Persistent Identification of Instruments

The RDA WG PIDINST

ENVRI-FAIR Training - February 4, 2021

A bit of history ...

- PIDapalooza, Reykjavik, November 2016
  - First presentation of Persistent Identification of Instruments
  - https://doi.org/10.6084/m9.figshare.4246100.v1
  - Yay! … Y-A-PID or http://yapid.org/
- FREYA proposal
  - Included effort to develop a concept for persistent identification of instruments
- THOR-ENVRIplus Bootcamp, Helsinki, March 2017
  - ORCID Integrations in Environmental Research Infrastructures
  - Met Louise Darroch and Alessandro Oggioni, mulling over the idea of RDA WG PIDINST

A bit of history ...

- September 2017: BoF at P10, Montreal
- December 2017: Case statement submission
- March 2018: RDA endorsement and kick-off at P11, Berlin
- October 2019: Wrap-up at P14, Helsinki
- December 2019: Submitted DSJ manuscript (WG Deliverable)
- May 2020: Published DSJ article
- July 2020: DSJ article is RDA Supporting Outputs
- August 2020: Published RTD Living Document (WG Deliverable)
- November 2020: RDA vP16

What begins in Finland ends in Finland!

About the logo ...

Why PID of instruments matters

- Instruments play an essential role in creating research data
- Instrument metadata needed to assess data quality and reuse
  - Borgman (*): “To interpret a digital dataset, much must be known about the hardware used to generate the data, whether sensor networks or laboratory machines.”
- Persistent linking of research data and instruments
- Instrument models typical mentioned in literature: could be cited
- Inventory, funding, etc.

(*) Borgman, CL. 2015. Big Data, Little Data, No Data. MIT Press. https://doi.org/10.7551/mitpress/9963.001.0001

What PIDINST did

- Collect use cases
- Identify common metadata about instrument *instances*
- Develop and publish the schema
- Implement community feedback to schema versions
- Catalyse schema implementation by *existing* PID infrastructure
- Prototype adoption by existing institutional instrument providers
- Engage the wider community at RDA Plenaries
- Hold regular biweekly virtual meetings

Use cases, metadata analysis and schema

- 15 use cases collected between November 2017 and February 2019
- 60% in Earth Sciences
- Of 15 were 14 complete and 10 timely for October 2018 metadata analysis
- Metadata analysis resulted in first version of the schema
- https://github.com/rdawg-pidinst/schema/blob/master/schema.rst
- Schema was revised to account for community feedback

PIDINST Schema essentials

Owner	Institution(s) responsible for the management of the instrument
Manufacturer	The instrument's manufacturer(s) or developer
Model	Name of the model or type of device as attributed by the manufacturer
MeasuredVariable	The variable(s) that this instrument measures or observes
Date	Dates relevant to the instrument (e.g., commissioned, de-commissioned)
RelatedIdentifier	Identifiers of related resources
AlternateIdentifier	Other identifiers pertaining to the same instrument instance

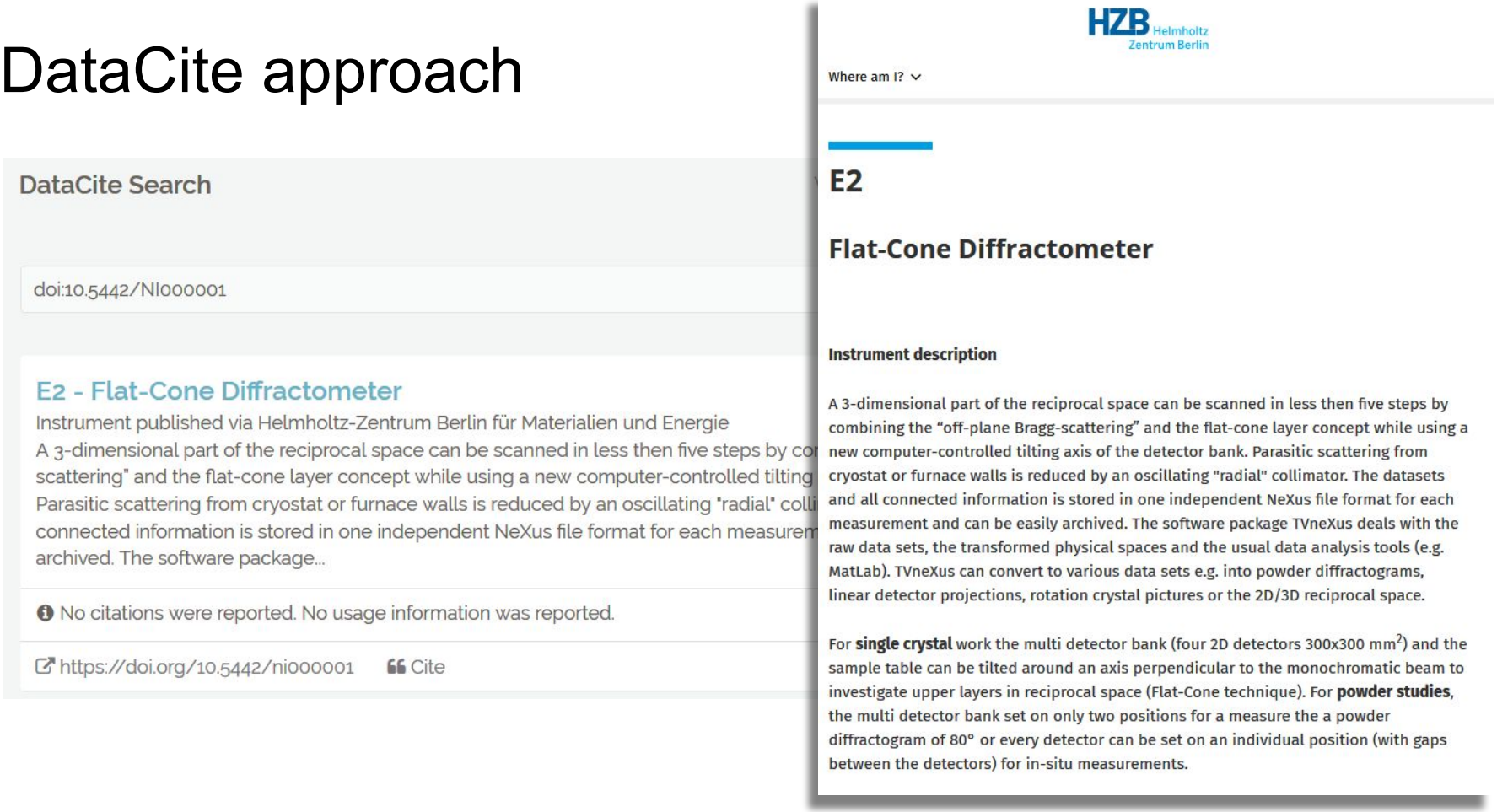
Schema implementations

- DataCite
  - Based on a PIDINST-DataCite Schema mapping
  - Only partial, e.g. no measured variable, model name not included
  - Bending of terminology needed, e.g. creator for manufacturer, publisher, publication year, …
  - Sure advantage, globally known PID infrastructure
- ePIC
  - Full PIDINST schema implementation (not in sync with the most recent schema changes)
  - Less well-known, more European-centric PID provider
- EUDAT (the new kid on the block)
  - B2INST service for registration and cataloguing
  - Particularly interesting for those lacking DataCite/ePIC membership
  - Result of a recent DataCite-ePIC-EUDAT collaboration
  - More about this later in this training ...

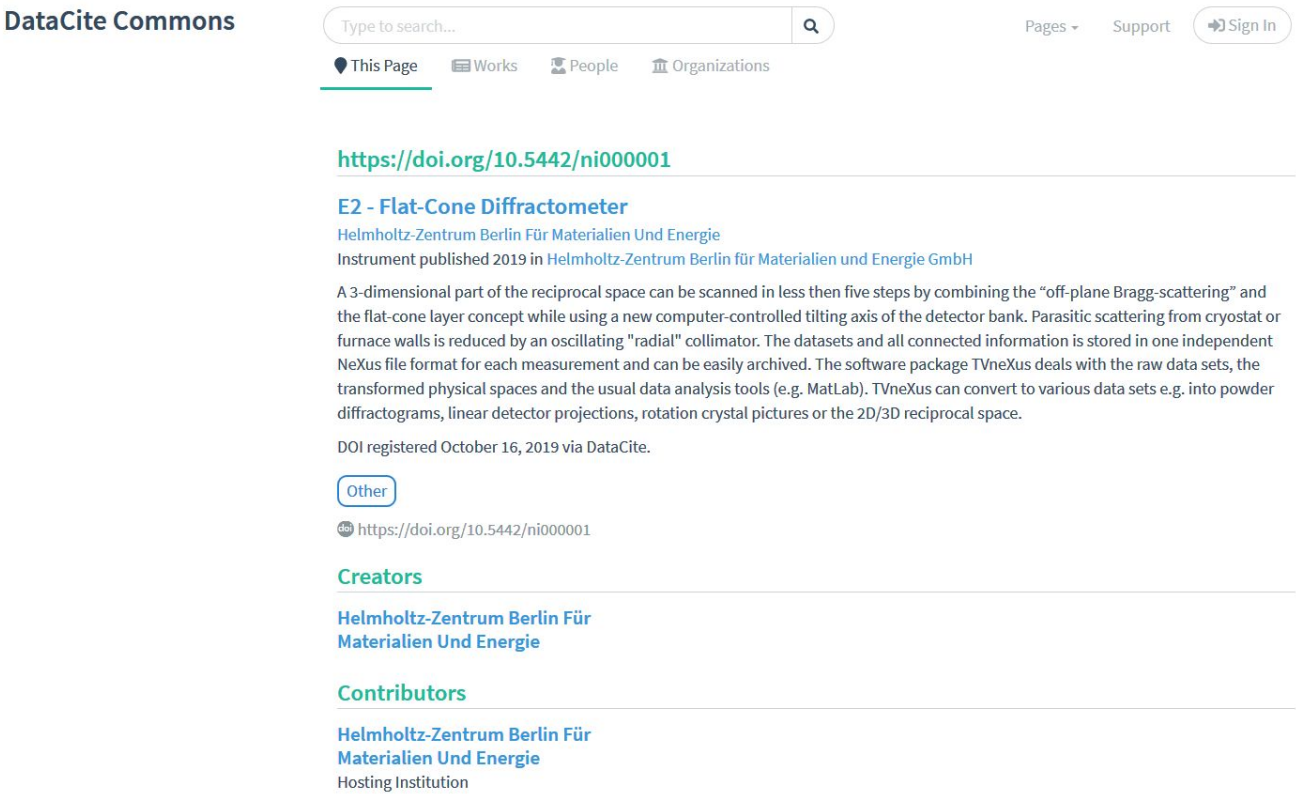
First (documented) tests

- Helmholtz-Zentrum Berlin (HZB)
  - Tested the DataCite implementation
  - https://doi.org/10.5442/NI000001
  - https://search.datacite.org/works?query=doi%3A10.5442%2FNI000001
  - https://commons.datacite.org/doi.org/10.5442/NI000001
  - https://api.datacite.org/dois/10.5442/NI000001
- British Oceanographic Data Centre (BODC)
  - Tested the ePIC implementation
  - https://doi.org/21.T11998/0000-001A-3905-F
  - http://hdl.handle.net/21.T11998/0000-001A-3905-F?noredirect

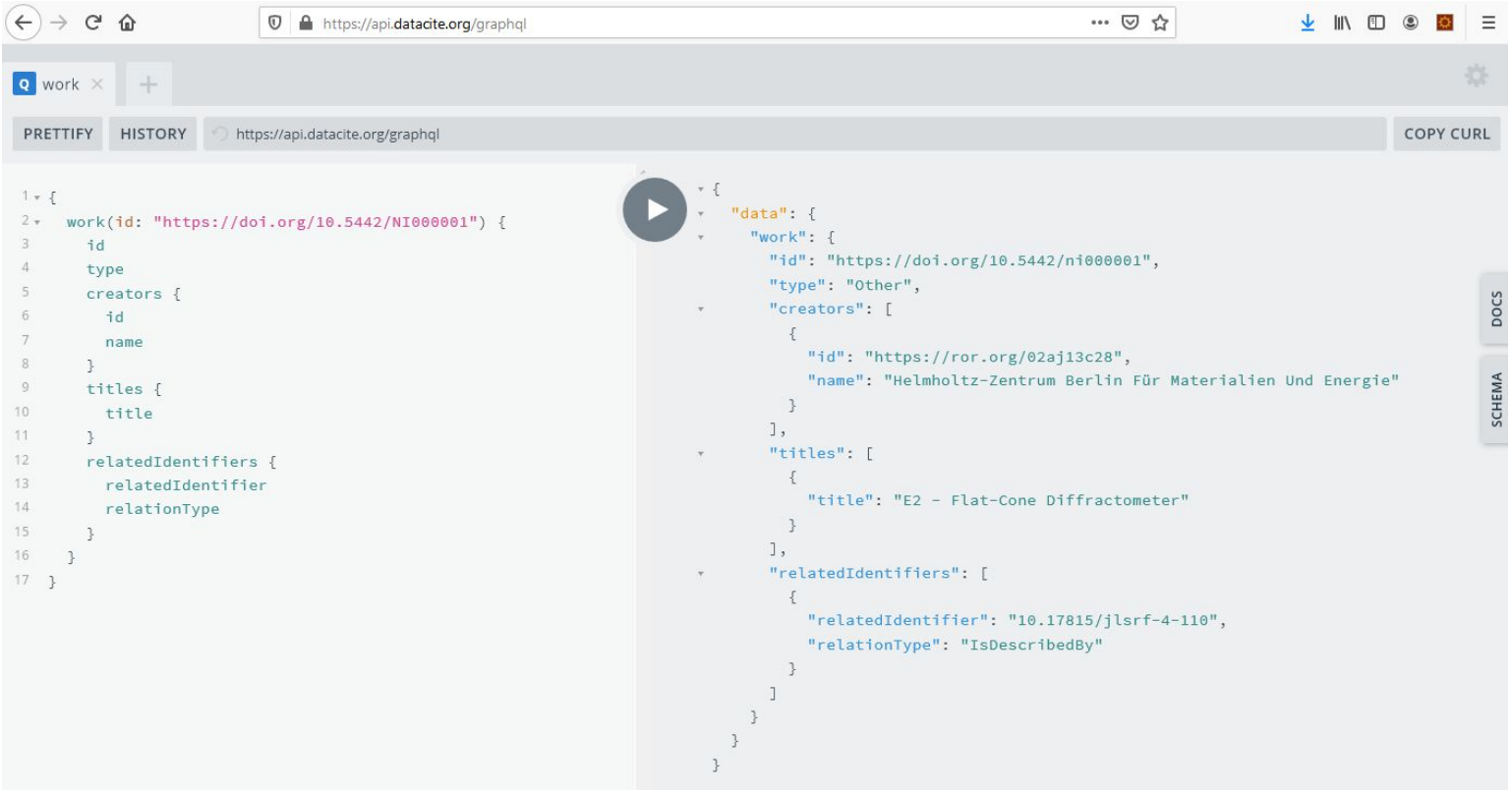
DataCite approach



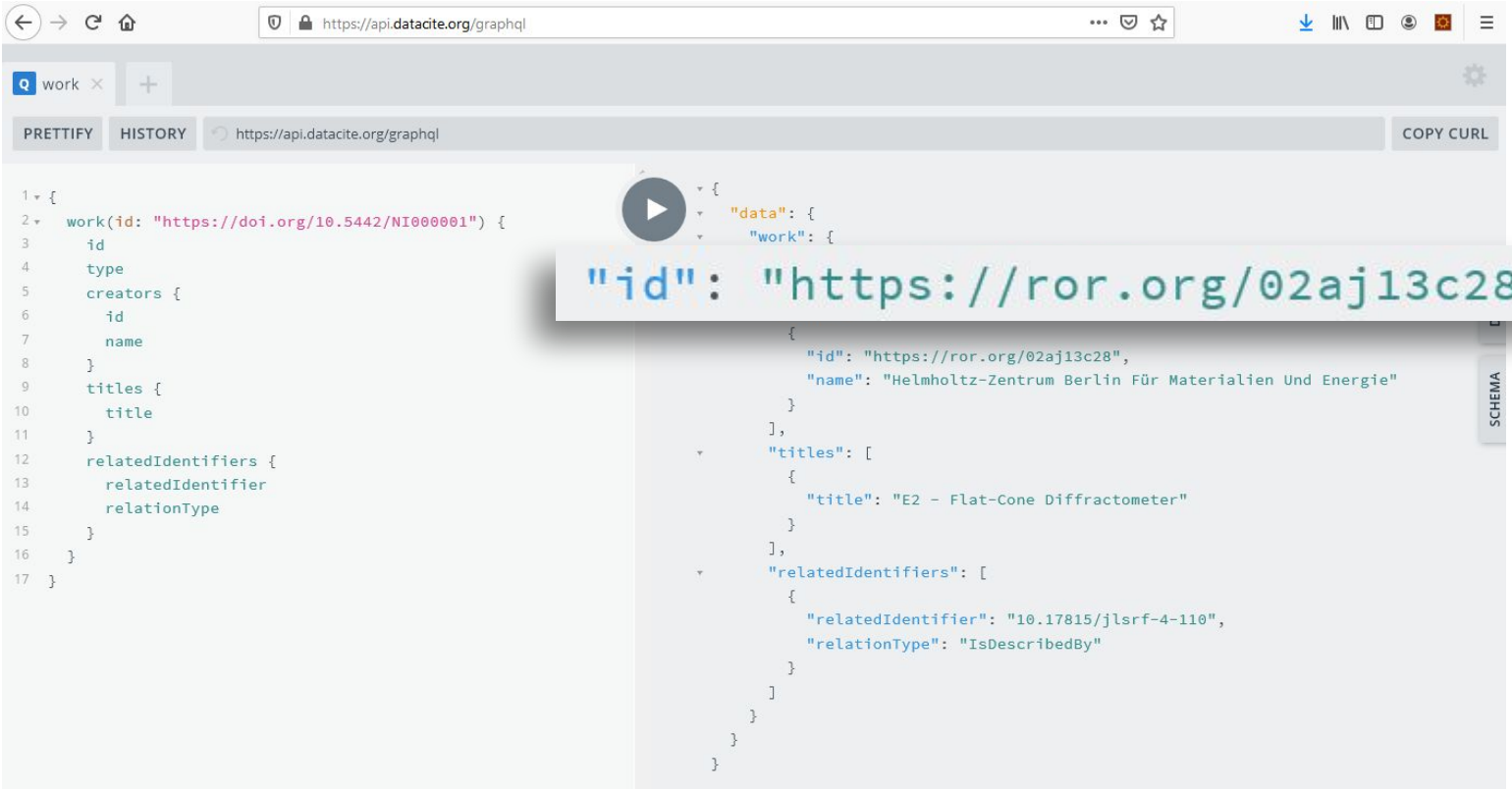
DataCite Commons



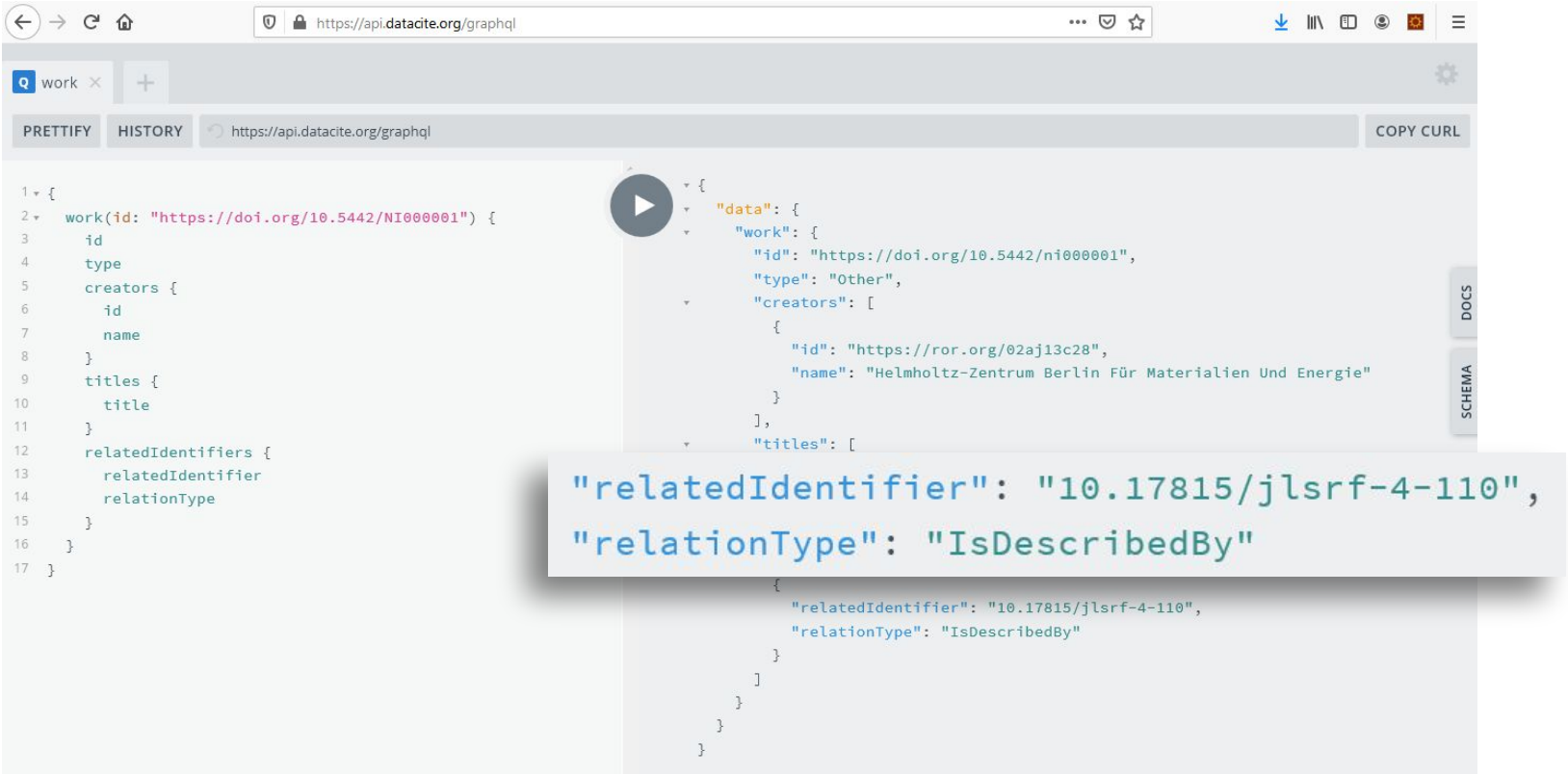
DataCite PID Graph



DataCite PID Graph



DataCite PID Graph



ePIC approach

Handle.Net®

Handle Values for: 21.T11998/0000-001A-3905-F

Index	Type	Timestamp	Data
1	URL	2019-12-10 12:27:48Z	https://linkedsystems.uk/system/instance/
2	21.T11148/8eb858ee0b12e8e463a5	2019-12-10 12:27:48Z	{"identifierValue":"http://hdl.handle.net/2
3	21.T11148/9a15a4735d4bda329d80	2019-12-10 12:27:48Z	https://linkedsystems.uk/system/instance/
4	21.T11148/709a23220f2c3d64d1e1	2019-12-10 12:27:48Z	Sea-Bird SBE 37-IM MicroCAT C-T Ser
5	21.T11148/4eaec4bc0f1df68ab2a7	2019-12-10 12:27:48Z	[{"Owner": {"ownerName":"National Oc {"ownerIdentifierValue":"http://vocab.ne
6	21.T11148/1f3e82ddf0697a497432	2019-12-10 12:27:48Z	[{"Manufacturer":{"manufacturerName" {"manufacturerIdentifierValue":"http://vo
7	21.T11148/55f8ebc805e65b5b71dd	2019-12-10 12:27:48Z	A high accuracy conductivity and temper IM model has an inductive modem for re
8	21.T11148/f76ad9d0324302fc47dd	2019-12-10 12:27:48Z	http://vocab.nerc.ac.uk/collection/L22/cu
9	21.T11148/72928b84e060d491ee41	2019-12-10 12:27:48Z	[{"MeasuredVariable":{"VariableMeasur {"VariableMeasured":"http://vocab.nerc.a {"VariableMeasured":"http://vocab.nerc.a {"VariableMeasured":"http://vocab.nerc.a
10	21.T11148/22c62082a4d2d9ae2602	2019-12-10 12:27:48Z	[{"date":{"date":"1999-11-01","dateType
11	21.T11148/eb3c713572f681e6c4c3	2019-12-10 12:27:48Z	[{"AlternateIdentifier":{"AlternateIdenti
12	21.T11148/178fb558abc755ca7046	2019-12-10 12:27:48Z	[{"RelatedIdentifier":{"RelatedIdentifier /37imbrochureju108.pdf","RelatedIdentif
100	HS_ADMIN	2019-10-14 16:31:53Z	handle=21.T11998/USER30; index=300; admin]

Handle Proxy Server Documentation
Handle.net Web Site

Please contact hdladmin@cnri.reston.va.us for your handle questions and comments.


```
−<sml:PhysicalSystem xsi:schemaLocation="http://www.opengis.net/swes/2.0 http://schemas.opengis.net
/swes/2.0/swesDescribeSensor.xsd http://www.opengis.net/sensorml/2.0 http://schemas.opengis.net/sensorML
/2.0/sensorML.xsd http://www.isotc211.org/2005/gmd http://schemas.opengis.net/iso/19139/20070417
/gmd/gmd.xsd http://www.isotc211.org/2005/gco http://schemas.opengis.net/iso/19139/20070417/gco/gco.xsd
http://www.opengis.net/gml/3.2 http://schemas.opengis.net/gml/3.2.1/gml.xsd" gml:id="TOOL0022_2490">
  <gml:description>SBE 37-IM MicroCAT-2490</gml:description>
  <gml:identifier codeSpace="uniqueID">TOOL0022_2490</gml:identifier>
 −<sml:keywords>
  −<sml:KeywordList>
     <sml:keyword>salinity sensor</sml:keyword>
     <sml:keyword>water temperature sensor</sml:keyword>
     <sml:keyword>Temp</sml:keyword>
     <sml:keyword>P_sal</sml:keyword>
     <sml:keyword>InSituCond</sml:keyword>
     <sml:keyword>Pres_MCat</sml:keyword>
   </sml:KeywordList>
  </sml:keywords>
 −<sml:identification>
  −<sml:IdentifierList>
   −<sml:identifier>
    −<sml:Term definition="TOOL0022_2490">
       <sml:label>UUID</sml:label>
       <sml:value>TOOL0022_2490</sml:value>
     </sml:Term>
    </sml:identifier>
   −<sml:identifier>
    −<sml:Term definition="http://vocab.nerc.ac.uk/collection/W07/current/IDEN0006/">
       <sml:label>Short Name</sml:label>
       <sml:value>SBE 37-IM MicroCAT 2490</sml:value>
     </sml:Term>
    </sml:identifier>
   −<sml:identifier>
    −<sml:Term definition="http://vocab.nerc.ac.uk/collection/W07/current/IDEN0002/">
       <sml:label>Long Name</sml:label>
       <sml:value>Sea-Bird SBE 37-IM MicroCAT C-T Sensor 2490</sml:value>
     </sml:Term>
    </sml:identifier>
   −<sml:identifier>
    −<sml:Term definition="http://vocab.nerc.ac.uk/collection/W07/current/IDEN0005/">
       <sml:label>Serial Number</sml:label>
       <sml:value>2490</sml:value>
     </sml:Term>
    </sml:identifier>
```

Deliverables

DATA SCIENCE JOURNAL

Reading: Persistent Identification of Instruments Share:

Special Collection: Research Data Alliance Results

Research Papers

Persistent Identification of Instruments

Authors: Markus Stocker, Louise Darroch, Rolf Krahl, Ted Habermann, Anusuriya Devaraju, Ulrich Schwardmann, Claudio D'Onofrio, Ingemar Häggström

Abstract

Instruments play an essential role
instruments and associated metac
reuse, globally unique, persistent
crucial. The Research Data Allianc
Instruments (PIDINST) developed
identification of instruments whic
analysis of 10 use cases, PIDINST
schema implementation with Data
identifier infrastructures and with HZB (Helmholtz-Zentrum Berlin für Materialien und Energie) and BODC (British Oceanographic Data Centre) as representative institutional instrument providers. These implementations demonstrate the viability of the proposed solution in practice. Moving forward, PIDINST will further catalyse adoption and consolidate the schema by addressing new stakeholder requirements.

Keywords: Persistent Identification, Instruments, Metadata, DOI, Handle

Markus Stocker, Louise Darroch, Rolf Krahl, Ted Habermann, Anusuriya Devaraju, Ulrich Schwardmann, Claudio D'Onofrio, and Ingemar Häggström (2020). **Persistent Identification of Instruments**. *Data Science Journal*, 19:18.

https://doi.org/10.5334/dsj-2020-018

Special Collection: Research Data Alliance Results

Research Papers

Persistent Identification of Instruments

Authors: Markus Stocker, Louise Darroch, Rolf Krahl, Ted Habermann, Anusuriya Devaraju, Ulrich Schwardmann, Claudio D'Onofrio, Ingemar Häggström

Abstract

Instruments play an essential role
instruments and associated meta
reuse, globally unique, persistent
crucial. The Research Data Allianc
Instruments (PIDINST) developed
identification of instruments whic
analysis of 10 use cases, PIDINST
schema implementation with Data
identifier infrastructures and with HZB (Helmhol erialien
und Energie) and BODC (British Oceanographi tative
institutional instrument providers. These im trate the viability
of the proposed solution in practice. Moving for ill further catalyse
adoption and consolidate the schema by addressing keholder requirements.

Keywords: Persistent Identification, Instruments, Metadata, DOI, Handle

Markus Stocker, Louis olf Krahl, Ted Habermann, Anusuriya Devaraju wardmann, Claudio D'Onofrio, and Ingemar Häggstr **ersistent Identification of Instruments**. e *Journal*, 19:18.

https://do 4/dsj-2020-018

Persistent Identification of Instruments

The Persistent Identification of Instruments WG (PIDINST) seeks to explore a community-driven solution for globally unique identification of measuring instruments operated in the sciences.

Measuring instruments, such as sensors used in environmental science, DNA sequencers used in life sciences or laboratory engines used for medical domains, are widespread in most fields of applied sciences. The ability to link an active instrument (instance) with an instrument type and with the broader context in which the instrument operates, including generated data, other instruments and platforms, people and manufacturers, etc., is critical, especially for automated processing of such contextual information and for the interpretation of generated data.

PIDINST is a working group in the Research Data Alliance (RDA). It aims to establish a cross-discipline, operational solution for the unique and lasting identification of measuring instruments actively operated in the sciences.

The group produced the following outputs:

- Stocker, M, Darroch, L, Krahl, R, Habermann, Häggström, I. 2020. Persistent Identification 1–12. DOI: https://doi.org/10.5334/dsj-2020 This paper provides an overview of the work
- PIDINST White Paper
  This white paper provides recommendations for the use of instrument PIDs and gives technical details that go beyond the overview provided in the Data Science Journal paper. It is expected to evolve with new user requirements and working group activities.
- ePIC Cookbook
  Detailed instructions on how to create instrument PIDs using the ePIC infrastructure.

https://rda-pidinst.readthedocs.io/

Adoption

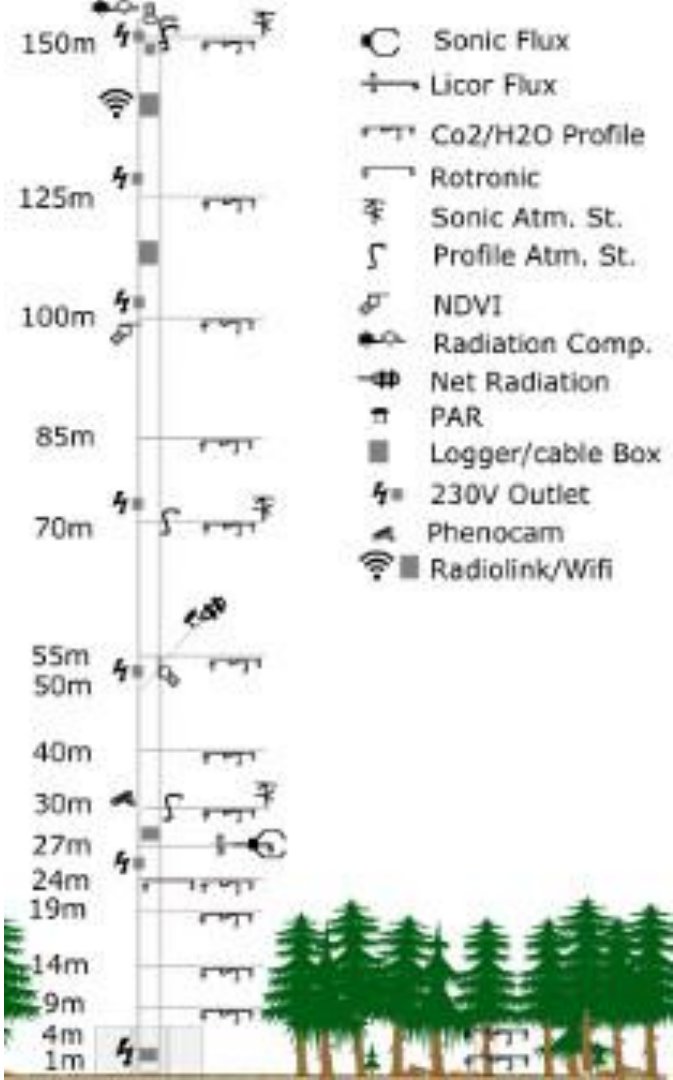
HZB Helmholtz Zentrum Berlin

ICOS INTEGRATED CARBON OBSERVATION SYSTEM

SENSOR.COMMUNITY

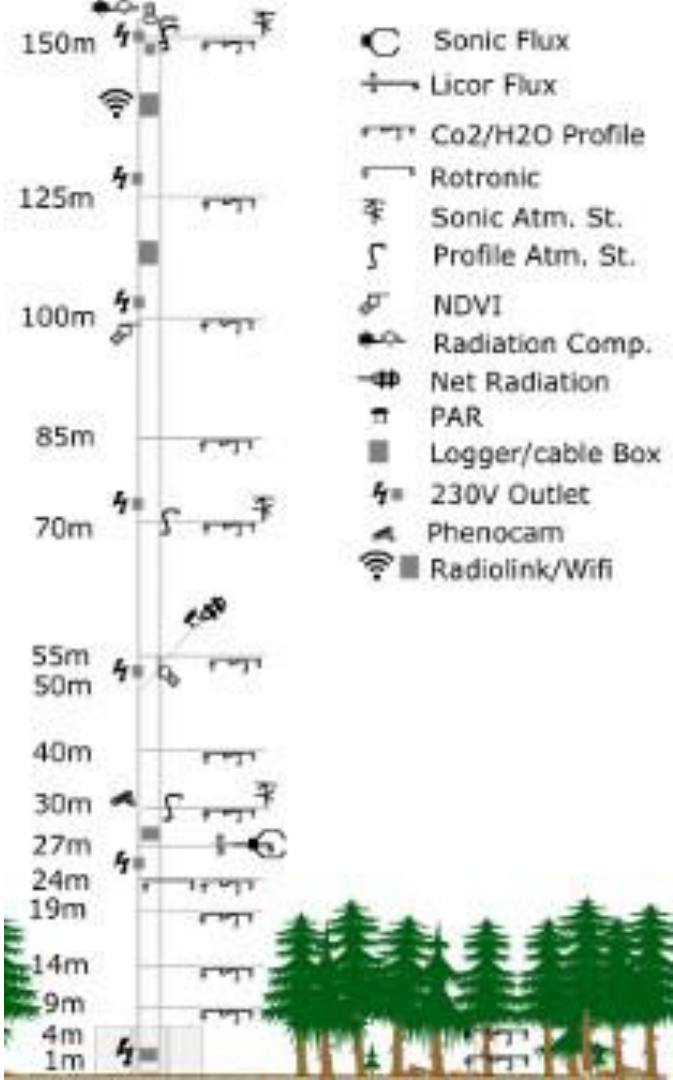
ICOS Carbon Portal

- ICOS operates ca 150 measurement stations in 12 countries
- GHG concentrations & fluxes, meteo, environmental monitoring
- Atmosphere, Ecosystem, Marine domains
- Instrument metadata include type, serial number, settings, location, calibration coefficients, …



ICOS Carbon Portal

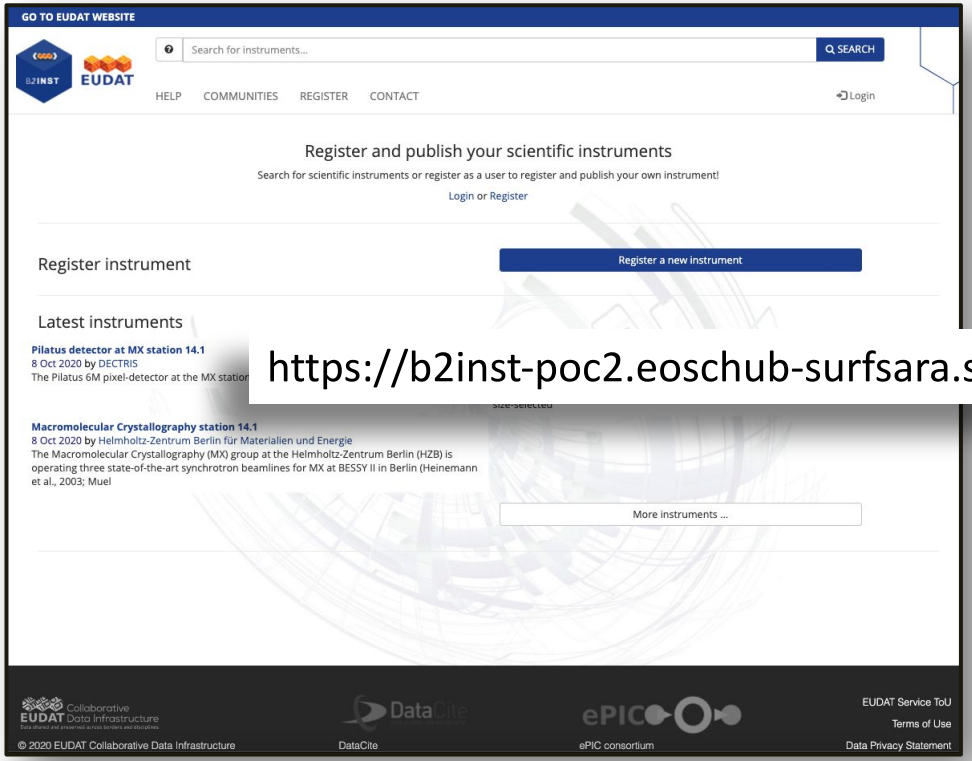
- Ensure PIDINST schema compatibility
- Mint PIDs with suitable and *operational* service
- Create informative Landing Pages



PANGAEA.

- PIDINST is relevant also for data publication
- PANGAEA is including PIDs for instruments in metadata
- Will be sharing links between data publications and instruments

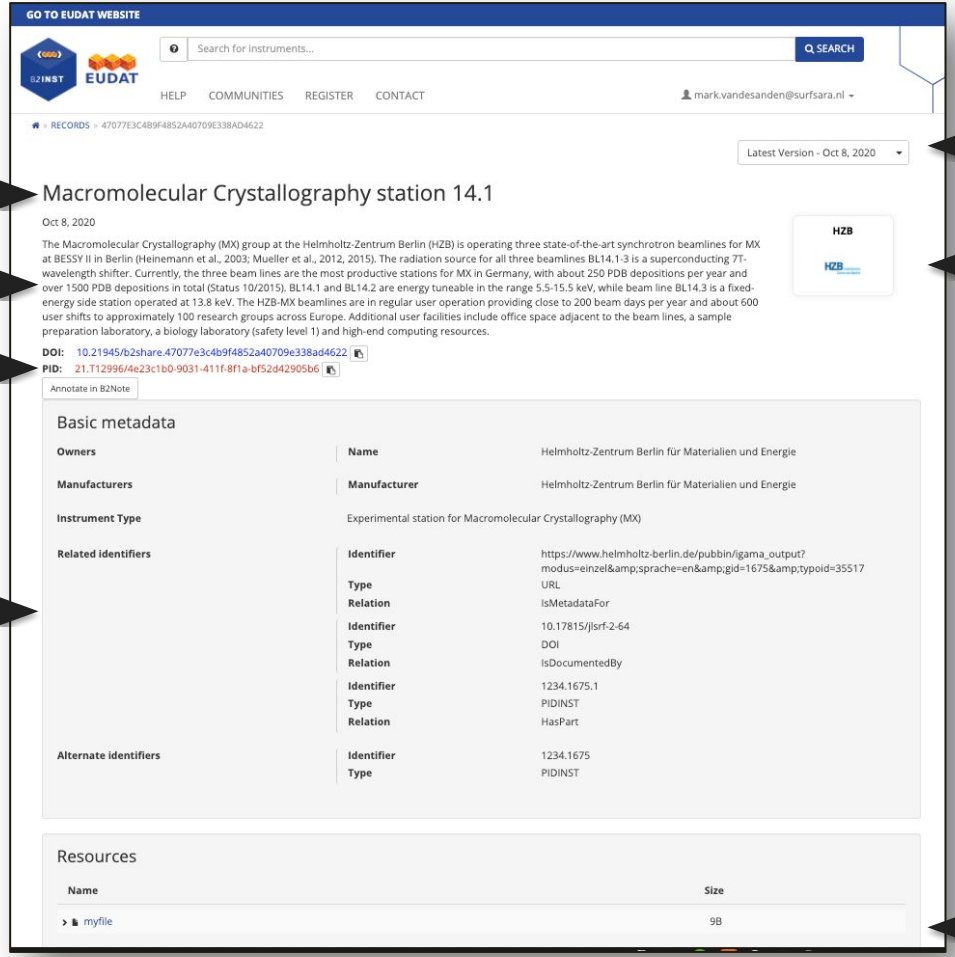
Collaborative
EUDAT Data Infrastructure



https://b2inst-poc2.eoschub-surfsara.surf-hosted.nl/

DataCite
FIND, ACCESS, AND REUSE DATA

ePIC
Persistent Identifiers for eResearch



Name

Description

PID

Basic Metadata

Versioning

Community Domain

Optional Files

Remarks

- RDA PIDINST WG has created a first schema for PIDs for instruments
- Remember: Instrument *instances* not models, instruments used in research
- With DataCite, ePIC and EUDAT several implementation choices
- There are first signs of schema adoption and field testing
- We need more of this; hence, also this ENVRI training