Enhancing Metadata Quality through Persistent Identifiers

Insights from PID4NFDI

Dr. Torsten Kahlert
TIB Hannover

CiVers Workshop, Berlin, DAI, 14. Oktober 2025

base4nfdi Basic Services for NFDI

In cooperation with nfdi

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Agenda

Wer wir sind (PID4NFDI / Base4NFDI)

Warum Metadatenqualität wichtig ist

- Metadatenvollständigkeit
- Beziehungsmetadaten
- PID Metadaten für Interoperabilität

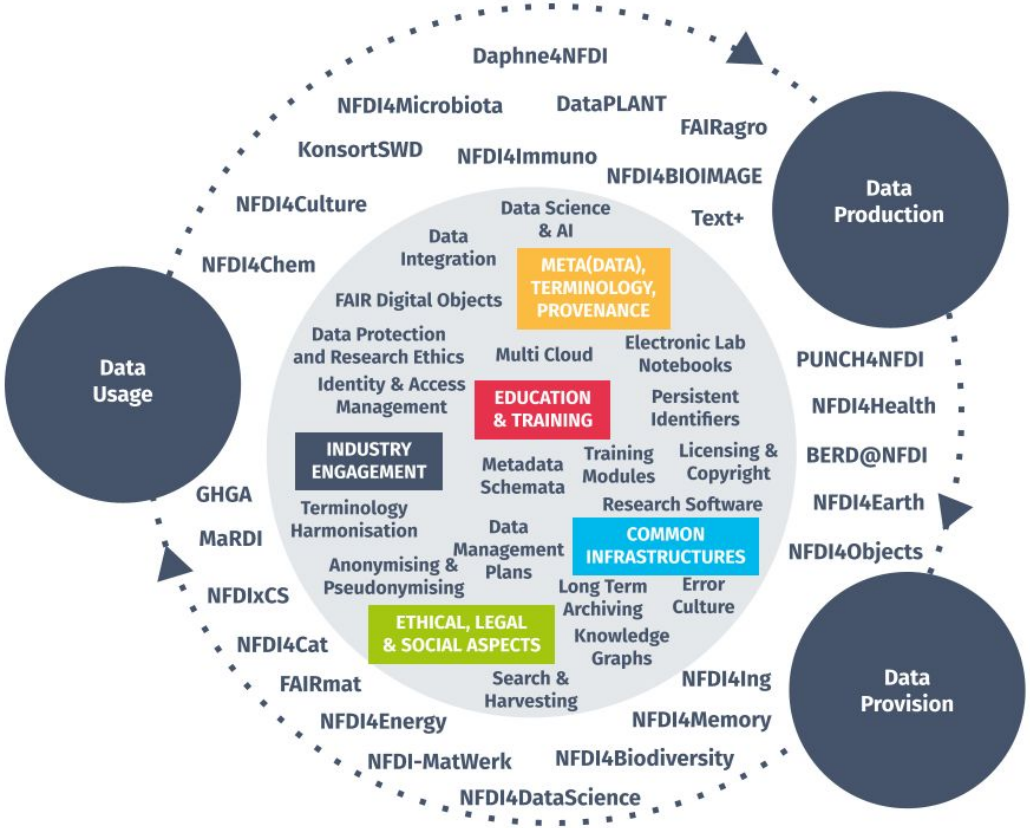
Basic Services innerhalb der NFDI

Base4NFDI ist eine gemeinsame Initiative aller 26 NFDI-Konsortien.

Ziel ist es, **verlässliche** Basisdienste für ein **FAIR**es 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

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

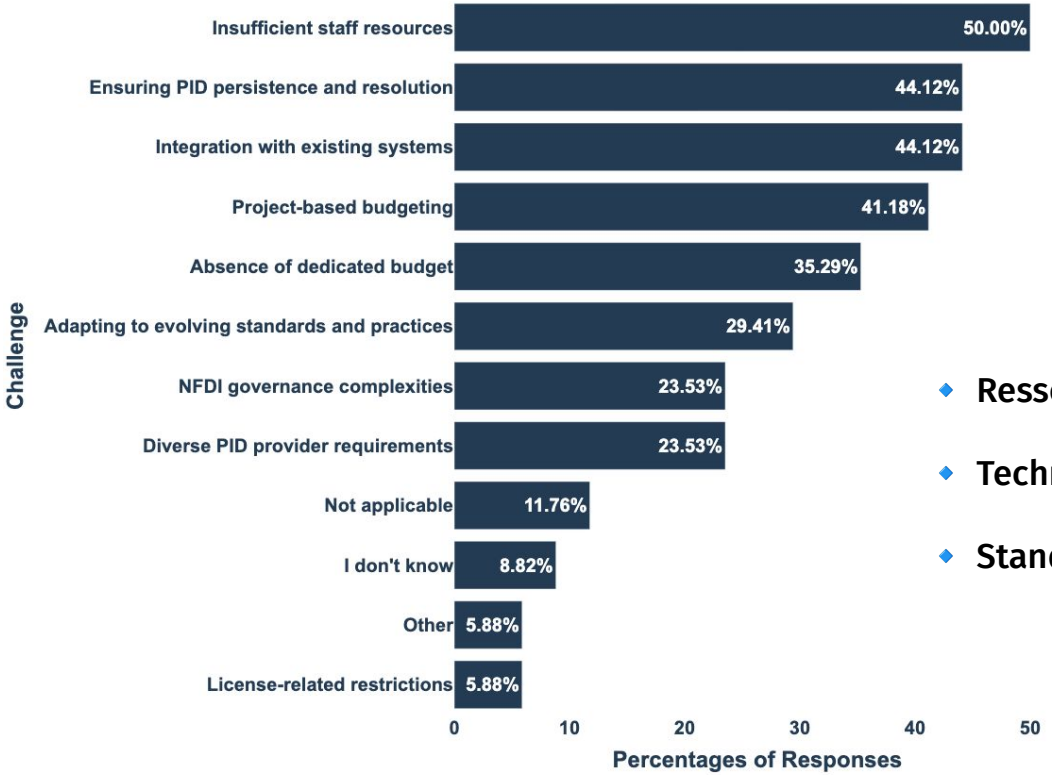
Antonia Schrader
Project team member

Frederik Springer
Project team member

Sonja Sternkopf
Project team member

Markus Stocker
Project team member

NFDI-Landschaftsumfrage – Herausforderungen



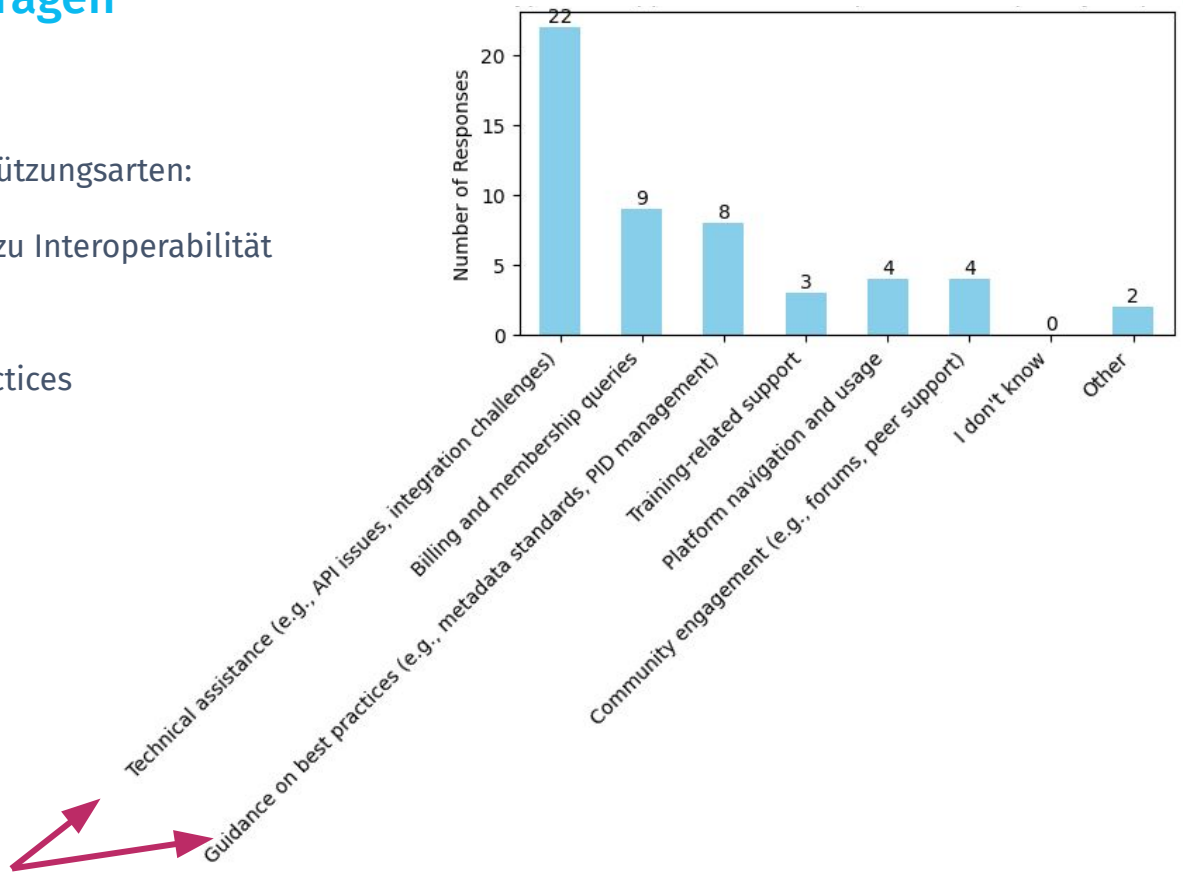
- Ressourcen & Finanzierung
- Technische Zuverlässigkeit & Integration
- Standards, Governance & Vielfalt der PID-Anbieter

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

base4nfdi

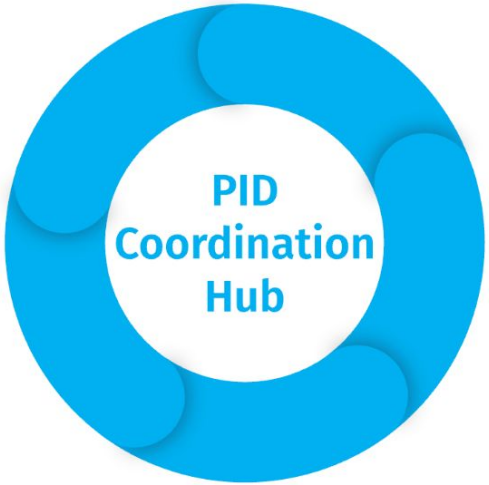
PID4nfdi

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



Metadata and Interoperability

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

Services

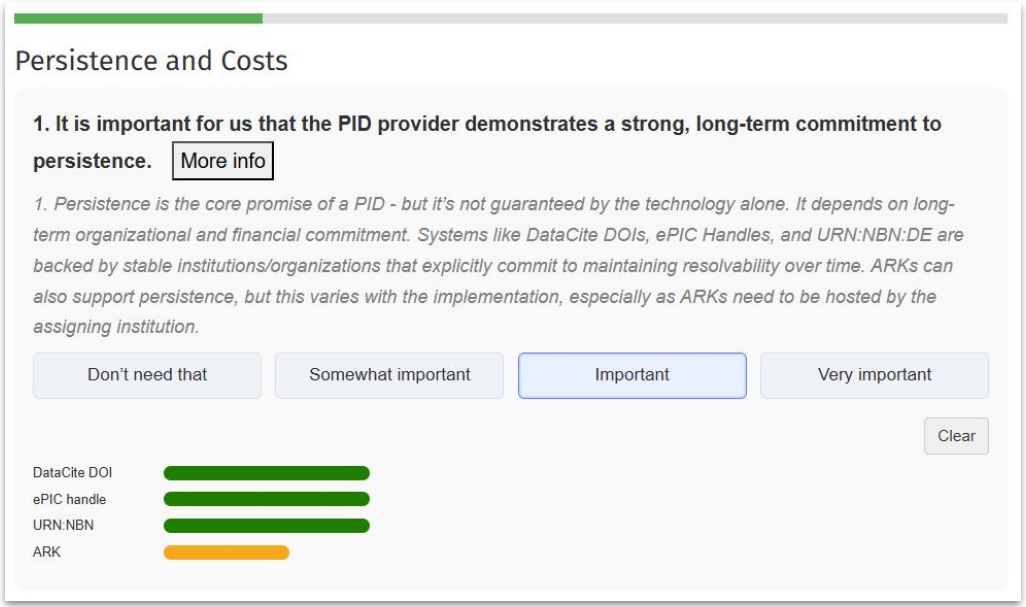
- Prefix registration service for PID use-cases
- B2Inst
- Data Type Registry (EOSC)
- PID Metaresolver (EOSC)
- Compliance Assessment Toolkit (EOSC)

Governance

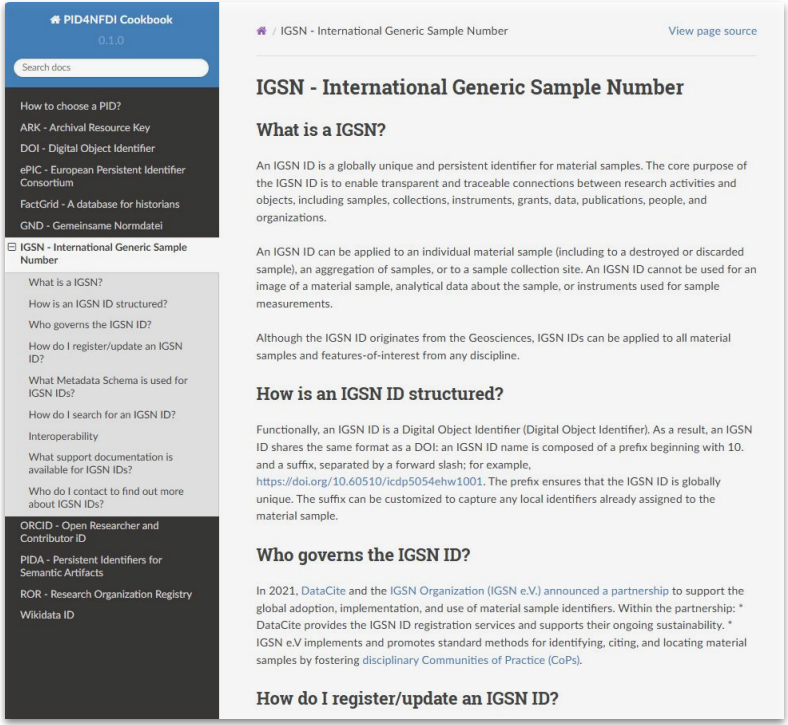
- Tool for PID provider selection
- NFDI-wide PID guidelines and compliance testing

Use Cases | **Focus Groups** | **Outreach**

PID Selection Tool & PID4NFDI Cookbook



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



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

Metadata Insights

1- Metadatenvollständigkeit

DataCite-Metadatenschema-Struktur

6 Mandatory

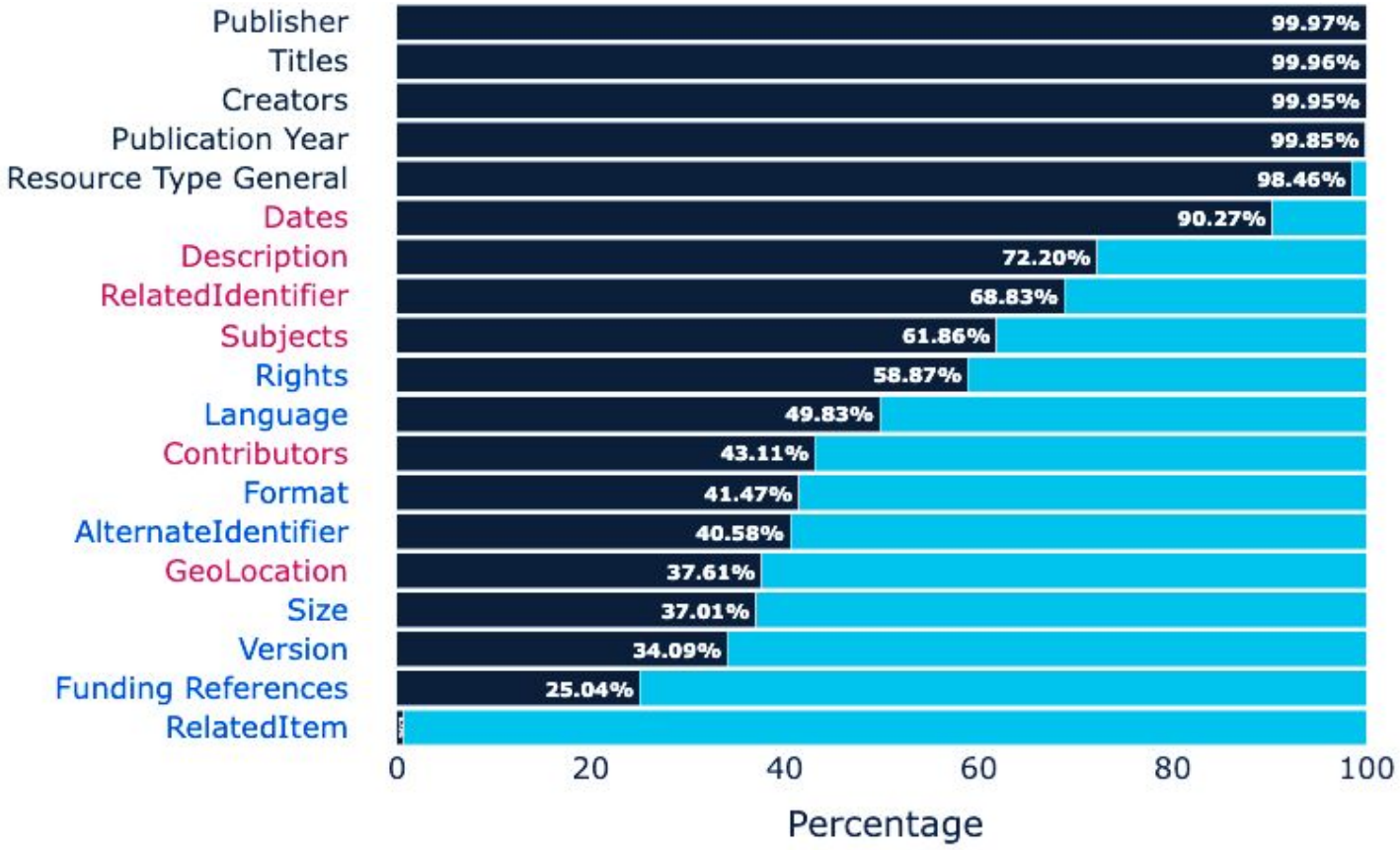
ID	Property
1	Identifier
2	Creator
3	Title
4	Publisher
5	PublicationYear
10	ResourceType

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

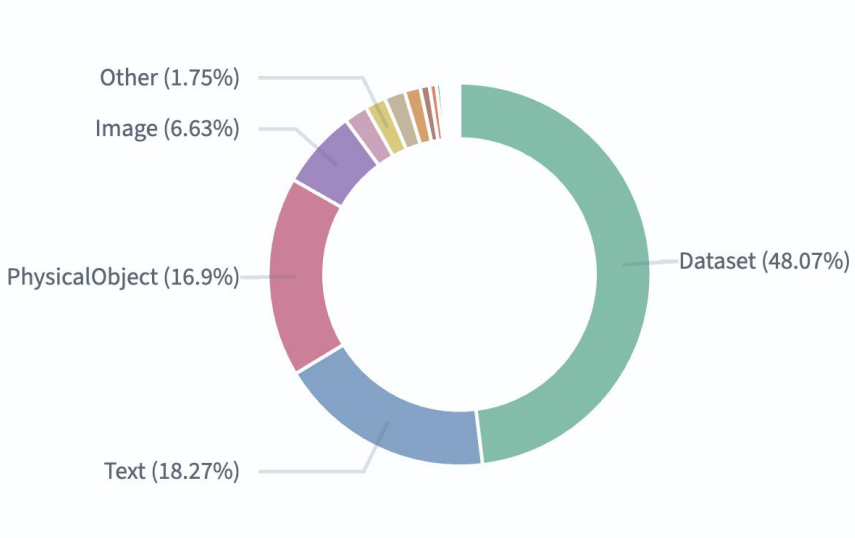
https://schema.datacite.org/

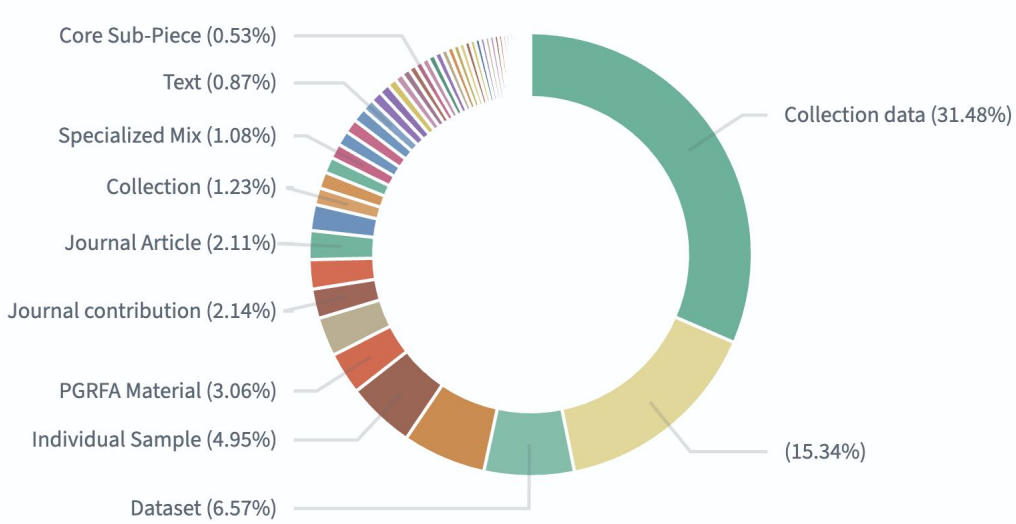
DataCite-Metadatenvollständigkeit



Mandatory
Recommended
Optional

Welche Ressourcentypen werden derzeit am häufigsten in DataCite publiziert?





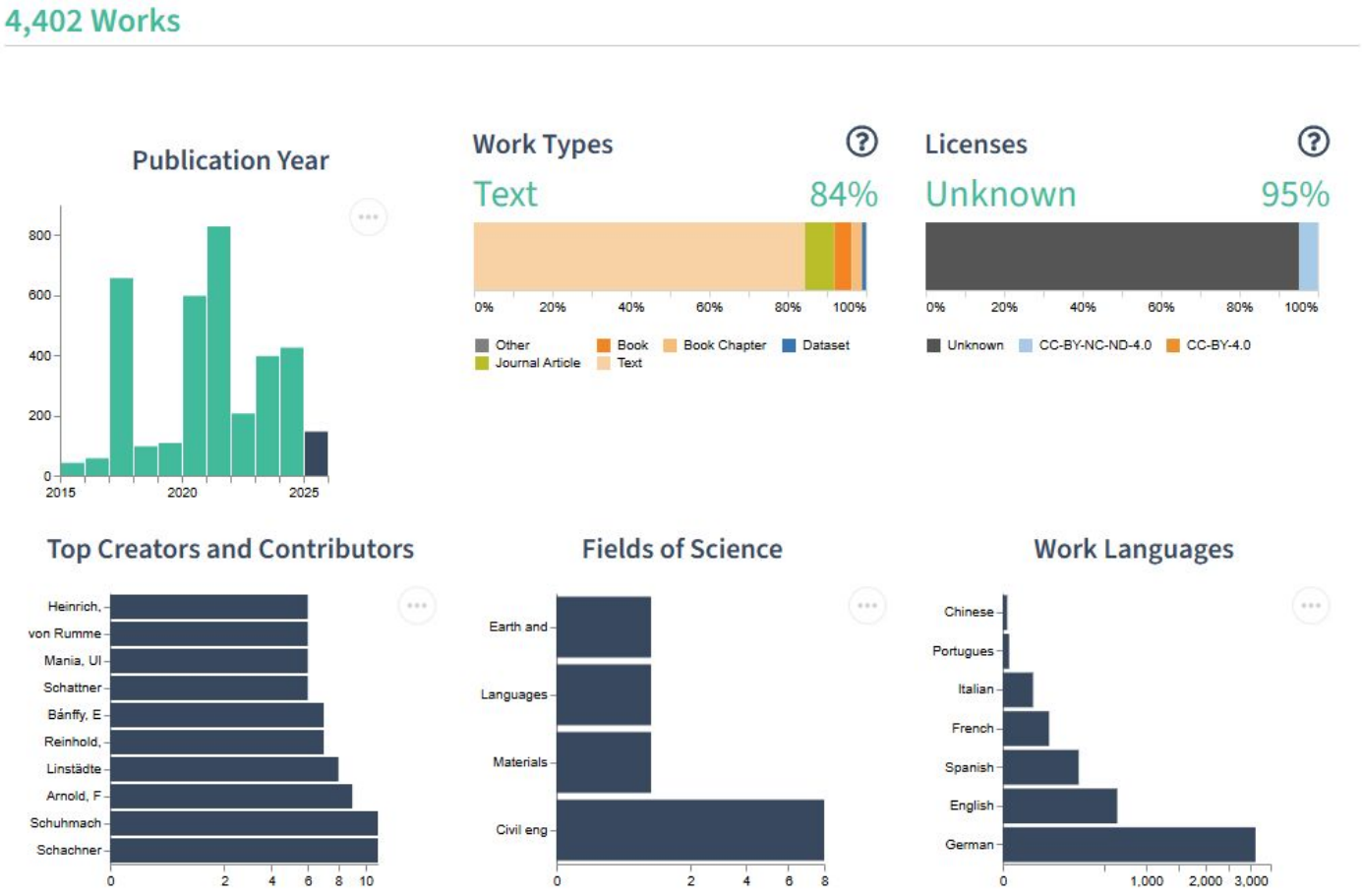
resourceTypeGeneral = kontrolliertes Vokabular

resourceType = Freitextangaben durch Nutzer*innen

Licensing Information

How much research is being made openly available for reuse?

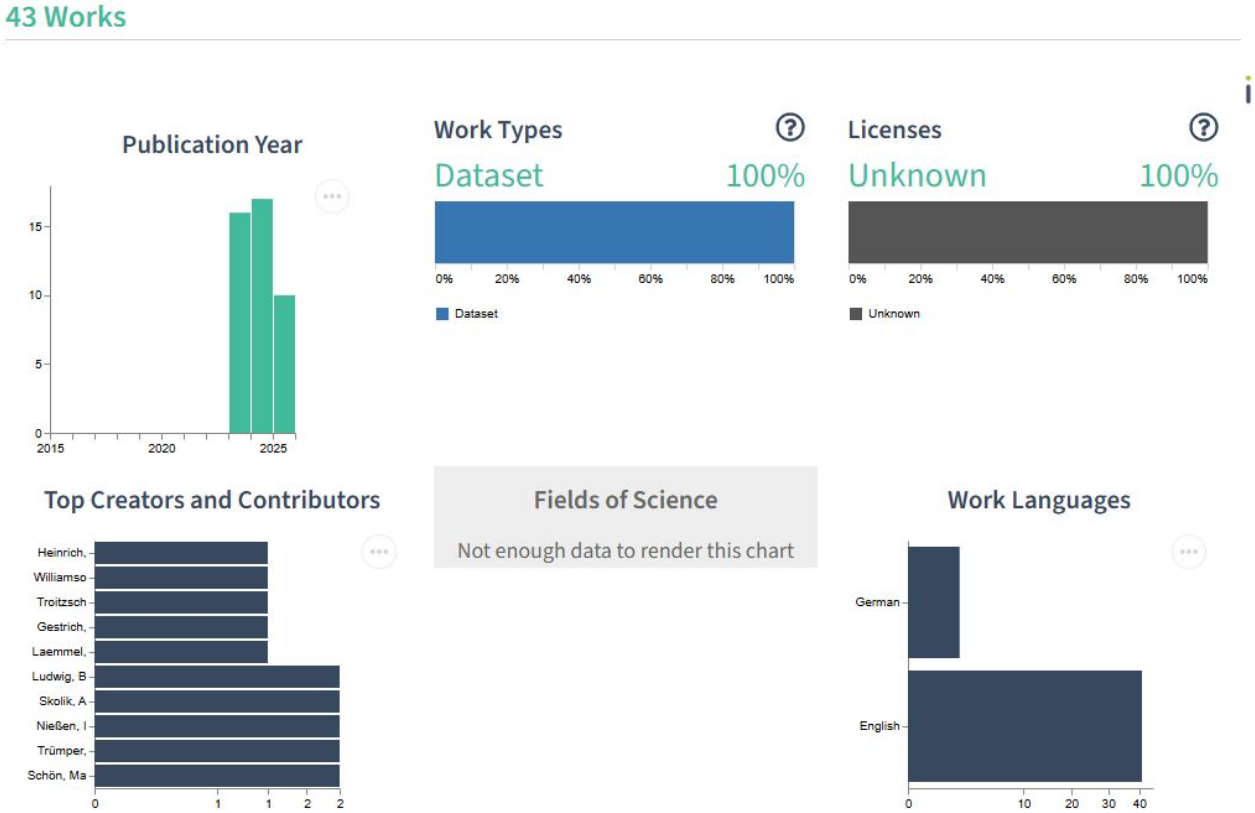
DAI Repo bei DataCite



https://commons.datacite.org/repositories/8orczey

DAI Repo bei DataCite

nur resource-type: Datensatz



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

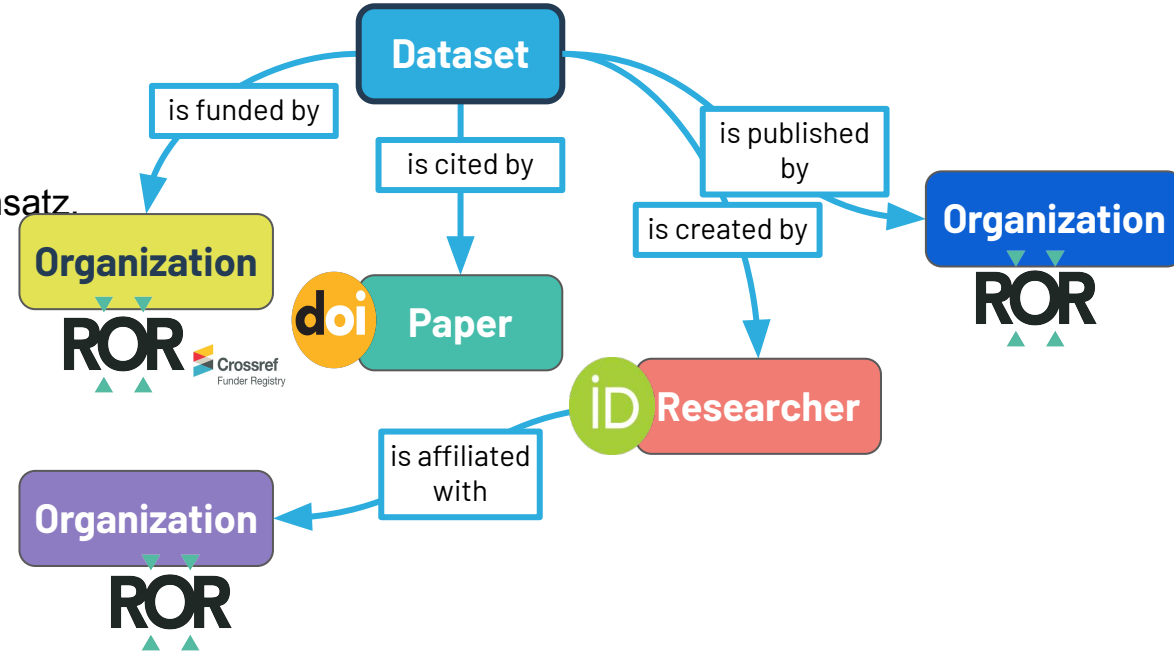
2- PID Beziehungsmetadaten

Was sind Beziehungsmetadaten?

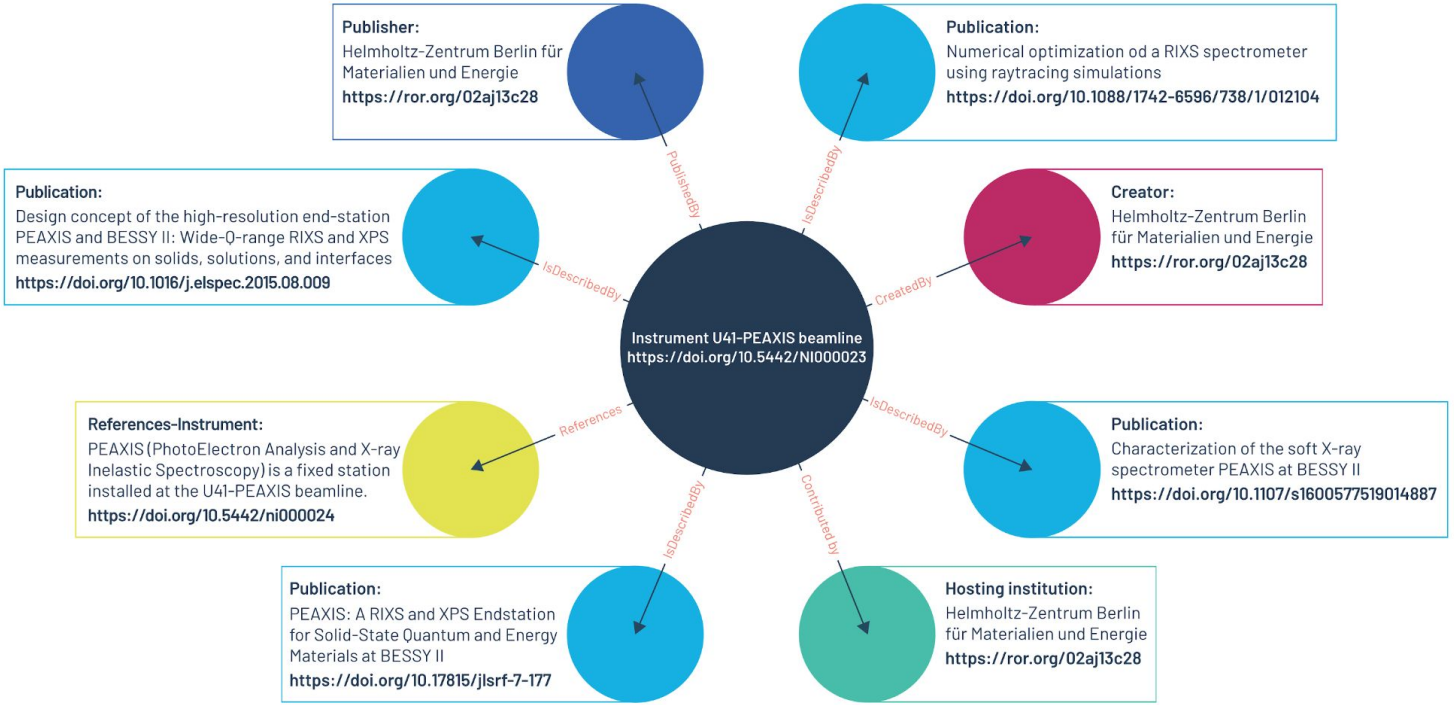
Metadaten, die Beziehungen – also Verbindungen – zwischen Entitäten abbilden

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?



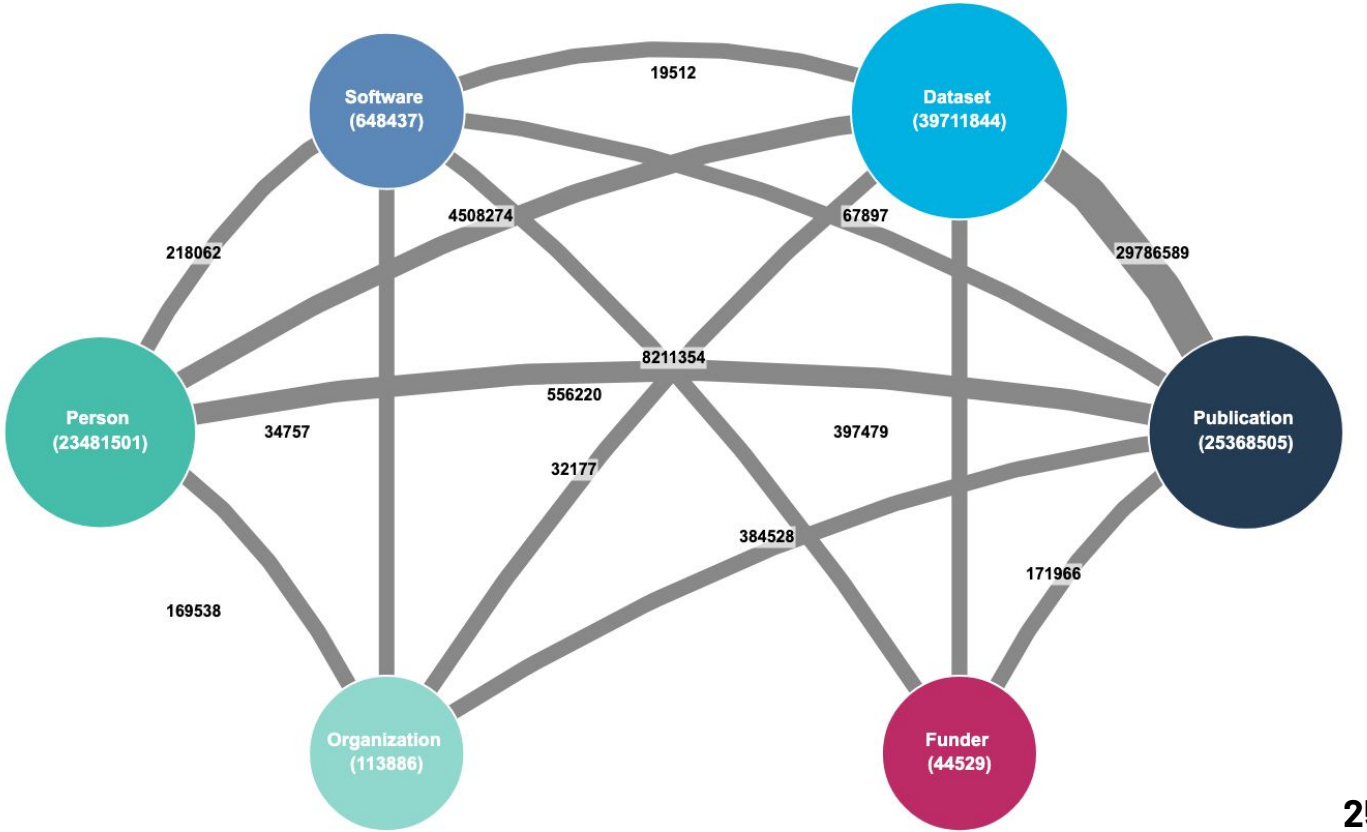
https://api.datacite.org/dois?resource-type-id=instrument
https://api.datacite.org/dois/10.5442/ni000023

Forschung ist ein Graph

Verbunden durch die Möglichkeiten von Metadaten

base4nfdi

PID4nfdi



25th of March 2025

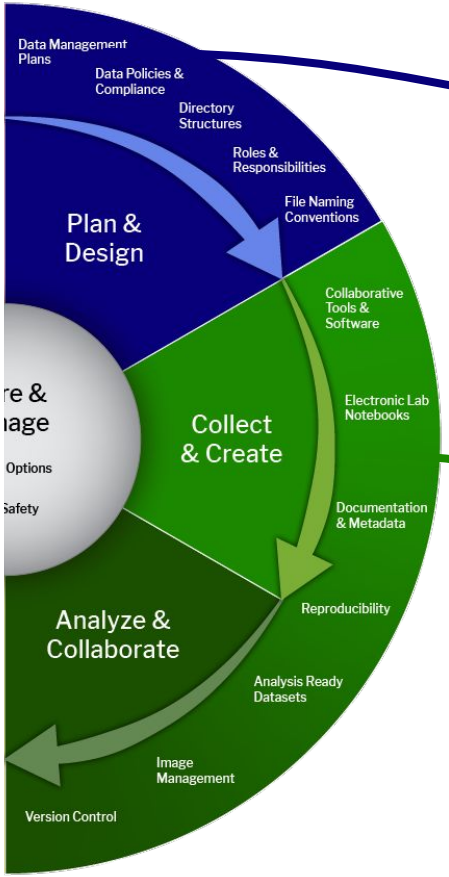
3- PID Metadaten für Interoperabilität

Research Life cycle

https://datamanagement.hms.harvard.edu/plan-design/biomedical-data-lifecycle

- **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://pid.services.base4nfdi.de