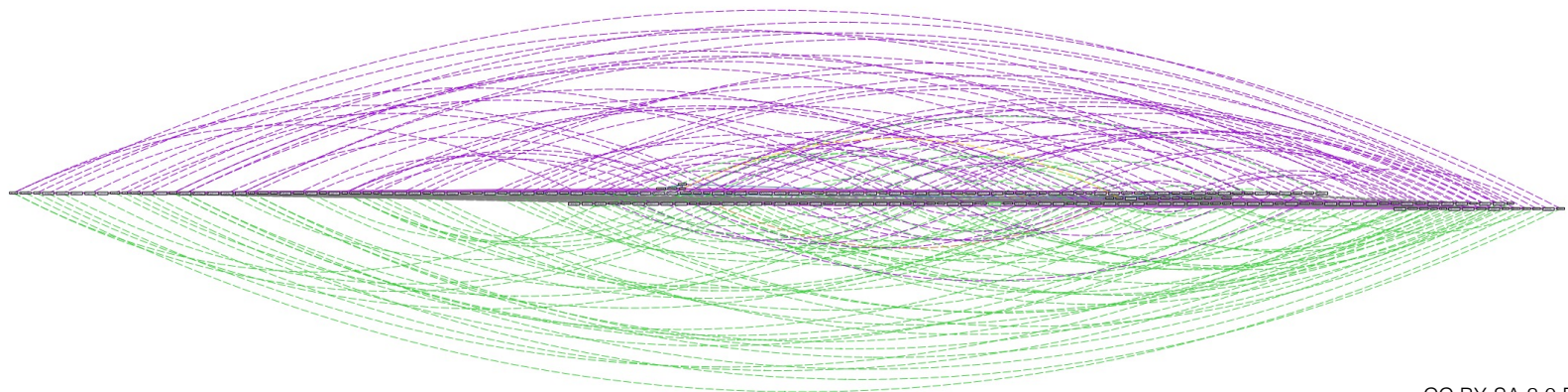# A Lightning Introduction to the NFDI4Ing Data Collections Explorer

TS4NFDI Workshop – 15.02.2024

Philipp Ost   philipp.ost@kit.edu

Why Build the Data Collections Explorer?

The NFDI4Ing proposal mandates to build „a collection of existing community specific repositories".[1]

- Developed as part of NFDI4Ing Base Services Measure S-4 "Repositories and Storage"

- There are a lot of repositories available – the challenge is finding the right one
  - re3data provides a lot of information
  - However, there is more…

[1]https://nfdi4ing.de/base-services/s-4

What is the Data Collections Explorer?

The Data Collections Explorer is an information system for repositories and data collections focussed on the engineering sciences.
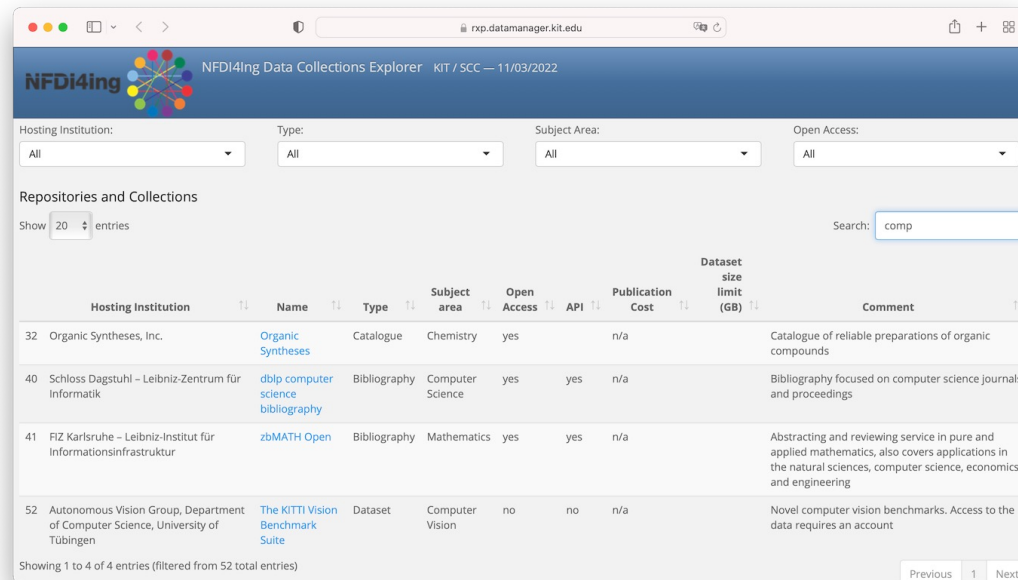
Typical use cases:
- Scientists searching for data sets
  - Example: An engineer is interested in material properties
    - Are there other materials that might fit the requirements?

- Scientists looking for repositories to publish their research data
  - Example: PhD students create data sets as part of their research
    - Where to publish them?
    - Are there size limits or costs involved?

Repositories and Data Collections

The Data Collections Explorer
- Currently: 91 repositories and data sets listed
- Access it here: https://data-collections.nfdi4ing.de

Data Collections Explorer
data-collections.nfdi4ing.de

Current Architecture

- Based on R Markdown, served as a Shiny application
- CSV table as data source → rendered as a HTML table
- Drop-down menus and full-text search are provided by R

- Very small: only ~56k in size (code including table, graphics)
- Source code is available on GitHub
  - https://github.com/kit-data-manager/Data-Collections-Explorer

- There is room for improvement:
  - CSV is not the most flexible format
  - No easy API access possible
  - And more…

A Look Forward – Current And Future Work

- CSV table replaced by a nowledge graph
- Addresses almost all shortcomings
  - One-to-many mappings come naturally
  - Access via SPARQL
  - Easier integration with other projects
  - Easy machine accessibility
  - More flexibility

- Currently missing:
  - New, user-friendly interface
  - Possibility to recommend new entries
  - Integration of controlled vocabularies and other ontologies

- Possibility to make use of a terminology service?