









NFDI4Earth

NFDI4Earth White Paper

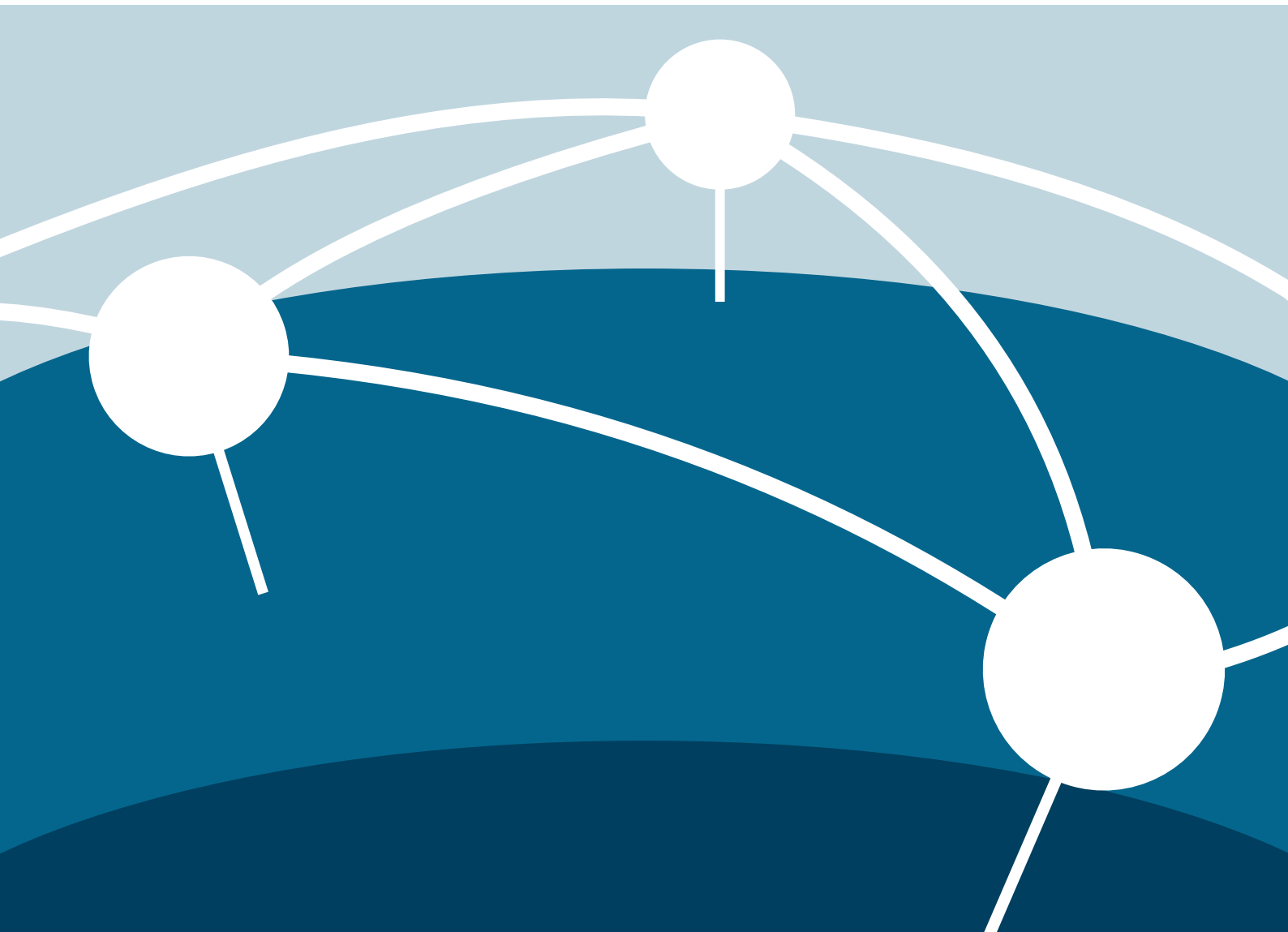
The NFDI4Earth Label - Concept Overview

Tim Schäfer  (tim.schaefer@senckenberg.de), Ronny Gey , Jonas Grieb ,
Christin Henzen , Claudia Müller , Markus Stocker , Claus Weiland ,
Stephan Frickenhaus 

2024-09

DOI: [10.5281/zenodo.13711459](https://doi.org/10.5281/zenodo.13711459)

nfdi4earth.de



Citation

Tim Schäfer, Ronny Gey, Jonas Grieb, Christin Henzen, Claudia Müller, Markus Stocker, Claus Weiland, Stephan Frickenhaus. 2024. *The NFDI4Earth Label - Concept Overview (NFDI4Earth White Paper)*. NFDI4Earth Community on Zenodo. <https://doi.org/10.5281/zenodo.13711459>

License

This work is licensed under a [Creative Commons “Attribution 4.0 International”](#) license.



Acknowledgement

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://www.nfdi.de/>).

Executive summary

The research data infrastructures in the Earth System Sciences (ESS) are highly diverse, featuring numerous data repositories and metadata aggregators with varying standards, which complicates the provision of essential cross-infrastructure services. The NFDI4Earth Label aims to enhance the interoperability and trustworthiness of these repositories by developing clear repository guidelines and a certification process tailored to the ESS community. Unlike existing certifications, the NFDI4Earth Label emphasizes ESS-specific metadata standards and is achievable for smaller repositories, ensuring relevance and practical applicability. Developed in collaboration with the ESS community and aligned with European Open Science Cloud (EOSC) initiatives, the NFDI4Earth Label seeks to foster a more integrated and trustworthy research data landscape in the ESS.

Contents

| | |
|--|----------|
| 1 Motivation | 1 |
| 2 Relation of the Label to other Certifications | 1 |
| 3 Approach | 2 |
| 4 Outlook | 6 |

1 Motivation

The landscape of research data infrastructures in the Earth System Sciences (ESS) is very diverse and contains a large number of data repositories and metadata aggregators of different sizes and from different sub disciplines. These data providers use a variety of interfaces, metadata exchange protocols and metadata schemes to facilitate FAIR¹ research data.

However, the degree to which standards are used and the exact choice of standards often differ between repositories, which makes it challenging to integrate infrastructures into a common framework and provide crucial services across infrastructures. Examples for such services required by scientists include searching across many repositories, the ability to filter by geographic location and time, and the ability to combine datasets from different sources and sub-disciplines to answer new research questions.

Scientists also need a means to identify trustworthy and suitable repositories, but information on research data repositories is often incomplete or out-of-date in public registries.

Clear guidelines for repositories that suggest a set of standards that have been agreed upon in the ESS community, in combination with readily available information on which repositories implement these guidelines, e.g. via a certification, a label or a graphical badge, would serve this purpose, but most currently available certifications are not specific to the ESS and hard to obtain for smaller data repositories.

Therefore, NFDI4Earth is developing the NFDI4Earth Label². The project aims to assess and subsequently improve the level of interoperability and trustworthiness of research data repositories in the ESS.

2 Relation of the Label to other Certifications

Research data repositories play a critical role in preserving and providing access to scientific data, i.e., in making research data more findable, accessible, interoperable and reusable (FAIR³). A key component in the relationship between repositories and their users is trust⁴. Users trust repositories which can demonstrate their ability to serve the needs of their

¹ Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

² Grieb, J., Schäfer, T., Frickenhaus, S., Gey R. and Weiland, C. (2023). First Status Report on the NFDI4Earth Label. Zenodo. <https://doi.org/10.5281/zenodo.13711420>

³ Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

⁴ Yoon, A. End users' trust in data repositories: definition and influences on trust development. *Arch Sci* **14**, 17–34 (2014). <https://doi.org/10.1007/s10502-013-9207-8>

research community, e.g. through continued operation in combination with external audits and a resulting certification. A variety of official certifications for data repositories exist, like the Core Trust Seal⁵ (CTS), the Nestor Seal, and certification according to ISO16363. However an evaluation of the current research infrastructure landscape in the ESS indicates that very few repositories hold a certification. For instance, as of July 2024, less than five percent of the geosciences repositories⁶ hold the CTS, the Data Seal of Approval (DSA), or the World Data System of the International Science Council (WDS) regular member certification according to re3data⁷ and the CTS website⁸. According to a recent study, providing the technical, legal, financial, and organizational resources for digital preservation is a significant challenge for the repository providers and, hence, the reason for a low rate of certification among repositories.⁹ To make the Label applicable to a significant share of the ESS infrastructures, and also include smaller, discipline-specific and institutional repositories in the intended interoperability improvements, we intend to keep the requirements for obtaining the label low compared to existing certifications, while recognizing any *additional* repository features and achievements.

Another key difference between the Label and existing certifications like the CTS, the Nestor SEAL, ISO 16363 and others is that the metadata standards suggested in the guidelines of the NFDI4Earth Label are specific to the requirements of the ESS community, where standardized information on geolocations or on the spatial extent of datasets is of high importance.

Repositories that have obtained an official certification, such as the CTS, have demonstrated a high level of trustworthiness¹⁰, e.g., they are known to have suitable organizational infrastructure backup infrastructure and a preservation plan. We recognize this achievement by automatically considering specific requirements for obtaining the NFDI4Earth Label which are directly equivalent to those already verified by the CTS as fulfilled.

3 Approach

To assess and subsequently improve the level of interoperability and trustworthiness of research data repositories in the ESS, we have developed a semi-automated process for the

⁵ <https://www.coretrustseal.org/>

⁶ These are repositories classified as DFG category '34 Geosciences (including Geography)' according to re3data.

⁷ <https://www.re3data.org/>

⁸ <https://www.coretrustseal.org/maps/fullscreen/10/>

⁹ Donaldson, D.R., Russell, S.V. (2023). Trustworthy Digital Repository Certification: A Longitudinal Study. In: Sserwanga, I., *et al.* Information for a Better World: Normality, Virtuality, Physicality, Inclusivity. iConference 2023. Lecture Notes in Computer Science, vol 13972. Springer, Cham. https://doi.org/10.1007/978-3-031-28032-0_42

¹⁰ Corrado, E. M. (2019). Repositories, trust, and the CoreTrustSeal. *Technical Services Quarterly*, 36(1), 61-72.

evaluation of research data repositories that consists of:

- The mandatory registration of the repository at re3data¹¹, which provides a unique identifier and ensures the public availability of metadata on the repository itself in machine-readable form.
- A set of guidelines for repositories, each one related to a FAIR principle, which define common standards that should be implemented by all repositories in the ESS to facilitate interoperability and increase trustworthiness. The guidelines which refer to metadata provisioning are in line with published metadata guidelines of the NFDI4Earth.¹²
- Specific evaluation metrics that can be used to programmatically evaluate repositories and measure the degree to which a repository conforms to the guidelines.
- A semi-automated assessment procedure that consists of (1) automated retrieval of information on the repository from re3data, combined with (2) the automated FAIR assessment of a sample of the datasets stored in the repository using the F-UJI tool^{13,14}, and (3) a self evaluation form to be filled out by repository representatives that queries information on the repository, which is not available from re3data.
- A transparent workflow that will be used to guide repository representatives through the process of applying for, obtaining, and renewing the Label.

To ensure that the Label is accepted by, and relevant for, the ESS community, it is developed in collaboration with the community. To discuss the plans for a Label and gather feedback, the Label group of the NFDI4Earth held a workshop with repository representatives in October 2023. The updated concept was presented at an internal NFDI4Earth meeting in Frankfurt in May 2024, and at the 3rd NFDI4Earth Plenary 2024 in Dresden¹⁵. The label team is taking part in the FAIR-IMPACT Support Action 'Recommendations for Trustworthy and FAIR-enabling Data Repositories'¹⁶ to align the guidelines of the NFDI4Earth Label with similar approaches of the European Open Science Cloud (EOSC) on the European level.

In the following section, we give an overview of the Label and the planned process that will be implemented to allow repository representatives to obtain the Label. The workflow is

¹¹ <https://www.re3data.org/>

¹² Bernard, L., Degbelo, A., Grieb, J., Henzen, C., Heß, R., Klammer, R., Koppe, R., Lorenz, C., Müller, C., & Weiland, C. (2024). Recommendations for Earth System Sciences Metadata Provision. Zenodo. <https://doi.org/10.5281/zenodo.10604587>

¹³ <https://www.f-uji.net/>

¹⁴ Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool. Zenodo. <https://doi.org/10.5281/zenodo.6361400>

¹⁵ <https://www.nfdi4earth.de/nfdi4earth-plenary-2024>

¹⁶ <https://fair-impact.eu/support-offer-3-recommendations-trustworthy-and-fair-enabling-data-repositories>

visualized in Fig. 1. It starts with the applicant providing the re3data identifier of the repository and applying for the Label. The identifier is used to automatically retrieve required information on the repository from re3data. Since re3data does not yet have fields for all required information, and some of the required information contains personal data that is not suitable for public display at re3data, an additional self-assessment form will need to be filled out on the NFDI4Earth website. Some technical capabilities of the repository are then queried automatically using the F-UJI tool. The F-UJI software is a FAIR assessment tool for datasets. Upon execution, a script samples several datasets stored in a repository and runs the F-UJI evaluation on them, and then aggregates the results. This provides evidence¹⁷ for claims made on re3data, by testing the metadata standards and other capabilities of the data repository. The information from re3data, the self-assessment form, and F-UJI is then used in combination with the metrics to perform the assessment.

The assessment result consists of a numeric score, recommendations on how to improve the FAIRness of the repository and thus the score, and, if the score exceeds a certain threshold, the achieved Label badge. The badge is planned as an image that holds the repository name and a text stating whether the repository has achieved the NFDI4Earth Label. The image is clickable and links to a URL on the NFDI4Earth website, where it can be verified by users. Here, users can also see the exact score (e.g., 17 out of 20 points) and details on the assessment process and the results. Internally, the badge will be implemented by using a service like shields.io¹⁸.

This preliminary assessment result is communicated to the repository representatives, who applied for the Label for review, to ensure that the results of the evaluation are representative of the capabilities of the repository, and that the automated tests did not miss any capabilities of the repository, e.g., due to software problems, a service down time or other technical issues. If consultancy is needed, members of the Label team will help to investigate any issues. Once the result is considered final, it will be communicated again, and will subsequently be made public. This includes updating the NFDI4Earth database with the information that the repository has obtained the Label, so that it is automatically displayed at relevant web pages and can be verified by users. The repository should also update its website to include the Label badge in a prominent location.

¹⁷ Ross, P. S., & McHugh, M. A. (2006). The role of evidence in establishing trust in repositories. <https://doi.org/10.1045/july2006-ross>

¹⁸ <https://shields.io/badges>

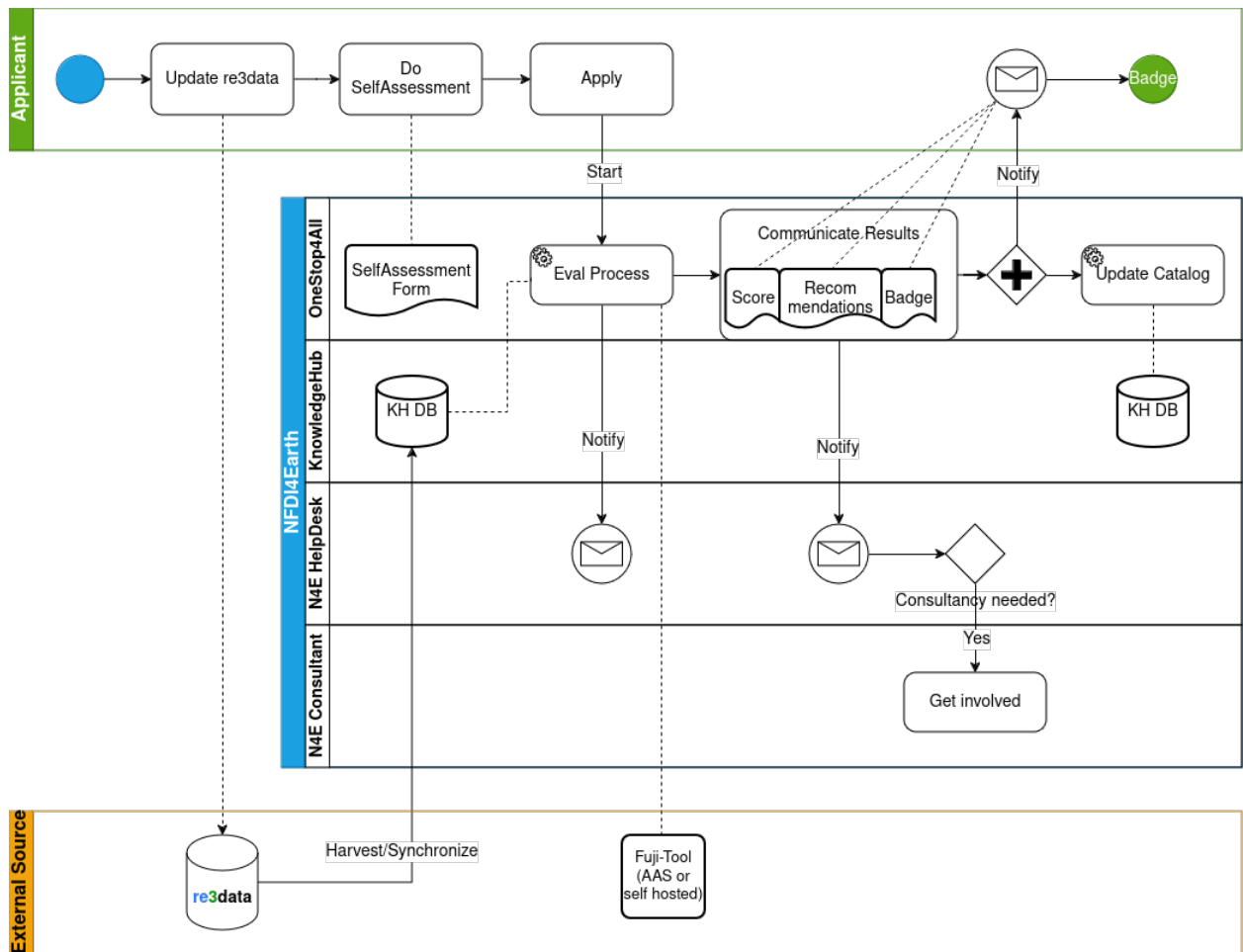


Figure 1: The process of obtaining the NFDI4Earth Label. Applicants provide information on the repository via re3data and a self-assessment form. More information is obtained via the F-UJI tool. The Label metrics are used to perform the assessment, and the results are discussed with the applicant and subsequently made public.

4 Outlook

We are currently working on the implementation of the first iteration of the remaining software components required for the Label workflow. This includes the self-assessment form, a means for repository representatives to start an automated test evaluation by themselves, the badge generation, and the verification web page. Once the software components are ready, the internal testing of the audit process will start. Afterwards, we will start pilot audits with a small but diverse set of manually selected and invited data repositories. These audits will help us to identify weak points and open issues in the audit procedure. The analysis of the resulting pilot assessment results will also help us to evaluate our metrics, and adapt them if needed. Finally, the Label process will become available to all interested repositories. The participation of many repositories and the re-evaluation of repositories on a regular basis will enable us to continually monitor and improve the trustworthiness and interoperability of research data repositories in the ESS.