HZB Helmholtz Zentrum Berlin

Persistent Identifiers

Rolf Krahl

NOBUGS 2024,
FAIR Data Management Collaboration,
24 September 2024

Persistent Identifiers (PIDs): Definition

- A persistent identifier (PID) is a long-lasting reference to a document, file, web page, or other object. (Wikipedia)
- A PID should:
  - identify some well defined thing,
  - be globally unique,
  - be resolvable, and
  - be persistent.
- There are PIDs for many things: publications, data, people, research organizations, samples, software, and instruments and lot more.
- In short: whenever you want to reference something in an unambiguous, reliable and lasting manner, you want to use a PID.

Persistent Identifiers (PIDs): Definition

- A persistent identifier (PID) is a long-lasting reference to a document, file, web page, or other object. (Wikipedia)
- A PID should:
  - identify some well defined thing,
  - be globally unique,
  - be resolvable, and
  - be persistent.
- There are PIDs for many things: publications, data, people, research organizations, samples, software, and instruments and lot more.
- In short: whenever you want to reference something in an unambiguous, reliable and lasting manner, you want to use a PID.

Important side remark

The "persistence" is not a technical property, but a social contract!

Persistent Identifiers (PIDs): Metadata

- A PID is usually registered along with some metadata (the PID record).
- The PID record should provide at least enough information about the thing it refers to, so one can understand what it is.
- The PID record should be machine readable.
- Very useful information to put in the PID record: references (e.g. PIDs) to other related resources, see next slide for examples.

Persistent Identifiers (PIDs): Purpose

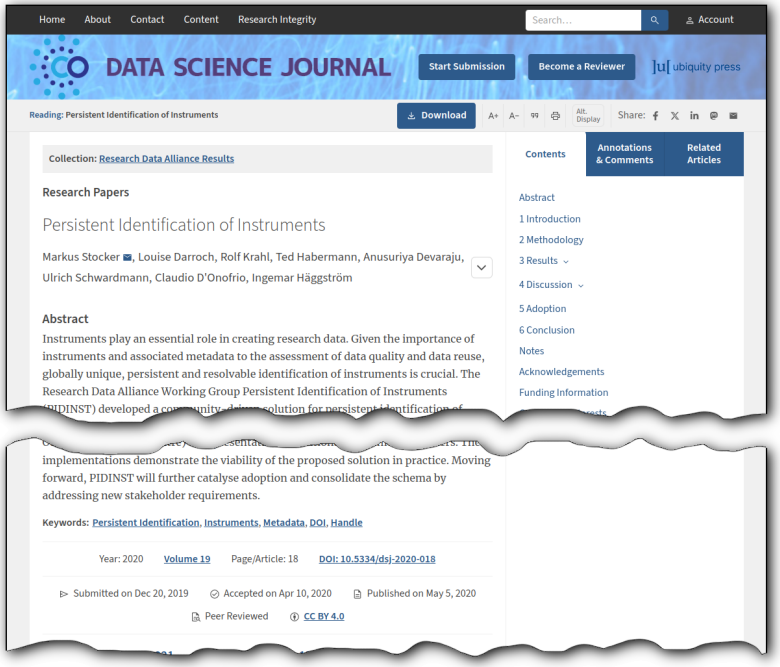
Use cases for PIDs:

- Resolve to the object, e.g. for a journal article from the entry in a reference to the article's landing page.
- Identify and disambiguate things.
- Make robust, machine understandable statements about the identity of things and their related things:
  - "*These* are the authors of the journal article."
  - "This journal article is based on *that* dataset."
  - "The dataset has been obtained from measuring *that* sample."
  - "The measurement has been performed at *that* instrument."
  - "The measurement was using *that* PaN experimental technique."
  - "*This* electronic lab notebook documents what was done for the sample preparation."

  These relations form the so called PID graph.

Relevant PID Types: Research Articles

- Probably the most common and best established PID type.
- Commonly a DOI.



Relevant PID Types: Raw Data

- FAIR Data Principles require that "(meta)data are assigned a globally unique and persistent identifier".
- Needed for being able to unambiguously reference the dataset.
- The raw dataset is the starting point to create most research outputs. Obviously we want to be able to trace it back from those outputs.
- Various PID types: DOI, Handle.

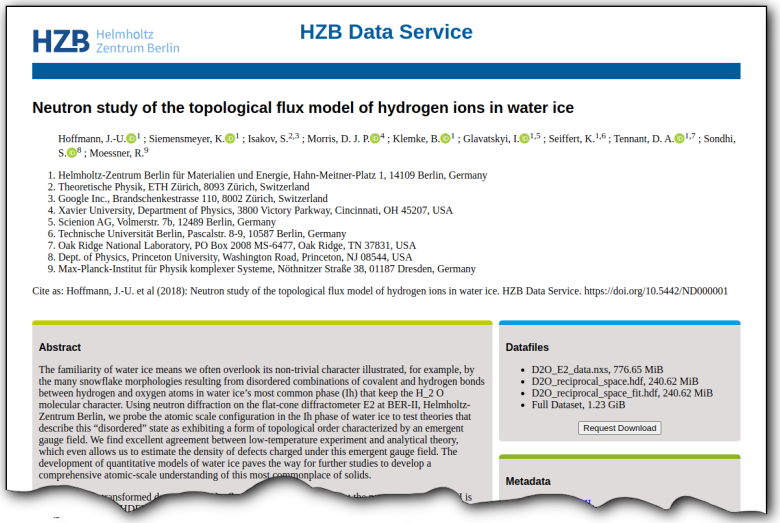
Handle.Net®

Handle Values for: 21.11151/zirq-2z3k

Index	Type	Timestamp	Data
1	URL	2023-01-20 14:50:13Z	https://data.helmholtz-berlin.de/permalink?pid=21.11151%2Fzirq-2z3k
2	identifier	2023-01-20 14:50:13Z	{"identifier": "21.11151/zirq-2z3k", "identifierType": "Handle"}
3	creators	2023-01-20 14:50:13Z	[{"creatorName": ":unav"}]
4	titles	2023-01-20 14:50:13Z	[{"title": "High-temperature investigation on structural changes during order-disorder processes in sanidine megacrystals, K1-xNaxAlSi3O8 (x ≈ 0.17), from the Eifel"}]
5	publisher	2023-01-20 14:50:13Z	Helmholtz-Zentrum Berlin für Materialien und Energie
6	publicationYear	2023-01-20 14:50:13Z	2022
7	subjects	2023-01-20 14:50:13Z	[{"subject": "neutron diffraction", "subjectScheme": "The Photon and Neutron Experimental Techniques Ontology", "schemeURI": "http://purl.org/pan-science/PaNET/", "valueURI": "http://purl.org/pan-science/PaNET/PaNET01217"}]
8	contributors	2023-01-20 14:50:13Z	[{"contributorType": "Proposer", "contributorName": "Schreuer, Jürgen", "nameType": "Personal", "givenName": "Jürgen", "familyName": "Schreuer"}, {"contributorType": "Co-Proposer", "contributorName": "Kähn, Johannes", "nameType": "Personal", "givenName": "Johannes", "familyName": "Kähn"}, {"contributorType": "Experimentator", "contributorName": "Schreuer, Jürgen", "nameType": "Personal", "givenName": "Jürgen", "familyName": "Schreuer"}, {"contributorType": "Experimentator", "contributorName": "Hirschle, Christian", "nameType": "Personal", "givenName": "Christian", "familyName": "Hirschle"}, {"contributorType": "HostingInstitution", "contributorName": "Helmholtz-Zentrum Berlin für Materialien und Energie", "nameType": "Organizational", "nameIdentifiers": [{"nameIdentifier": "02aj13c28", "nameIdentifierScheme": "ROR", "schemeURI": "https://ror.org/"}]}]
9	dates	2023-01-20 14:50:13Z	[{"date": "2016-01-13T08:00:00+01:00/2017-02-28T10:16:02+01:00", "dateType": "Created"}, {"date": "2022-02-28T10:16:02+01:00", "dateType": "Available"}]
11	resourceType	2023-01-20 14:50:13Z	{"resourceType": "Investigation", "resourceTypeGeneral": "Collection"}
12	alternateIdentifiers	2023-01-20 14:50:13Z	[{"alternateIdentifier": "141-00585-1.1-N", "alternateIdentifierType": "GATE proposal number"}]
13	relatedIdentifiers	2023-01-20 14:50:13Z	[{"relatedIdentifier": "10.5442/NI000001", "relatedIdentifierType": "DOI", "relationType": "IsCompiledBy"}]
14	sizes	2023-01-20 14:50:13Z	["5.25 GiB"]
15	formats	2023-01-20 14:50:13Z	["application/octet-stream", "application/x-hdf5", "application/xml", "application/zip"]
17	rights	2023-01-20 14:50:13Z	[{"rights": "CC0 1.0 Universal", "rightsUri": "http://creativecommons.org/publicdomain/zero/1.0/", "rightsIdentifier": "CC0-1.0", "rightsIdentifierScheme": "SPDX"}]
18	descriptions	2023-01-20 14:50:13Z	[{"description": "Monoclinic sanidine is the high-temperature modification of potassium rich alkalifeldspars. The crystal structure consists of a framework of corner-sharing TO4 tetrahedra (T = Si, Al) with a temperature dependent disorder of Si and Al on the tetrahedral sites. Among the sanidines those from the volcanic fields of the Eifel play an exceptional role as they show unique optical anomalies at temperatures above about 750 °C. Further, ultrasonic investigations revealed strong evidence for the onset of dynamic processes between 900 °C and 1000 °C on laboratory time-scale. The proposed in-situ high-temperature neutron diffraction experiments on a sanidine megacrystal from Volkesfeld, Eifel, will provide insight into the role of Al/Si mobility and disorder for the anomalies observed.", "descriptionType": "Abstract"}]

Relevant PID Types: Data Publications

- We want recognition of data as scientific output on its own right.
- Therefore it must be properly published with good bibliographic metadata and a PID to be citable.
- Journals start to require that the underlying data to be published independently from the journal article.
- Commonly a DOI.



HZB Helmholtz Zentrum Berlin

HZB Data Service

Neutron study of the topological flux model of hydrogen ions in water ice

Hoffmann, J.-U.[1] ; Siemensmeyer, K.[1] ; Isakov, S.[2,3] ; Morris, D. J. P.[4] ; Klemke, B.[1] ; Glavatskyi, I.[1,5] ; Seiffert, K.[1,6] ; Tennant, D. A.[1,7] ; Sondhi, S.[8] ; Moessner, R.[9]

1. Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany
2. Theoretische Physik, ETH Zürich, 8093 Zürich, Switzerland
3. Google Inc., Brandschenkestrasse 110, 8002 Zürich, Switzerland
4. Xavier University, Department of Physics, 3800 Victory Parkway, Cincinnati, OH 45207, USA
5. Scienion AG, Volmerstr. 7b, 12489 Berlin, Germany
6. Technische Universität Berlin, Pascalstr. 8-9, 10587 Berlin, Germany
7. Oak Ridge National Laboratory, PO Box 2008 MS-6477, Oak Ridge, TN 37831, USA
8. Dept. of Physics, Princeton University, Washington Road, Princeton, NJ 08544, USA
9. Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Straße 38, 01187 Dresden, Germany

Cite as: Hoffmann, J.-U. et al (2018): Neutron study of the topological flux model of hydrogen ions in water ice. HZB Data Service. https://doi.org/10.5442/ND000001

Abstract

The familiarity of water ice means we often overlook its non-trivial character illustrated, for example, by the many snowflake morphologies resulting from disordered combinations of covalent and hydrogen bonds between hydrogen and oxygen atoms in water ice's most common phase (Ih) that keep the H_2 O molecular character. Using neutron diffraction on the flat-cone diffractometer E2 at BER-II, Helmholtz-Zentrum Berlin, we probe the atomic scale configuration in the Ih phase of water ice to test theories that describe this "disordered" state as exhibiting a form of topological order characterized by an emergent gauge field. We find excellent agreement between low-temperature experiment and analytical theory, which even allows us to estimate the density of defects charged under this emergent gauge field. The development of quantitative models of water ice paves the way for further studies to develop a comprehensive atomic-scale understanding of this most commonplace of solids.

Datafiles

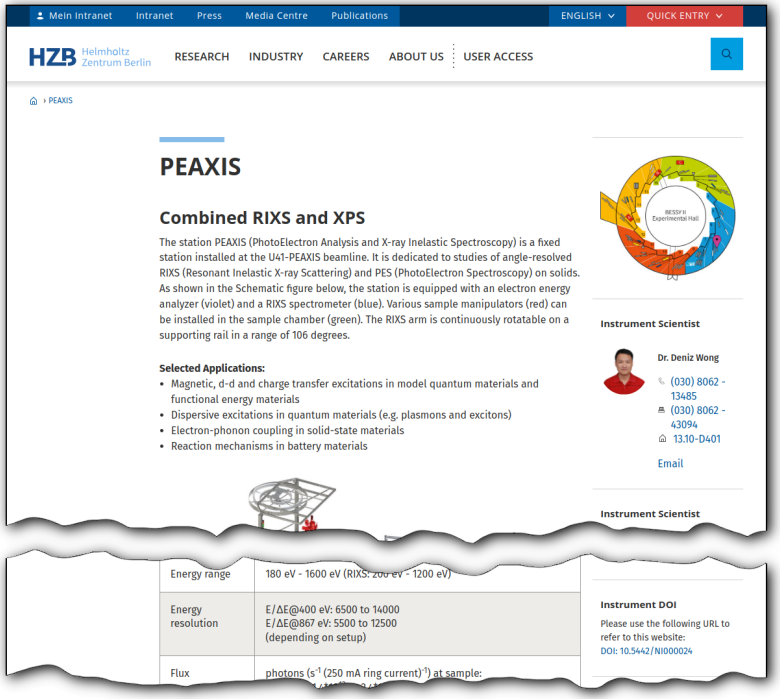
- D2O_E2_data.nxs, 776.65 MiB
- D2O_reciprocal_space.hdf, 240.62 MiB
- D2O_reciprocal_space_fit.hdf, 240.62 MiB
- Full Dataset, 1.23 GiB

Request Download

Metadata

Relevant PID Types: Instruments

- Being able to reference the instrument that collected the data from the dataset's PID record.
- Document the provenance of the data.
- Track the scientific output of the instrument.
- Various PID types: DOI, Handle.



Relevant PID Types: Samples

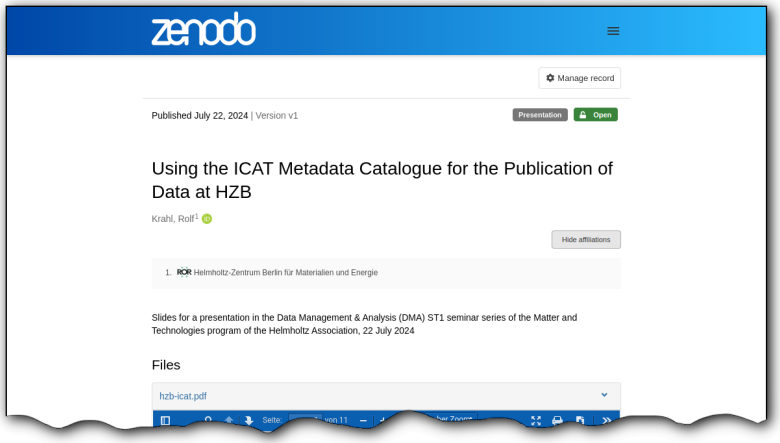
- Need to track the sample history from the creation to the measurement.
- Being able to reference the sample that has been measured from the dataset's PID record.
- Document the provenance of the data.
- Give credit to researchers being involved in the sample preparation.
- Dedicated PID type: IGSN.

Relevant PID Types: Software

- Again, software should be considered a scientific output on its own right.
- Being able to reference the software being used to create the data.
- Document the provenance of the data.
- Not yet well established, various PID types, often DOIs.

Relevant PID Types: People & Organizations

- Who did create this research output?
- And what organization he or she is affiliated with?
- Give credit where credit is due. And be accurate in addressing those people.
- Track the scientific output of the research organization.
- Dedicated PID types: ORCID iD for people, ROR for organizations.



Relevant PID Types: Research Activities

- A new emerging PID type at the horizon: RAiD, the Research Activity Identifier.
- Identify research projects and activities.
- Creates a single place for storing and retrieving project information.
- Could be useful as a "PID hub" to link everything (by the respective PID) related to the activity: people, organizations, inputs, outputs, resources being used ...
- Could be created for a proposal for beam time at the facility.

Relevant PID Types: Others

Many more things to PIDify:

- funding,
- data managements plans,
- terms in a controlled vocabulary,
- licenses,
- standards,
- ...

Conclusions

- PIDs are not only for journal articles and data: many different things may (and should) be attributed a PID.
- Also DOIs can be used for many different things.
- There are other PID types than DOIs: in some cases, there are good reasons, not to use a DOI.
- PIDs are most useful when they refer to each other, e.g. when the PID record includes the PIDs of other related entities.
- This forms a PID graph, where information can be automatically aggregated.