HMC Earth and Environment - Using PIDs in Helmholtz Earth and Environment Data Infrastructures

Emanuel Söding[1] // Martin Weinelt[1] // Pier Luigi Buttigieg[2] // Andrea Pörsch[3]

[1] GEOMAR Helmholtz Centre for Ocean Research
[2] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI)
[3] Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences

Interconnected PID Systems

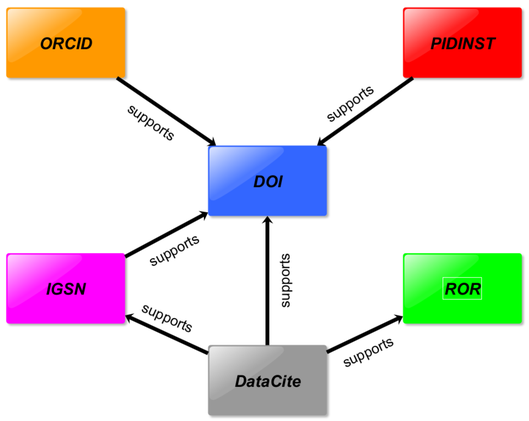
We looked into six established PID systems (ROR, ORCID, PIDINST, IGSN, DataCite DOI, Crossref DOI) to map the interconnection (graph) and overlap between systems. This was carried out by inspecting and comparing the metadata schemas of these PID systems in their current version to find out, to what extent they support each other and how this is done.

The number of external PID systems supported varies considerably for the six PID systems investigated, with ROR at 4 and up to 49 systems at ORCID. The system mostly implemented as a reference is DOI (4 other systems do refer to DOI in their metadata schema), while ROR is only referenced by DataCite DOIs yet.

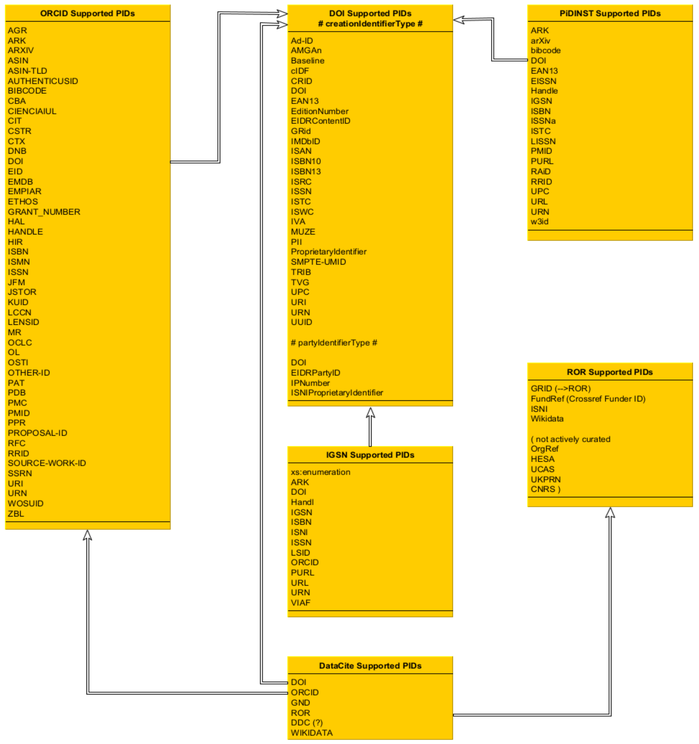
First step: identify the set of external PIDs that is supported for each of the systems (comprehensive list →)

PID System	Number of external PID systems supported
ROR	4
ORCID	49
PIDINST	20
IGSN	13
DOI	33

Interconnection between PID systems:



Detailed graph of PID systems supporting other systems

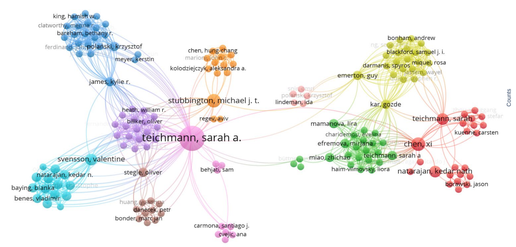


Potential of PID graphs

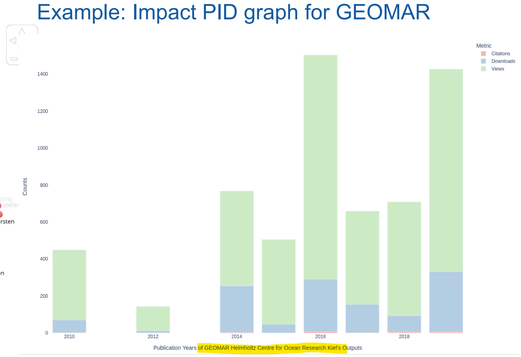
Interconnected PID systems can be visualized as graphs of relationships (PID graphs) between for instance scientists, datasets, publications, institutions etc. They can be machine actionable and thus be tailored to specific questions or fields of interest, as shown by EU program FREYA (https://www.project-freya.eu).

Our findings show, that PIDs should act as an important part of the data space we strive to construct. They allow to link meta information of different data sets in a uniform manner. Consistently implementing PIDs will allow a high level of informational data interoperability among distributed data sets, which should complement other interoperability measures, e.g. the semantic interoperability

FREYA* PID Graphs
Example: researcher co-authors



Example: Impact PID graph for GEOMAR



*PID Graphs establishes connections between different entities within the research landscape, thereby enabling researchers and institutions to access new information. Through the PID Graph, the infrastructure is in place to answer new questions about connections within the research world. [They are established by programmatically performing a] crosswalk through connected PID systems. **

* Cousijn, H. et al.: Connected Research: The Potential of the PID Graph (2021) (https://doi.org/10.1016/j.patter.2020.100180)