1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.288

Published: 07 Sept. 2023

NFDI4Earth: Improving Research Data Management in the Earth System Sciences

Lars Bernard[https://orcid.org/0000-0002-3085-7457], Christin Henzen[https://orcid.org/0000-0002-5181-4368], Auriol Degbelo[https://orcid.org/0000-0001-5087-8776], Daniel Nüst[https://orcid.org/0000-0002-0024-5046], and Jörg Seegert[https://orcid.org/0000-0001-9357-2830]

Chair of Geoinformatics, TU Dresden, Germany

Keywords: Geospatial resources, Research Data Management, Earth System Sciences

The increasing availability of digital research products – typically encompassing data and software – in the Earth System Sciences (ESS) calls for new approaches to facilitate the management and reuse of these digital resources. NFDI4Earth addresses this gap and targets the harmonization of services related to digital research products in the ESS. In our initial phase, we aim to support researchers in (i) discovering and exploring relevant data and software sources, (ii) exploratory spatial data analysis, (iii) solving research data management problems and (iv) creating and publishing information products [1, 2]. This abstract briefly presents concepts and first results of the harmonization efforts within NFDI4Earth. It touches upon four points: the harmonization of resource descriptions, the harmonization of ESS-specific support for research data management (RDM), joint training activities for researchers in the ESS, and the harmonization of discovery and publishing workflows.

Harmonizing resource descriptions: A common thread of the various disciplines within the ESS is the collection, processing and sharing of spatial data spanning over broad range of topics (e.g., climate, geology, marine, land use and property, habitats, etc.). Spatial data refers data to which a spatial position can be directly/indirectly assigned through the use of spatial references. These spatial data come in different formats, and so do the interfaces of the services used for their storage, processing and analysis. This diversity demands strategies for harmonizing and linking to established ESS standards for metadata, data and services such as the APIs from the Open Geospatial Consortium (https://www.ogc.org/). NFDI4Earth proposes two key innovations in this context: 1) the integration of various metadata from and for the ESS through the use of Semantic Web concepts (e.g., the Resource Description Framework) and bridging towards concepts such as FAIR Digital Objects; and 2) the NFDI4Earth Label as a digital badge to inform about the level of FAIRness of ESS services. As of this writing, the NFDI4Earth Knowledge Hub is implemented as an open-source software stack using a data management system and a tightly coupled triplestore. (Meta) data is harvested from different external sources (e.g., re3data for ESS repositories, Research Organization Registry and Wikidata for organizations, Digital Curation Centre for ESS-specific metadata standards, GitLab for software projects…) and enriched with internally collected information (Figure 1). Also, community-driven tools developed within the NFDI4Earth Pilots and Incubators (e.g., for data cube visualization) are made accessible through the Knowledge Hub. We thereby enable the ESS community to answer previously unanswered and challenging questions such as ‘list all ESS services related to Oceanography published within the last two years, along with their service types’.

Harmonizing RDM support: Researchers and research data managers face several challenges across the research data management life cycle (see e.g., [3]). These include, for

instance, questions related to what data can be legally and ethically collected and stored, compliance with funding agencies' requirements, and dilemmas related to the sharing of sensitive research data. While efforts at specific universities and university libraries are underway to facilitate RDM in-house services for the institutes' researchers, there is still a lack of platforms in the ESS where researchers can ask questions and receive help from their peers directly. This gap is addressed in NFDI4Earth through the User Support Network, which is currently envisioned as a single point of access to expert knowledge and experts for RDM-related issues. As the experts are distributed across the NFDI4Earth community, the most suitable person is matched to the inquiry independent from organizational membership, personal contacts, or previous public record on a topic.

Joint training activities: Through two-year training programs for doctoral and postdoctoral researchers, the NFDI4Earth Academy offers a learning environment within the ESS that goes beyond institutional boundaries. As of this writing, the Academy offers training to 39 Fellows from 24 institutions in Germany. The training features, among others, research data management and Data Science skills, "think tank" events to brainstorm and develop synergies between the fellows' projects, and cross-consortia activities (e.g., Hackathons).

Harmonizing discovery and publication workflows: Scientific (geo)data infrastructures, despite being sometimes established for years and using open standardized APIs, still lack shared policies and a common organizational framework for the sharing of scientific geospatial resources [4]. In ESS, we can build upon various well-known and accepted services. However, with the increasing number of diverse services, e.g., for specific tasks or formats, researchers lack an overview of existing services and support in choosing proper services according to their needs. Here, NFDI4Earth proposes two solutions for harmonizing discovery and publishing workflows: 1) a single access point, called OneStop4All, for the discovery of distributed resources and 2) a concept to publish (meta) data using best-fitting existing repositories or platforms. The custom open-source OneStop4All solution requests resource descriptions from the Knowledge Hub via an API to generate a user-friendly view of harmonized resource descriptions with particular respect to spatio-temporal characteristics (Figure 1). With the OneStop4All, we also implement the NFDI4Earth concept of data publishing, that is, recommending existing repositories where researchers can publish (meta) data for their resources as opposed to offering a new publishing platform. By doing so, we harmonize publication workflows building on community-accepted and established services, foster the visibility/transparency of relevant data outside NFDI4Earth, reduce implementation efforts in the project and support consolidation and efficient use of resources. Moreover, we envision fostering a cultural change in ESS, e.g., by focusing on FAIR, open and sustainable ESS services, supported by a NFDI4Earth Commitment Statement [5].

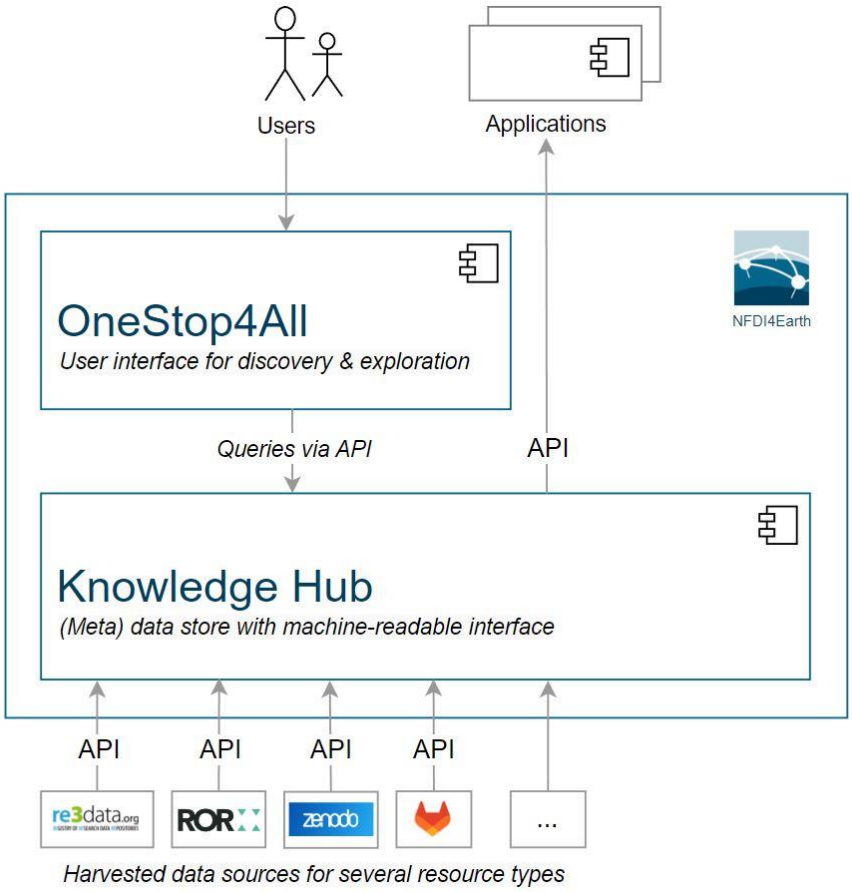

Figure 1. Simplified version of the NFDI4Earth Software Architecture

Data availability statement

The submission is not based on data.

Author contributions

L.B. contributed conceptualization, funding acquisition, supervision, review & editing. C.H. contributed conceptualization and writing – original draft. A.D. contributed conceptualization and writing – original draft. D.N. contributed conceptualization and writing – review & editing. J.S. contributed conceptualization, funding acquisition, project administration.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been funded by the German Research Foundation (DFG) through the project NFDI4Earth (DFG project no. 460036893, https://www.nfdi4earth.de) within the German National Research Data Infrastructure (NFDI, https://nfdi.de/).

Acknowledgement

The authors are grateful for the valuable input of the NFDI4Earth Consortium in the process of shaping and implementing these ideas.

References

1. L. Bernard, P. Bräsicke, R. Bertelmann, S. Frickenhaus, H. Gödde, C. Keßler, S. Lorenz, M. Mahecha, H. Marschall, D. Hezel, W. Nagel, M. Reichstein, M. Sester, H. Thiemann, C. Weiland, A. Wytzisk-Arens, NFDI4Earth Consortium, NFDI Consortium Earth System Sciences - Proposal 2020 revised, https://doi.org/10.5281/zenodo.5718944, 2021
2. C. Henzen, A. Degbelo, D. Nüst, Challenges in Developing a Software Architecture for a National Research Data Infrastructure in Earth System Sciences, EGU General Assembly Vienna, Austria, 24–28 Apr 2023, EGU23-17491, https://doi.org/10.5194/egusphere-egu23-17491, 2023
3. J. Bossaller, A.J. Million, The research data life cycle, legacy data, and dilemmas in research data management, Journal of the Association for Information Science and Technology, 1-6, https://doi.org/10.1002/asi.24645, 2022
4. L. Bernard, S. Mäs, M. Müller, C. Henzen, J. Brauner, Scientific geodata infrastructures: challenges, approaches and directions, International Journal of Digital Earth, 7(7), 613-633, https://doi.org/10.1080/17538947.2013.781244, 2014
5. J. Seegert, D. Nüst, L. Bernard, Initiating Cultural Change in the German Earth System Sciences Community with a Commitment Statement, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-14456, https://doi.org/10.5194/egusphere-egu23-14456, 2023.