Persistent Identifiers for and in NFDI –PID4NFDI Stakeholder Workshop

Report of Contributions

Contribution ID: **1**   Type: **not specified**

Welcome and Introduction

Monday 11 November 2024 09:00 (30 minutes)

After welcoming the participants and introducing the workshop agenda, PID4NFDI will give a short project overview and present its current status and activities, results from a PID landscape analysis and future plans.

Presenters: LANGE, Marc (PID4NFDI / Helmholtz Open Science Office); BÖHM, Jana (PID4NFDI / GWDG); KAHLERT, Torsten (PID4NFDI / German National Library of Science and Technology (TIB))

Contribution ID: **2**   Type: **not specified**

Wrap-Up

Monday 11 November 2024 12:20 (10 minutes)

Contribution ID: 4   Type: **not specified**

DataCite: Connecting Research, Advancing Knowledge

Monday 11 November 2024 09:30 (10 minutes)

This presentation provides an introduction to DataCite, a global non-profit organization dedicated to supporting the research community through persistent identifiers (PIDs). DataCite plays a key role in making research outputs—including datasets, software, and other non-traditional research materials—findable, citable, and interconnected. The talk will highlight the importance of PIDs and rich metadata in enhancing research transparency, visibility, and reusability, and will include real-world examples of the impact of our PID infrastructure.

Presenter: EL-GEBALI, Sara (PID4NFDI / DataCite)

Session Classification: PID Provider

Contribution ID: **5**   Type: **not specified**

European Persistent Identifier Consortium (ePIC)

Monday 11 November 2024 09:40 (10 minutes)

The European Persistent Identifier Consortium (ePIC) is a collaborative initiative established in 2009 to enhance the management of research data through the use of PIDs. With nine founding members, ePIC aims to provide a robust and user-friendly PID system that supports researchers in creating, processing, and resolving identifiers for their digital resources. As the volume and complexity of research data continue to grow, ePIC addresses the critical need for stable references that ensure long-term accessibility and citation of scientific materials. By leveraging the Handle system, ePIC facilitates reliable linking between data and related resources, thereby fostering a sustainable research environment. The GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen), which operates parts of the global PID system, is a founding member of ePIC.

Presenter: BINGERT, Sven (PID4NFDI / GWDG)

Session Classification: PID Provider

Contribution ID: **6**　　　　　　　　　　　　　　　　　　　Type: **not specified**

PIDA: Empowering your semantic artifacts with reliable PURLs

Monday 11 November 2024 09:50 (10 minutes)

PIDA is a lightweight PURL service that provides unique persistent URLs (PURLs) that can be used as internationalized resource identifiers (IRIs) within semantic artifacts such as glossaries, thesauri, ontologies and knowledge graphs. The main objective is to support and facilitate the findability and accessibility of semantic artifacts published on the Web in the long term. PIDs provided by PIDA are functional URLs/IRIs; instead of pointing directly to the location of an internet resource, a PID points to an intermediate-resolution service. The resolution service associates the PID with the actual URL/IRI and returns that location on the web to the client. The client can then complete the transaction in a normal fashion. PIDA also supports content negotiation. This allows specifying which representation of a resource (e.g., HTML, TTL, etc.) is required by semantic agents via the same identifier.Furthermore, this also ensures that: 1) a resource can be reliably referenced for future access; 2) links to resources do not break; 3) support findability—one of the key elements in the FAIR principles; and 4) entities can be identified unambiguously. In this talk, I will introduce the new features offered by PIDA and how to obtain PIDA PURLs free of charge. These features include usage statistics dashboards, regular system health checks, and automatic notifications for broken PURLs.

Presenter: FATHALLA, Said (Helmholtz Metadata Collaboration / Forschungszentrum Jülich)

Session Classification: PID Provider

Contribution ID: **7** Type: **not specified**

ORCID: Enhancing Research Interoperability through PIDs

Monday 11 November 2024 10:00 (10 minutes)

This presentation will provide a brief overview of ORCID, as standard Persistent Identifier (PID) for people, explaining what it is, how it works, and its role in the research ecosystem. We will emphasize the importance of interoperability and how ORCID contributes to it by connecting researchers with other PIDs for research outputs, funding and research organizations, or even other PIDs for people. Through ORCID's API active integrations, a network of trusted assertions is created, enhancing data reliability and collaboration across the research landscape. The session will also include a quick demo of an ORCID integration, showcasing its practical application for researchers and institutions.

Presenter: ALSINA, Francisco (ORCID)

Session Classification: PID Provider

Contribution ID: **8** Type: **not specified**

StrainInfo: A central database for resolving microbial strain identifiers

Monday 11 November 2024 10:50 (10 minutes)

Microbial strains can be known by a myriad of different strain designations, culture collection numbers and sequence accessions, which poses a challenge to the communication of research findings, as well as the comparison and reuse of data. Culture collection numbers have the advantage of being unique, stable and subject to high quality standards. Nevertheless, each collection receiving a culture of the same strain assigns their own number at deposition. Different designations are thus used throughout publications and databases, making it difficult for scientists to draw connections between them. Here we present the StrainInfo database, a service that collects and curates culture collection numbers as well as their relations and links them with different sources of information, such as publications and sequence accession numbers. This facilitates the connecting of data describing the same strain. The information is provided through a modern and intuitive web user interface, which enables users to easily find corresponding strain identifiers and links to associated data, and through a web API, that allows for direct integration of strain identity resolution into workflows and other databases. In the future, StrainInfo additionally aims to provide a central registry service for cultures, allowing microbiologists to register strain designations and receive persistent identifiers prior to deposition and publication.

Presenter: REIMER, Lorenz (NFDI4Microbiota / Leibniz Institute DSMZ)

Session Classification: Use Cases

Contribution ID: **9** Type: **not specified**

RSpace: PIDs in an early phase of research data management

Monday 11 November 2024 11:00 (10 minutes)

PIDs are often considered in the late research phases, e.g. upon publishing a manuscript or publicizing a dataset. RSpace is a digital research data management platform used in the early, active research phase. RSpace integrates with a wide range of research data infrastructure, e.g. data repositories and data management planning tools, to facilitate the passage of data and metadata through a connected research ecosystem. Here, we share learnings from a recent integration of a PID registration workflow into RSpace's physical sample management solutions using International Generic Sample Numbers (IGSNs), as well as discuss opportunities for how PIDs can be incorporated early in the development of metadata of research objects. We'll overview further PID related initiatives to improve the FAIRness of research data and metadata generated and passing through RSpace.

Presenter: MATHES, Tilo (RSpace)

Session Classification: Use Cases

Contribution ID: **10** Type: **not specified**

PIDs below the study level: Advancing fine-grained data citation by PIDs for dataset elements

Monday 11 November 2024 11:10 (10 minutes)

Persistent Identifiers (PIDs) at the study or dataset level are insufficient for addressing the complexity of data management in research. The lack of granularity citation at the level of inline data objects, such as individual survey variables, qualitative data files, and even smaller data points, leads to ambiguities in data citation, inadequate metadata, and data discovery and reusability challenges. The PID registration service introduced by KonsortSWD, part of the German National Research Data Infrastructure (NFDI), significantly advances the granularity of PIDs. This service supports the assignment of PIDs to these finer dataset elements and ensures accurate data citation since researchers usually use only a subset of the elements in a dataset. The service boosts data referencing practices and also ensures adherence to FAIR principles by enabling precise referencing of individual data elements. A single data point, the PID, enhances data citation, reuse, and direct access for automated access (i.e., by a computer program and under some requirements). In terms of implementation, the technical solution employs the ePic API and relies on the Handle standard. Tests PIDs have been successfully applied to diverse datasets, including survey variables from GESIS and the German Institute for Economic Research (DIW). Tests are also currently taking place at the German Center for University and Science Research (DZHW) in 2024 and are planned at Qualiservice in 2025. By serving as a base service under the umbrella of PID4NFDI, KonsortSWD's PID service provides a scalable framework that can adapt to many domains across the NFDI.

Primary authors: SALDANHA BACH, Janete (KonsortSWD / GESIS –Leibniz Institute for the Social Sciences); MUTSCHKE, Peter (KonsortSWD / GESIS –Leibniz Institute for the Social Sciences)

Presenter: SALDANHA BACH, Janete (KonsortSWD / GESIS –Leibniz Institute for the Social Sciences)

Session Classification: Use Cases

Contribution ID: **11**   Type: **not specified**

Archival Resource Key (ARK): The decentralized non-paywalled PID

Monday 11 November 2024 10:10 (10 minutes)

This talk introduces ARKs (Archival Resource Keys), decentralized non-paywalled persistent identifiers (PIDs) used for decades to identify scientific and cultural heritage of any type, digital, physical, or abstract. Since 2001, 8.2 billion ARKs have been created by over 1400 organizations –data centers, publishers, libraries, archives, museums, and government agencies. With highly flexible metadata and resolution options, ARKs are well-suited for citation, linked-data, and identifying component parts, such as IIIF image details. In 2018, multiple organizations partnered to form the ARK Alliance in order to sustain the ARK infrastructure and guide its future.

Presenter: KUNZE, John (ARK Alliance / Drexel University, Metadata Research Center)

Session Classification: PID Provider

Contribution ID: **12**   Type: **not specified**

Q & A

Monday 11 November 2024 10:20 (15 minutes)

Session Classification: PID Provider

Contribution ID: **13**   Type: **not specified**

Q & A

Monday 11 November 2024 11:20 (10 minutes)

Session Classification: Use Cases