

Abstract

Persistent Identifiers (PIDs) have become indispensable in the world of research information. Standards are beginning to be established for publications, persons, or organizations – the best known being DOI, ORCID, and ROR. Initiatives for additional PIDs, such as for devices, events, or scientific awards, are working toward their implementation.

The main purpose of PIDs is to uniquely identify objects. Persons with identical names become distinguishable, variations in the spelling of journal article titles lose their relevance, and authors' affiliations are clearly referenced.

But what is the informational added value of PIDs? This question is easily answered with regard to DOIs, because they generally provide direct access to the corresponding information resource—such as a journal article—when referenced online (provided the corresponding access rights exist).

However, what about access to information that is of value for research reporting, or more generally for empirical science studies, or even simply for scientific statistics?

Such access exists, but is often little known. The presentation therefore primarily aims to make the range of possibilities visible. The central question is how PIDs can help to gain access to corresponding data sources. The presentation also highlights the special importance of translation or concordance tables between different PIDs.

Our presentation focuses on data resources that have so far remained largely hidden. Special emphasis is placed on the knowledge base Wikidata, the Integrated Authority File (GND) of the German National Library, the Directory of German Research Institutions (GERiT), and the emerging PID system for scientific awards and prizes (ORAP) of the Commission for Research Information in Germany (KFiD) [1].

After a general introduction to the topic, practical examples will illustrate the fundamental possibilities that arise from the skillful use of PID systems for accessing original data sources.

Only briefly mentioned at the beginning is the well-known use case for DOIs: if analyses of bibliometric indicators—such as citation frequency, international co-authorships, or thematic focuses—are to be carried out for a defined set of publications with known DOIs, this is easily possible through access to relevant databases, provided the necessary knowledge is available. The presentation particularly highlights the possibilities offered by the growing spread of non-commercial bibliographic databases such as OpenAlex [2].

A main focus of the presentation is on Wikidata. Wikidata is an open, collaboratively editable knowledge base operated by the Wikimedia Foundation that provides machine-

readable, structured data for all Wikimedia projects and beyond. Wikidata is a central source for facts such as birth dates and geographical information, which can be used in different languages and edited collaboratively by users worldwide.. Such data are provided in a structured format that allows easy use for various purposes—including, as discussed here, for in-depth analysis for research statistical purposes.

With regard to Wikidata, the topics of PIDs and data access are closely intertwined—Wikidata directly provides access to the relevant data. The second main focus of the presentation concerns a PID whose particular appeal lies in its broad use across a wide range of data collections. This refers to the identifier of the Integrated Authority File (GND) of the German National Library, focusing here on its use as a person identifier. In the CRIS and PID communities, the GND-ID operates somewhat in the shadow of the ORCID identifier, and only a few use cases are widely known [3]. The introductory presentation of the GND identifier will therefore first outline how it is assigned, what international equivalents exist (VIAF, ISNI), and what efforts are being made to develop interfaces between ORCID and GND. The focus then shifts to examples of data collections that are accessible via GND for statistically meaningful analyses.

As a third use case, we then turn to a PID for research institutions. The focus of this analysis, which concentrates on German institutions, demonstrates the possibilities offered by the institutional identifier developed by the German Research Foundation (DFG), the DFG-INST-ID—both on its own and in combination with the ROR identifier, which covers the global research landscape, as well as the institution identifier used in Wikidata. Emphasis is also placed on the possibilities for detailed analysis arising from the fact that the DFG-INST-ID is assigned not only to institutions as such but also to their subunits (as can be seen in the information system GERiT, www.gerit.org/de).

The presentation concludes with an outlook on a new data source currently being developed by the Commission for Research Information in Germany (KFiD). The goal of this collection is to make data on scientific awards and prizes accessible via a website administered by the office of the Research Core Data Set (KDSF) as an “Open Register for Awards and Prizes (ORAP).” The presentation outlines the genesis of this data collection, its objectives, and the perspectives for its further development. Finally, it sketches the analytical possibilities that arise from the data. Wikidata is also one of the main sources for ORAP. As an example of how data on scientific awards and prizes can be used for empirical science research, it is shown—using examples of historical research personalities —how data extracted from Wikidata can be used to develop a typology of laureates.

The models presented are largely participatory in nature. They show success when communities come together that are willing to share data and to participate in its quality assurance and ongoing updating. The presentation is therefore also intended as an impetus to invest jointly in the development of meaningful data collections.

References

- [1] See Biesenbender, Sophie, 2025: An update from the Commission for Research Information in Germany-KFiD: the FIS Landkarte portal and version 2 of the KDSF standard. EuroCris Membership Meeting 2025, Leuven. <http://hdl.handle.net/11366/2712>.
- [2] See Mrsulja, Ivan; Popvic, Milos and Dragan Ivanovic, 2024: Supervised research information import into CRISs from global sources. EuroCris Strategic Membership Meeting 2024, Paris. <http://hdl.handle.net/11366/2658>.
- [3] See Walther, Marcus and Bastian Melsheimer, 2019: Linking normative data in Converis. EuroCris Strategic Membership Meeting 2019, Münster. <http://hdl.handle.net/11366/1257>.

Authors

Juergen Guedler, Deutsche Forschungsgemeinschaft, Bonn, Germany juergen.guedler@dfg.de

Richard Heidler, Deutsche Forschungsgemeinschaft, Bonn, Germany richard.heidler@dfg.de