



# Use Case Analysis

**KonsortSWD Persistent Identifier (PID)  
Registration Service for Dataset Elements**

**Jana Böhm and Janete S. Bach**

20 December 2024

## Imprint

<b>Authored by</b>	Jana Böhm ( <a href="https://orcid.org/0009-0004-9802-113X">https://orcid.org/0009-0004-9802-113X</a> ) Janete S. Bach ( <a href="https://orcid.org/0000-0001-9011-5837">https://orcid.org/0000-0001-9011-5837</a> )
<b>Contributions by</b>	Peter Mutschke ( <a href="https://orcid.org/0000-0003-3517-8071">https://orcid.org/0000-0003-3517-8071</a> )
<b>Reviewed by</b>	Stephanie Hagemann-Wilholdt ( <a href="https://orcid.org/0000-0002-0474-2410">https://orcid.org/0000-0002-0474-2410</a> )  Torsten Kahlert ( <a href="https://orcid.org/0009-0003-3264-5006">https://orcid.org/0009-0003-3264-5006</a> ) Marc Lange ( <a href="https://orcid.org/0000-0002-7742-3867">https://orcid.org/0000-0002-7742-3867</a> ) Jana Tatscheck ( <a href="https://orcid.org/0009-0002-1285-2027">https://orcid.org/0009-0002-1285-2027</a> )
<b>Date of publication</b>	20 December 2024
<b>DOI</b>	<a href="https://doi.org/10.5281/zenodo.14327770">https://doi.org/10.5281/zenodo.14327770</a>
<b>License</b>	This work is licensed under <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
<b>Deliverable</b>	Part of D1.2 and D2.1

### About

PID4NFDI (<https://base4nfdi.de/projects/pid4nfdi>) is the basic service for persistent identifiers in development for the German National Research Data Infrastructure (NFDI). PID4NFDI is part of and funded through Base4NFDI.

Funded by DFG as part of NFDI. DFG Grant Number: 521453681

## Contents

<b>Background</b>	<b>1</b>
Criteria for Use Case Selection	1
Knowledge Base for Use Case Analysis	2
<b>The PID Registration Service</b>	<b>3</b>
Motivation: PIDs for Dataset Elements	4
Registration Workflow and Software	6
PID Maintenance Workflow and Responsibilities	7
Number of PIDs	7
Metadata	8
Schema and Alignment	8
Validation and Metadata Quality	9
Enhancing Interoperability	9
PID Provider Considerations	9
Support & Training	10
Licensing	10
Cost Model	10
PID Policy and Governance	11
Outreach	11
Wishlist for PID4NFDI	12
<b>Conclusion</b>	<b>13</b>
<b>References</b>	<b>14</b>

## Background

The initial phase of the PID4NFDI project is dedicated to establishing a comprehensive understanding of the persistent identifier (PID) landscape within the German National Research Data Infrastructure (NFDI). This foundational phase encompasses use case analysis, requirements engineering, and concept development. These activities are not only designed to address immediate project requirements but also to ensure that the insights gained will facilitate seamless integration and ongoing development in subsequent project phases.

As part of our deliverables for the initialization phase work packages 1 and 2, this use case analysis serves several key purposes:

- **Work package 1 (WP1):** This encompasses **D1.1**, which focuses on exploring the landscape of PID practices across different NFDI services, and **D1.2**, which involves a requirements analysis of selected use cases to derive practical and relevant insights. These analyses set the stage for a broader understanding of how PIDs are implemented and managed across diverse research contexts within NFDI.
- **Work package 2 (WP2):** Within WP2, deliverable **D2.1** aims to develop a conceptual framework for mapping selected use cases to existing PID services. The goal here is to ensure that our understanding of use cases translates into actionable strategies for effectively integrating PID services across NFDI consortia.

## Criteria for Use Case Selection

The selection of use cases was a crucial step in ensuring a diverse and representative analysis. We established the following criteria for choosing which use cases to examine:

1. **Diversity in duration of operation and PID service providers:** We aimed to reflect a range of maturities, encompassing both well-established and newer initiatives. This variety helps us understand PID usage at different stages of project development and their engagement with different stakeholders (i.e. PID providers).
2. **Disciplinary breadth:** The use cases needed to span multiple disciplines to assess the adaptability and versatility of PID adoption across different scientific fields and sectors.

3. **Active engagement:** The use cases had to demonstrate active contributions and engagement.

## Knowledge Base for Use Case Analysis

This use case analysis is primarily based on an extensive survey designed by the PID4NFDI project team, which was distributed in March 2024 and completed by two KonsortSWD<sup>1</sup> representatives with responsibilities related to PID management. In addition, a guideline-based interview was conducted in October 2024.

The survey yielded valuable insights into the experiences, challenges, and potential areas for improvement related to PID management within the KonsortSWD consortium. This allows for a comparison of the responses from KonsortSWD with those from other NFDI consortia, thus enabling an assessment of whether experiences and challenges are commonly shared between consortia or if they are specific to particular use cases.

The interview provided an opportunity to gain a deeper understanding of the technical aspects of the PID registration service, including system structure, PID registration workflows, PID maintenance, metadata schemas, and validation. Additionally, we inquired about expectations and wishes regarding the PID4NFDI basic service in the long term.

---

<sup>1</sup> KonsortSWD provides research infrastructures for the social, behavioral, educational, and social sciences. Webpage: <https://www.konsortswd.de/>.

## The PID Registration Service

This chapter presents the PID registration service<sup>2</sup> for dataset elements (in short, PID registration service), which is currently being developed by KonsortSWD. The aim is to describe the integration of PIDs within this service as well as identify requirements and challenges.

The PID registration service was selected as a use case based on the following reasons:

1. **PID service in development:** The PID registration service is currently in the final stage of development, with a fully functional prototype already published and operational for testing (a productive system is planned for KonsortSWD's second funding phase). This offers PID4NFDI the chance to gain insight into the processes of service development and PID integration.
2. **PIDs for dataset elements within a parent resource:** PIDs are assigned at a lower level of granularity for dataset elements within parent resources. This could be exemplified by variables within a dataset, a subset of questions within a survey, or fragments of a video. The use case also encompasses both the Digital Object Identifier (DOI) standard and the Handle standard, as the parent resources are frequently referenced by DOI, whereas PIDs for dataset elements within the parent resource are based on the Handle system (in particular, ePIC<sup>3</sup> PIDs are used).
3. **Confirmed Engagement and Contributions:** Representatives of KonsortSWD<sup>4</sup>, particularly those engaged in Task 5 Measure 1 (PIDs), have demonstrated a high level of commitment to the project, including the following contributions:
  - Contributions to the PID Working Group concept of the NFDI Section Common Infrastructures<sup>5</sup> [1]
  - Contributions to the PID4NFDI basic service proposal [2]

---

<sup>2</sup> The service is available at <https://labs.da-ra.de/nfdi/>.

<sup>3</sup> European Persistent Identifier Consortium, <https://www.pidconsortium.net/>.

<sup>4</sup> Janete S. Bach (<https://orcid.org/0000-0001-9011-5837>) and Peter Mutschke (<https://orcid.org/0000-0003-3517-8071>).

<sup>5</sup> In the NFDI Section Common Infrastructures (<https://www.nfdi.de/section-infra/?lang=en>), representatives of NFDI work together to identify components for common use and to work out proposals on how these can be technically structured and implemented.

- Contributions to the analysis of a first survey on PID practices within NFDI conducted in 2023 by the PID Working Group [3]
- Presentation at the community workshop on PIDs within NFDI, promoted by the PID Working Group [4]
- Contributing as a PID4NFDI use case partner for the emerging PIDs of dataset elements in KonsortSWD, also for PIDs needs in NFDI4DS, and BERD@NFDI
  - i. Providing information to this use case analysis through survey [5], interview, and potential follow-up meetings
  - ii. Presentation at the PID4NFDI talk [6]
  - iii. Presentation at the PID4NFDI Stakeholder Workshop [7]

## Motivation: PIDs for Dataset Elements

PIDs are typically associated with specific resources, including datasets, studies, publications, and numerous others. When researchers work with resources, they usually refer to specific elements within the resources they are utilising. To illustrate, they may wish to cite specific variables within a dataset, a subset of questions and answers within a survey, fragments of video or audio material, image areas, or text excerpts [8]. At present, the only citation that can be made is that of the entire resource. To clarify which of the elements researchers refer to within the resource, they include textual descriptions in their research, indicating the variable, survey question, or video fragment they have utilized. However, such descriptions are often semantically ambiguous. For example, there may be multiple variables with the same name when working with several data files or several waves of a study [9]. Further difficulties may arise when datasets are particularly large, containing, for instance, several hundred questions and variables [10]. As a result, it may be challenging to reproduce the specific elements that have been used, potentially jeopardizing the FAIR principles [11].

KonsortSWD is developing a technical solution to reference elements below the dataset, study, or publication level. Many studies, particularly in the social sciences, are conducted as surveys repeated across multiple waves at different points in time. These waves often capture the same variables to allow for longitudinal analysis. For this reason, a single study in the social sciences may consist of multiple datasets, each of which might include between one and many

files. For example, a dataset can be a collection or mix of text files, audiovisual files, and rectangular data (data in columns and rows format). Each file within a dataset is composed of smaller granular elements, such as text excerpts in a text file, audio/video segments, or variables in rectangular data. The term *parent resource* might refer to either a dataset or a file, depending on the context. When the parent resource refers to the dataset, all resources registered at a lower granularity level under the dataset are called dataset elements (files and elements within files). Conversely, if the parent resource is a file, then all lower granular elements within the file are the dataset elements. The parent resource is typically identified via a PID assigned in advance (but not necessarily). Additional PIDs can then be assigned at a lower granular level within the parent resource. Each element PID resolves to a landing page which contains metadata describing the element.

The PIDs of dataset elements are based on ePIC Handles provided by GWDG<sup>6</sup>. The prefix associated with the PID registration service is 21.T11998, a test prefix which is temporary and will compound the test PID syntax assigned during the test phase. A fully functional PID registration service prototype will be provided until September 2025.

Key features of the developed service are:

- **FAIRness:** Such a PID registration service would enable the sustainable identification of dataset elements and thus ease reliable citations, unambiguous descriptions, and improve the reusability of dataset elements. Additionally, links between variables, for example, across different waves of a study, can be stored in a standardized way and across institutions [10]. To this end, the initial motivation and demand for the PID registration service was to better comply with the FAIR principles.
- **Scalability:** The registration of PIDs for elements within resources entails the potential registration of a significant number of PIDs in relatively short time frames. Consequently, the infrastructure must handle a substantial volume of requests in an efficient and timely manner (asynchronous bulk registration organized in a queue) [12]. Additionally, the registration of numerous PIDs could result in considerable costs, particularly if the cost model of PID providers scales with the number of PIDs being registered [9].

---

<sup>6</sup> Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, <https://gwdg.de/>.



While initially focused on KonsortSWD's needs, the system offers a scalable and adaptable framework that can be interconnected with other NFDI consortia, such as NFDI4DS<sup>7</sup> and BERD@NFDI<sup>8</sup>. The PID system can also be integrated with data repositories managed by KonsortSWD partner institutions such as SOEP-DIW<sup>9</sup>, DZHW<sup>10</sup>, University of Bremen<sup>11</sup>, and GESIS<sup>12</sup>. The service will be in line with the EOSC<sup>13</sup> PID Policy and has already been assessed using the Compliance Assessment Toolkit (CAT)<sup>14</sup>.

## Registration Workflow and Software

We provide a high-level description of the software and the registration workflow for PIDs. The service is based on open-source software. Detailed technical information can be found in [9], [12], and the documentation<sup>15</sup>.

The PID registration service is not a repository solution in itself. Rather, it is a pure PID registration service and pass-through system for metadata information. This means that the parent resource and the dataset elements can be located in any repository and do not need to be uploaded to the PID registration service in order to receive PIDs for elements. However, metadata information describing the elements must be submitted to the service. In particular, the workflow to receive PIDs for dataset elements via the service is outlined below.

Users registering PIDs are research data centers (RDCs) or, in general, any data provider on organizational level (not single researchers). The user needs to have a parent resource (e.g., dataset) available which can be located in any repository. To receive PIDs for elements (e.g., variables) within the parent resource, the user must first authenticate via login at the PID registration service. Then he needs to choose the correct element type and fill in the required metadata fields for the element by

---

<sup>7</sup> NFDI4DS is the consortium for data science and artificial intelligence, <https://www.nfdi4datascience.de/>.

<sup>8</sup> Berd@NFDI is the consortium for business, economic, and related data, <https://www.berd-nfdi.de/>.

<sup>9</sup> The Socio-Economic Panel Study (SOEP) is one of the largest and longest-running multidisciplinary panel studies in the world, for which around 30,000 people in almost 22,000 households are currently surveyed each year. Webpage: [https://www.diw.de/de/diw\\_01.c.615551.de/forschungsbasierte\\_infrastruktureinrichtung\\_sozio-oekonomisches\\_panel\\_soep.html](https://www.diw.de/de/diw_01.c.615551.de/forschungsbasierte_infrastruktureinrichtung_sozio-oekonomisches_panel_soep.html).

<sup>10</sup> Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, <https://www.dzhw.eu/>.

<sup>11</sup> University of Bremen, <https://www.uni-bremen.de/>.

<sup>12</sup> Leibniz Institut für Sozialwissenschaften, <https://www.gesis.org/home>.

<sup>13</sup> European Open Science Cloud, <https://eosc.eu/>.

<sup>14</sup> Compliance Assessment Toolkit, <https://eosc.eu/roadmap/eosc-compliance-assessment-toolkit/>.

<sup>15</sup> API Documentation: <https://labs.da-ra.de/nfdi/api/swagger-ui/index.html>.

providing a JSON file and sending that information to the system. The metadata are extracted and validated against the corresponding schema. For each element PID to be registered, a job is scheduled and jobs are processed based on a registration queue. Hence, PIDs are registered asynchronously, which enhances scalability and enables bulk registration [9]. The registration process uses the ePIC API. The service offers a web-based user interface for monitoring registration jobs and statuses, and a REST API to carry out the PID registration process.

## PID Maintenance Workflow and Responsibilities

If some content information about the element changes (for example, the URL which the PID resolves to, or the metadata), the element PID needs to be updated. An update can only be performed by the owner of a PID. As the RDCs log into the PID registration service, PIDs are created by the PID registration service on behalf of the RDCs. The ownership of the PID therefore lies with the PID registration service. However, all the content information is provided by the RDCs. They have to ensure the availability of the landing page and correct metadata. The process of updating a PID, therefore, has to be initiated by the RDC, and updates must pass through the PID registration service.

The workflow for updating a PID is as follows: The RDC submits a new JSON file containing the updated metadata to the PID registration service. To ensure that no new PID is generated, the *PID proposal* field in the metadata must contain the PID that previously identified the dataset element. The update process follows the same steps as the PID registration workflow.

## Number of PIDs

In the first year of the PID registration service's operation, a minimum of one million PIDs will need to be minted to fulfil the requirements of KonsortSWD's use case partners. To this end, some initial datasets have already been defined. For some resources, estimating the number of PIDs that need to be registered is straightforward. For example, in rectangular or tabular data, the number of variables is easy to count. For this resource type, GESIS itself currently requires 500,000 PIDs for variables. This number may also increase until the system becomes productive. SOAP-DIW requires 100,000 variables, and DZHW requires 30,000 variables [14]. The number of PIDs required for qualitative data cannot be fully estimated in advance. For example, the number of PIDs that are registered for video fragments depends on how many specific fragments are actually required by

the user. There are metadata fields to include the start and end point (in minutes and/or seconds) to inform the video fragments.

As a proof-of-concept, the service has already been successfully applied to register variables during the current test phase. Hereby, the use case partners registered PIDs under a test prefix for a smaller amount of variables than actually required in the long run. However, during this phase, the goal was to establish the method and extract the metadata. The amount of registered PIDs was not crucial for this purpose.

After the first productive year, the targeted number of PIDs to be registered will depend on the future needs of the use case partners.

## Metadata

### Schema and Alignment

Together with the PID registration service for dataset elements, a set of metadata schemas is being developed that aligns with some commonly applied metadata standards, for example, the DataCite Metadata Schema<sup>16</sup>, the Data Documentation Initiative (DDI)<sup>17</sup>, and schema.org<sup>18</sup>. The PID service has also aligned its metadata schemas with CESSDA<sup>19</sup> to enhance cross-archive interoperability. One of the PID service outcomes is a controlled vocabulary published on the CESSDA website, the *GESIS controlled vocabulary for variables relations for social sciences research data* [13]. The schemas for variables can be found in [14]. Schemas will be further updated and extended with additional dataset element types of different granularity levels, such as files within a dataset, text excerpts, video/audio segments, and image areas.

A suggestion from PID4NFDI is to register the metadata schemas in the Data Type Registry (DTR)<sup>20</sup>. An advantage of this approach is that the system is machine-actionable, with the ability to define types and schemas, and then validate content against the schema. Furthermore, an API can be used to retrieve the schemas in JSON format.

---

<sup>16</sup> DataCite Metadata Schema, <https://schema.datacite.org/>.

<sup>17</sup> Data Documentation Initiative, <https://ddialliance.org/Specification/>.

<sup>18</sup> Schema.org, <https://schema.org/>.

<sup>19</sup> Consortium of European Social Science Data Archives, <https://www.cessda.eu/>.

<sup>20</sup> The Data Type Registry, available at <https://typeregistry.lab.pidconsortium.net/>, is a type definition system that allows users to create their own schema based on a combination of subtypes. The system is hierarchical. Users would start from the basics, for example by defining that an author is a string, and add more and more levels to provide greater complexity.

## Validation and Metadata Quality

The RDCs have to choose the element type (e.g. variables, text excerpts) they want to register the PIDs for. Hereby, they have to provide the metadata for the specified element type as a JSON file and send it to the system. The system performs several internal checks to validate the metadata against the schema before registering a PID. If the parent resource is identified by a PID, the initial step is to validate that the PID actually resolves. The next step is to validate the user-provided metadata against the metadata schema. Should any errors arise during the validation process (for instance, if required information is missing or values are in incorrect format), the user will be informed via an on-screen message detailing the nature of the issue. Once the input has been amended, the user will then need to restart the process [9].

The service does not include any metadata curation, as the RDCs are responsible for selecting the appropriate element type and providing accurate semantic information for the metadata fields.

## Enhancing Interoperability

From the PID provider side, the inclusion of metadata and controlled vocabularies for resource types at lower levels of granularity and metadata to document attribute elements within variables, audio/video data, and text would be useful improvements to enhance interoperability.

## PID Provider Considerations

The PID registration service registers ePIC PIDs. There are several reasons for the selection of ePIC, with GWDG serving as the responsible partner in the consortium, as the PID provider.

- **Cost considerations:** An initial cost analysis for this specific use case indicated that assigning ePIC PIDs is approximately a magnitude more cost-effective than assigning DOIs. The cost analysis included factors such as the planned number of PIDs to be registered by each use case partner, the number of required prefixes, whether use case partners already have PID/DOI membership in advance or not, and the cost of the GWDG infrastructure. Given the significant number of PIDs that will be registered, the cost factor has a considerable impact on the selection of a PID provider.
- **PID profiles:** GWDG allows users to register a PID profile, offering the possibility to validate the metadata of the Handle record at creation step.

The general advice for KonsortSWD's use case partners is that all metadata displayed on the landing page should also be conveyed into the PID record.

- **API availability:** It is very important that a well-curated and documented API is available.
- **Reliability:** ePIC is a well-established PID provider. The policies provided by ePIC apply to each member. To this end, GWDG guarantees quality of services<sup>21</sup> and resolvability of PIDs for at least 10 years.

## Support & Training

GESIS has occasionally reached out to ePIC regarding technical assistance and billing or membership queries. GESIS itself provides training on the importance of comprehensive and accurate metadata for PIDs, specifically aimed at staff (i.e., PID managers and IT staff implementing the API).

## Licensing

The license of a dataset element typically follows the license of the parent resource it relates to. Therefore, no metadata field is required to licence the dataset element. The decision of whether data and dataset elements are openly available is at the discretion of the use case partners owning the data.

An interesting issue is the public availability of metadata within the PID record of elements that are part of a closed dataset. In particular, when a PID for an element within a closed parent resource is being registered, the metadata within the PID is publicly accessible, unless it is actively hidden. As the data holder is responsible for providing metadata, they need to define which fields should be hidden or not. There is a lot of concern about data protection laws here. KonsortSWD can suggest which fields to hide in terms of potential personal information. However, the RDCs may need to hide some fields for other purposes as well.

## Cost Model

The test prefix for the prototype PID registration service was free of charge. In the future, several productive prefixes will be required for the use case partners. A cost model design will be carried out in KonsortSWD's second funding phase to tackle financial issues.

---

<sup>21</sup> ePIC Quality of Service and Policies: [https://www.pidconsortium.net/?page\\_id=904](https://www.pidconsortium.net/?page_id=904).

It would be advantageous in the future if costs were to be borne more at the NFDI level, as the service could prove beneficial for numerous NFDI consortia or NFDI member institutions.

## PID Policy and Governance

The staff of the PID registration service has participated in a support action<sup>22</sup> in 2024 to better understand the requirements of EOSC-compliant PID policies. They learned how to manage, write, and evaluate a policy for their own service and which requirements need to be addressed by such a policy. The support action included workshops, reading support, and tasks to write policies and assess policy compliance using the CAT tool. A similar support action promoting a PID policy for NFDI could be beneficial for all consortia.

GESIS, as an organisation, has a policy or set of guidelines regarding the use of licenses for research outputs, for the implementation and use of PIDs, and to promote open science. Hence, GESIS is generally well equipped on the policy level. Primary challenges faced by the organization are project-based budgeting (e.g., reliance on project-specific funding for PID management without a strategy for long-term sustainability) and NFDI governance complexities (issues arising from the distributed responsibility and consortium-based governance structure of NFDI affecting PID strategies).

## Outreach

There have been many outreach activities related to KonsortSWD's PID registration service. The target groups are RDCs as use case partners and social scientists in the broadest sense because they should use and adopt these PID practices and get the most out of the new service. Outreach formats have included conferences, talks, and meetings, with a total of 29 events attended. A cornerstone of outreach is also the provision of publications via Zenodo<sup>23</sup>. Around 40 research outputs are already available within the Zenodo community for Task 5 Measure 1<sup>24</sup> with 2,500 downloads, informing many people about PID practices within KonsortSWD.

---

<sup>22</sup> The FAIR-IMPACT support offer #2, the Support Program for Creating EOSC-Compliant PID Policies.

<sup>23</sup> Zenodo: <https://zenodo.org/>.

<sup>24</sup> Zenodo Community KonsortSWD Task Area 5 - Measure 1 - PIDs, <https://zenodo.org/communities/konsortswd-ta5-m1/records?q=&l=list&p=1&s=10&sort=newest>.

## Wishlist for PID4NFDI

The following wishes and expectations for the PID4NFDI basic service were named during the interview or within the survey:

- **Provision of a unified NFDI gateway:** A centralized platform to register various entities and ensure efficient PID allocation. It would be beneficial if all the NFDI consortia could rely on the same infrastructure to assign PIDs. This would save resources by preventing every consortium from having to build its own infrastructure and would support alignment in terms of metadata and interoperability.
- **Best practices:** Providing access to well-defined guidelines and best practices on PIDs assignment and usage would be helpful.
- **PID Policy:** There are also high expectations at the policy level. PID4NFDI should advocate a common harmonised PID policy across NFDI for consistent and interoperable PID management. This policy should be aligned with EOSC PID Policy<sup>25</sup>. Such a supranational and inter-institutional policy could be very good guidance on requirements for PID integration.

---

<sup>25</sup> The EOSC PID policy is available at <https://doi.org/10.2777/926037>.

## Conclusion

The use case *PID registration service for dataset elements* provides valuable insights into service development and PID integration at a fine-grained level. The findings from this project phase are crucial for understanding PID practices and identifying areas for improvement, gaps as well as challenges within NFDI. Future steps (in the context of this use case and also for PID4NFDI and GWDG in general) could target the following directions:

- Better communication by PID4NFDI and GWDG of tools, such as the DTR, for registering metadata schemas. Improving accessibility to the DTR via a connection to IAM4NFDI is also planned for the next funding phase.
- KonsortSWD will discuss whether it is feasible to register the schemas of the PID registration service in the DTR. They stated general interest in this, but no resources were planned in the original proposal for this.
- A potential direction of collaboration could be to better align PIDs for dataset elements with the FAIR Digital Object (FDO) specification<sup>26</sup>, as the initial motivation for the PID registration service was to increase the FAIRness of dataset elements. A starting point could be to assess the gap between PIDs registered by the service and the FDO specification.

---

<sup>26</sup>FDO specifications: <https://fairdo.org/specifications/>.



## References

- [1] Bingert, S., Brase, J., Burger, F., Dreyer, B., Hagemann-Wilholt, S., Vierkant, P., & Wieder, P. (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>.
- [2] Bertelmann, R., Buys, M., Sens, I., Wieder, P., Bingert, S., Dreyer, B., Hagemann-Wilholt, S., Schrader, A. C., Stathis, K., Vierkant, P., & PID4NFDI Working Group. (2024). PID4NFDI – Persistent Identifier Services for the German National Research Data Infrastructure: Proposal for the Initialisation Phase of Base4NFDI. Zenodo. <https://doi.org/10.5281/zenodo.14281250>.
- [3] Hagemann-Wilholt, S. (2023, February 13). PID4NFDI: Survey on PID Practices. Main results. Zenodo. <https://doi.org/10.5281/zenodo.7635791>.
- [4] Saldanha Bach, J., Klas, C.-P., & Mutschke, P. (2023, January 27). Workshop on PIDs within NFDI: KonsortSWD PID Registration Service. Zenodo. <https://doi.org/10.5281/zenodo.7575522>.
- [5] El-Gebali, S., Hagemann-Wilholt, S., Böhm, J., & Kahlert, T. (2024). D1.1 Landscape of PID practices within NFDI services. Zenodo. <https://doi.org/10.5281/zenodo.14277479>.
- [6] Hagemann-Wilholt, S., Wieder, P., Martini, D., & Lange, M. (2024, February 20). PID4nfdi – Example Use Cases and Services for NFDI. Zenodo. <https://doi.org/10.5281/zenodo.10685209>.
- [7] PID4NFDI (2024). PID4NFDI Stakeholder Workshop. Report. <https://doi.org/10.5281/zenodo.14232461>.
- [8] Saldanha Bach, J., Klas, C.-P., & Mutschke, P. (2023). Use cases and benefits of persistent identifiers for dataset elements. 3rd Workshop on Metadata and Research (objects) Management for Linked Open Science (DaMaLOS 2023). 15p. <https://doi.org/10.4126/FRL01-006444991>.
- [9] Klas, C.-P., Zloch, M., Saldanha Bach, J., Baran, E., & Mutschke, P. (2022). KonsortSWD Measure 5.1: PID Service for variables report (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6397367>.

- [10] Saldanha Bach, Janete, and Claus-Peter Klas. 2024. "Enhancing FAIR compliance: a controlled vocabulary for mapping Social Sciences survey variables [Paper]." IASSIST Quarterly (48(2)). <https://doi.org/10.29173/iq1118>.
- [11] 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>.
- [12] Baran, E., Zloch, M., Saldanha Bach Estevao, J., Klas, C.-P., Zhang, Y., & Mutschke, P. (2023). KonsortSWD Measure 5.1: extended report on the architecture of the PID registration service. In KonsortSWD Measure 5.1: Extended report on the architecture of the PID registration service (1.0, pp. 1–18). Zenodo. <https://doi.org/10.5281/zenodo.10401641>.
- [13] Saldanha Bach, Janete, Claus-Peter Klas, and Peter Mutschke. 2023. "Breaking down hurdles of current data citation practices: use cases and benefits of persistent identifiers for dataset elements." In E-Science-Tage 2023: Empower Your Research – Preserve Your Data, edited by Vincent Heuveline, Nina Bisheh, and Philipp Kling, 41–55. Heidelberg: Heidelberg: heiBOOKS. doi: <https://doi.org/10.11588/heibooks.1288.c18063>.
- [14] Saldanha Bach, Janete, Klas, Claus-Peter, & Mutschke, Peter. (2023). KonsortSWD Measure 5.1: use cases description extended report (1.0). Zenodo. <https://doi.org/10.5281/zenodo.7588944>.
- [15] Saldanha Bach, J., Klas, C.-P., & Mutschke, P. (2023). KonsortSWD Measure 5.1: metadata schema extended report (1.0). Zenodo. <https://doi.org/10.5281/zenodo.758890>.