



Enhancing Metadata Quality through Persistent Identifiers

Insights from PID4NFDI

Dr. Torsten Kahlert
TIB Hannover

CiVers Workshop, Berlin, DAI, 14. Oktober 2025

In cooperation with



Agenda

Wer wir sind (PID4NFDI / Base4NFDI)

Warum Metadatenqualität wichtig ist

- Metadatenvollständigkeit
- Beziehungsmetadata
- PID Metadaten für
Interoperabilität

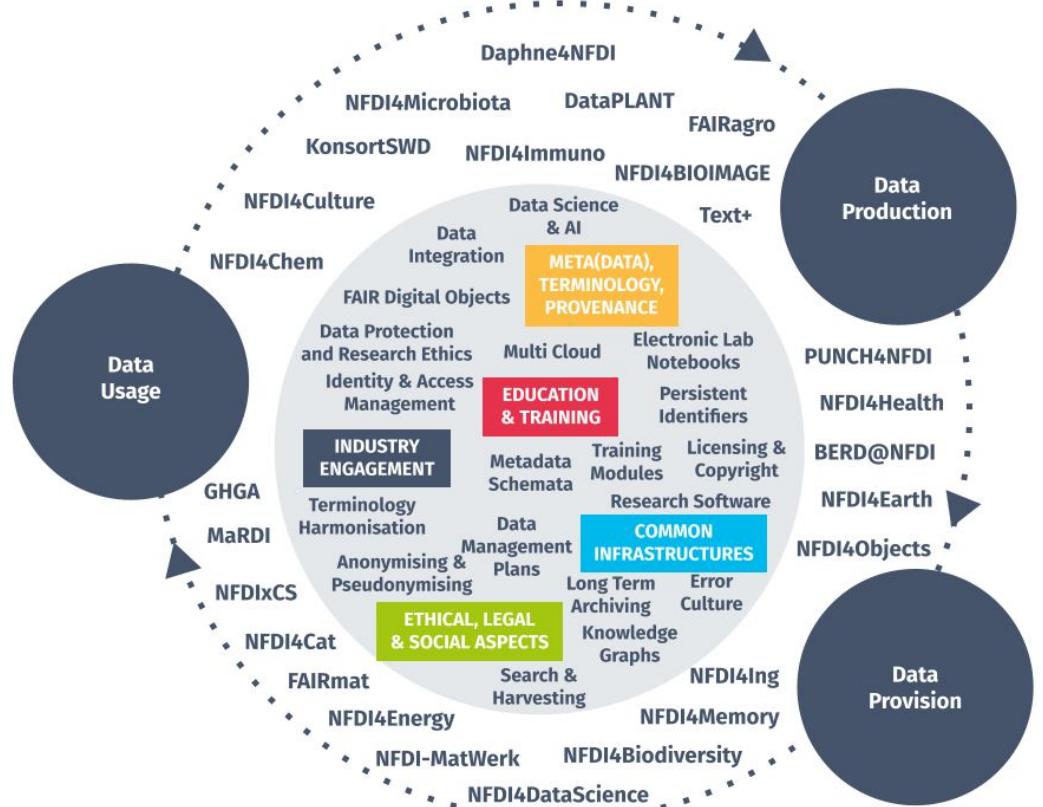
Basic Services innerhalb der NFDI

Base4NFDI ist eine gemeinsame Initiative aller 26 NFDI-Konsortien.

Ziel ist es, **verlässliche** Basisdienste für ein **FAIRes** Forschungsdatenmanagement aufzubauen.

Basisdienste sollen ...

- möglichst **einfach** von vielen Konsortien nutzbar sein,
- **interoperabel** mit internationalen Standards und Diensten, und
- **Synergien** zwischen bestehenden Lösungen nutzen.



PID4NFDI - Service und technisches Konzept

Zielsetzung: Weiterentwicklung der PID-Service-Landschaft innerhalb der NFDI auf allen Ebenen

- Integration bestehender PID-Infrastrukturen
- Verbesserung der Metadatenqualität und -mappings
- Schulungen & Outreach
- Unterstützung der PID-Integration in den Konsortien entsprechend ihrer Bedarfe
- Abstimmung mit nationalen und internationalen PID-Strategien und -Policies



Initialisation

Integration

Ramp-Up

Project Team und Partner Institutionen

Team

Project Team Members



Jana Böhm

Project team member



Sara El-Gebali

Project team member



Steffi Genderjahn

Project team member



Stephanie Hagemann-Wilholt
Project team member



Torsten Kahlert

Project team member



Antonia Schrader

Project team member



Frederik Springer

Project team member



Sonja Sternkopf

Project team member



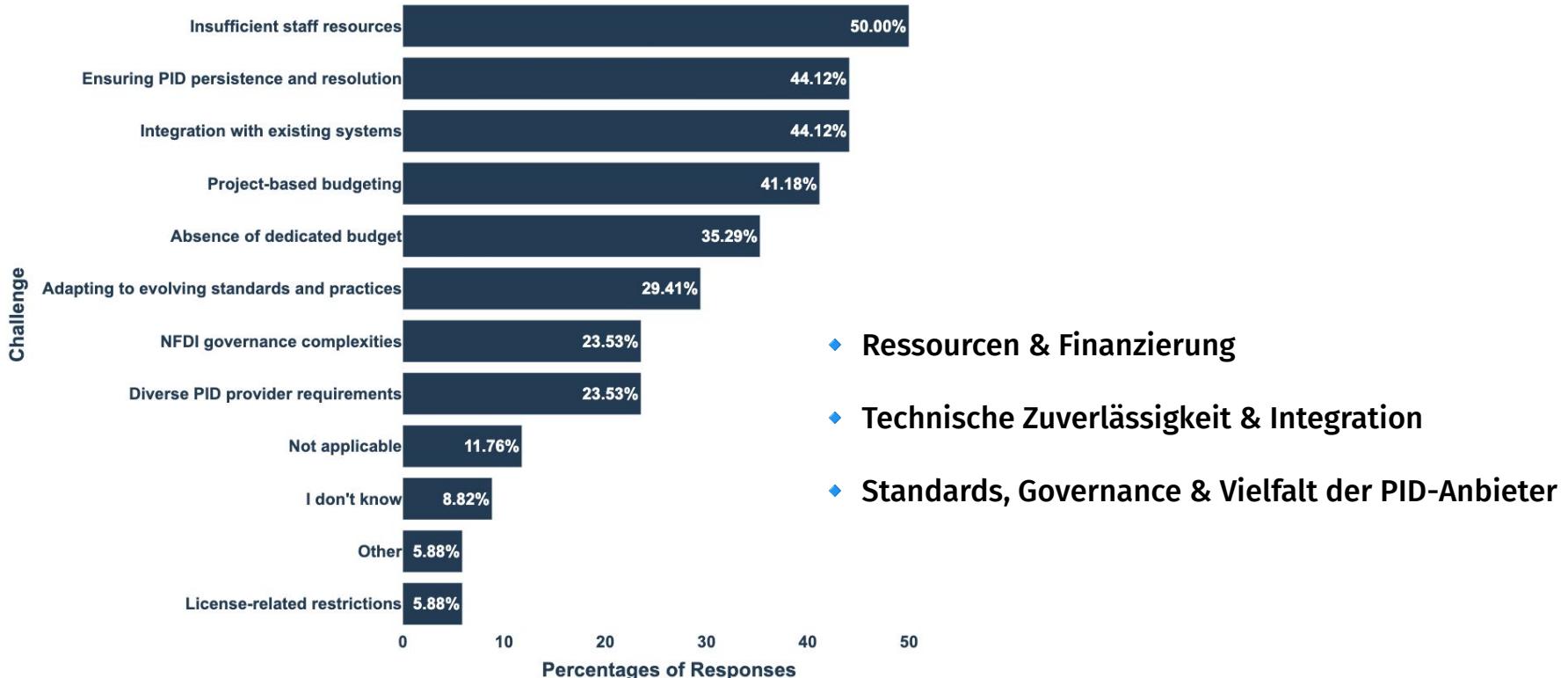
Markus Stocker

Project team member



HELMHOLTZ
Open Science

NFDI-Landschaftsumfrage – Herausforderungen

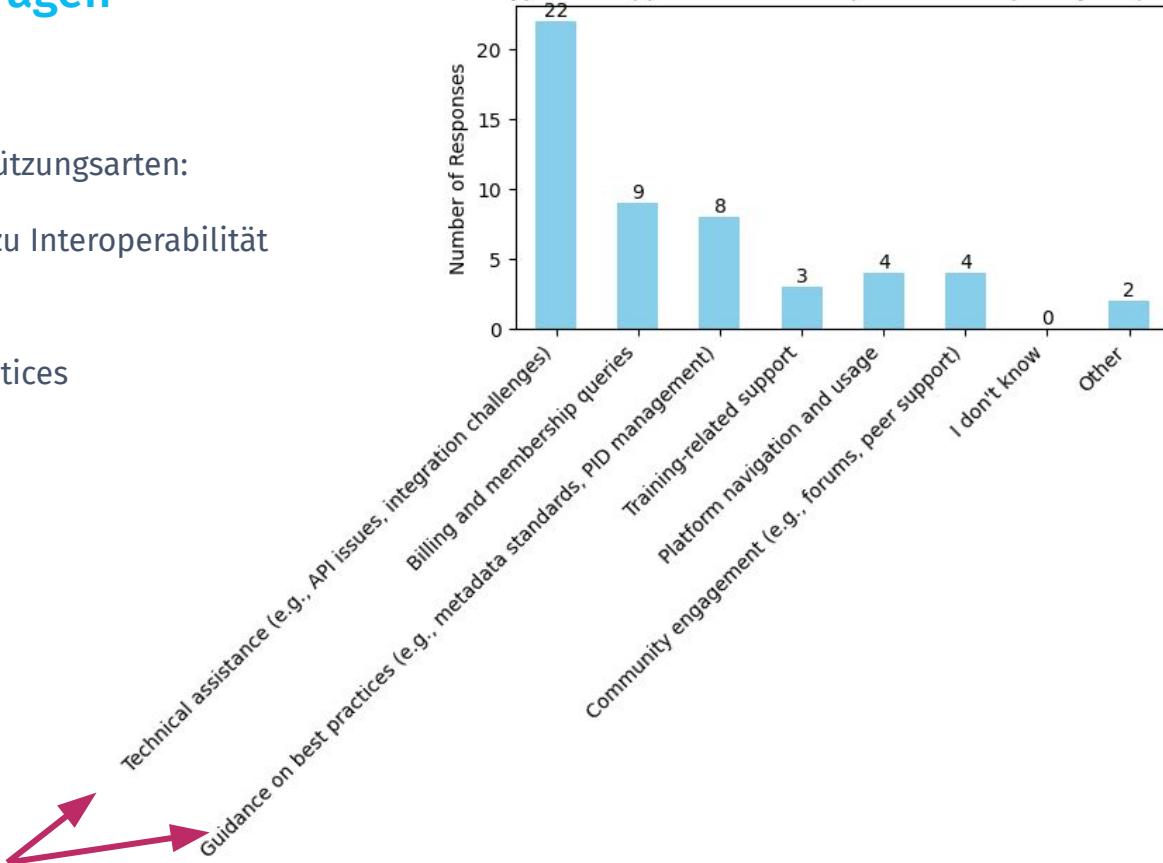


Ergebnisse der NFDI-Landschaftsumfrage

Unterstützungsanfragen

Häufig nachgefragte Unterstützungsarten:

- Technischer Support zu Interoperabilität und Integration
- Leitlinien zu Best Practices



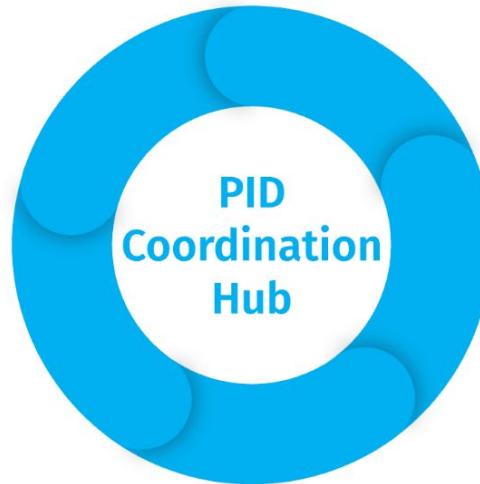
PID Coordination Hub - A Vision for PID4NFDI

Support and Information

- Linking existing PID resources
- Display best practices
- Helpdesk and open hours
- Training

Target group-specific community engagement

- Repository managers
- Researchers
- Organizations
- Decision makers



Use Cases

Focus Groups

Outreach

Metadata and Interoperability

- Support for metadata quality assessment
- Guidelines for metadata harmonization
- Focus on resources from use cases

Services

- Prefix registration service for PID use-cases
- B2Inst
- Data Type Registry (EOSC)
- PID Metaresolver (EOSC)
- Compliance Assessment Toolkit (EOSC)

Governance

- Tool for PID provider selection
- NFDI-wide PID guidelines and compliance testing

PID Selection Tool & PID4NFDI Cookbook

Persistence and Costs

1. It is important for us that the PID provider demonstrates a strong, long-term commitment to persistence.

[More info](#)

1. Persistence is the core promise of a PID - but it's not guaranteed by the technology alone. It depends on long-term organizational and financial commitment. Systems like DataCite DOIs, ePIC Handles, and URN:NBN:DE are backed by stable institutions/organizations that explicitly commit to maintaining resolvability over time. ARKs can also support persistence, but this varies with the implementation, especially as ARKs need to be hosted by the assigning institution.



<https://pid.services.base4nfdi.de/pidtool/>

PID4NFDI Cookbook

Search docs

How to choose a PID?

- ARK - Archival Resource Key
- DOI - Digital Object Identifier
- ePIC - European Persistent Identifier Consortium
- FactGrid - A database for historians
- GND - Gemeinsame Normdatei
- IGSN - International Generic Sample Number

What is a IGSN?

How is an IGSN ID structured?

Who governs the IGSN ID?

How do I register/update an IGSN ID?

What Metadata Schema is used for IGSN IDs?

How do I search for an IGSN ID?

Interoperability

What support documentation is available for IGSN IDs?

Who do I contact to find out more about IGSN IDs?

ORCID - Open Researcher and Contributor ID

PIDA - Persistent Identifiers for Semantic Artifacts

ROR - Research Organization Registry

Wikidata ID

/ IGSN - International Generic Sample Number

[View page source](#)

IGSN - International Generic Sample Number

What is a IGSN?

An IGSN ID is a globally unique and persistent identifier for material samples. The core purpose of the IGSN ID is to enable transparent and traceable connections between research activities and objects, including samples, collections, instruments, grants, data, publications, people, and organizations.

An IGSN ID can be applied to an individual material sample (including to a destroyed or discarded sample), an aggregation of samples, or to a sample collection site. An IGSN ID cannot be used for an image of a material sample, analytical data about the sample, or instruments used for sample measurements.

Although the IGSN ID originates from the Geosciences, IGSN IDs can be applied to all material samples and features-of-interest from any discipline.

How is an IGSN ID structured?

Functionally, an IGSN ID is a Digital Object Identifier (Digital Object Identifier). As a result, an IGSN ID shares the same format as a DOI: an IGSN ID name is composed of a prefix beginning with 10. and a suffix, separated by a forward slash; for example, <https://doi.org/10.60510/icdp5054ehw1001>. The prefix ensures that the IGSN ID is globally unique. The suffix can be customized to capture any local identifiers already assigned to the material sample.

Who governs the IGSN ID?

In 2021, DataCite and the IGSN Organization (IGSN e.V.) announced a partnership to support the global adoption, implementation, and use of material sample identifiers. Within the partnership: * DataCite provides the IGSN ID registration services and supports their ongoing sustainability. * IGSN e.V. implements and promotes standard methods for identifying, citing, and locating material samples by fostering disciplinary Communities of Practice (CoPs).

How do I register/update an IGSN ID?

<https://pid4nfdi-training.readthedocs.io/en/latest/>

Metadata Insights

1- Metadatenvollständigkeit

DataCite-Metadatenschema-Struktur

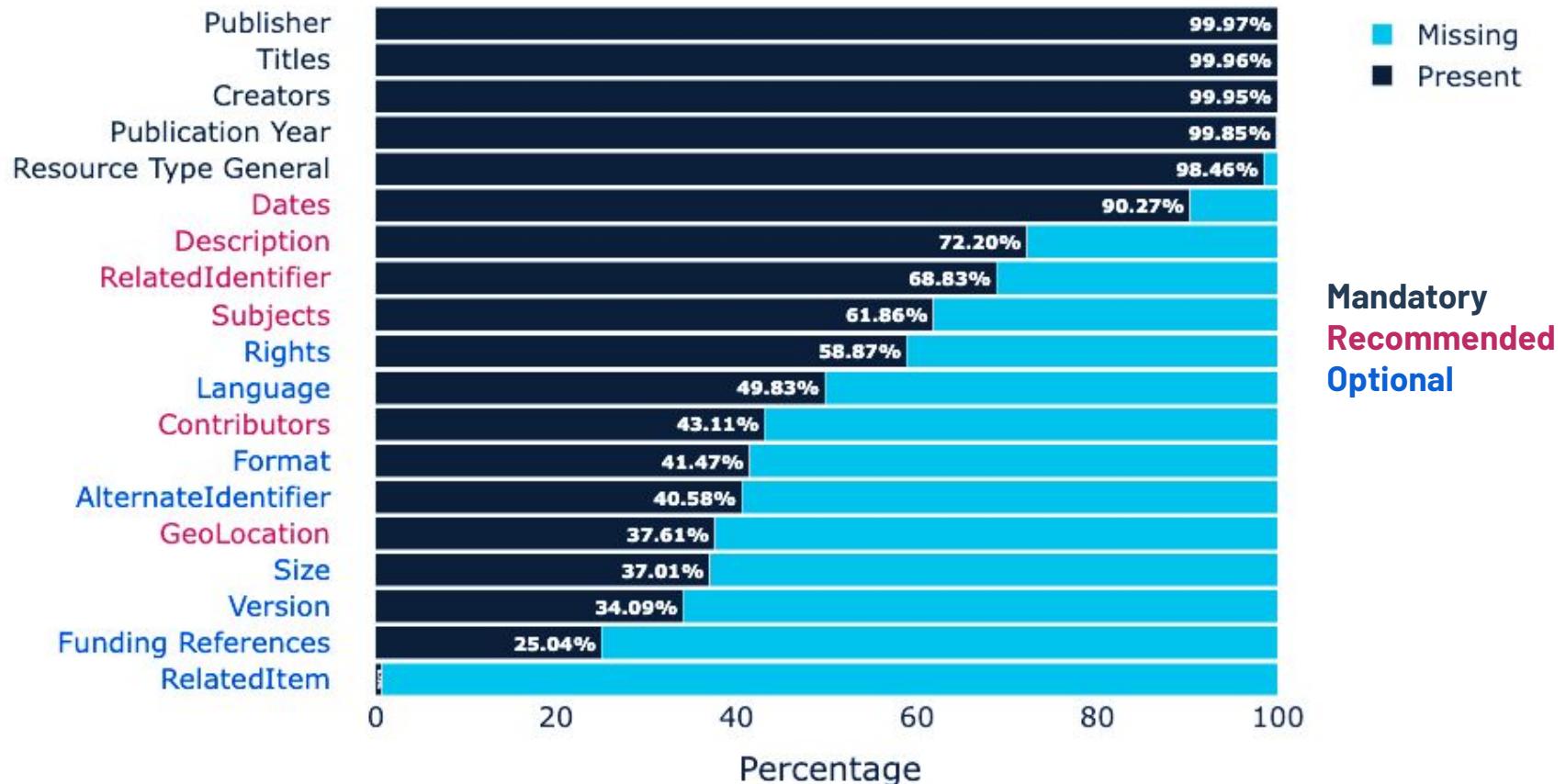
6 Mandatory

ID	Property
1	Identifier
2	Creator
3	Title
4	Publisher
5	PublicationYear
10	ResourceType

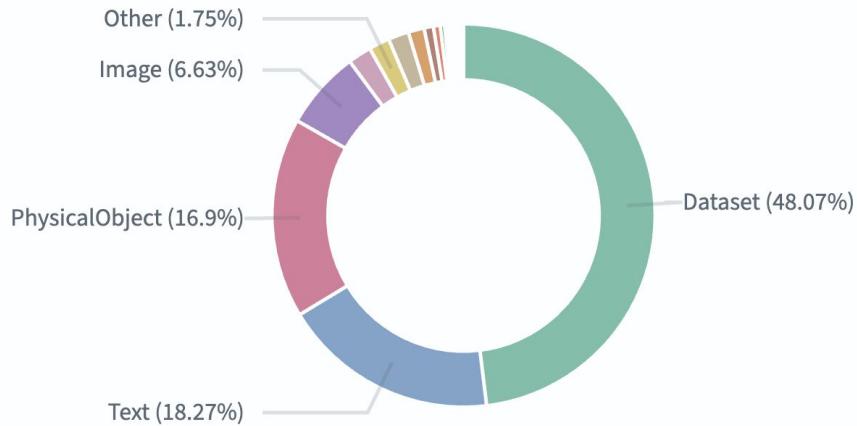
6 Recommended and 8 Optional

ID	Property	Obligation
6	Subject	R
7	Contributor	R
8	Date	R
9	Language	O
11	AlternateIdentifier	O
12	RelatedIdentifier	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description	R
18	GeoLocation	R
19	FundingReference	O
20	RelatedItem	O

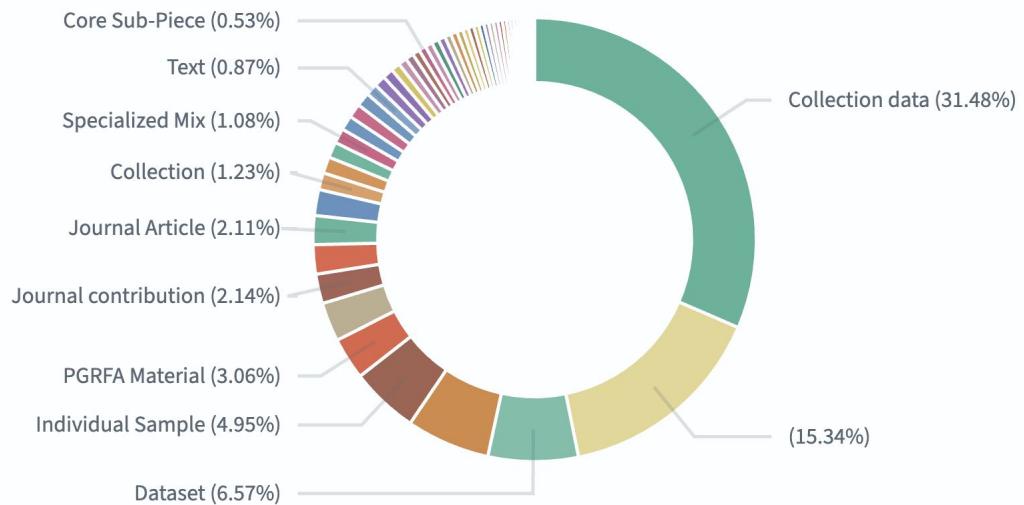
DataCite-Metadatenvollständigkeit



Welche Ressourcentypen werden derzeit am häufigsten in DataCite publiziert?



resourceTypeGeneral = kontrolliertes Vokabular



resourceType = Freitextangaben durch Nutzer*innen

Licensing Information

How much research is being made openly available for reuse?

Refer to CDS usage	Default License	Creative Commons CC0 for images UseAndReproduction: Glenbeigh Records Management	cc-by-sa-4.0	Creative Commons Attribution
Domaine public	KITopen License	Creative Commons Attribution Share Alike 4.0 International		CC BY-NC 4.0
Sabine Fourrier free	Open License Not Specified	arXiv.org perpetual, non-exclusive license	Closed Access None	CC BY Public CC BY SA 4.0

Attribution-NonCommercial (CC BY-NC)

Open Access

cc-by

Creative Commons Attribution 4.0

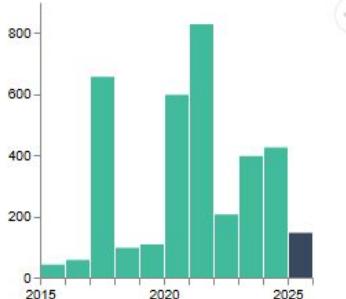
Creative Commons Attribution Non Commercial 4.0 International

Public Data restricted	Creative Commons Attribution Non Commercial Share Alike 4.0 International	pdm	open	Autre CC BY + CC0
Licensing unknown: Please contact principal investigator/authors to gain access and request licensing terms				
Etalab Open License 2.0	TK Open to Collaboration (TK CB)	openAccess	ADS Terms and Conditions of reuse apply	

DAI Repo bei DataCite

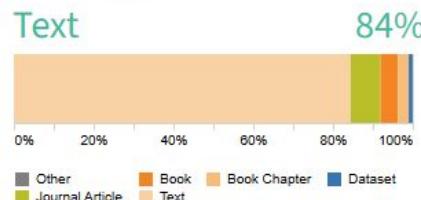
4,402 Works

Publication Year



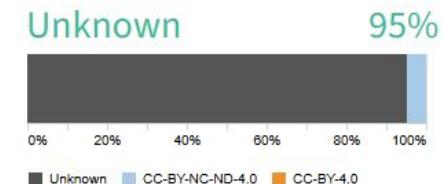
Work Types

Text

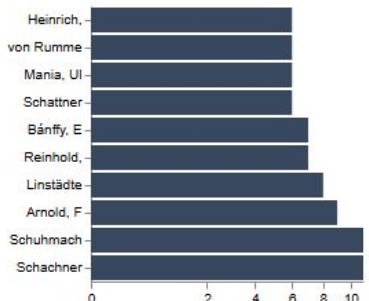


Licenses

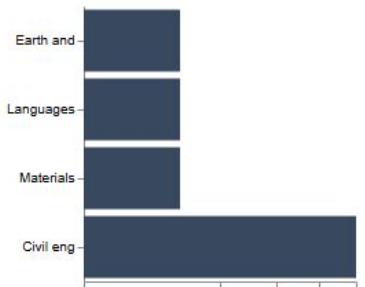
Unknown



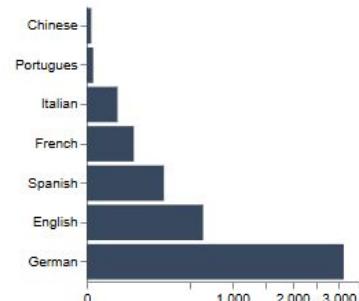
Top Creators and Contributors



Fields of Science



Work Languages

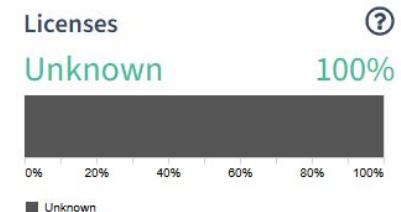
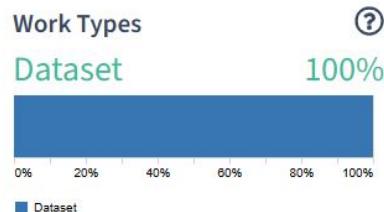
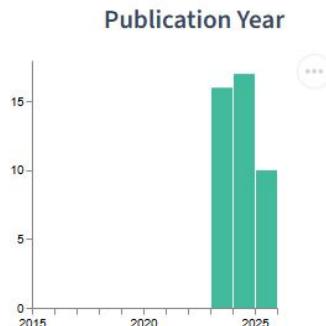


[https://commons.datacite.org/
repositories/8orczech](https://commons.datacite.org/repositories/8orczech)

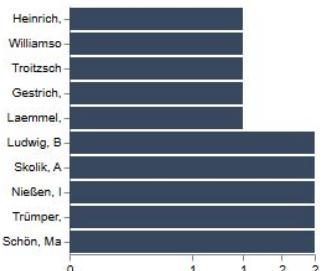
DAI Repo bei DataCite

nur resource-type:
Datensatz

43 Works



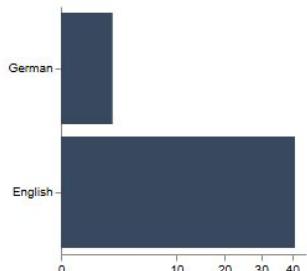
Top Creators and Contributors



Fields of Science

Not enough data to render this chart

Work Languages



<https://commons.datacite.org/repositories/dai.avnrkz?resource-type=dataset>

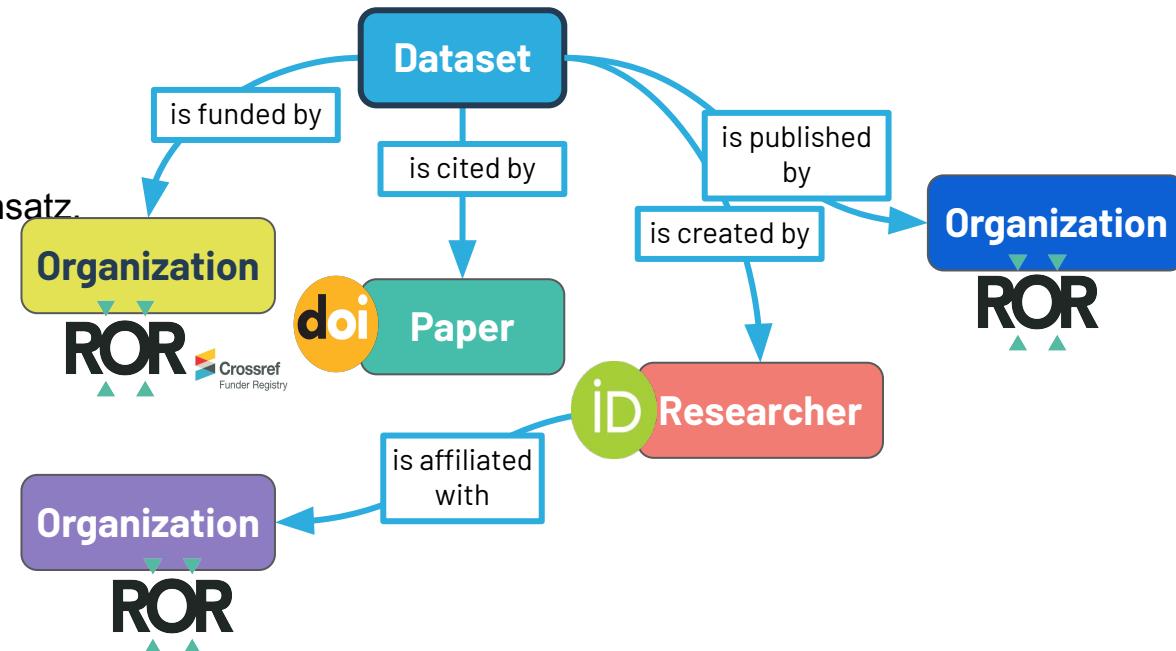
2- PID Beziehungsmetadaten

Was sind Beziehungsdaten?

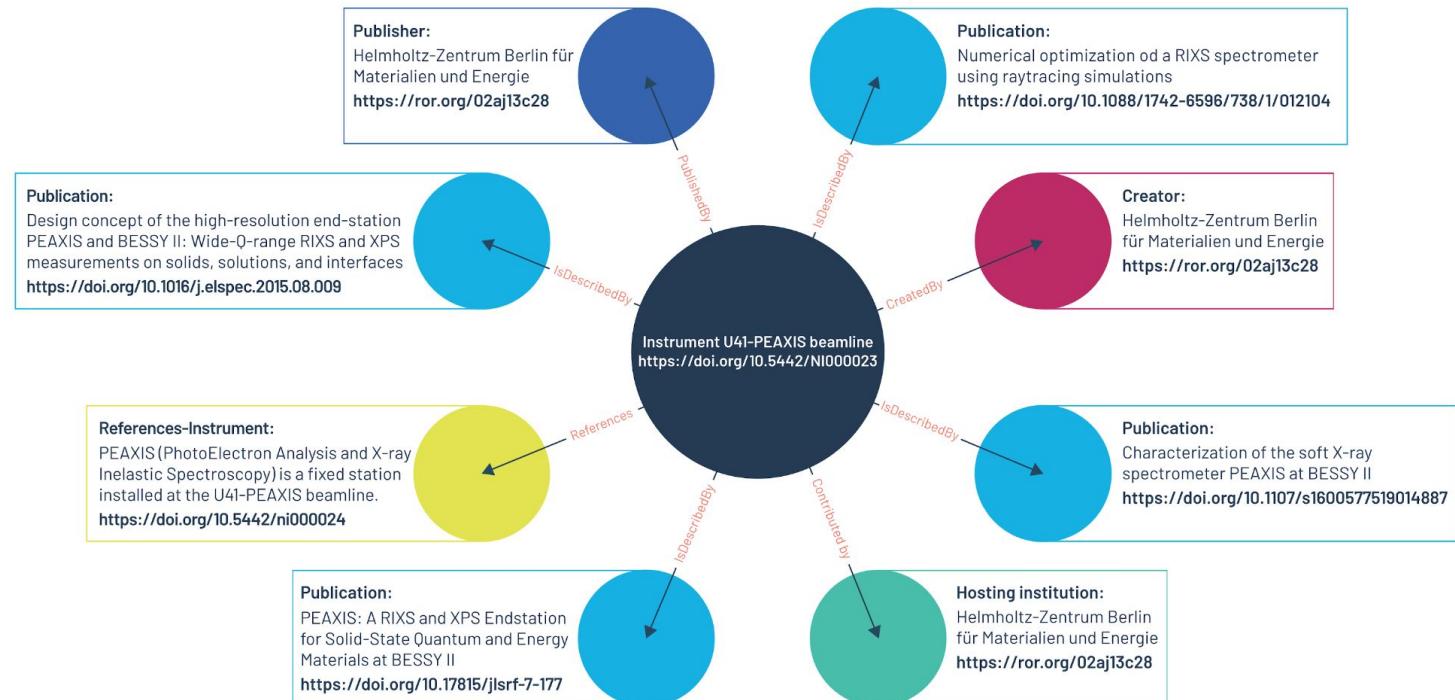
Metadaten, die Beziehungen – also Verbindungen – zwischen Entitäten abbilden

Beispiele:

- Eine Publikation zitiert einen Datensatz.
- Eine Person ist Autor*in einer Publikation.
- Eine Person ist einer Institution zugeordnet.
- Eine Institution fördert ein Forschungsergebnis.
- Ein Datensatz wird durch Software erstellt oder zusammengestellt.



Wie stellen Forschende Verbindungen zwischen Daten, Software, Publikationen und weiteren Outputs her?

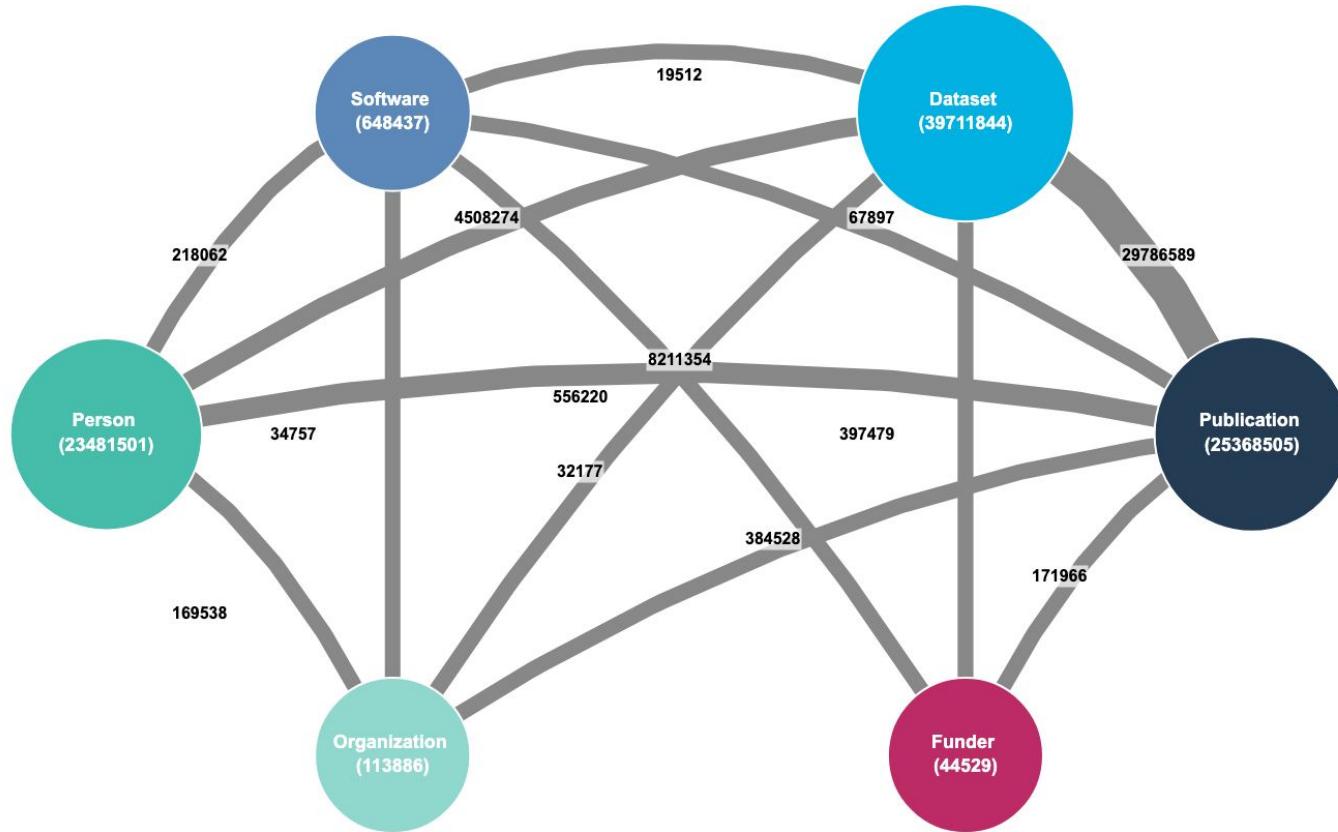


<https://api.datacite.org/dois?resource-type-id=instrument>

<https://api.datacite.org/dois/10.5442/ni000023>

Forschung ist ein Graph

Verbunden durch die Möglichkeiten von Metadaten



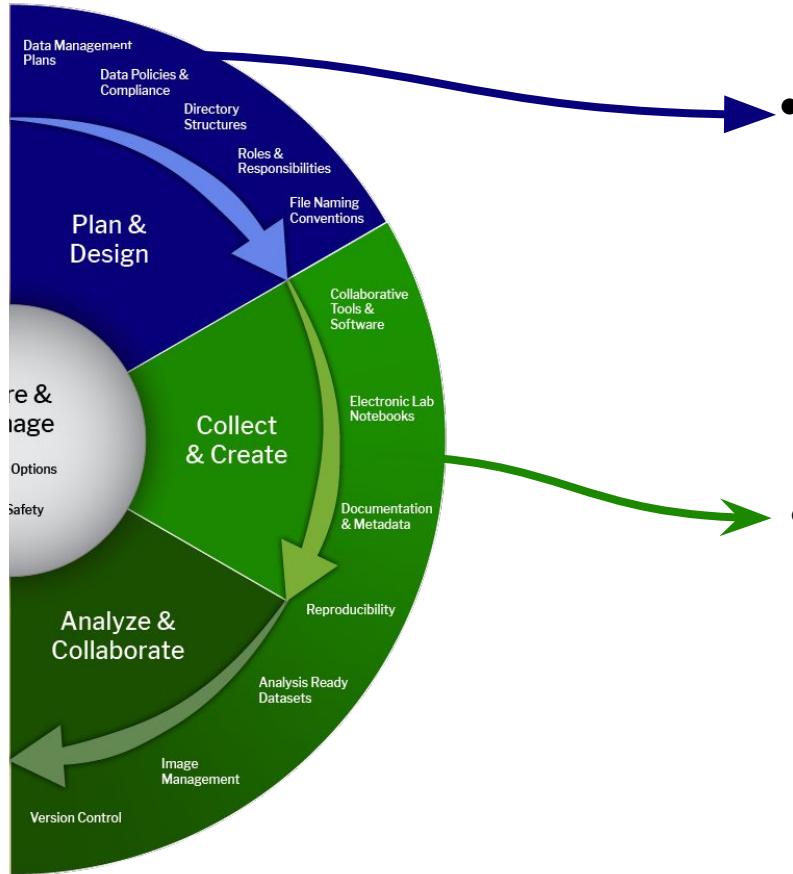
3- PID Metadaten für Interoperabilität

Research Life cycle



- **Nahtloser Datenfluss:** Sicherstellung eines reibungslosen Datenflusses zwischen verschiedenen Systemen.
- **Auffindbarkeit und Nachnutzung:** Verlässlicher Austausch und Harmonisierung von Metadaten zu Forschungsobjekten.
- **Von Beginn an FAIR:** Verankerung der FAIR-Prinzipien bereits bei der Datenerstellung, nicht erst nachträglich.

Research Life cycle



- **Daten Management Pläne (DMPs):** Von Förderorganisationen gefordert; sie erfassen, wer das Projekt finanziert, welche Ergebnisse erwartet werden und wie Speicherung, Austausch und Langzeitarchivierung geplant sind.

- **Erfassungssysteme zur digitalen Grabungsdokumentation (Field Notebooks):** Alltägliche Werkzeuge zur Dokumentation von Feldarbeit, Beobachtungen oder Ausgrabungen – häufig einschließlich Kontextbeschreibungen, stratigrafischer Beziehungen, Objektfunden, Fotografien, GPS-Daten und Survey-Informationen.

Warum PIDs und PID Metadaten in der Praxis wichtig sind

- **Schnellere Ablage** – Metadaten werden automatisch übernommen, weniger Nacharbeit für Repositorien.
- **Weniger Dubletten** – Personen und Institutionen sind eindeutig über ORCID und ROR identifizierbar.
- **Bessere Auffindbarkeit** – strukturierte Metadaten auf Landing Pages erhöhen Sichtbarkeit in Suchmaschinen.
- **Nachvollziehbare Herkunft** – Datensätze, Objekte und Publikationen bleiben eindeutig miteinander verknüpft.
- **Anerkennung & Berichte** – Rollen (CRediT) und Förderinformationen fließen automatisch in Reports ein.
- **Flexibler Zugriff** – ein PID liefert wahlweise menschen- oder maschinenlesbare Informationen.
- **Vernetzte Daten** – explizite Beziehungen bauen Netzwerke statt isolierter Datensätze.
- **Metadatenflüsse** – Informationen wandern mit dem PID durch den Forschungszyklus, ohne doppelte Eingaben.

Vielen Dank! Fragen?



pid4nfdi@lists.nfdi.de



pid4nfdi.bsky.social



<https://www.linkedin.com/showcase/pid4nfdi>



[@pid4nfdi@nfdi.social](https://pid4nfdi@nfdi.social)



<https://pid.services.base4nfdi.de>

