

PID4NFDI

Persistent Identifier Services for the German National Research Data Infrastructure

Roland Bertelmann ^a, Matthew Buys ^b, Irina Sens ^c, Philipp Wieder ^d

with contributions from

Sven Bingert ^d, Britta Dreyer ^{b,c}, Stephanie Hagemann-Wilholt ^c,
Antonia C. Schrader ^a, Kelly Stathis ^b, Paul Vierkant ^b

and the PID4NFDI Working Group members

^a Helmholtz Association, Helmholtz Open Science Office

^b DataCite

^c German National Library of Science and Technology (TIB) – Leibniz Information Centre for
Science and Technology

^d Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG)

Proposal for the Initialisation Phase of Base4NFDI

Submitted May 15, 2023

On behalf of WG Persistent Identifier Services, Section Common Infrastructures

Published December 18, 2024

<https://doi.org/10.5281/zenodo.14281250>

This work is licensed under <https://creativecommons.org/licenses/by/4.0/>

1 General Information

Name of the proposed basic service

PID4NFDI

Service "subtitle" explaining key functionality

Persistent Identifier Services for the German National Research Data Infrastructure

Lead institutions

Technische Informationsbibliothek (TIB) –
German National Library of Science and
Technology
Welfengarten 1B, 30167 Hannover

Gesellschaft für wissenschaftliche
Datenverarbeitung (GWDG)
Burckhardtweg 4, 37077 Göttingen

Home: <https://www.tib.eu/de/>

ROR ID: <https://ror.org/04aj4c181>

Home: <https://www.gwdg.de/>

ROR ID: <https://ror.org/00cd95c65>

Member in consortium: NFDI4Ing

Member in consortium: Text+

Name of lead institutions principal investigators

Dr. Irina Sens

<https://orcid.org/0000-0001-9190-8628>

Prof. Dr. Philipp Wieder

<https://orcid.org/0000-0002-6992-1866>

Participating institutions

Principal investigator	Institution & Location	Member in consortium
Roland Bertelmann	https://orcid.org/0000-0002-5588-0290 Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences Helmholtz Open Science Office https://ror.org/04z8jg394	NFDI4Earth
Matthew Buys	https://orcid.org/0000-0001-7234-3684 DataCite – International Data Citation Initiative e.V. https://ror.org/04wxnsj81	NFDI4Ing

Table 1: List of participating institutions

Use case partners

Main Contact	Institution & Location	Member in consortium
Matthias Lange	https://orcid.org/0000-0002-4316-078X Leibniz Institute of Plant Genetics and Crop Plant Research https://ror.org/02skbsp27	FAIRagro, NFDI4Biodiversity
Peter Mutschke	https://orcid.org/0000-0003-3517-8071 GESIS - Leibniz Institute for the Social Sciences (www.gesis.org) https://ror.org/018afyw53	KonsortSWD, NFDI4DS, BERD@NFDI
Jan Brase	https://orcid.org/0000-0002-8250-6253 Staats- und Universitätsbibliothek Göttingen https://ror.org/01y9bpm73	NFDI4Biodiversity, NFDI4Culture, NFDI4Earth, Text+,
Lorenz Reimer	https://orcid.org/0000-0002-7805-0660 Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen https://ror.org/02tyer376	NFDI4Microbiota, NFDI4Biodiversity

Table 2: List of associated use case partners (no funding requested)

Planned duration of the project

From 01.10.2023 - 30.09.2024 (12 months)

Summary of the proposal in English

Persistent identifiers (PIDs) are central to FAIR research data management and have become an essential component of applications and services. However, the demand for PIDs for different resources results in diverse requirements: Different disciplines and NFDI consortia have different levels of maturity in PID implementation. Also, PID-related metadata are central to the FAIRness of research resources and support their reusability and the reproducibility of research. Initial analysis of the use cases, requirements and plans of the NFDI consortia has revealed that these and other challenges need to be addressed by the NFDI as a whole. PID4NFDI will address these challenges and design the work programme to build an NFDI foundation service on established PID infrastructures. Core players of the national and international PID service provider and developer community make up the PID4NFDI consortium. Within the initialisation phase it is planned to lay the foundation for further phases through use cases analyses, requirements engineering, and concept development. Envisaged results can be used as is, but will seamlessly lead to future development and integration activities. Regarding technical service implementation, PID4NFDI will follow a clear

path of promoting the uptake by PID infrastructures and service providers and will only develop and operate NFDI-specific services if necessary.

Summary of the proposal in German

Persistent Identifiers (PIDs) sind zentral für ein FAIRes Forschungsdatenmanagement und ein wesentlicher Bestandteil von Anwendungen und Diensten geworden. Die Nachfrage nach PIDs für verschiedene Ressourcen führt jedoch zu vielfältigen Anforderungen: Verschiedene Disziplinen und NFDI-Konsortien weisen unterschiedliche Reifegrade bei der PID-Implementierung auf. Auch sind PID-bezogene Metadaten zentral für die FAIRness von Forschungsressourcen und unterstützen deren Wiederverwendbarkeit und die Reproduzierbarkeit von Forschung. Erste Analysen der Anwendungsfälle, Anforderungen und Pläne der NFDI-Konsortien haben ergeben, dass diese und weitere Herausforderungen von der NFDI als Ganzes angegangen werden müssen. PID4NFDI wird sich diesen Herausforderungen stellen und das Arbeitsprogramm so gestalten, dass ein NFDI-Basisdienst auf bewährten PID-Infrastrukturen aufgebaut wird. Das PID4NFDI-Konsortium setzt sich aus den Hauptakteuren der nationalen und internationalen Community der PID-Dienstleister zusammen. In der Initialisierungsphase werden durch Fallanalysen, Requirements-Engineering und Konzeptentwicklung die Grundlagen für weitere Phasen gelegt. Die Ergebnisse der ersten Phase können bereits verwendet werden, fließen aber auch nahtlos in künftige Entwicklungs- und Integrationsaktivitäten ein. In der technischen Umsetzung wird PID4NFDI die Übernahme von PID-Infrastrukturen und -Dienstleistern fördern und nur bei Bedarf NFDI-spezifische Dienste entwickeln und betreiben.

2 State-of-the-Art of Proposed Basic Service

Background and Motivation

In recent years, persistent identifiers (PIDs) have been widely accepted in the general scientific community to identify and link research assets like data objects, general research outputs, or the researchers themselves. Because of this, members of all NFDI consortia use different kinds of PIDs already in their everyday work. Furthermore, reliable PID systems are the backbone for many additional services, like research data repositories, knowledge graphs, or portfolio analytics services. Thus, persistent identifiers are by definition a fundamental building block of research data management (RDM) and a mandatory element of FAIR data infrastructures.¹ Globally operating organisations and consortia like DataCite², the DOI Foundation³, or the ePIC consortium⁴ already offer trustworthy, mature, and well-established infrastructures that are used for identification of research assets by almost all NFDI consortia. Furthermore, a growing number of systems for the persistent identification of entities such as persons, organisations, places, events, or for general terms already exist.

Nevertheless, the use of identifiers in all NFDI consortia at the moment is scattered and heterogeneous in terms of actors, services, scope, quality, and costs involved. Furthermore, each PID provider has its own technical infrastructure, metadata scheme, and governance model. Having the core function of PIDs in RDM in mind, it is essential to analyse existing gaps and develop joint solutions in order to serve the highly-specific needs of the individual communities and the NFDI as a whole concerning 1) governance, 2) cross-disciplinary standardisation to foster findability and re-use, and 3) integration into different infrastructures with their own technical and organisational requirements. This implies finding answers to questions like: How can the existing infrastructure of PID services be optimally used and interoperability with global research infrastructures like EOSC be ensured? What interfaces, metadata and licences are needed? Which specific needs of consortia should be addressed? Where are central solutions – e.g. for support and training – needed? To answer these questions, PID4NFDI proposes a balanced approach to i) harmonise the requirements of the various stakeholders, e.g. of existing PID infrastructures as well as consortium needs and practices, ii) to support the integration of PIDs, and also iii) to develop (concepts of) additionally required services or service components. **The central objective of PID4NFDI is to consolidate and evolve the PID service landscape within NFDI at all levels: technical, organisational, methodological, and in communication.**

¹ Wilkinson et al 2016.

² <https://datacite.org/>

³ <https://www.doi.org/>

⁴ <https://www.pidconsortium.net/>

State-of-the-art

A persistent identifier is a persistent, unique, and globally resolvable identifier based on an openly specified PID scheme.⁵ The global PID ecosystem relies on many persistent identifier standards (like, e.g., ARK, DOI, Handle, or URN) to identify a referent: a digital object or a digital representation of an entity. Whatever the chosen scheme is, the PIDs' basic features ensure global uniqueness, persistent identification, and long-term resolvable function to locate the referent. Those functionalities permit unambiguously referencing and citation and improve the object's visibility and findability.⁶

The European Open Science Cloud (EOSC) is an integrated infrastructure to create a web of FAIR data. The development of EOSC is a significant and ongoing multi-stakeholder initiative with a large number of associated projects that build services integrated into the overall EOSC landscape. The EOSC Persistent Identifier (PID) policy establishes service and infrastructure requirements for potential services providers. Also, the PID technical architecture document presents guidelines on the implementation of compliant PIDs and related services with PID EOSC Policy. It also identifies opportunities for interoperability between PID services and the EOSC framework.⁷ In this sense, these policy-related and technical guidelines shape the high-level aligning of PID4NFDI with EOSC.

Status of work results in preparation for the Basic Service

The **Persistent Identifier Services Working Group** (PID WG) is located in the NFDI section Common Infrastructures and represents the stakeholders of the German PID landscape within the NFDI. It was founded in 2022 and since then integrates the PID-related activities within the NFDI and started to establish all necessary processes to best support its stakeholders. The WG published a concept on its strategy for implementation and expansion of community-based and established PID services that is closely aligned with the needs of the NFDI consortia. This includes necessary technical and organisational measures as well as information and training aspects.⁸ The WG conducted an initial survey among NFDI consortia to learn more about the status quo of PID use cases within the NFDI: Only 30% of the services mentioned by the respondents already integrate (P)IDs from other sources, 10% are planning to do so; 23% are integrating (P)IDs. As PIDs are fundamental for FAIR data management, we see considerable need here for optimisation in the reuse and allocation of PIDs.⁹ Exemplary deeper insights were provided by a first stakeholder workshop with representatives of selected advanced use cases, which was conducted by the working group on January 27, 2023. Questions that emerged from this workshop were, for example, how to optimise linking of PIDs and to improve their interoperability; how to cope with ephemeral objects and objects of different granularity; how to improve metadata for machine actionability; and in which way to consider metadata requirements for resources where the usage of PIDs is not mature yet (e.g.

⁵ European Commission 2020.

⁶ Ferguson et al. 2018.

⁷ European Commission 2020 and 2021.

⁸ Bingert et al. 2022.

⁹ See a summary of some main results in: Hagemann-Wilholt 2023.

terminologies).¹⁰ Feedback from both the survey and the workshop confirmed and reinforced the direction of the planned work program in this proposal, starting with a deeper analysis of the existing, scattered PID landscape in NFDI.

Since 2009, **DataCite** works worldwide primarily with research institutions that operate repositories and other publication services. The well-established DataCite Metadata Schema (V4.4.)¹¹ currently covers 28 different "Resource Types" and supports the integration of other PIDs, such as ORCID and ROR IDs. In addition, DataCite supports the ROR and re3data registries, facilitating the standardisation of their metadata schemas.

- Membership in working groups (selection): German Initiative for Network Information DINI e. V. (DINI), the EOSC PID Policy Task Force, RDA
- PID related projects (selection): FAIRCORE4EOSC, FAIR-IMPACT, DICE, PID Network Germany, Global Citation Corpus, NFDI, ORCID DE¹²

The **Persistent Identifier Consortium for eResearch (ePIC)**¹³ as an international consortium of partners provides various PID services for the research community – based on the Handle System – and a global infrastructure that is ready for FAIR Digital Objects (FDO). ePIC is an early adopter of new standardisation outputs from several RDA working groups. The Data Type Registry (DTR), e.g., fosters the use of machine actionable standardised PID metadata. This service is one of the key elements of the EOSC PID policy as well as a fundamental building block of Fair Digital Objects.

- Membership in working groups (selection): EOSC Task-Forces, RDA Working Groups
- PID related projects (selection): FAIR-Impact or FAIRCORE4EOSC, FDO Forum,¹⁴ NFDI

The **Helmholtz Open Science Office** is involved in national and international Open Science initiatives, e.g., "Digital Information"¹⁵, the German Initiative for Network Information DINI e. V. (DINI)¹⁶, German Reproducibility Network (GRN)¹⁷.

- Membership in working groups (selection): RDA, EOSC
- PID related projects (selection): ORCID DE, PID Network Germany, NFDI (Helmholtz Association¹⁸), re3data COREF, STD-DOI (root of the development of DataCite and IGSN – International Generic Sample Number)¹⁹

The **Technische Informationsbibliothek (TIB)** – German National Library of Science and Technology – leads the TIB DOI Consortium²⁰ and the ORCID Germany Consortium and represents its members in the international committees of DataCite and ORCID Inc.²¹ In this function, the TIB

¹⁰ Schrader et al. 2023.

¹¹ <https://schema.datacite.org/meta/kernel-4.4/>

¹² <https://www.orcid-de.org/home>

¹³ <https://www.pidconsortium.net/>

¹⁴ <https://fairdo.org/>

¹⁵ <https://www.allianz-der-wissenschaftsorganisationen.de/en/>

¹⁶ <https://dini.de/>

¹⁷ <https://reproducibilitynetwork.de>

¹⁸ <https://os.helmholtz.de/en/open-research-data/helmholtz-in-the-nfdi>

¹⁹ <https://www.forschungsdaten.org/index.php/STD-DOI>

²⁰ <https://projects.tib.eu/pid-service/en/tib-doi-konsortium/members-of-the-tib-doi-konsortium/>

²¹ <https://www.orcid-de.org/konsortium/ueber-das-konsortium>

PID Competence Center is the support and coordinator for most German universities and research institutions for the allocation and use of DOIs and ORCID iDs.

- Membership in working groups (selection): EOSC Task Force “PID Policy and Implementation”, KE Task Force “PID Risks & Trust”, DataCite Metadata Working Group, ROR Community Advisory Group²²
- PID related projects (selection): ORCID DE, PID Network Germany, ConfIDent, OPTIMETA, TAPIR, NFDI

FAIRCORE4EOSC²³ is a European project aiming at establishing nine core services for EOSC. supporting the EOSC PID infrastructure (e.g. through a Type Registry and a MetaResolver), the EOSC research software infrastructure, and the discovery service for all EOSC repositories (e.g. using the PID Graph and the ResearchGraph). The consortium consists of 22 partners, with GWDG being the second largest partner. FAIRCORE4EOSC works in close collaboration with FAIR-IMPACT²⁴ with participation of DataCite, which will support the implementation of FAIR-enabling practices, tools and services.

The **PID Network Germany**²⁵ is a DFG-funded project executed by DataCite, the German National Library, the Helmholtz Open Science Office, the German National Library of Science and Technology (TIB) and the Bielefeld University Library. The project aims to establish a network of stakeholders around the persistent identification of people, organisations, and resources in the field of digital communication in science and culture, which promotes the dissemination and connection of PID systems in Germany. The focus is also on identifying needs and optimization potential for existing PID systems and embedding them in international knowledge graphs. The project findings should lead to recommendations in the order to create a national PID roadmap for Germany. The PID Network Germany complements PID4NFDI in particular on the policy level and both projects agreed to closely collaborate. In contrast to PID4NFDI, PID Network Germany has a broader focus in terms of targeted use cases and networks. However, since PID4NFDI addresses the RDM community and can take their requirements into account in more detail, we see this project as an important and complementary building block to PID Network Germany.

A number of use case partners with strong PID-related requirements and developments within their own NFDI projects have expressed interest to support PID4NFDI as associated partners (without requesting funds through PID4NFDI):

1. **GESIS – Leibniz Institute for the Social Sciences** (KonsortSWD): PIDs for fine-grained research resources, e.g. dataset elements, survey variables;²⁶
2. The **IPK – Leibniz Institute of Plant Genetics and Crop Plant Research** (FAIRagro, NFDI4Biodiversity): PIDs for digital twins of plant genetic resources;²⁷

²² <https://projects.tib.eu/pid-service/en/pid-competence-center/projects-and-publications/>; <https://www.pidinst.org/>;
Stocker et al. 2020.

²³ <https://faircore4eosc.eu>

²⁴ <https://fair-impact.eu>

²⁵ Bertelmann et al. 2023.

²⁶ Klas et al. 2022 on KonsortSWD Measure TA.5-M.1.

²⁷ Ewert et al. 2021.

3. The **State and University Library (SUB) Göttingen** (Text+): PIDs and metadata requirements for the humanities;²⁸
4. The **Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures** (NFDI4Microbiota and NFDI4Biodiversity): PIDs to disambiguate and link microbial strains, establishment of central registry of PIDs for strains.²⁹

Current Technical Readiness Level (TRL) of the proposed Basic Service

The globally adopted PID infrastructure is the basis of the services to be developed by PID4NFDI. All of them are operational for years, globally available, and supported by strong and consistent governance frameworks. Furthermore, the maturity of the existing PID services ensures compliance with international standards and best practices, enabling the PID4NFDI services and thus the NFDI developments to connect to international RDM initiatives such as EOSC.

PID4NFDI will base its services primarily on the following technical baseline services, which all have a Technology Readiness Level of TRL9: (i) DataCite Commons, Fabrica, and APIs, (ii) ePIC Persistent Identifiers for Research and APIs, and (iii) ORCID API, Registry, and Member Portal. In addition, other mature PID infrastructure services such as ROR and IGSN ID registration are also highly relevant to the project and will be taken into account. PID4NFDI partners will ensure that developments within Base4NFDI will be brought upstream into the respective services to sustain the results.

The non-technical, organisational service offers, like training and support or networking, will be built reusing and strengthening existing, well-received channels, fora, and networks. Details can be found within the respective work package description

3 SWOT Analysis

The following table summarises an analysis of the strengths and weaknesses of the current PID-related service offers available to the NFDI consortia and reflects on the opportunities as well as potential threats related to the work plan and roadmap outlined in this proposal.

Internal	Strengths	Weaknesses
	<ol style="list-style-type: none"> 1. Service offers within the NFDI are based on already well-established and mature PID services with high TRL. 2. Consortia and research organisations within NFDI have already (technical) expertise related to the existing PID services. 3. Partners of PID4NFDI are involved in highly-visible and influential policy-making and infrastructure projects. 4. PID4NFDI partners already successfully supported PID service implementation in Germany. 5. PID4NFDI organisations are well connected within European and global PID service development and provider communities. 	<ol style="list-style-type: none"> 1. No consistent analysis of the PID-related requirements within the NFDI has been conducted. 2. Existing offers are based on different existing services with different infrastructures, standards and dependencies. 3. PID metadata is not formalised or standardised across services. 4. Tailored training and communication for the NFDI is missing. 5. No NFDI-wide governance exists to support projects and other working groups with their PID-related demand.

²⁸ <https://de.dariah.eu/>

²⁹ <https://straininfo.dsmz.de>

External	Opportunities <ol style="list-style-type: none"> 1. Services offer PID solutions for different use cases across different stages of the RDM lifecycle and promote interoperability across services and domains. 2. Service operates according to established international standards. 3. Training and support services help to raise awareness of the value of PIDs and metadata to operationalize FAIR principles and thus to unleash the full potential of PIDs. 	Threats <ol style="list-style-type: none"> 1. Specific requirements of NFDI consortia are not necessarily met by PID providers. 2. Lack of persistence of identifiers when responsibility for maintenance lies in short-term funded project contexts. 3. Lack of awareness for the value of PID metadata.
-----------------	--	---

Table 3: SWOT Analysis

PID4NFDI starts with experienced partners who already provide mature infrastructure services, some of which already interact with each other, and who collaborate with numerous institutions from the NFDI. However, the novel nature of the collaboration and the particular focus on explorative and highly specific research data management issues within the NFDI brings new challenges of governance, standardisation across disciplines, and their integration into different infrastructures, each with their own technical requirements. The desire for consolidation of infrastructures and the resulting synergy effects contrasts with the so far uncoordinated PID integration in the individual consortia. To counteract this fragmentation and to provide core concepts for PID integration is the central concern of PID4NFDI in the initialisation phase. The planned horizontal landscape analysis and vertical use case needs assessment (WP 1) form the starting point to better capture and develop the fragmented PID landscape in the NFDI. In order to be able to guarantee the persistence of identifiers, organisational measures are necessary (see also paragraph on "Risks and Challenges") in addition to the implementation of technical and methodological measures, i.e. the implementation of technical and metadata standards. For this purpose, we develop concepts that specifically address the needs of NFDI consortia and groups (see WP 2 and 4). A prerequisite for successful implementation of the concepts is the creation of awareness for the central building blocks of PID integration and efficient use. To this end, dedicated training and networking activities will be created (see WP 3 and 5).

4 Working Concept for the Development of the Basic Service

The envisaged future service offer of PID4NFDI after a potentially successful ramp-up phase encompasses the provision of the following technical and non-technical services: (i) development and operation of customized, technical PID-services for the NFDI based on existing baseline PID-services with high TRL, (ii) integration support for NFDI-consortia based on an evolving knowledge of the NFDI landscape, and (iii) communication and outreach portal within the NFDI as well as globally. To achieve this, the service initialisation phase sets the scene based on a community-driven process by providing landscape analyses, requirements engineering, detailed technical concepts, and business as well as license models. The integration phase builds upon these foundations and initiates the development tasks necessary to implement the customized technical services. Furthermore, first

integrations of baseline PID-services into the portfolios of individual consortia will be realized. Finally, the concept foresees a ramp-up phase that not only makes the technical services fit for operation, but also establishes integration support as well as communication on a professional level.

Service initialisation concept

The goal of this first phase is to develop an NFDI-wide PID strategy that captures the existing consortium- and community-focused solutions and requirements, and provides a blueprint for further development phases to roll out a portfolio of services that covers the various RDM use cases and the exceptional variety of requirements of the NFDI consortia. To achieve this, the concept foresees as first steps the analysis of the NFDI PID landscape and an in-depth review, collection, and evaluation of the requirements of all NFDI consortia, the sections, and other base services. These tasks will be conducted based on two already existing surveys and the results of an initial stakeholder workshop in close collaboration with the DFG-funded project PID Network Germany. The PID4NFDI service initialisation phase will be also used to identify and differentiate discipline-, research-, and resource-specific use cases for PIDs and to support NFDI consortia with their PID-related plans and demands. Technical, organisational, and community aspects of PIDs will be derived to support the integration and use of PIDs in service portfolios of NFDI consortia. Relevant technical and organisational concepts will be published as building blocks of the overall blueprint including governance and business models, training material, and a communications and outreach strategy. PID4NFDI plans to extensively take advantage of existing channels and networks of its partners to reach out to – in particular European – PID projects and networks to ensure connectivity and increase visibility.

Overall, the main outcome of the service initialisation phase is the provision of fundamental information to successfully implement the PID services concepts and solutions based on the consortia-driven requirements in the subsequent phases. All results will be delivered such that they can be also used in case PID4NFDI ends after the initial phase.

It is acknowledged that Base4NFDI will realise a community- and demand-driven process in close collaboration with the sections of the NFDI, which will also affect the PID4NFDI service offer. Service-specific development and decision processes will therefore be closely synchronised with the overall Base4NFDI process and will be adopted according to the evolving requirements. International developments and standards are taken into account and requirements/results from PID4NFDI are fed back to these PID communities.

Development and integration outlook

Based on the requirements and concepts of the service initialisation phase, PID4NFDI plans to support technical, organisational, and methodological implementations during the integration phase to meet the requirements of the consortia as well as the NFDI sections and related working groups. The aim is to enable integration of existing and future PID services in a functional NFDI architecture and highlight missing blocks in an iterative process. Planned service offerings of the partners and beyond will be reviewed periodically for their benefit to the NFDI community.

Wherever possible, PID4NFDI partners will promote the implementation of technical solutions as part of sustainable PID service infrastructures of institutions like DataCite, ePIC, or ORCID, and will ensure uptake and compatibility with developments of EOSC through projects like FREYA, FAIRCORE4EOSC or FAIR-IMPACT. Potential examples, which could become the home of a technical concept of PID4NFDI, include the Metadata Schema and Crosswalk Registry and the PID MetaResolver, which are developed by FAIRCORE4EOSC, as well as the PID Graph, developed by FREYA, or the Open Global Data Citation Corpus by DataCite, which can be used to improve provenance of research data. Other topics include the implementation of PIDs for specific resource types, the integration of community-specific, PID-related metadata, or the integration with services of individual consortia. Prioritisation and decisions on the actual implementation of the concepts and technical solutions will be made based on the results of the service initialisation phase. These developments will be complemented by continuous community involvement, requirements updates, and use case integration.

Further, complementing items on the roadmap for the integration phase, which cannot be detailed due to space limits, include (i) a concept to incentivise metadata curation, (ii) the development of use case specific metadata guidelines, and (iii) the evaluation of tools to maintain metadata quality. In addition, we will seek a dialogue with funding agencies and research institutions about financing PID registration and maintenance in a sustainable manner for NFDI resources. This also includes the costs for developing and using tools to improve PID integrations and for trained personnel to guarantee high level of metadata. These steps will be taken in co-operation with PID Network Germany.

Ramping up for Operation

In addition to scaling the technical PID services, which have been mainly developed in the previous phase, towards NFDI-wide operation, it is crucial to establish clear governance for cross-service synchronisation of PID services including aspects of payment, maintenance, and licensing. Therefore, the implementation of the respective concepts from the service initialisation phase is a key task of the ramp-up phase complemented by the development and publication of a NFDI PID Policy as common ground of PID use in research data management.

The website developed in the service initialisation phase will be expanded to a service-wide support structure. The aim is to build up a sustainable support infrastructure as a central point of contact for NFDI consortia: It will support the selection process of PIDs and provide access to relevant technical services like the PID type registration services, including information on service conditions and technical service descriptions. A helpdesk will also be established and trainings for managers of infrastructure services will be established. The overall goal is to professionalise the support of the NFDI consortia.

From the technical perspective, metadata quality tools will be added based on the evaluation from the previous phase. Furthermore, PID4NFDI will explore the need for additional APIs and/or integrations to facilitate cross-service and cross-consortia synchronisation and to enable the exchange of metadata.

The actual requirements to put the envisaged services into operation are difficult to quantify as this will depend largely on the on the use cases and the actually selected technical services and integrations as well as on the upstream integration of PID4NFDI services into the baseline services provided by ePIC, DataCite, et al. The long-term provision of the services, however, will most likely require primarily funding for specialised staff that can evolve the service development and provide the necessary expertise and support.

Risks and Challenges

To avoid dysfunctional or orphan PIDs, and ultimately ensure the availability of research resources, long-term maintenance of PID metadata must be ensured. For this purpose, the responsibilities for PID maintenance among the stakeholders of the NFDI network must be defined explicitly and embedded in the governance structure of the NFDI PID services beyond funding periods and contexts. Since PID assignment is a cross-cutting task, it requires a highly scalable service that can provide a cross-disciplinary technical infrastructure and metadata schema that enables the global resolution of PIDs, and already has a certain technology readiness level, to create the conditions for sustainable research data management. In addition, to ensure interoperability and secure provenance of data, a strategy must be devised to link discipline-specific metadata to generic schemas in a way that enables reuse by humans and machines.

Legal issues are relevant in the area of PIDs: It is necessary to clarify licensing options for metadata (e.g., free availability of metadata for patents, from publishers, etc.) and for the subsequent use of developed tools. A lack of awareness of the significance of PIDs and their metadata results in research data management not meeting FAIR criteria and superficial use of PIDs that do not realise their full potential (e.g., providing contextual information about the described object or related objects, provenance, citation information, etc.). Complete and consistent metadata is a prerequisite for a high-quality description of research outputs and resources. Meaningful use of metadata requires i.a. that metadata are mapped according to international standards or best practices and user requirements that interfaces between different systems are available, and that licences regulate their re-use.³⁰

The objectives and work packages described below address these risks and challenges, and the partners are able to manage them. However, the assignment and efficient use of PIDs requires the participation and engagement of all stakeholders in the NFDI. A close exchange with other actors within the NFDI is therefore a prerequisite for a functioning PID service.

³⁰ Burger et al. 2021; Strecker 2021.

5 Work Programme

Overview of work packages

Work package	Deliverables (D) and milestones (M)	Responsible partner
WP 1 - Analyse PID Landscape, Requirements of NFDI Consortia, and Services	D1.1 Landscape of PID practices within NFDI services D1.2 Requirement analysis of selected use cases M1 Advanced PID use cases identified and analysed within NFDI	Lead: GWDG Participation: DataCite, Helmholtz, TIB, Use Case Partners
WP 2 - Develop Concepts for Technical Implementation of PID Services and Metadata Interoperability	D2.1 Concept of mapping of use cases to existing PID services D2.2 Catalogue of relevant metadata standards M2.1 Concept for technical integration of PID infrastructure published M2.2 Concept for metadata interoperability and harmonisation published	Lead: GWDG Participation: DataCite
WP 3 - Support Service Providers within NFDI	D3 Cookbooks for simple PID registration in research workflows M3 Concept for training published	Lead: TIB Participation: GWDG
WP 4 - Establish PID-related Governance, Business Models and Licensing within the NFDI	D4 Overview and evaluation of relevant business, governance and licence models and their modalities M4 Concept for organisational integration of PID infrastructure (business, governance & licence model(s)) published	Lead: TIB Participation: GWDG
WP 5 - Outreach and Networking	D5.1 Communication strategy D5.2 Project website D5.3 Results of the stakeholder workshop M5.1 Project website launched M5.2 Stakeholder workshop held	Lead: Helmholtz Participation: TIB

Table 4: Overall work program with work packages, deliverables, milestones, and responsible partners

Detailed work programme

This paragraph describes the work packages listed in the table in detail and connects them to each other. The project management of PID4NFDI and its integration into the governance structure of Base4NFDI and the section Common Infrastructures are taken into account.

5.2.1 WP1 – Analyse PID Landscape, Requirements of NFDI Consortia, and Services

Task T1.1 Landscape and Analysis of PID Practices (Responsible: GWDG)

In T1.1, PID4NFDI will analyse how PIDs are currently used within the NFDI and evaluate how this can be optimised to support research workflows within NFDI consortia, while keeping a project- and discipline-specific focus. By using quantitative and qualitative survey methods in close collaboration with the PID Network Germany project, an overview of the type, scope, and use of PIDs in Germany with a special focus on NFDI should be gained. Based on our findings from our initial survey and workshop, we have already seen that many of the same challenges exist for certain types of resources, some of which are answered with individual solutions and repeatedly bind resources in

the individual consortia. In addition, there is a danger of producing isolated solutions here. The objectives of further analysis are 1) to identify common features of PID use cases across disciplinary boundaries, 2) to develop solutions that can be applied agnostically, but also take into account the specific requirements of the communities and research objects; and in this way to make the use of PIDs more efficient and to ensure interoperability between the various solutions.

Task T1.2 Requirements Analysis of Selected Use Cases (Responsible: GWDG)

T1.2 will, built on the results of T1.1 and the workshop organised by T5.3, provide an in-depth analysis of selected use cases from within the NFDI consortia (at least based on the requirements of the associated use case consortia partners of PID4NFDI) and provide a concrete grounding and feedback loop as a vertical perspective for the more general findings of T1.1 with a horizontal perspective. In detail, we want to find out which PID systems are applied to which resources and in what way. This deeper analysis should, for example, demonstrate whether the PIDs are already being used optimally in a specific application or this should be achieved; which stakeholders are involved in PID integration and usage and what are their needs in order to ensure findability, accessibility, interoperability and re-use; what expectations do these stakeholders have with regard to the added value of use (e.g. findability of the resources and their metadata in other databases); what technical, organisational or community-related restrictions, path dependencies and priorities exist; which aspects of PID integration and use could/have not been considered so far; how does the scalability of the existing solutions used/developed look like, etc.

5.2.2 WP2 – Develop Concepts for Technical Implementation of PID Services and Metadata Interoperability

WP2 will analyse which technical measures and metadata harmonisation processes are needed to improve the ease-of-use and integration of scalable PID services within the NFDI.

Task T2.1 Map Requirements to Service Capabilities (Responsible: GWDG, DataCite)

Taking the results of T1.1 and T1.2, this task will map the requirements to capabilities of existing PID infrastructures and services like those offered by organisations represented by PID4NFDI partners or ongoing projects like FAIRCORE4EOSC. Here, the options for use of services such as the PID meta-resolver (developed in FAIRCORE4EOSC) and the DataCite GraphQL API (PID Graph) (developed in the FREYA project) will be evaluated.³¹ T2.1 will put particular emphasis on the development of a concept that can be operationalised by PID service providers and will lead to services integrating NFDI-related requirements. T2.1 will also prioritise the different tasks in a way that the concept can serve as input to potential later phases of PID4NFDI. The aim is to allow existing PID provider services, existing consortia solutions, and identified requirement gaps to be harmonised and to create an ecosystem of PID integrations that supports research data reuse from NFDI and its impact evaluation (e.g. via citation analysis in PID Graph) on a large scale and across disciplines.

³¹ <https://faircore4eosc.eu/eosc-core-components>; Cousijn et al. (2021); <https://www.project-freya.eu/the-pid-graph.html>.

Task T2.2 Identify and Define Metadata Best Practices and Standards (Responsible: DataCite)

The heterogeneity of research assets used and produced by the NFDI consortia results in a variety of metadata requirements for the description of the respective assets and their referencing via PIDs. Based on the identified requirements in WP 1 and on an analysis of currently used metadata with selected use cases of NFDI consortia, T2.2 will identify the best practices and standards that align with the existing PID services infrastructure and international approaches for metadata harmonisation. The best practices will account for current metadata practices through an analysis of existing PID metadata (e.g., metadata in the DataCite Metadata Store) and seek to resolve gaps and inconsistencies. The task will initially focus on such use cases that have advanced PID registration workflows as well as sufficient maturity and coverage to act as blueprints for similar use cases within NFDI.

Task T2.3 Develop Concepts to Improve Metadata Interoperability and Quality (Responsible: DataCite, GWDG)

In order to achieve metadata interoperability, harmonisation of the metadata standards in use is necessary. We will develop a concept on how even the complex and specific metadata requirements of the NFDI consortia can be taken into account and at the same time be aligned with international standards to operationalize the FAIR principles. Here, the purposes and requirements of the metadata layers (local metadata, PID metadata, search engines/aggregators) need to be considered and seen how these different layers are aligned with each other. Metadata quality tools and dashboards will be used to improve the completeness and quality of the metadata as key drivers for data FAIRness. Both the DataCite GraphQL API and the DataCite Commons web interface can already be used to query research output (DOI), researcher (ORCID) affiliation (ROR) and repository (re3data) metadata and associated connections, and enable various metadata dashboards including usage metrics. The Data Type Registry (DTR)³², already in operation and a core service of FAIRCORE4EOSC, will assist in the standardisation and interoperability of PID metadata. The DTR allows to register basic as well as complex data types and offers APIs to validate single elements as well metadata schemas. In addition, there are other tools with different focuses and scopes for capturing the FAIRness of metadata.³³ The selection and application of these tools will serve as a basis for a conceptual approach to further develop incentives for metadata curation (quality improvement via metadata completeness) in the integration phase (see above: integration outlook). Here, we will work in close collaboration with the working group “Search & Harvesting” (applied Base4NFDI project: HaDES – Harvesting and Discovery Enhancing Service) to benefit from synergy effects in this joint endeavour.

³² <http://dtr-pit.pidconsortium.net/>

³³ E.g. FUJ-I (<https://www.f-uj-i.net/?action=test#>), FAIR Checker (<https://fair-checker.france-bioinformatique.fr/>), FAIR Evaluation Services (<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/#%2F!>), ARDC FAIR Self Assessment Tool (<https://ardc.edu.au/resource/fair-data-self-assessment-tool/>) etc. In November 2022, TIB, together with partners from HU Berlin and SUB Göttingen, submitted an application to the DFG to develop a metadata checker tool for institutional repositories (IRMA – Improve your Repository Metadata). In this context, a detailed analysis of the current FAIR Checker tool market took place.

5.2.3 WP3 – Support Service Providers in NFDI

Task T3.1 Develop a Concept for Training (Responsible: TIB)

Raising awareness of the importance of PIDs and supporting their implementation and management – taking into account domain and resource specific characteristics – is fundamental to research data management according to the FAIR principles. Based on the outcome of the landscape and requirement analysis of WP 1, we will develop a concept for training and guidance aimed at infrastructure managers and data stewards of NFDI services. Researchers are addressed with training formats focused on the added value of PIDs for their research. We will ensure that the content and formats of the future training material aligns and extends existing material from the ongoing support and adoption work of the project partners. The material will encompass the needs of the different development stages of the existing and emerging infrastructure services as well as their subject-specific demands. This will be done in coordination with the Section EduTrain, the work area of the section Common Infrastructure on materials and workshops/training, and PID Network Germany, whose guidance and training offers will be complemented by NFDI-specific perspectives and materials to promote a knowledge transfer into and out of NFDI.

Task T3.2 Create Cookbooks for PID Registration (Responsible: TIB)

In a second step and based on the training concept in T3.1, we will create low-threshold instructions (*cookbooks*) based on subject and resource specific use cases as well as best practice guidelines to disseminate common international standards and recommendations for PID registrations. The need for this offer has already been communicated to us by representatives of the consortia in the previous stakeholder workshop.³⁴ We will provide guidance and best practices to optimise PID registration processes (simple access) within the research workflows of the project partners PID infrastructures, while ensuring compliance with the FAIR principles. The aim is to support metadata interoperability with other services e.g. aggregators and addressing the importance of metadata standards and quality (addressed by WP 2). This task also supports the planned base service proposal Knowledge Graph Infrastructure (KGI), which will also (re)use PIDs. In the subsequent integration phase, cookbooks and training offers are developed with a focus on subject and resource specific requirements. The training opportunities e.g. webinars, workshops, training courses and material shall be communicated and provided on the PID4NFDI website, which will be built in WP 5, and via information offers of the section EduTrain. These training resources will be archived persistently in repositories such as TIB AV Portal,³⁵ twillo³⁶ and the Open Educational Resources platform being developed in the DALIA project in cooperation with the section EduTrain.

³⁴ Schrader et al. 2023, p. 12.

³⁵ <https://av.tib.eu/>

³⁶ <https://www.twillo.de/oer/web/>

5.2.4 WP4 – Establish PID-related Governance and Licencing within the NFDI

Task T4.1 Evaluation and Overview of Relevant Governance, Business and Licence Models
(Responsible: TIB, GWDG)

Different PID providers address various use cases and operate on the basis of different governance, business and licence models. The persistence of PIDs is strongly dependent on sustainable infrastructures and an actively engaged community that takes responsibility for the PID adoption and maintenance. To meet the PID needs of the various NFDI consortia use cases it is essential that the NFDI consortia participate in the existing governance of the PID providers shaping the strategic roadmap via the established communication channels and feedback processes. In WP4 we will analyse, document and provide an overview of the relevant governance, business and licence models of the various PID providers as well as the risks and associated mitigation strategies (POSI³⁷). We will focus on the models of the project partners and other frequently used PID services as identified in the landscape analysis (WP 1). Furthermore, we will provide a decision model that assesses the advantages, disadvantages and costs associated with the different approaches, and match these to use cases that cover the wide variety of applications of PIDs.

Task T4.2 Develop a Concept for Sustainable PID Registration Workflows
(Responsible: TIB, GWDG)

Based on the evaluation of the governance, business, and licence models (T4.1) of the various PID providers, we will develop a concept to adapt and apply these models according to the governance and funding structure of the NFDI. The aim is to identify organisational and funding opportunities and gaps to enable sustainably financed PID registration based on the landscape and requirement analysis (WP1).

5.2.5 WP5 – Outreach and Networking

Task T5.1 Develop Communication Strategy (Responsible: Helmholtz)

In order to raise awareness of the importance of PIDs, the need of interoperability, and the services of PID4NFDI basic service within NFDI and to ensure that the activities of PID4NFDI are complementary to international endeavours, such as of the EOSC PID strategy, a communication strategy will be developed. This includes an analysis of stakeholders, the establishment of suitable communication channels (such as mailing list, forum etc.) and the creation of dialogue platforms for the exchange with international research communities and PID providers. In this context we are eager to align with the undertakings of PID Network Germany and to join forces.

Task T5.2 Set up Project Website (Responsible: Helmholtz)

In order to enable community building and networking as well as knowledge transfer (see WP 3) we will set up a project website, which will be publicly available. Via the website we will inform about new developments in the PID landscape, providing information about existing PID standards, and a

³⁷ <https://openscholarlyinfrastructure.org/>

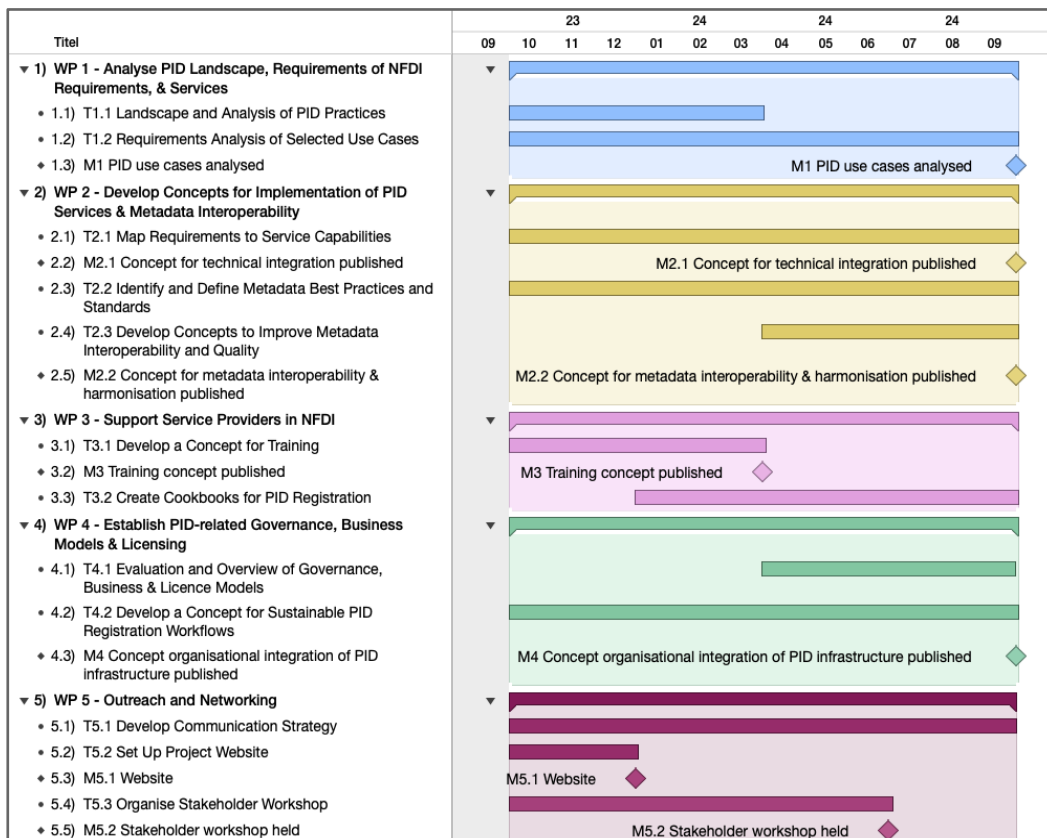
collection of reports on best practices on the use and implementation of PIDs in NFDI consortia. We plan to set up an independent website in the environment of the NFDI e.V., which will have references on the dialogue platform of PID Network Germany³⁸ and vice versa, amongst others.

Task T5.3 Organise Stakeholder Workshop (Responsible: Helmholtz)

We plan to organise a one-day hybrid workshop with representatives of NFDI Consortia, NFDI sections & (inter)national stakeholders and projects at the end of the first phase of PID4NFDI. On the one hand the workshop will provide an overview of the results of the first phase of PID4NFDI (What was achieved? What has been done?), and on the other hand, the workshop will be a platform to discuss further steps and help to identify further needs and challenges the NFDI consortia are confronted with in the context of the implementation of PIDs to build a bridge to the subsequent integration phase.

The workshop will take place in Berlin or Potsdam at a location of the Helmholtz Association or jointly with a larger conference. We plan to conduct a hybrid workshop with around 50 participants on site.

Gantt Chart



³⁸ Bertelmann, R., et al (2023), p 15.

Bibliography and list of references

- Bertelmann, R.**, et al. (2023). PID Network Deutschland. Netzwerk für die Förderung von persistenten Identifikatoren in Wissenschaft und Kultur. Helmholtz Open Science Office. DOI: <https://doi.org/10.48440/os.helmholtz.059>.
- Bingert, S., Brase, J., Burger, F., Dreyer, B., Hagemann-Wilholt, S., Vierkant, P., and Wieder, P.** (2022). Concept for Setting up the Persistent Identifier Services Working Group in the NFDI Section "Common Infrastructures" (1.0). Zenodo. DOI: <https://doi.org/10.5281/zenodo.6507760>.
- Burger, M. Cordts, A.**, und Habermann, T. (2021). Wie FAIR sind unsere Metadaten?: Eine Analyse der Metadaten in den Repositorien des TIB-DOI-Services, in: Bausteine Forschungsdatenmanagement, 3, S. 1-13. DOI: <https://doi.org/10.17192/bfdm.2021.3.8351>.
- Cousijn, H.**, Braukmann, R., Fenner, M., Ferguson, C., van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected Research: The Potential of the PID Graph. Patterns, 2(1), 1–7. <https://doi.org/10.1016/j.patter.2020.100180>.
- European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R., et al. (2020). A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC), Publications Office. DOI: <https://data.europa.eu/doi/10.2777/926037>.
- European Commission, Directorate-General for Research and Innovation, **Schwardmann, U.**, Fenner, M., Hellström, M., et al. (2021). PID architecture for the EOSC: report from the EOSC Executive Board Working Group (WG) Architecture PID Task Force (TF), Publications Office. DOI: <https://data.europa.eu/doi/10.2777/525581>.
- Ewert, Frank, Arend, Daniel, Asseng, Senthold, Boehm, Franziska, Feike, Til, Fluck, Juliane, Gackstetter, David, Gonzalez-Mellado, Aida, Hartmann, Thomas, Haunert, Jan-Henrik, Hoedt, Florian, Hoffmann, Carsten, König, Patrick, Lange, Matthias, Lesch, Stephan, Lindstädt, Birte, Lischeid, Gunnar, Martini, Daniel, Möller, Markus, ... Weiland, Claus. (2023). FAIRagro FAIR Data Infrastructure for Agrosystems - Proposal 2021 (1.1). Zenodo. <https://doi.org/10.5281/zenodo.7528172>.
- Ferguson, Ch., McEntrye, J., Bunakov, V., Lambert, S., van der Sandt, S., Kotarski, R.I, Stewart, S., MacEwan, A., Fenner, M., Cruse, P., van Horik, R., Dohna, T., Koop-Jacobsen, K., Schindler, U., and McCafferty, S. (2018). D3.1 Survey of Current PID Services Landscape (Version 1). Zenodo. DOI: <https://doi.org/10.5281/zenodo.1324296>.
- Hagemann-Wilholt, S.** (2023). PID4NFDI: Survey on PID Practices. Main results. With contributions from the members of the NFDI working group Persistent Identifier Services. Zenodo. DOI: <https://doi.org/10.5281/zenodo.7635791>.
- Klas, C.-P., Zloch, M., **Bach, J. S.**, Baran, E., and **Mutschke, P.** (2022). KonsortSWD Measure 5.1: PID Service for variables report (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6397367>.
- Schrader, A. C., Arend, D., Bach, J. S., Elger, K., Göller, S., Hagemann-Wilholt, S., Krahl, R., Lange, M., Linke, D., Mayer, D., Mutschke, P., Reimer, L., Scheidgen, M., Selzer, M., and Wieder, P. (2023). Workshop on PIDs within NFDI (Version 1). Internal Workshop on PIDs within NFDI, Online. Zenodo. DOI: <https://doi.org/10.5281/zenodo.7635905>.
- Stocker, M.**, Darroch, L, Krahl, R. Habermann, T., Devajaru, A., **Schwardmann, U.**, D'Onofrio, C., and Häggström, I. (2020). Persistent Identification of Instruments. Data Science Journal, 19(1). DOI: <http://doi.org/10.5334/dsj-2020-018>.
- Strecker, D. (2021). Quantitative assessment of metadata collections of research data repositories. Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft. , 470, <https://doi.org/10.18452/22916>.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>.