

There will be FAIR. Insights into the PID landscape analysis

Authors:

Sara El-Gebali^{1*}, Stephanie Hagemann-Wilholt², Torsten Kahlert² and Marc Lange³

*Lead presenter

¹sara.elgebal@datacite.org, Datacite

²TIB Hannover

³Helmholtz Open Science Office

Abstract:

This presentation will demonstrate how a strategic, PID-based framework can transform research data management into FAIR, paving the way for enhancements in the reliability and accessibility of scientific research.

One of the main goals of the basic service PID4NFDI is gaining insights from extensive surveys to optimize PID application across varied research outputs, addressing current challenges within NFDI services according to diverse requirements and different maturity levels of PID usage. We will also explore how PIDs do more than ensure persistence and identification; they significantly enhance metadata quality and research integrity.

By integrating Persistent Identifiers (PIDs) with diverse research elements – such as instruments, methodologies, datasets, DMPs and publications – we enhance both the replication of experiments and the reuse of research data. This connectivity enables researchers to access all necessary components for comprehensive studies, thereby increasing the reliability and utility of research findings. Additionally, by making the provenance and production/creation process of diverse research outputs more transparent and by providing accessible licensing information through PIDs, we foster a transparent research environment that maintains high standards of integrity and promotes open science. In addition, PIDs offer context information by providing metadata on protocols and methods, software and input/output datasets, creators and contributors including affiliation organizations. They enable us to trace other related publications e.g. citations and references where an instrument or physical sample is used. Ultimately, leading to reduced duplicate efforts, and allows higher re-use of research data, making it more transparent and impactful.

PID4NFDI is committed to providing clear guidance on choosing and implementing the right PIDs for different types of research outputs. We are developing a comprehensive matrix that identifies suitable PID services for various research entities at different stages of the data lifecycle. This information will help researchers identify the most effective PIDs for their needs. The proposed solution/outcome is in direct response to the challenges identified by NFDI service providers, as highlighted in our survey responses and further feedback from consortia.

We will present the results of our landscape analysis and the resulting outcomes to guide subsequent implementation phases of a PID4NFDI basic service. Additionally we will highlight preliminary results from our cooperation with selected use case partners, where we examine use case-specific challenges, interoperability issues with existing PID infrastructures, and metadata quality insights. Comprehensive, open and linked metadata of research resources and entities is crucial for an ecosystem of well-functioning PIDs offering the above mentioned benefits to support academic collaboration through data sharing and trustworthiness of scientific processes. However, these can vary greatly depending on the discipline, subject of research and methods, and accordingly require different approaches to documentation and referencing.

At the end of the first project phase, a first collection of best practices should be available that can be reused by the consortia for typical use cases in NFDI service infrastructures. Referring

to our evolving training and support concept, we will outline in our presentation how the PIDs4NFDI basic service can support consortia in the future.

Keywords: Persistent Identifier, Basic Service, Metadata, Interoperability