



Persistente Identifikatoren in Deutschland – Verbreitung an Kultur- und Wissenschaftseinrichtungen –

.: Monitoring von PIDs - der erste Prototyp .:

.: PID-Monitor .:

Kolloquium Wissensinfrastruktur

17. Januar 2025, Universitätsbibliothek Bielefeld

Andreas Czerniak <andreas.czerniak@uni-bielefeld.de> , 0000-0003-3883-4169



- Einführung in das Projekt
- Einblick in die quantitative Erhebung zu PIDs an deutschen Kultur- und Wissenschaftseinrichtungen
- PID - Monitor (Prototyp)
 - Konzept
 - Live-Demo
 - Technischer Überblick
- Ausblick

Einführung in das Projekt

- Projekt: PID Network Deutschland
 - gefördert von der DFG für 36 Monate, Projekt-Nr. [506475377](#)
 - Start: 01. März 2023
 - Antrag: <https://doi.org/10.48440/os.helmholtz.059>
 - Webseite: <https://www.pid-network.de>
- Ziele
 - Etablierung eines Netzwerkes aus bereits bestehenden und sich aktuell formierenden Akteuren in Wissenschaft und Kultur.
 - Förderung und Konsolidierung der Anwendung Implementierung, Standardisierung und internationale Anschlussfähigkeit von PID-Systemen auf lokaler, nationaler, europäischer und internationaler Ebene.
 - Entwicklung einer nationalen PID-Roadmap.
- Projektpartner-Institutionen:



10 PID-Anwendungsgebiete



Forschungsdaten



Instrumente



Kulturelle
Objekte



Organisationen
& Projekte



Personen



Physische
Objekte



Publikations-
dienste
& FIS



Software



Text-
publikationen



Veran-
staltungen

Einblick in die quantitative Erhebung zu PIDs

an deutschen Kultur- und Wissenschaftseinrichtungen

- Laufzeit:
 - Start: 11. März 2024 , Ende: 31. Mai 2024
- 5 Abschnitte:
 - Allgemeines
 - Verbreitung und Nutzung von PIDs
 - Informations- und Schulungsbedarfe
 - Bedarfe und Lücken
 - Einsatz von PIDs nachhaltig fördern
- Anzahl persistenter Identifikatoren: 21
- Fragenanzahl: 21 + kontext-basierende Unterfragen
- benötigte Zeit zur Beantwortung: ca. 15-20 Minuten

- Zielgruppen
 - Hochschulen & Universitäten (HK: öffentlich-rechtlich)
 - 274
 - Außeruniversitäre Forschungseinrichtungen
 - inkl. deutsche Knoten von europäischen Infrastrukturen
 - 295
 - Ressortforschung des Bundes
 - 51

Gesamt: 620 Einrichtungen

- Erreichbarkeit: Digital

- Versendet an
 - Hochschulen & Universitäten (öffentlich-rechtlich)
 - 264 (>96%)
 - Außeruniversitäre Forschungseinrichtungen, inkl. deutsche Knoten von europäischen Infrastrukturen
 - 205 (>69%)
 - Ressortforschung des Bundes
 - 32 (>64%)

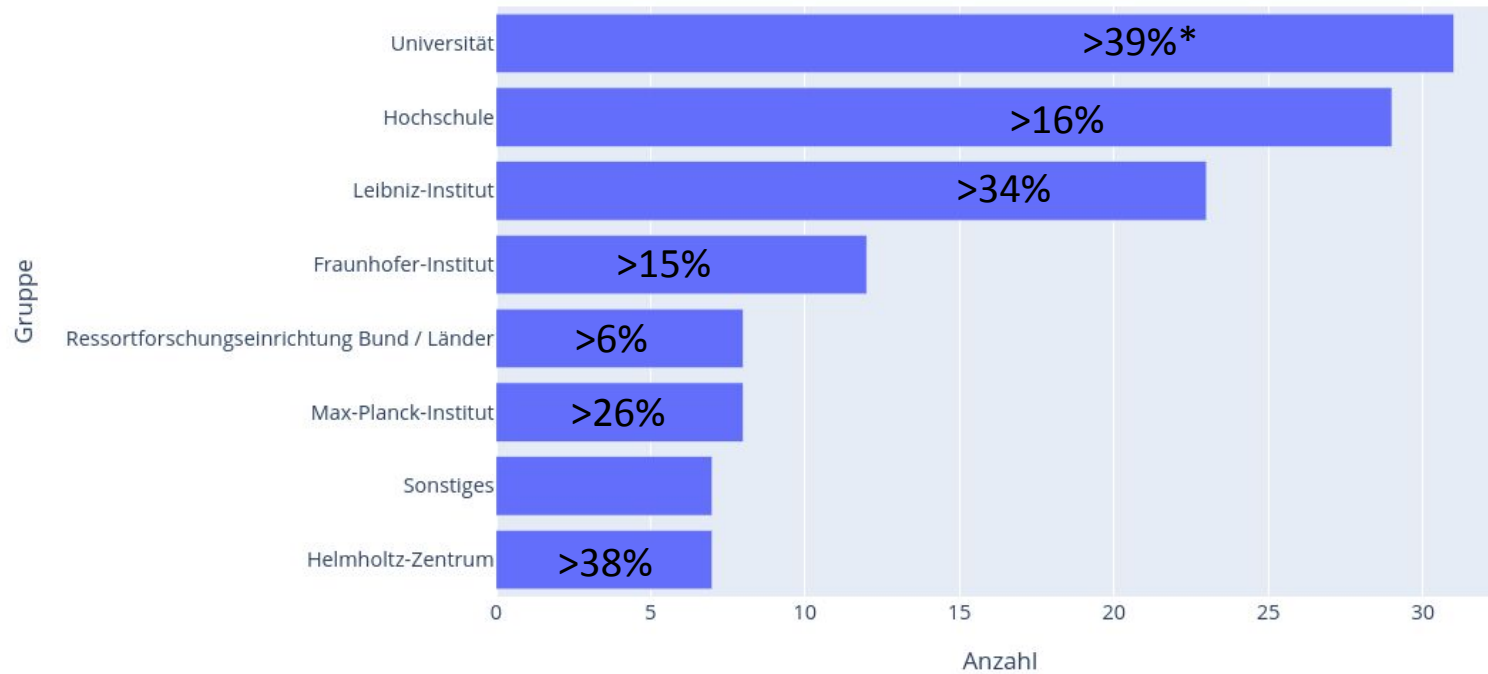
Gesamt: 501 (>80%)

- Vollständig ausgefüllte Fragebögen: 125 (~25%)

Auswertung: Allgemeines, Verbreitung und Nutzung

Zugehörigkeit

Welche Zugehörigkeit hat ihre Einrichtung? [n=125]



Angeschrieben	Geantwortet (* korrigiert)
79	31
180	30**
67	24**
79	12
34	8
18	7

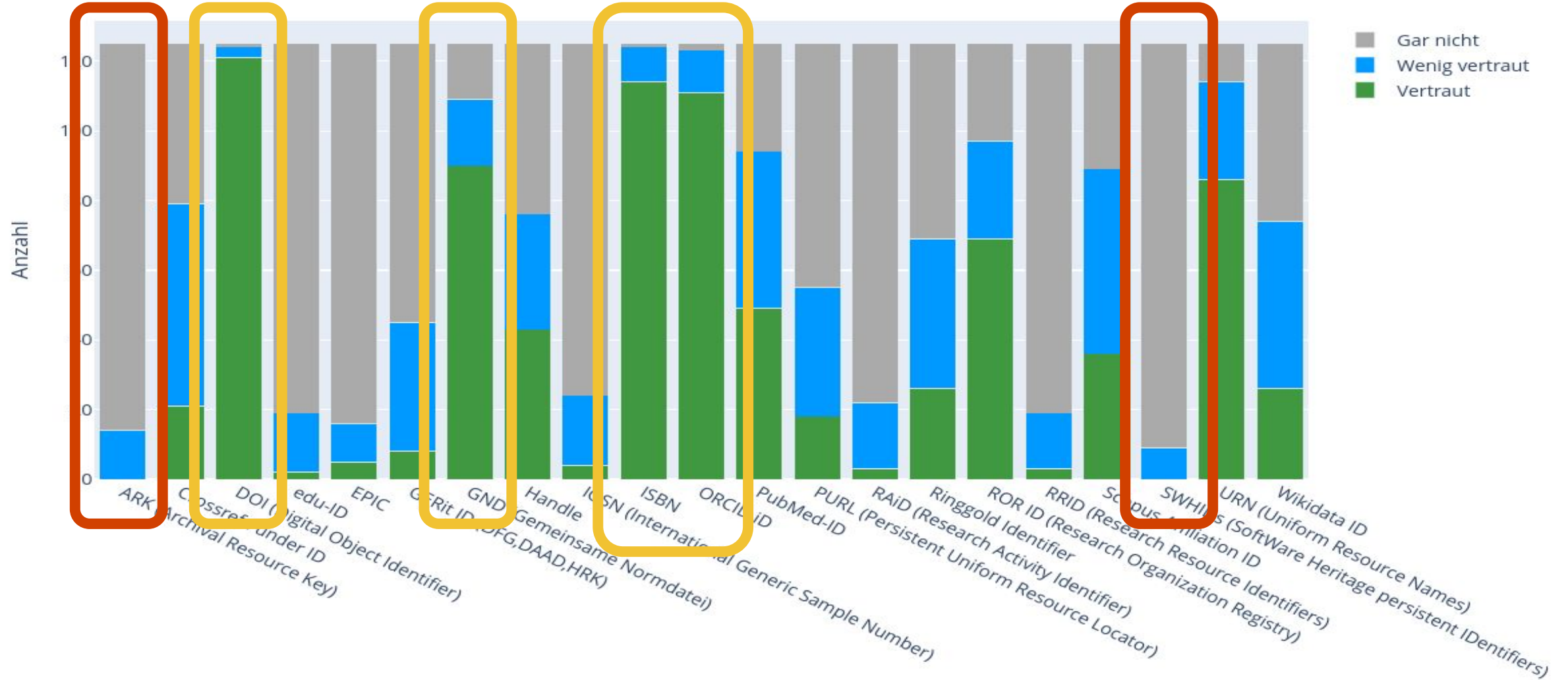
* Prozentuales Verhältnis zu den erreichten Einrichtungen.

** "Sonstige" Zuordnung von "Leibniz-Institut und Universitätsbibliothek" zu "Leibniz-Institut".

Vertrautheit mit PIDs

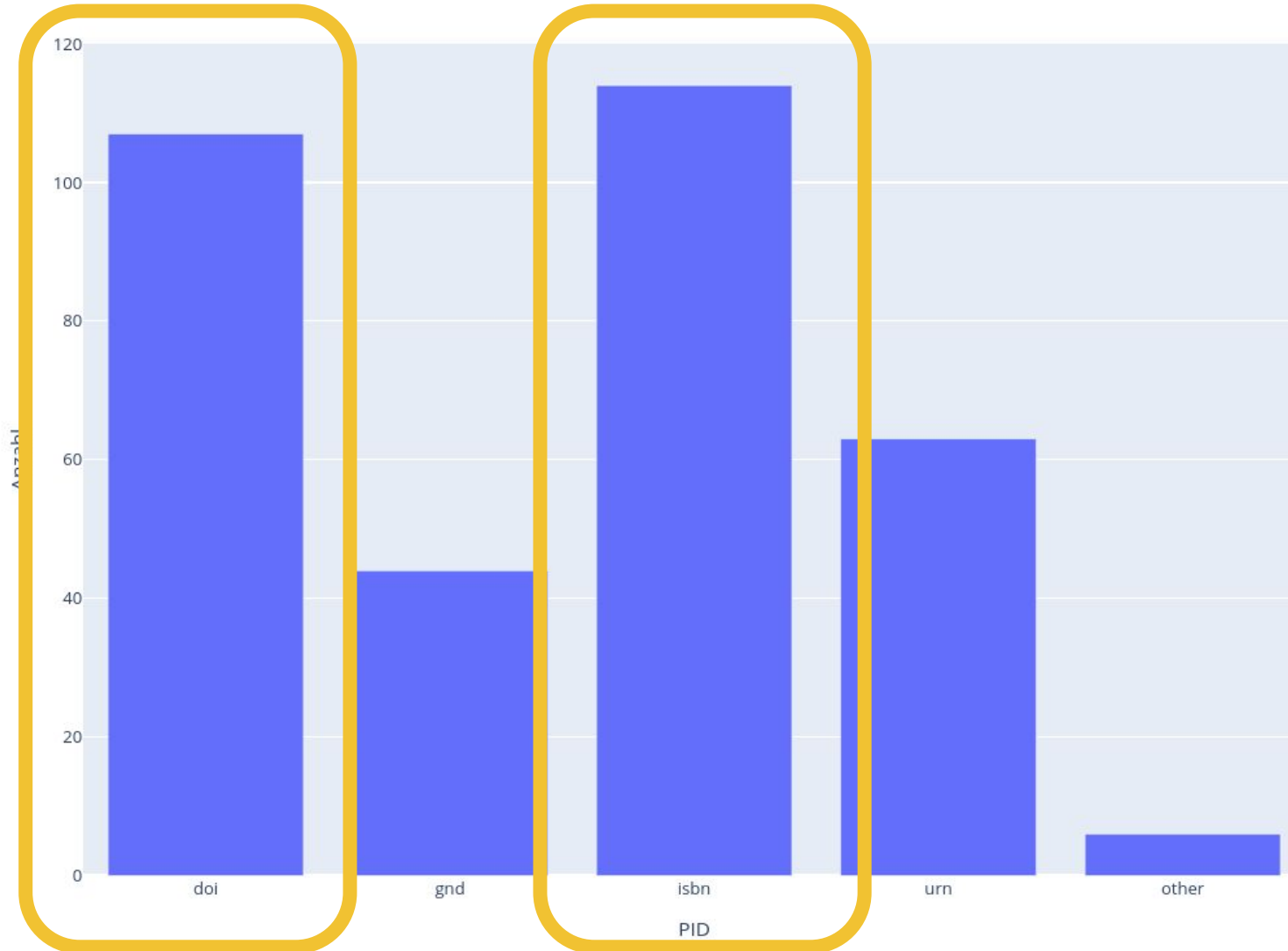


Vertraut mit PIDs? [n=125]



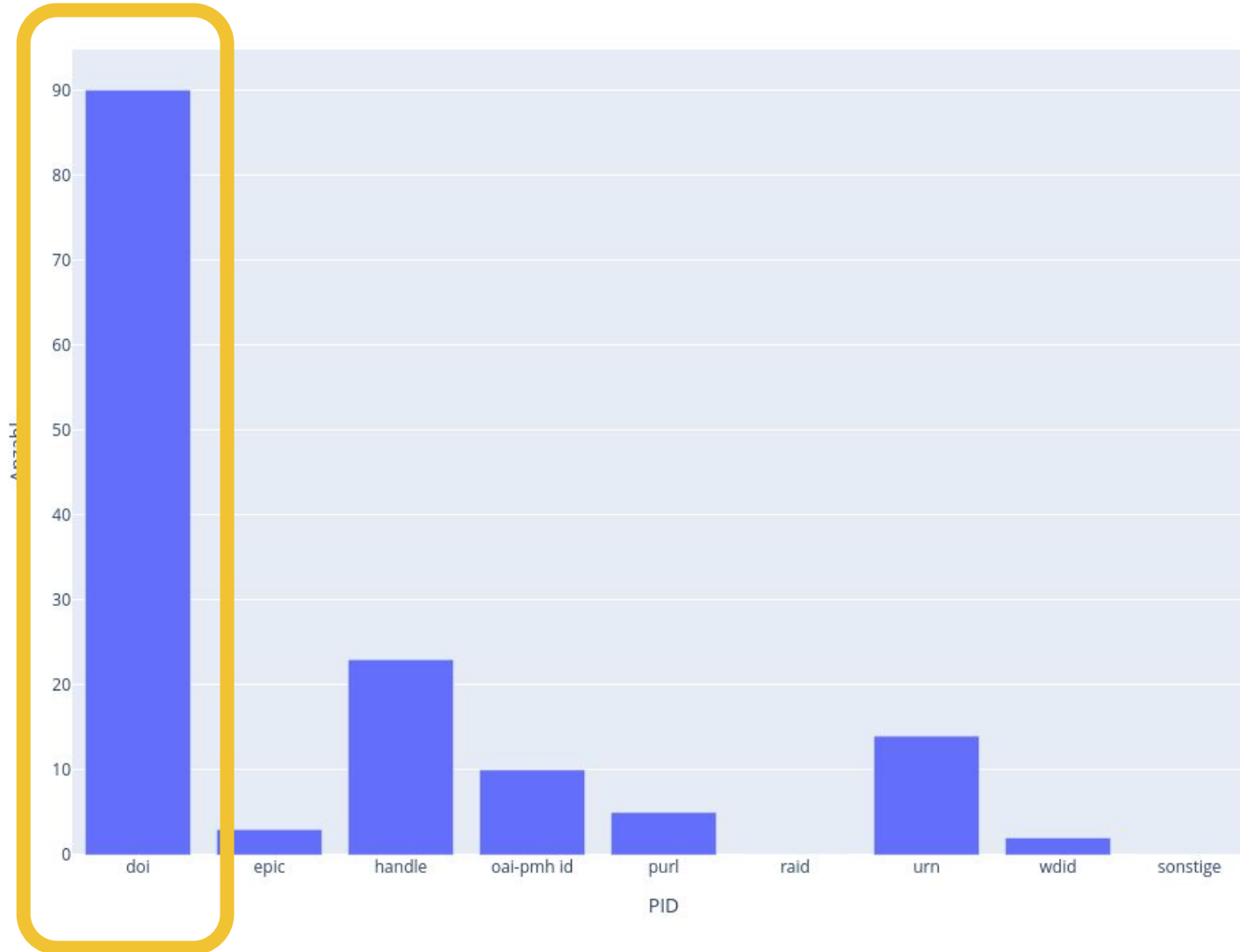
Welche PIDs werden genutzt für: Bücher

Welche PIDs werden für Bücher an Ihrer Einrichtung genutzt? (n=125)



Welche PIDs werden genutzt für: Forschungsdaten

Welche PIDs werden für Forschungsdaten an Ihrer Einrichtung genutzt? (n=125)



Der PID-Monitor

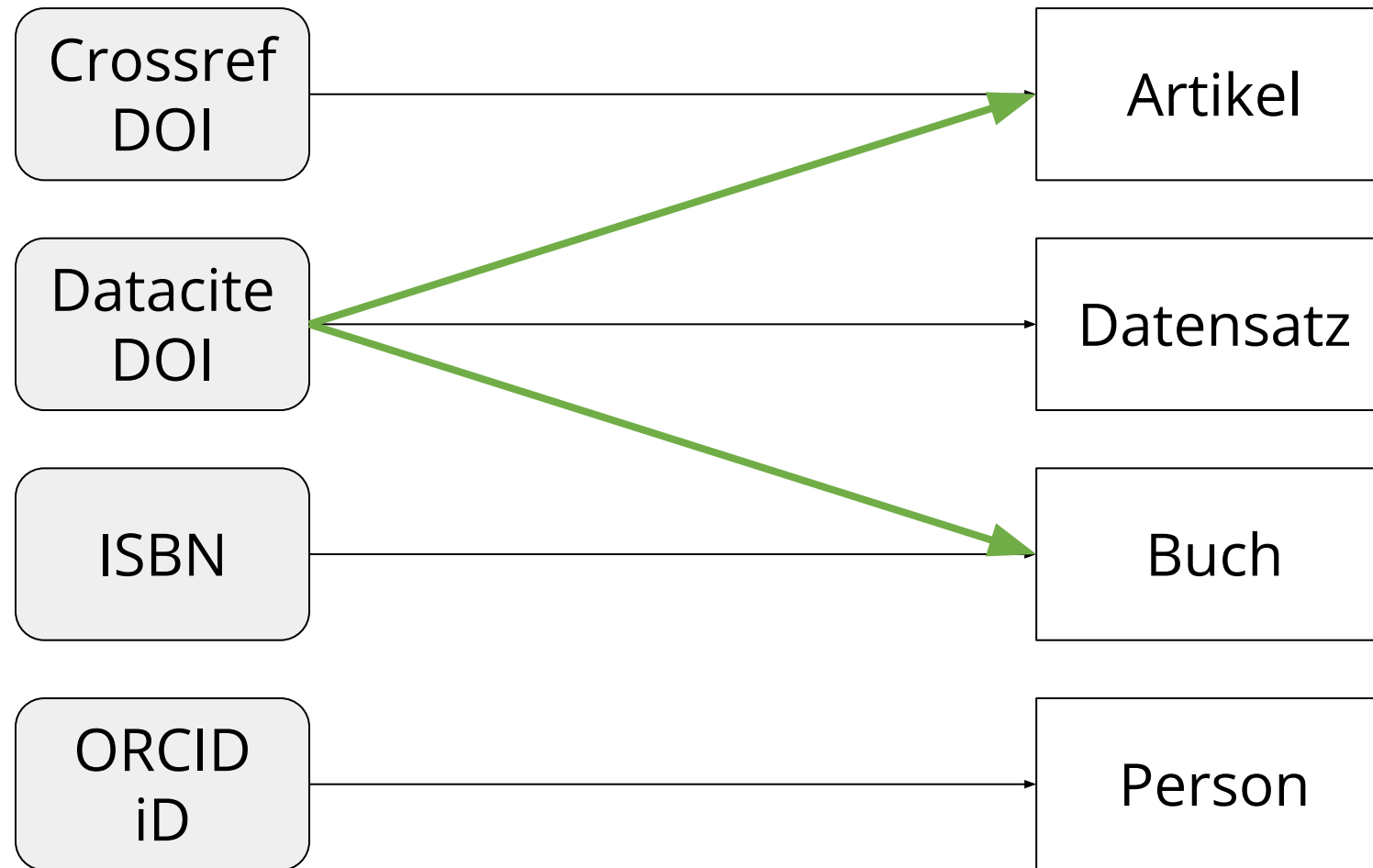
laut Projekt-Beschreibung:

Aufgabe und Ziel des Werkzeugs ist **das kontinuierliche Monitoring über die Verbreitung der verschiedenen PID-Typen** in Publikations-, Informations- und Forschungsinfrastrukturen und die **Verknüpfung von PIDs im zeitlichen Verlauf.**

PID - Sichtweisen



PID - Nutzung



(Teil-)Übersicht PID-Typ & Entität

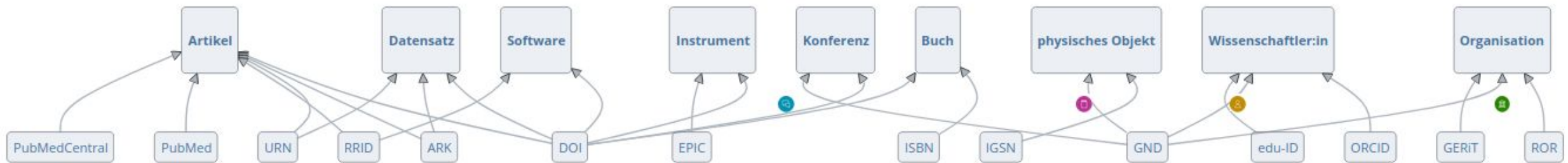
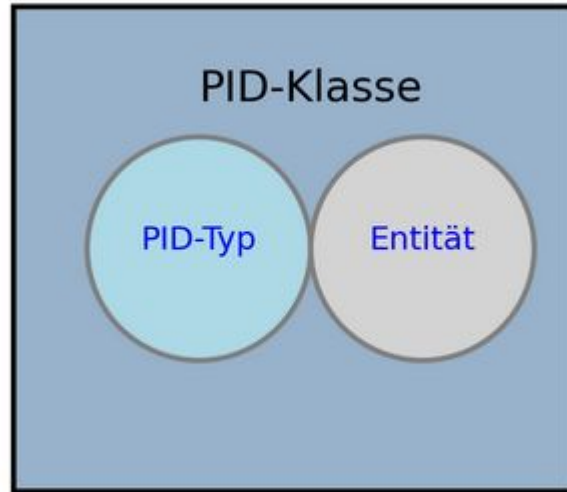


Figure 1: Beispielhafte Übersicht von PID-Typ und Ressourcety/Entität

Die Kombination aus **PID-Typ und Entität/Ressourcety** wird zu einer **PID-Klasse** verknüpft. Abbildung [2](#) illustriert diese Kombination.

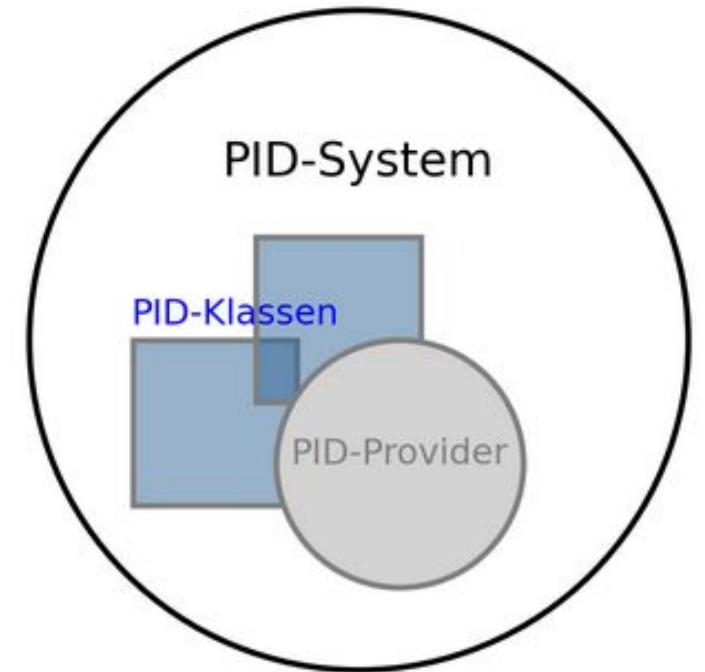
https://pid-monitor.org/pid_class.html



PID-Typ: Repräsentiert persistente Identifikatoren, welche ihre eigene Semantik, Syntax und Komplexität haben, wie DOI, ORCID iD, ROR.

Entität: Beschreibt allgemein etwas, das eine eigenständige Existenz besitzt, unabhängig davon, ob diese materieller oder immaterieller Natur ist.

Entität = Ressourcentyp



https://pid-monitor.org/pid_class.html

- Aufbau (multitier)
 - Frontend: Dashboard
 - Backend:
 - Abruf von Daten an Endpunkten im Batch-Verfahren
 - Aufbereitung der Zeitreihen
 - Aggregation
 - Statistik-Analysen
 - Bereitstellung per API
- Domain: **pid-monitor.org**

.: Live-Demo :.

Technische Aspekte

- Dashboard
 - die Implementierung basiert auf dem Framework [Quarto](#), einem Open Source Tool für wissenschaftliche und technische Veröffentlichungen.
 - Quarto ermöglicht die Einbettung von Code-Snippets, zB. Jupyter-Notebooks in Python, R und Julia.
 - verschiedene Mechanismen für die Ausführungsumgebung und verschiedene Ausgabeformate
 - Ergebnis: HTML

- Einbindung in die UB-IT Landschaft
 - Entwicklung mit IDEs, wie VS-Code oder R-Studio
 - Versionierung im GitLab.UB Repository
 - unterschiedliche Branches für Produktiv und BETA
 - automatische Erstellung der Container
 - Orchestrierung im Kubernetes-Cluster der UB
 - einheitliche Bereitstellung von robusten, skalierbaren Diensten
 - Faktoren der “benutzerfreundlichen Bereitstellung” **nicht zu Ungunsten** von Sicherheits- und Wartungs-Aspekten

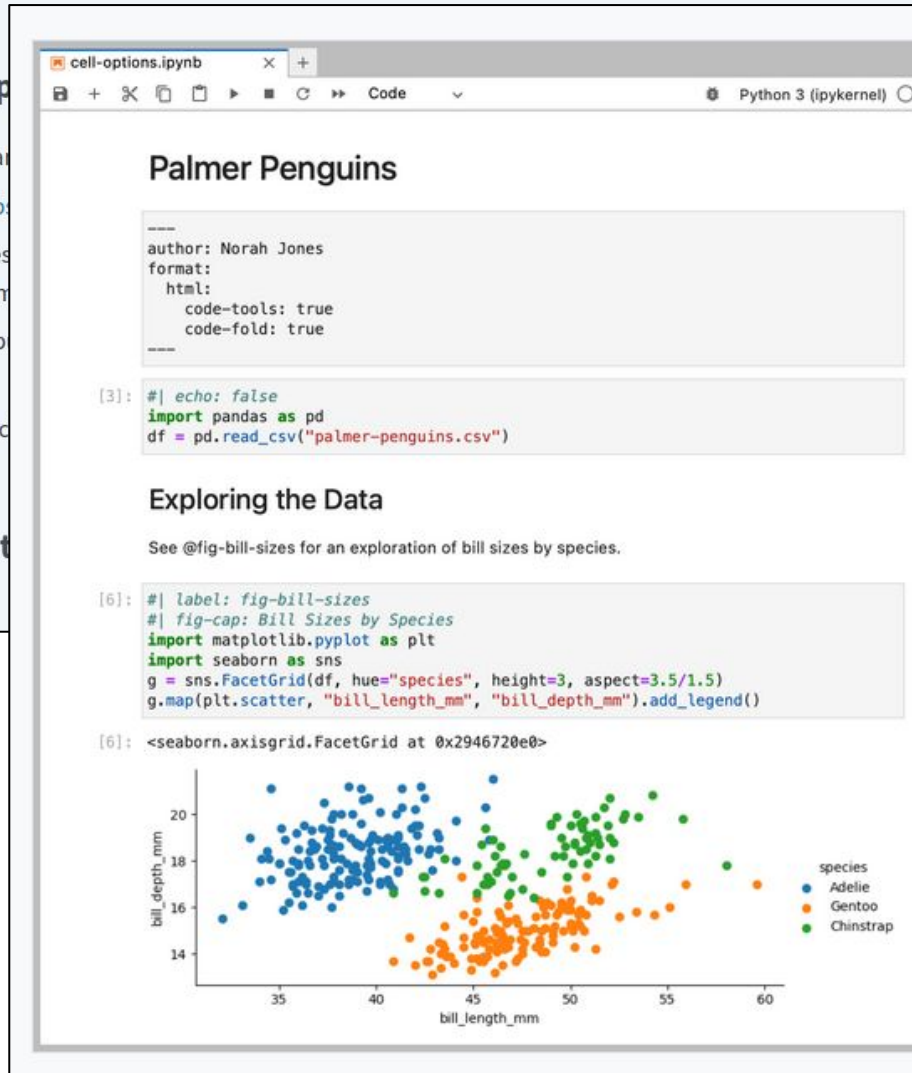
Welcome to Quarto

An open-source scientific and technical publishing platform

- Author using [Jupyter](#) notebooks or with plain text markdown
- Create dynamic content with [Python](#), [R](#), [Julia](#), and [Observable](#)
- Publish reproducible, production quality articles, presentations, blogs, and books in HTML, PDF, MS Word, ePub, and more
- Share knowledge and insights organization-wide by publishing to [Confluence](#), or other publishing systems.
- Write using [Pandoc](#) markdown, including equations, code callouts, advanced layout, and more.

Analyze. Share. Reproduce. You have a story to tell. Tell it with Quarto.

<https://quarto.org>



```
cell-options.ipynb Python 3 (ipykernel)
author: Norah Jones
format:
  html:
    code-tools: true
    code-fold: true
[3]: #| echo: false
import pandas as pd
df = pd.read_csv("palmer-penguins.csv")

Exploring the Data
See @fig-bill-sizes for an exploration of bill sizes by species.
[6]: #| label: fig-bill-sizes
#| fig-cap: Bill Sizes by Species
import matplotlib.pyplot as plt
import seaborn as sns
g = sns.FacetGrid(df, hue="species", height=3, aspect=3.5/1.5)
g.map(plt.scatter, "bill_length_mm", "bill_depth_mm").add_legend()
[6]: <seaborn.axisgrid.FacetGrid at 0x2946720e0>
```

Palmer Penguins

AUTHOR
Norah Jones

PUBLISHED
March 12, 2023

Show All Code

Hide All Code

Exploring the Data

See [Figure 1](#) for an exploration of bill sizes by species.

Code

```
import matplotlib.pyplot as plt
import seaborn as sns
g = sns.FacetGrid(df, hue="species", height=3, aspect=3.5/2)
g.map(plt.scatter, "bill_length_mm", "bill_depth_mm").add_legend()
```

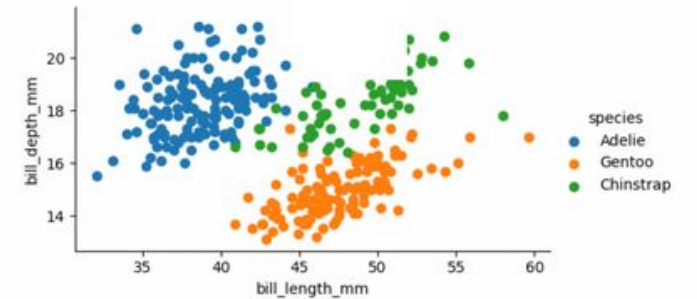


Figure 1: Bill Sizes by Species

Ausblick

- weitere Einbindung von PID-Typen <-> Entitäten aus BASE.
- Funktionserweiterung bei Visualisierungen.
- Ausarbeitung des Konzeptes um den PID-Typ und die Entität, sowie die Dokumentation der Methoden.
- Etablierung von Relationen und deren Auswertung.
- Nachnutzung der Daten durch Bereitstellung via API.
- Launch des PID-Monitors: **Ende Januar 2025.**

Abschluss - PID-Community





Mitwirkende

Barbara Fischer (DNB), Steffi Genderjahn (Helmholtz Open Science Office),
Stephanie Hagemann-Wilholt (TIB), Marc Lange (Helmholtz Open Science Office),
Antonia C. Schrader (Helmholtz Open Science Office), Paul Vierkant (DataCite),
Frauke Ziedorn (TIB), Stephanie Glagla-Dietz (DNB), Vitali Peil (UB Bielefeld)





Vielen Dank für eure und Ihre Aufmerksamkeit!

<https://pid-monitor.org>

www.pid-network.de

Kontakt: pid-monitor@uni-bielefeld.de

Folgen Sie uns gerne auch auf Social Media:

Mastodon ([@PIDNetworkDE@openbiblio.social](https://openbiblio.social/@PIDNetworkDE))

LinkedIn (<https://www.linkedin.com/showcase/pid-network-de>)