

Isn't a Number and a URL Enough?

Why PIDs Matter and Technical Solutions Alone are not Sufficient.

Stephanie Hagemann-Wilholt¹[\[https://orcid.org/0000-0002-0474-2410\]](https://orcid.org/0000-0002-0474-2410), Antonia Schrader²[\[https://orcid.org/0000-0001-7080-634X\]](https://orcid.org/0000-0001-7080-634X), and Andreas Czerniak³[\[https://orcid.org/0000-0003-3883-4169\]](https://orcid.org/0000-0003-3883-4169)

¹ Technische Informationsbibliothek (TIB), Germany

² Helmholtz Association, Germany

³ Bielefeld University Library, Germany

Abstract. Persistent identifiers (PIDs) are an integral part of research data management and can be found throughout the entire lifecycle of research data. However, their ability to function – to ensure persistence – depends on numerous factors: technical infrastructure, international standards and best practices and their dissemination, agreements on long-term governance of infrastructures, etc. Their applicability is diverse and requires adaptation to the resources and entities referenced by them. The paper describes two projects – PID4NFDI and PID Network Germany – that aim to address these challenges.

Keywords: Persistent Identifier, Research Data Management, Metadata, Base Services

1. FAIR Research Data Management – are PIDs an Additional Extra or Mandatory?

Imagine you get a dataset from a colleague to reuse for your own research. You know the name of the data producer, the data set has a working title and a versioning. The data producer also gives you some contextual information that is not in the dataset itself. This is all very helpful. But how do you cite this dataset? What if others would also like to find and reuse this dataset? Wouldn't it be nice to make the contextual information persistently available? These questions are already addressed by the endeavours of research data management (RDM), and its central efforts are to make data and its metadata findable, accessible, interoperable and re-usable. The backbone of all this effort is Persistent Identifiers (PIDs). The FAIR principles ask for the assignment of PIDs, the description of data with rich, standardised and machine-readable metadata and its long-term availability, referencing to other resources via PIDs, licensing and provenance information [1].

Therefore, PIDs have a significant impact on the entire life cycle of research data (RD) from the application for funding to the re-use of data and offer numerous advantages. They enable unique identification of research resources such as data, publications, or code; of persons, institutions, projects, grants etc [2]. PIDs support publication processes as well as discoverability and citability of research results. As widely accepted means to identify and link research entities, they are embedded in already established globally operating infrastructures and existing communities, which provide standards and best practices as a prerequisite for the harmonisation and thus the interoperability of RD metadata.

PIDs are an integral part of RDM services such as research data repositories, research information systems or knowledge graphs. A key aspect is that PIDs are established at the global level. They reduce administrative overhead, and the error rate of metadata ingestions and updates. PID metadata standards allow the linking of research entities and increase the global visibility and reusability of research resources as well as their accountability. Therefore, they are an important contribution to the quality assurance of research data and contribute to the reproducibility and reusability of research results.

Type and scope of entities described with PIDs in the context of RDM are continuously growing and gaining more and more maturity, but the landscape of actors, services, infrastructures and use cases is scattered. To leverage synergies within the existing structures, *PID4NFDI* and *PID Network Germany* aim to capture and analyse this landscape and to identify ways to consolidate it, to build and intensify networks and to establish international best practice.

2. PID4NFDI, PID Network Germany and their key objectives

PID4NFDI applies for funding within Base4NFDI to build a NFDI-wide PID service in line with the strategic goals of NFDI. It was initiated by the PID Working Group [3] in the section Common Infrastructures. DataCite, the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG), the Helmholtz Open Science Office, and the German National Library of Science and Technology (TIB) are partners in the proposed project. Discipline-specific NFDI members will support the project as use case partners.

The project focuses on enhancing and optimising PID use in NFDI infrastructure services as a basic element of FAIR data management. To this end, technical, organisational, metadata, training and network aspects are tackled to serve the needs of existing and emerging services and communities as well as the NFDI as a whole. The aim is to harmonise requirements, to promote the integration of PIDs and to develop individual as well as central solutions and support for identified gaps. *PID4NFDI* aims to address these different use cases that are highly dependent on the type of research and nature of the data produced and used. Different research fields and methodologies each have their own requirements for the uses and benefits of PIDs: Which entities should be described? How granular are they? How long should they actually be available? Does the data have an impact on other research?

In the application phase, we already identified initial needs and requirements for an NFDI-wide PID infrastructure via a survey [4] and a stakeholder workshop [5]. Among other things, needs for an overview of PIDs and their scopes, the integration of subject-specific metadata, PIDs for entities such as instruments, ontologies, code, physical objects, projects, recommendations for dealing with ephemeral resources, granularity etc. have emerged.

PID Network Germany [6] is a DFG-funded project by DataCite, the German National Library (DNB), the Helmholtz Open Science Office, the German National Library of Science and Technology (TIB) and the Bielefeld University Library. The project started in March 2023 and 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 will also be on identifying needs and optimization potential for existing PID systems and on embedding them in international knowledge graphs. The project findings will lead to recommendations in the order to create a national PID roadmap for Germany.

The endeavour of *PID4NFDI* is in line with *PID Network Germany* and will make the German scientific landscape more robust, transparent and accessible. *PID Network Germany* pursues an overarching approach across the boundaries of individual scientific institutions and research disciplines. In this context, *PID4NFDI* represents an important building block for addressing the specific requirements of the NFDI. Constant and intensive exchange between the

projects will ensure that synergies are used and strengths are bundled, and that the activities are carried out in a target-oriented and at the same time target-group-specific manner.

Author contributions

The listed authors have prepared and written this extended abstract (role: Writing – original draft according to [CReDiT guidelines](#), Contributor Roles Taxonomy).

Competing interests

The authors declare that there are no competing interests.

Funding

PID Network Germany is funded by the German Research Foundation (DFG).

Acknowledgement

Thanks to Britta Dreyer, Lars G. Svensson and the members of the PID Working Group for helpful comments on the initial draft of this abstract.

References

- [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] Ferguson, Christine, McEntrye, Jo, Bunakov, Vasily, Lambert, Simon, Sandt, Stephanie van der, Kotarski, Rachael, Stewart, Sarah, MacEwan, Andrew, Fenner, Martin, Cruse, Patricia, Horik, René van, Dohna, Tina, Koop-Jacobsen, Ketil, Schindler, Uwe, & McCafferty, Siobhan. (2018). D3.1 Survey of Current PID Services Landscape (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.1324296>.
- [3] Bingert, Sven, Brase, Jan, Burger, Felix, Dreyer, Britta, Hagemann-Wilholt, Stephanie, Vierkant, Paul, & Wieder, Philipp. (2022). Concept for Setting up the Persistent Identifier Services Working Group in the NFDI Section "Common Infrastructures" (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6507760>.
- [4] Hagemann-Wilholt, Stephanie. (2023, February 13). PID4NFDI: Survey on PID Practices. Main results. Zenodo. <https://doi.org/10.5281/zenodo.7635791>.
- [5] Schrader, Antonia C., Arend, Daniel, Bach, Janete, Elger, Kirsten, Göller, Sandra, Hagemann-Wilholt, Stephanie, Krahl, Rolf, Lange, Matthias, Linke, David, Mayer, Desiree, Mutschke, Peter, Reimer, Lorenz, Scheidgen, Markus, Selzer, Michael, & Wieder, Philipp. (2023). Workshop on PIDs within NFDI (Version 1). Internal Workshop on PIDs within NFDI, Online. Zenodo. <https://doi.org/10.5281/zenodo.7635905>.
- [6] Bertelmann, R., Buys, M., Kett, J., Pampel, H., Pieper, D., Scholze, F., Sens, I., Burger, F., Dreyer, B., Glagla-Dietz, S., Hagemann-Wilholt, S., Hartmann, S., Schrader, A. C., Schirrwagen, J., Summann, F., Vierkant, P. (2023): PID Network Deutschland. Netzwerk für die Förderung von persistenten Identifikatoren in Wissenschaft und Kultur, Potsdam : Helmholtz Open Science Office. <https://doi.org/10.48440/os.helmholtz.059>.