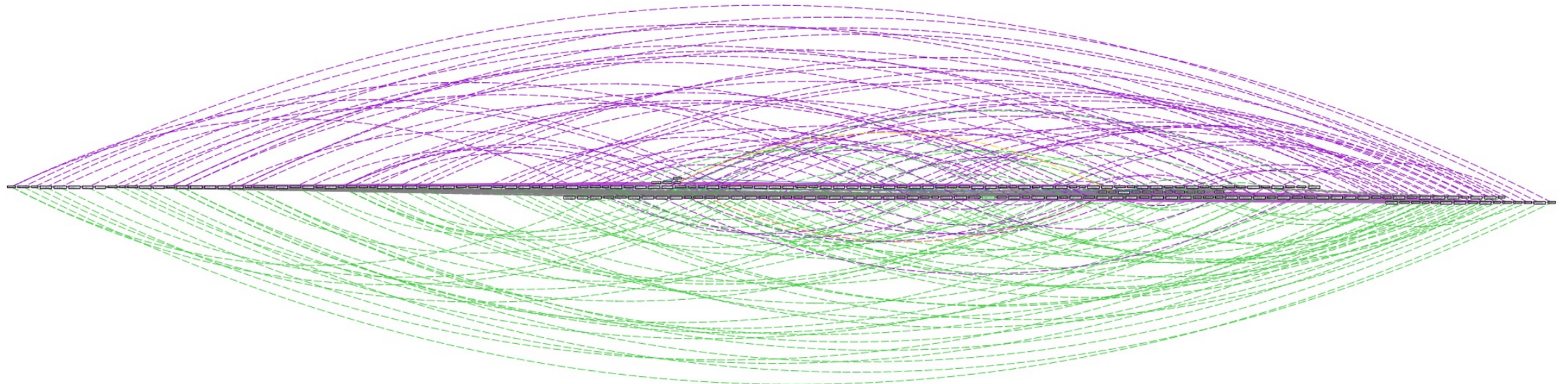


A Short Introduction to the Data Collections Explorer

NFDI4ing SIG RDM Training & Education – 6. October 2023

Philipp Ost philipp.ost@kit.edu



The big Picture: Data Repositories

What is a data repository?

A data repository registers, archives, and provides access to data collections.

- They (usually) provide unique identifiers for data sets
- They help making data FAIR

There are a lot of repositories available – the challenge is finding the right one

- re3data provides a lot of information
- However, there is more...

What is the Data Collections Explorer?

The Data Collections Explorer is an information system for repositories and data collections focussed on the engineering sciences.

Primary use cases include

- Scientists searching for datasets
 - Example: An engineer is interested in material properties
 - Are there other materials that might fit the requirements?

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

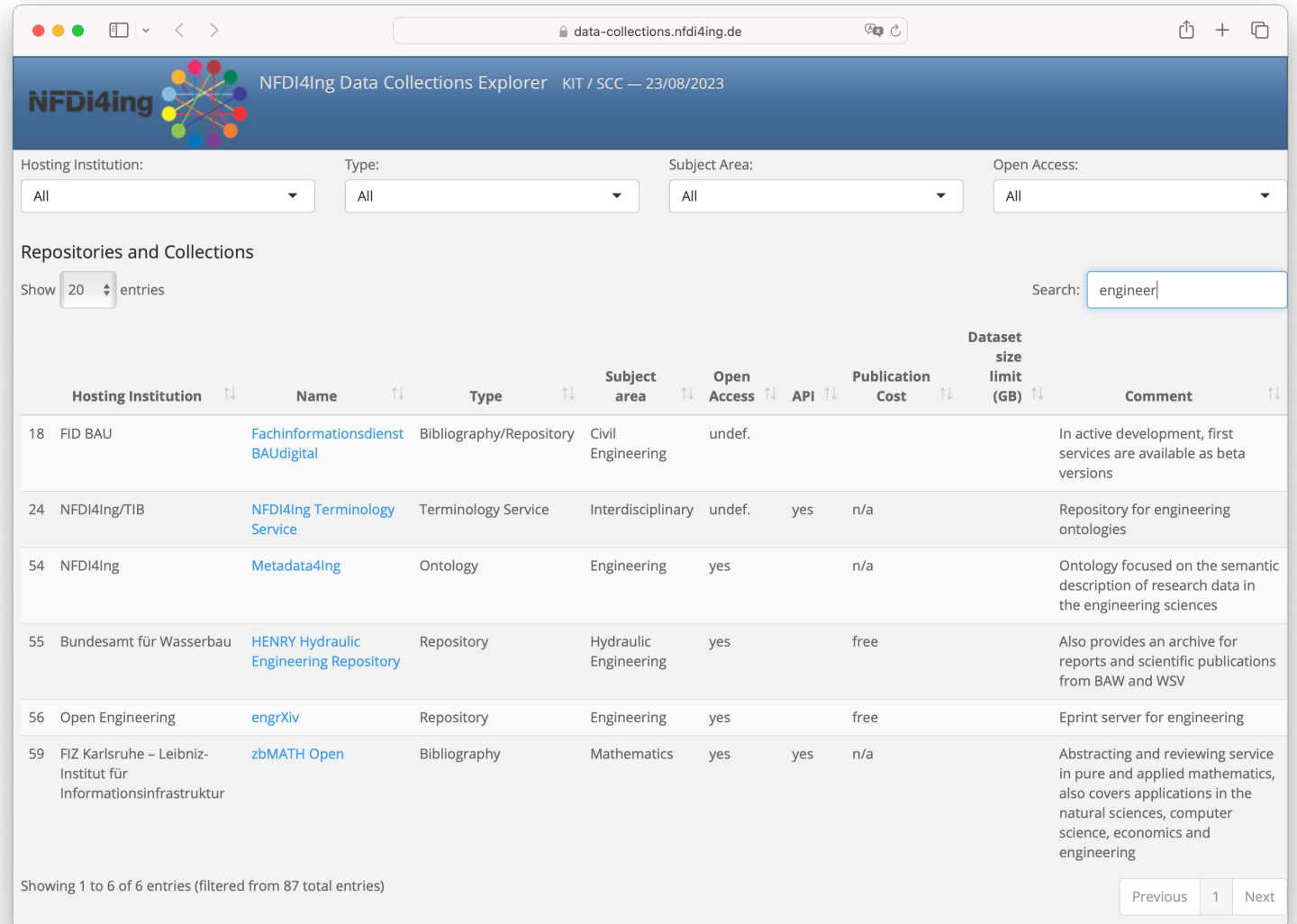
What is the Data Collections Explorer?

The Data Collections Explorer

- Allows filtering and searching for entries
- Complements re3data
 - Individual data sets (e.g., The KITTI Vision Benchmark)
 - Privately or community curated repositories and databases (e.g., The Wood Database, The Phase Field Hub)
 - Repositories not yet listed in re3data
- Receives continuous updates based on community feedback

What Does the Data Collections Explorer Offer?

- Currently 87 repositories and data sets listed
- Access it here: data-collections.nfdi4ing.de

NFDI4ing Data Collections Explorer KIT / SCC — 23/08/2023

Hosting Institution: All Type: All Subject Area: All Open Access: All

Repositories and Collections

Show 20 entries Search: engineer

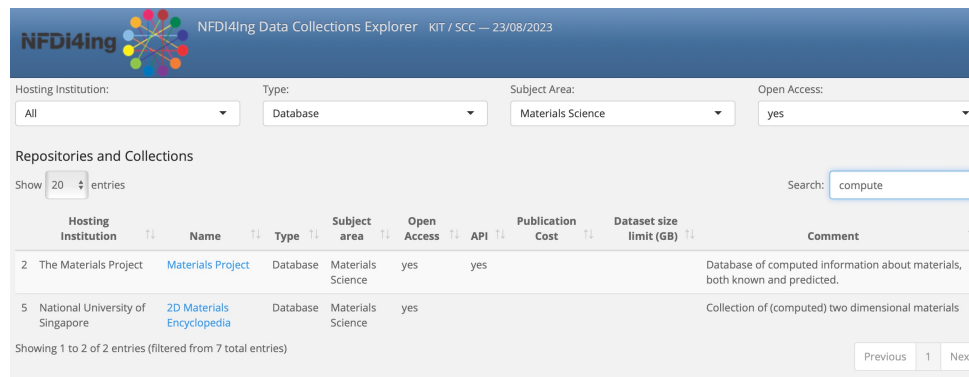
Host	Name	Type	Subject Area	Open Access	API	Publication Cost	Dataset size limit (GB)	Comment
18 FID BAU	Fachinformationsdienst BAUdigital	Bibliography/Repository	Civil Engineering	undef.				In active development, first services are available as beta versions
24 NFDI4ing/TIB	NFDI4ing Terminology Service	Terminology Service	Interdisciplinary	undef.	yes	n/a		Repository for engineering ontologies
54 NFDI4ing	Metadata4Ing	Ontology	Engineering	yes		n/a		Ontology focused on the semantic description of research data in the engineering sciences
55 Bundesamt für Wasserbau	HENRY Hydraulic Engineering Repository	Repository	Hydraulic Engineering	yes		free		Also provides an archive for reports and scientific publications from BAW and WSV
56 Open Engineering	engrXiv	Repository	Engineering	yes		free		Eprint server for engineering
59 FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur	zbMATH Open	Bibliography	Mathematics	yes	yes	n/a		Abstracting and reviewing service in pure and applied mathematics, also covers applications in the natural sciences, computer science, economics and engineering

Showing 1 to 6 of 6 entries (filtered from 87 total entries)

Previous 1 Next

What Does the Data Collections Explorer Offer?

- The drop down menus allow filtering for entries
 - Hosting Institution
 - Type
 - Subject Area
 - Open Access
- The full text search allows to refine results further:



NFDI4Ing Data Collections Explorer KIT / SCC — 23/08/2023

Hosting Institution: All Type: Database Subject Area: Materials Science Open Access: yes

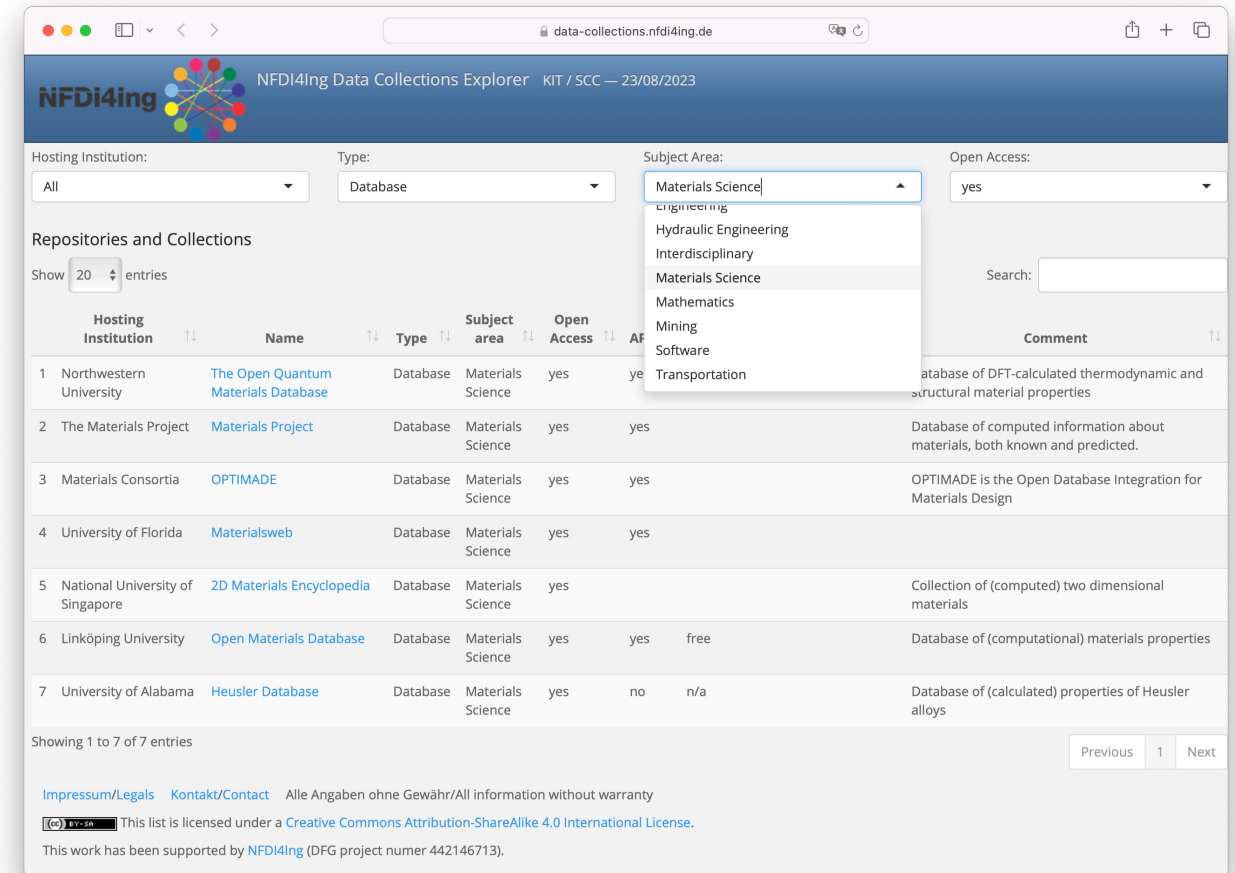
Repositories and Collections

Show 20 entries

Search: compute

Hosting Institution	Name	Type	Subject area	Open Access	API	Publication Cost	Dataset size limit (GB)	Comment
2	The Materials Project Materials Project	Database	Materials Science	yes	yes			Database of computed information about materials, both known and predicted.
5	National University of Singapore 2D Materials Encyclopedia	Database	Materials Science	yes				Collection of (computed) two dimensional materials

Showing 1 to 2 of 2 entries (filtered from 7 total entries)



NFDI4Ing Data Collections Explorer KIT / SCC — 23/08/2023

Hosting Institution: All Type: Database Subject Area: Materials Science Open Access: yes

Repositories and Collections

Show 20 entries

Search:

Subject Area dropdown menu:

- Materials Science
- Engineering
- Hydraulic Engineering
- Interdisciplinary
- Materials Science
- Mathematics
- Mining
- Software
- Transportation

Hosting Institution	Name	Type	Subject area	Open Access	API	Publication Cost	Dataset size limit (GB)	Comment
1	Northwestern University The Open Quantum Materials Database	Database	Materials Science	yes	yes			Database of DFT-calculated thermodynamic and structural material properties
2	The Materials Project Materials Project	Database	Materials Science	yes	yes			Database of computed information about materials, both known and predicted.
3	Materials Consortia OPTIMADE	Database	Materials Science	yes	yes			OPTIMADE is the Open Database Integration for Materials Design
4	University of Florida Materialsweb	Database	Materials Science	yes	yes			
5	National University of Singapore 2D Materials Encyclopedia	Database	Materials Science	yes				Collection of (computed) two dimensional materials
6	Linköping University Open Materials Database	Database	Materials Science	yes	yes	free		Database of (computational) materials properties
7	University of Alabama Heusler Database	Database	Materials Science	yes	no	n/a		Database of (calculated) properties of Heusler alloys

Showing 1 to 7 of 7 entries

Impressum/Legals Kontakt/Contact Alle Angaben ohne Gewähr/All information without warranty

This list is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

This work has been supported by NFDI4Ing (DFG project number 442146713).

Conclusions

- The Data Collections Explorer is an easy-to-use and flexible tool
 - Share and find data from the engineering domain
 - Low barrier of entry
- The Data Collections Explorer is a service for the community
 - Open for contributions – please contact us!
 - You can also create at pull request at <https://github.com/kit-data-manager/Data-Collections-Explorer>
 - Continuously updated based on community feedback