

Data Collections Explorer

An easy-to-use tool for sharing and discovering research data

Philipp Ost¹[\[https://orcid.org/0000-0002-7198-0566\]](https://orcid.org/0000-0002-7198-0566), Yusra Shakeel¹[\[https://orcid.org/0000-0001-5135-4325\]](https://orcid.org/0000-0001-5135-4325), and Philipp Tögel¹[\[https://orcid.org/0000-0002-1791-0268\]](https://orcid.org/0000-0002-1791-0268)

¹ Karlsruhe Institute of Technology, Germany

Abstract

There is a wide variety of archives, databases, and repositories currently available that provide access to research data. However, basic information about these systems is often difficult to gather, such as whether there are limits to the size of data sets that can be published or whether there is any publication fee that applies. In addition to that, there are plenty of research groups publishing their research data sets independently of these infrastructures, making it difficult for scientists to find them since they are not centrally registered. Research data must be easily discoverable and accessible for scientists to use it effectively. The Data Collections Explorer [1, 2], developed within the national research data infrastructure for the engineering sciences NFDI4Ing, is an easy-to-use information system addressing these needs. It is a low threshold information system that provides an overview of research data repositories, archives, databases as well as individually published data sets. Similar systems exist in other subject areas, for example the Data Repository Finder [3] focusing on the medical, life and social sciences. Contrary to the Data Collections Explorer, the Data Repository Finder only lists repositories.

By providing a low entrance barrier, the Data Collections Explorer makes it easy to share and discover repositories as well as data sets. Furthermore, scientists get a quick overview of the most important facts about services and data sets, such as access rights or usage restrictions. To further elaborate, we consider two practical use case scenarios:

1. Scientists searching for data sets: are there data sets available to aid in my research? Are there benchmarks available to check my results? Are these data sets available under an open access license? Are there usage or access restrictions?
2. Scientists aiming to publish data sets: among community-specific repositories, which ones are suitable to publish my research data? Do repositories restrict the size of the data sets that can be uploaded, and if so, what are the limits? Is there any publication fee charged and if so, how much is it?

To answer these questions, the Data Collections Explorer provides a free text search and filters for the type of service, subject area, and access license. Wherever appropriate and available, information on data size limits and publishing fees are given. Updates to the content are mostly initiated by the input of scientists, thus making sure the service fits their needs. Individual data sets, i.e., from the engineering sciences, are listed only if they are published outside the established infrastructure. Unlike re3data [4], services are listed irrespective of whether the operator is an individual, a community, or a legal entity. As a widely accepted authority [5], re3data is imposing strict criteria to list new entries. While smaller-scale projects,

such as databases or repositories run by individuals or communities, might still provide value to researchers, their lack of a legal entity makes them ineligible for listing in re3data [6]. In contrast to this, owing to the low-threshold approach, the Data Collections Explorer is thus capable to complement such established systems, as it provides coverage of entries beyond their scope, or which are not listed there.

Since its launch in March 2022, the Data Collections Explorer is being updated regularly based on the constructive feedback received from the engineering community, namely NFDI4Ing, but also NFDI4Chem and NFDI-MatWerk. After its launch, the Data Collections Explorer was presented in each NFDI4Ing Task Area. These task areas cover all aspects of engineering according to the classification by the German Research Foundation (DFG) [7]. Scientists provided feedback on their requirements of such a service, for instance on usability and access methods; contributions also included the repositories, databases, and data sets which are relevant to their work. Motivated by this positive feedback received during the past year, we are developing a new and improved version to overcome some limitations of the current approach. Currently, the Data Collections Explorer is a human-centered information system and not readily machine-accessible. To achieve the latter, we build a knowledge graph, c.f. Figure 1, that would not only allow easier machine-accessibility, but also a smoother integration with existing efforts within NFDI4Ing and the wider scientific community. Future work comprises improving the technical aspects, i.e., the development of user interfaces and APIs, but also maintenance and curation processes, i.e., to ensure the long-term sustainability of the service.

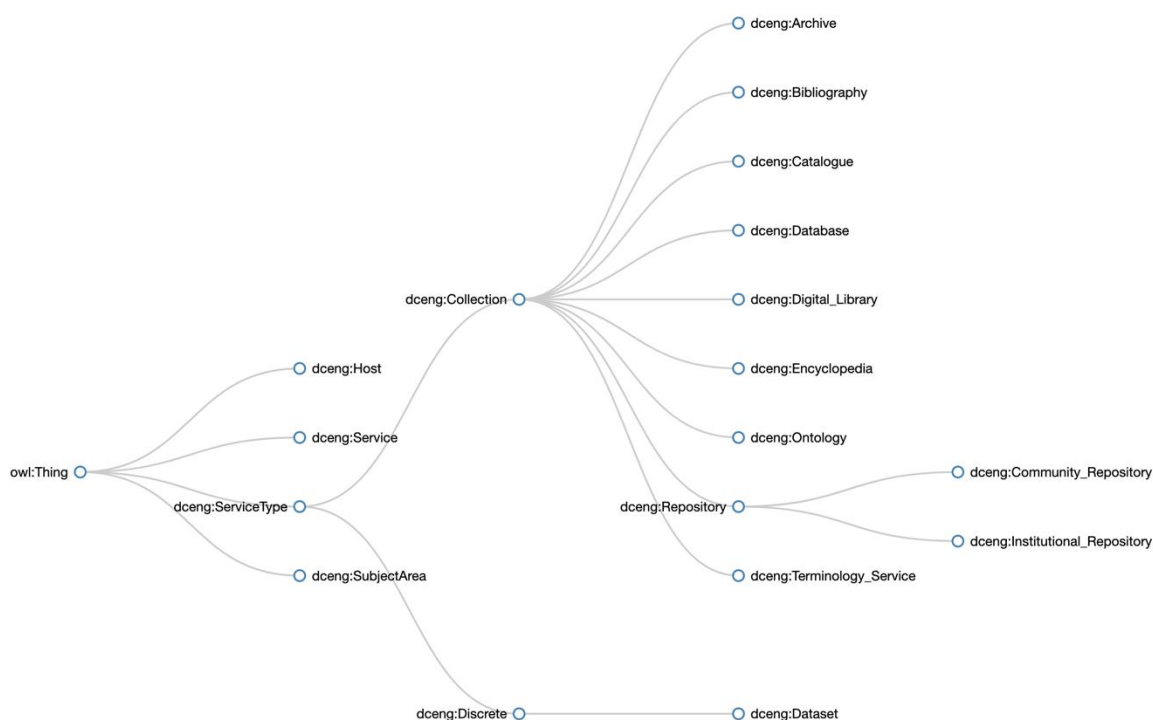


Figure 1 The new graph structure of the Data Collections Explorer.

Our proposed concept is not solely limited to the engineering sciences. The Data Collections Explorer has sparked attention in various NFDI consortia, as well as projects outside of NFDI. To broaden the impact of our concept, as a first step we are working on expanding the Data Collections Explorer to the materials science and engineering community within NFDI-MatWerk.

Competing interests

The authors declare that they have no competing interests.

Contributions

- Philipp Ost – Conceptualization, Investigation, Writing – original draft
- Yusra Shakeel – Writing – review & editing
- Philipp Tögel – Writing – review & editing

Funding

This work has been supported by NFDI4Ing (DFG – project number 442146713), NFDI-Mat-Werk (DFG – project number 460247524) and the Helmholtz Metadata Collaboration (HMC) platform.

References

- [1]. NFDI4Ing Data Collections Explorer, data-collections.nfdi4ing.de, last accessed: 2023-04-25
- [2]. Ost, P. (2022). Data Collections Explorer. Helmholtz Metadata Collaboration | Conference 2022, Online. <https://doi.org/10.5445/IR/1000151429>
- [3]. Data Repository Finder, data-repository-finder.ll.mit.edu, last accessed: 2023-07-17
- [4]. re3data.org - Registry of Research Data Repositories. <https://doi.org/10.17616/R3D> last accessed: 2023-04-25
- [5]. Kindling, M., Pampel, H., van de Sandt, S., Rücknagel, J., Vierkant, P., Kloska, G., Witt, M., Schirnbacher, P., Bertelmann, R., Scholze, F. (2017). The Landscape of Research Data Repositories in 2015: A re3data Analysis. D-Lib Magazine, 23(3/4). <https://doi.org/10.1045/march2017-kindling>
- [6]. Strecker, D., Bertelmann, R., Cousijn, H., Elger, K., Ferguson, L. M., Fichtmüller, D., Goebelbecker, H.-J., Kindling, M., Kloska, G., Nguyen, T. B., Pampel, H., Petras, V., Schabinger, R., Schnepf, E., Semrau, A., Trofimenko, M., Ulrich, R., Upmeier, A., Vierkant, P., Weisweiler, N. L., Wang, Y., Witt, M. (2021). Metadata Schema for the Description of Research Data Repositories: version 3.1, re3data, 37 p. <https://doi.org/10.48440/re3.010>
- [7]. The DFG subject area structure, https://www.dfg.de/en/research_funding/proposal_funding_process/interdisciplinarity/subject_area_structure/index.html, last accessed: 2023-07-18