



PIDs for Instruments - the B2INST Service

Tibor Kálmán

tibor.kalman@gwdg.de

*Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen*

B2INST

ePICO 
Persistent Identifiers for eResearch

 DataCite
FIND, ACCESS, AND REUSE DATA


EUDAT



SCAN ME

 The DICE project has received funding from the European Union's Horizon 2020 project call H2020-INFRA-EOSC-2018-2020 under Grant Agreement no. 101017207

- Setting the stage (B2INST)
- Information modelling (PID4INST)
- Management & Strategy
- Metadata indexing & Discovery
- Technology & Development

Register and publish your scientific instruments

Search for scientific instruments or register as a user to register and publish your own instrument!

[Login or Register](#)

[Register a new instrument](#)

0 XR (EDYTEM CNRS)

500 µm Laser diffraction plus
Scattering (PIDS) technology
reporting of real data

für Materialien und Energie
combines a cryogenic linear
magnetic field for x-ray magnetic
selected

Uwitec Pilot 90 (EDYTEM CNRS) (example)

5 May 2021 by Uwitec Sampling Equipments
Interface sediment corer - diam 90mm id190 U-PILOT 90 (EDY)
<https://www.cybercarotheque.fr/refoutil.php>

Pilatus detector at MX station 14.1 (example)

19 Apr 2021 by DECTRIS
The Pilatus 6M pixel-detector at the MX station 14.1

Macromolecular Crystallography station 14.1 (example)

19 Apr 2021 by Helmholtz-Zentrum Berlin für Materialien und Energie
The Macromolecular Crystallography (MX) group at the Helmholtz-Zentrum Berlin (HZB) is operating three state-of-the-art synchrotron beamlines for MX at BESSY III in Berlin (Heinemann et al., 2003; Muel

[More instruments ...](#)

Motivation:

Create a public service to describe, register, and reference instruments

Trends:

- Registries for INST (community level → MD issues)
- Assign PIDs for their INST (emerging new type?)

GO TO EUDAT WEBSITE

Search for instruments...

mark.vandesanden@surfara.nl

Latest Version - Oct 8, 2020

Macromolecular Crystallography (MX)

Oct 8, 2020

The Macromolecular Crystallography (MX) group at the Helmholtz-Zentrum Berlin (HZB) is operating three state-of-the-art synchrotron beamlines for MX at BESSY II in Berlin (Heinemann et al., 2003; Mueller et al., 2012, 2015). The radiation source for all three beamlines BL14.1-3 is a superconducting 7T-wavelength shifter. Currently, the three beam lines are the most productive stations for MX in Germany, with about 250 PDB depositions per year and over 1500 PDB depositions in total (Status 10/2015). BL14.1 and BL14.2 are energy tuneable in the range 5.5-15.5 keV, while beam line BL14.3 is a fixed-energy side station operated at 13.8 keV. The HZB-MX beamlines are in regular user operation providing close to 200 beam days per year and about 600 user shifts to approximately 100 research groups across Europe. Additional user facilities include office space adjacent to the beam lines, a sample preparation laboratory, a biology laboratory (safety level 1) and high-end computing resources.

DOI: 10.21945/92share.47077e3c4b9f4852a40709e338ad9622

PID: 21.1129964e23c1b0-9031-411f8f1a-bf52d42905b6

Annotate in B2Wiki

Basic metadata		
Owners	Name	Helmholtz-Zentrum Berlin für Materialienforschung
Manufacturers	Manufacturer	Helmholtz-Zentrum Berlin für Materialienforschung
Instrument Type	Instrument Type	Experimental station for Macromolecular Crystallography (MX)
Related identifiers	Identifier	https://www.helmholtz-berlin.de/en/modus-reinzelkammer
	Type	URL
	Relation	ISMetadataFor
	Identifier	10.17815/gurf-2-64
	Type	DOI
	Relation	ISDocumentedBy
	Identifier	1234.1675-1
	Type	PIDINST
	Relation	HasPart
Alternate identifiers	Identifier	1234.1675
	Type	PIDINST

Resources

Name	Size
myfile	98

B2INST service – landing page

Instrument Name

Versioning

Instrument Description

Scientific Domain (Community)

Persistent Identifier (ePIC / DOI)

Basic Metadata

Optional Files (manuals, pictures of the instrument, etc)

Docs » PIDINST White Paper » 4. PIDINST metadata schema

4. PIDINST metadata schema

The metadata that is to be registered with an instrument PID need to contain enough information to unambiguously identify the instrument across networks and infrastructures. It furthermore allows to link resources related to the instrument, such as documents describing the instrument or external metadata records, possibly using other metadata standards to provide more details about the instrument.

Currently, two variants of the metadata schema are in use. The schema described here is the one that is used in the PIDINST metadata properties in the original PIDINST metadata schema.

Manufacturer
The organization or individual that built the instrument. In the case of an off the shelf product, this will probably be a commercial company that put the instrument on the market. In the case of a custom built instrument, the manufacturer may be the same as the owner. In the latter case, they would be registered as both, owner and manufacturer. In case of doubt, the manufacturer would be the instance that defined the technical specification of the instrument.

Identifier
The PID of the instrument. Handle or DOI in the case of a DOI.

LandingPage
The URL of the landing page of the instrument.

Name
The name by which this instrument is known within the organization.

Owner
The organization or individual that owns the instrument. It could also be the manufacturer.

Model
The name of the model of the instrument. It may not have a model value can be obtained, but it should have a persistent identifier for the model.


Description
A textual description of the instrument and should provide a new textual description of the instrument.

InstrumentType
A classification of the instrument that would apply to instruments of this type. If applicable, it should be a textual description of the instrument.

RelatedIdentifier
This can be used to establish links to related resources, such as documents describing the instrument or external metadata records, possibly using other metadata standards to provide more details about the instrument. Another application might be, if an instrument has been substantially modified, it would make sense to issue a new PID for the modified instrument with a new metadata record. In this case both PIDs should relate to each other to indicate that one is a new version of the other. Furthermore, in the case of a complex instrument, it can make sense to issue PIDs for individual components, such as an individual detector in a larger experimental station. In this case, the relation between the complex instrument and its components should be established by creating links between the respective PIDs. The links established using this property are particularly useful as they allow the automatic aggregation of a rich set of information about the instrument. Each *RelatedIdentifier* needs to have subproperties *relatedIdentifierType* and *relationType* to specify the type of the related PID and the type of the relation respectively.

AlternateIdentifier
If the instrument instance is also registered elsewhere, aside from the persistent identifier, *AlternateIdentifier* is the place to store a reference to these register entries. Common use cases are the serial number attributed by the manufacturer or inventory number used by the owner. But also other instrument databases or access portals may hold an entry for the instrument that should be referenced from the PIDINST metadata. The subproperty *alternateIdentifierType* needs to specify the kind of the alternate identifier. Standardized values should be used where applicable. For serial and inventory numbers, the suggested values are *serialNumber* and *inventoryNumber* respectively.

Basic fields

Community 

Name *

Description

Owners * Owner Name *

Manufacturers * Manufacturer Name *

Instrument Types Instrument Type Name *

Measured Variables

Open Access True

Instrument Identifier * Identifier * IdentifierType *

Dates Date Date Type

LandingPage *

Models Model Name Model Identifier Model Identifier Type

Schema Version *

Instrument Name

Instrument Description

INST specific fields

Schema version & Optional Files (manuals, pictures of the instrument)

Community

Owner

Basic Metadata

Docs » PIDINST White Paper » 4. PIDINST metadata schema [Edit on GitHub](#)

4. PIDINST metadata schema

The metadata that is to be registered with an instrument PID need to contain enough information to unambiguously identify the instrument across networks and infrastructures. It furthermore allows to link resources related to the instrument, such as documents describing the instrument or external metadata records, possibly using other metadata standards to provide more details about the instrument.

Currently, two variants of evaluation of use cases of metadata properties in the PIDINST metadata properties in the original properties in the original

Identifier
The PID of the instrument. The PID is the instrument's Handle or DOI. In the case of a custom built instrument, the manufacturer may be the same as the owner. In case of doubt, the manufacturer would be the instance that defined the technical specification of the instrument.

LandingPage
The URL of the landing page of the instrument. The name by which the instrument is known within the organization or inventory number used by the owner. It could also be the name of the instrument.

Owner
The organization or individual that built the instrument. In the case of an off the shelf product, this will probably be a commercial company that put the instrument on the market. In the case of a custom built instrument, the manufacturer may be the same as the owner. In the latter case, they would be registered as both, owner and manufacturer. In case of doubt, the manufacturer would be the instance that defined the technical specification of the instrument.

Manufacturer
The organization or individual that built the instrument. In the case of an off the shelf product, this will probably be a commercial company that put the instrument on the market. In the case of a custom built instrument, the manufacturer may be the same as the owner. In the latter case, they would be registered as both, owner and manufacturer. In case of doubt, the manufacturer would be the instance that defined the technical specification of the instrument.

Model
The name of the model of the instrument. It may not have a model value can be obtained. A persistent identifier for the model.

InstrumentType
A textual description of the instrument and should provide a new instrument type.

RelatedIdentifier
This can be used to establish links to related resources, such as documents describing the instrument or external metadata records, possibly using other metadata standards to provide more details about the instrument.

Another application might be, if an instrument has been substantially modified, it would make sense to issue a new PID for the modified instrument with a new metadata record. In this case both PIDs should relate to each other to indicate that one is a new version of the other.

Furthermore, in the case of a complex instrument, it can make sense to issue PIDs for individual components, such as an individual detector in a larger experimental station. In this case, the relation between the complex instrument and its components should be established by creating links between the respective PIDs.

The links established using this property are particularly useful as they allow the automatic aggregation of a rich set of information about the instrument. Each *RelatedIdentifier* needs to have subproperties *relatedIdentifierType* and *relationType* to specify the type of the related PID and the type of the relation respectively.

AlternateIdentifier
If the instrument instance is also registered elsewhere, aside from the persistent identifier, *AlternateIdentifier* is the place to store a reference to these register entries. Common use cases are the serial number attributed by the manufacturer or inventory number used by the owner. But also other instrument databases or access portals may hold an entry for the instrument that should be referenced from the PIDINST metadata.

The subproperty *alternateIdentifierType* needs to specify the kind of the alternate identifier. Standardized values should be used where applicable. For serial and inventory numbers, the suggested values are *serialNumber* and *inventoryNumber* respectively.



Search for Instruments...

Q SEARCH

HELP COMMUNITIES REGISTER CONTACT

Login

RECORDS - NEW

Please login. A new instrument can only be registered by authenticated users.

Title

Community

AWI

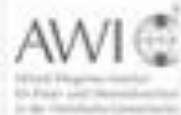
BODC

EISCAT

EUDAT

HZB

ICOS



Create draft instrument

You can also update the data of an existing instrument by creating a new version of that instrument. Search for the 'Create new version' button on the instrument's landing page.





Information modeling

1. *Persistent Identification of Instruments*, doi:10.5334/dsj-2020-018
2. *PIDINST White Paper*

1. Identify use cases

Collected real[*] use cases (PIDs for instruments)
[*] we deal with researchers!

2. Gather requirements

What information is needed to describe instruments?

3. Design an information model

RDA
(PID4INST)

Define a schema and the necessary attributes to store the information about the instruments

4. Define a data model

Explored potential PID service providers & their PIDs.
Identified two candidates: ePIC (Handles) / DataCite (DOIs)

5. Implement a solution

???

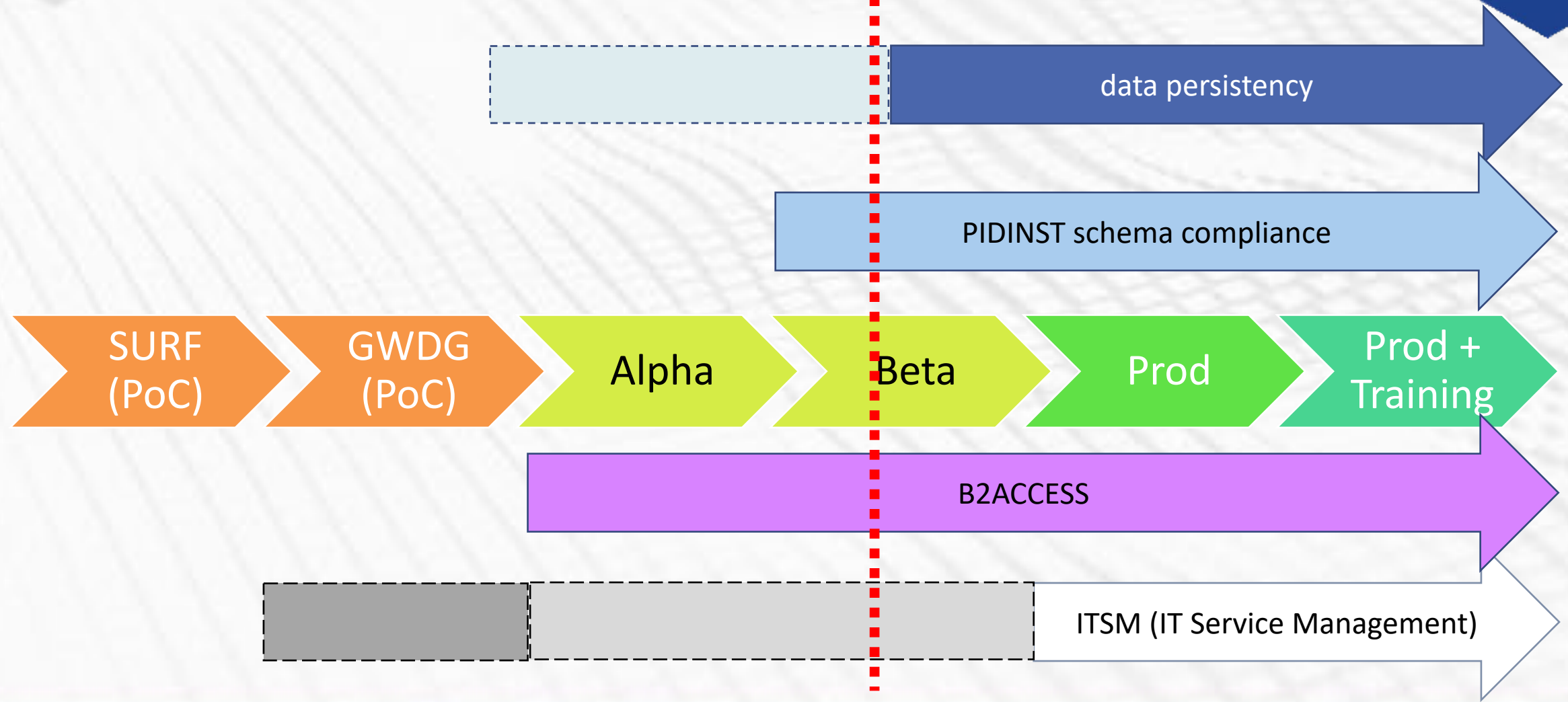


Roadmap, ongoing activities



Roadmap

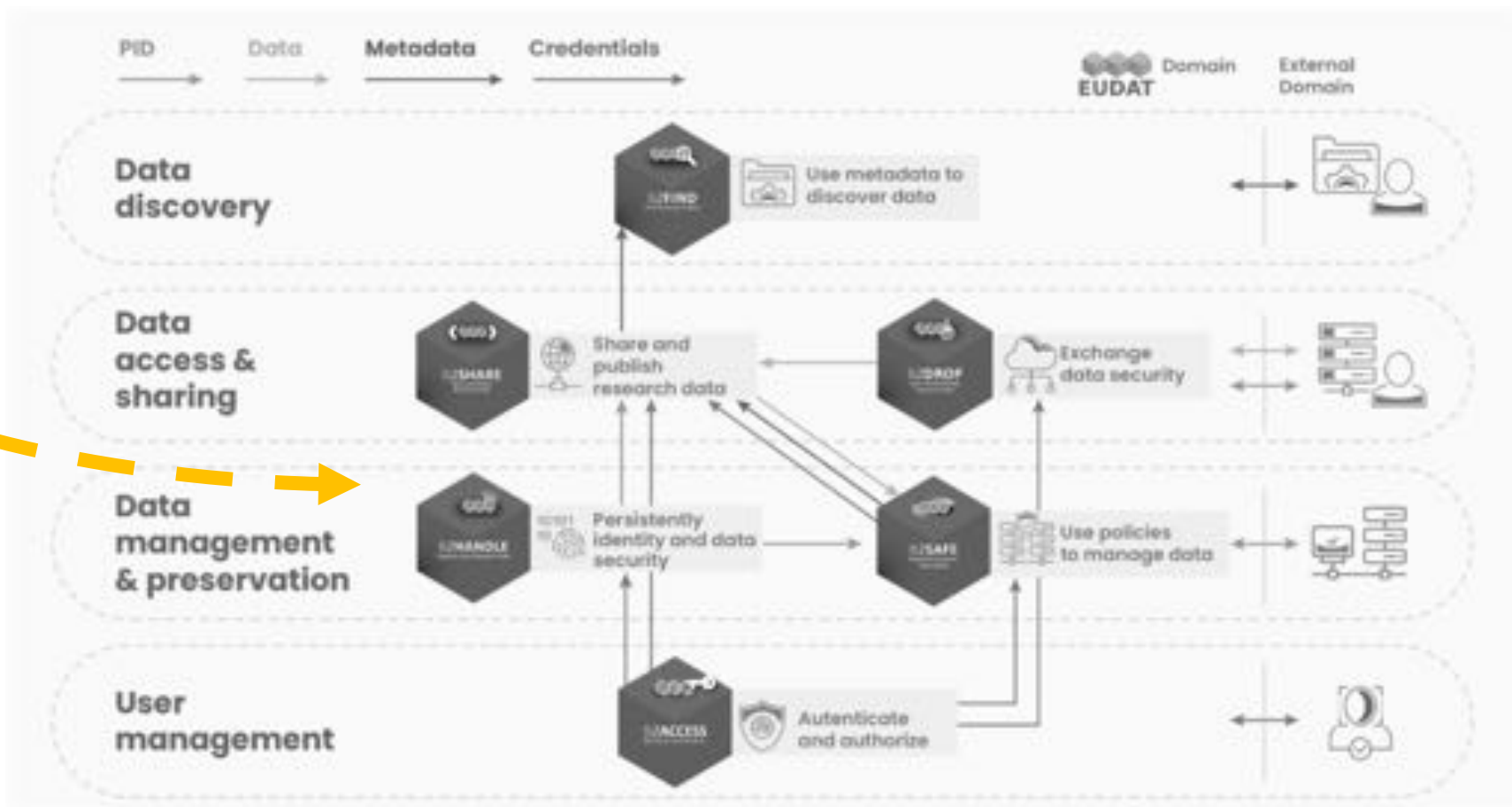
now





Service Management

The EUDAT Service Suite



ITSM process

- ▶ We are implementing a new service under the umbrella of EUDAT
 - ▶ This requires us to follow some ITSM processes
- ▶ B2INST is now in Service Portfolio Management (SPM) process
 - ▶ SPM is one of the most strategic SM process
 - ▶ It defines how new services are introduced & created
 - ▶ Describes a managed way how EUDAT implements its B2INST service
- ▶ Working on a business model template
 - ▶ Example on the right →
- ▶ *Not all actions are obvious for us*
 - ▶ *"It was quite some time ago when a new service was added to the EUDAT portfolio."*

Business Model Template

- ▶ [...]
- ▶ *Customers and Users*
- ▶ [...]
- ▶ *Value proposition and service description*
- ▶ [...]
- ▶ *Required partners*
- ▶ *Technology and Architecture*
- ▶ *Risk analysis*
- ▶ [...]
- ▶ *Communication and Marketing*

B2INST meets B2FIND



Goal:

- make use of registered instruments' metadata

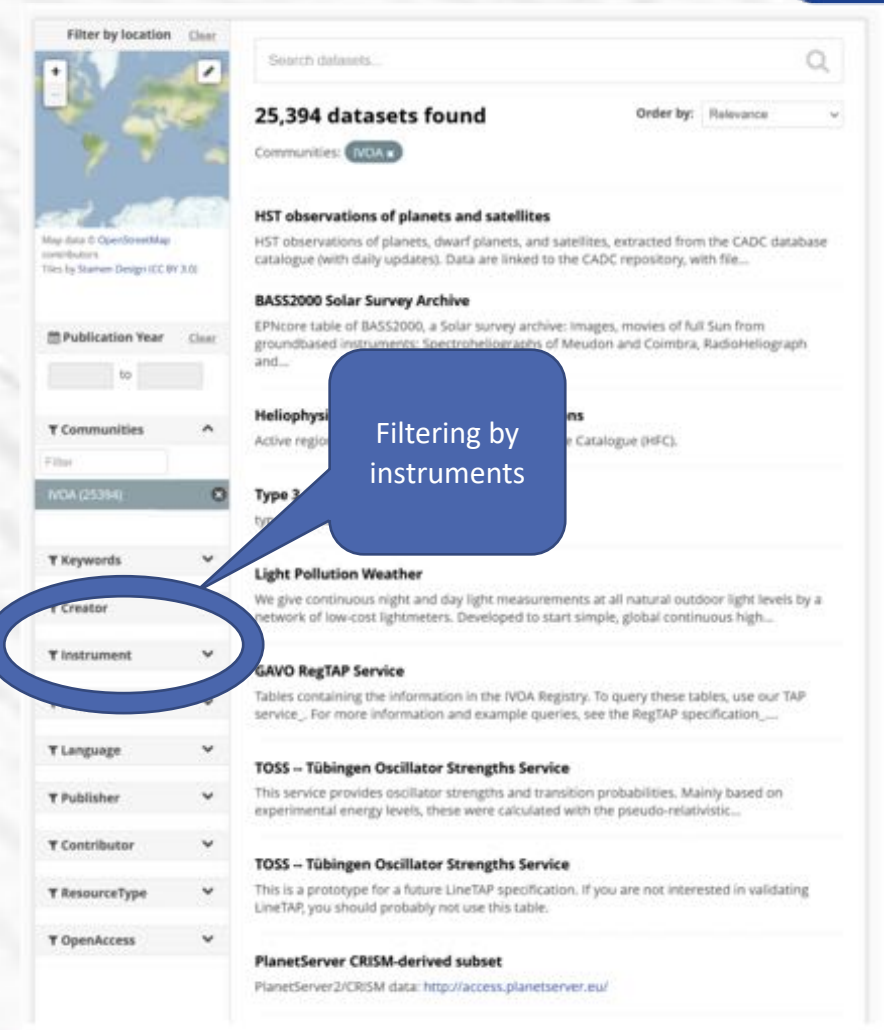
B2FIND

- Provides metadata indexing service and joint metadata catalogue
- Powerful discovery portal, which allows users to find data collections, objects, etc based on relations
- Possible gateway to EOSC

- Prerequisite: Objects (data sets) in data repositories are linked to instruments (PID)

What to show? Approaches being evaluated:

- Link to instruments' landing page
- Search in B2INST



- ▶ Service provider partner has been changed
 - ▶ SURF → GWDG
- ▶ ITSM
 - ▶ PoC, alpha and beta successful
 - ▶ Towards a production service
 - ▶ Still waiting for B2SHARE v3
 - ▶ Providing a Production and a Test (Training) instance
 - ▶ Refine Service Level & agree on Terms of Use
- ▶ Schema
 - ▶ DataCite v4.5
 - ▶ Checking full compliance with the PIDINST metadata schema
 - ▶ Options for community extensions
- ▶ Keeping data persistent from Beta onwards
- ▶ PIDs
 - ▶ DOI prefix provided by DataCite
 - ▶ PID prefix provided by ePIC
- ▶ Discovery service
- ▶ Types & Type registries

- Current B2INST is based on B2SHARE (“Invineo”, “Zenodo”)
 - Still committed to B2SHARE
- But... our development roadmap & timeline differs from B2SHARE’s
 - Handling of metadata & schemas
 - Supporting testbeds

Requirements

- AAI
- Payloads / additional objects
- OAI-PMH
- Central / distributed approach

- [DOIP]



JOIN OUR COMMUNITY



DICEosc



/company/diceosc



Tibor Kálmán

tibor.kalman@gwdg.de

*GWDG - Gesellschaft für wissenschaftliche
Datenverarbeitung mbH*

Göttingen

PID (Handle):

21.11101/0000-0007-F961-3

Resolvable URL:

<https://hdl.handle.net/21.11101/0000-0007-F961-3>



SCAN ME



Backup slides

Research Data Life Cycle

