

## PIDs for instruments: B2INST perspective

Dr. Tibor Kálmán

[tibor.kalman@gwdg.de](mailto:tibor.kalman@gwdg.de)

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Persistent Identifiers (PIDs) for Facility Research

ExPaNDS Workshop

22. October 2021.



## Agenda

- Motivation for a public service to reference instruments
- Information modeling:
  - adapt a “mature” schema
  - support community specific requirements
- Sustainability
  - through Persistent Identifiers (PIDs)
  - through organizational structures
  - through technology
  - through data re-use
- Direct involvement of users to co-design & co-create the service

*(Slides also from:*

*Sanden, Piggelen, Fenner, Schwardmann, Kalman: Update on B2INST @RDA #17  
EUDAT Secreteriat & DICE project)*

# **A PUBLIC SERVICE TO DESCRIBE AND REFERENCE INSTRUMENTS**

## Motivation

- Idea: Create a public service to describe, register, and reference instruments
- Register persistent identifiers (PIDs) for instruments.
- Possible impacts:
  - Add the instrument-PID to research outputs (like journal articles, datasets, etc), this enables to **reference the instrument**, which created the data
    - Instrument-PIDs can **track** the **instruments** of datasets
    - Instrument-PIDs can help to **track** the scientific **output** of instruments
  - **Aggregation of metadata** is possible (view dataset → see also metadata about the instrument, which generated the data)
  - PID Graphs: instruments could be an additional **node in the graph**

The screenshot shows the EUDAT website interface. At the top left are the logos for EUDAT and I2INSY. A search bar is located at the top right with the text "Search for instruments..." and a "SEARCH" button. Below the search bar are navigation links: "HELP", "COMMUNITIES", "REGISTER", "CONTACT", and a "Login" button with a user icon. The main heading is "Register and publish your scientific instruments", followed by the subtext "Search for scientific instruments or register as a user to register and publish your own instrument!". A "Login or Register" link is centered below. A large blue button labeled "Register a new instrument" is positioned to the right of the "Register instrument" heading. The "Latest instruments" section lists three examples: 1) Particle Size Analyzer Beckman LS 13 320 XR (EDYTEM CNRS) (example), dated 5 May 2021 by Beckman Coulter, with a description of its expanded measurement range and PIDS technology. 2) Uwitec Pilot 90 (EDYTEM CNRS) (example), dated 5 May 2021 by Uwitec Sampling Equipments, described as an interface sediment corer. 3) Pilatus detector at MX station 14.1 (example), dated 19 Apr 2021 by DECTRIS, described as a 6M pixel-detector. A fourth example, Macromolecular Crystallography station 14.1 (example), dated 19 Apr 2021 by Helmholtz-Zentrum Berlin, is partially visible. A "More instruments ..." link is at the bottom.

**Register instrument** [Register a new instrument](#)

### Latest instruments

**Particle Size Analyzer Beckman LS 13 320 XR (EDYTEM CNRS) (example)**  
5 May 2021 by Beckman Coulter  
Expanded measurement range: 10 nm – 3,500  $\mu\text{m}$  Laser diffraction plus advanced Polarization Intensity Differential Scattering (PIDS) technology enable high-resolution measurement & reporting of real dat

**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 ...](#)

GO TO EUDAT WEBSITE

Search for instruments... SEARCH

HELP COMMUNITIES REGISTER CONTACT

mark.vandesanden@urfzara.nl

RECORDS - 47377534894852440709E338204622

Latest Version - Oct 8, 2020

**Name** → **Macromolecular Crystallography station 14.1**

**Description** → 

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.

**PID** → 

DOI: 10.21945/zenodo.4707763.46394852440709E338204622

PID: 21.1129964e23c160-9031-411f-8f1a-bf52d42905b6

Annotate in R2Home

**Basic Metadata**

<b>Owners</b>	<b>Name</b>	Helmholtz-Zentrum Berlin für Materialien und Energie
<b>Manufacturers</b>	<b>Manufacturer</b>	Helmholtz-Zentrum Berlin für Materialien und Energie
<b>Instrument Type</b>		Experimental station for Macromolecular Crystallography (MX)
<b>Related identifiers</b>	<b>Identifier</b>	https://www.helmholtz-berlin.de/pubbin/gama_output?modus=einzel&amp;sprache=en&amp;id=1675&amp;typoid=35517
	<b>Type</b>	URL
	<b>Relation</b>	ISMetadataFor
	<b>Identifier</b>	10.17815/gurf-2-64
	<b>Type</b>	DOI
	<b>Relation</b>	ISDocumentedBy
	<b>Identifier</b>	1234.1675.1
	<b>Type</b>	PIDINST
	<b>Relation</b>	HasPart
<b>Alternate identifiers</b>	<b>Identifier</b>	1234.1675
	<b>Type</b>	PIDINST

**Resources**

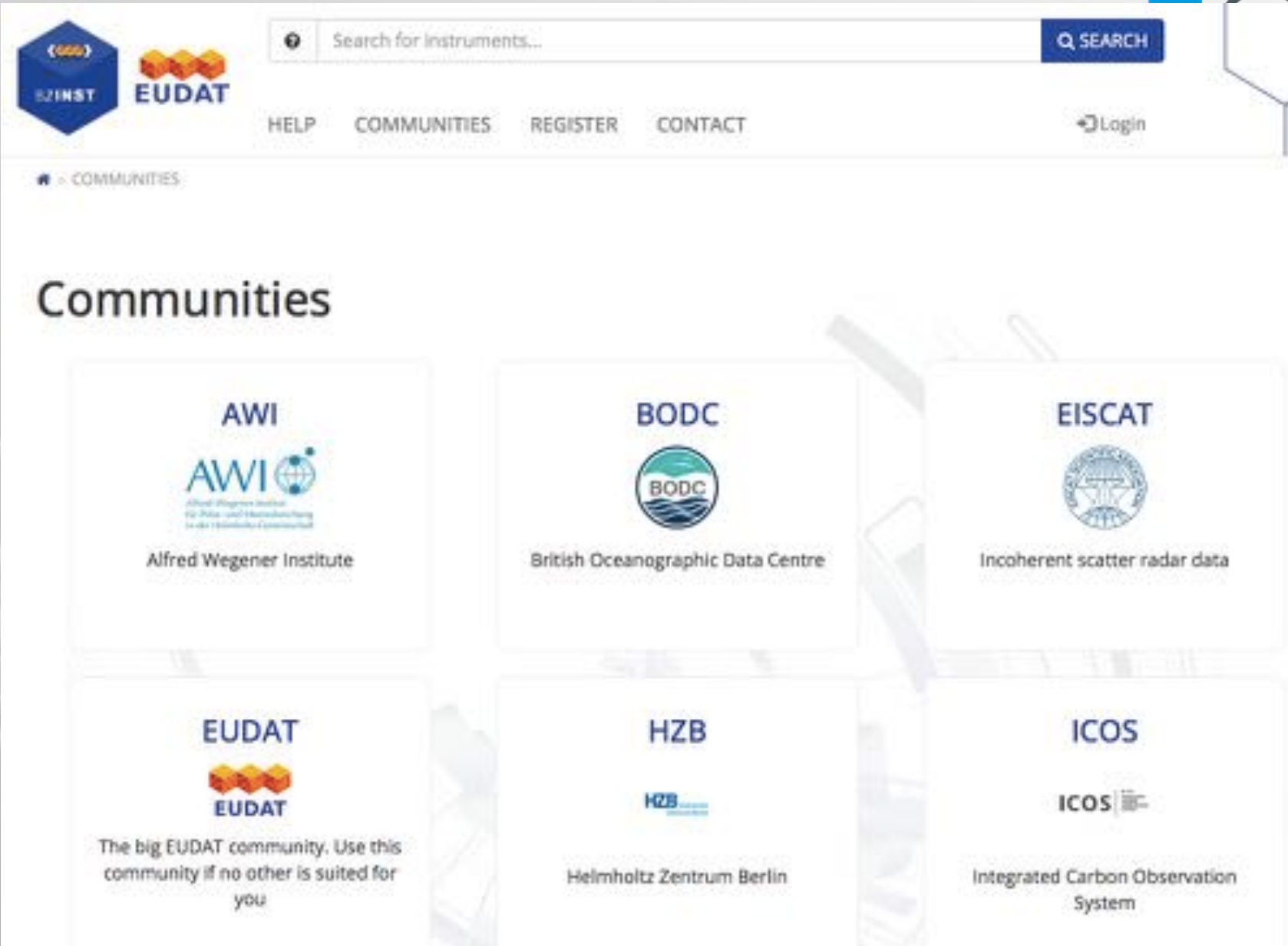
Name	Size
myfile	98



Versioning

Community Domain

Basic Metadata

Optional Files




  Search for instruments... [Q SEARCH](#)


[HELP](#) [COMMUNITIES](#) [REGISTER](#) [CONTACT](#) [Login](#)

[HOME](#) > [COMMUNITIES](#)


## Communities




**AWI**  
Alfred Wegener Institute




**BODC**  
British Oceanographic Data Centre




**EISCAT**  
Incoherent scatter radar data



**EUDAT**  
The big EUDAT community. Use this community if no other is suited for you



**HZB**  
Helmholtz Zentrum Berlin



**ICOS**  
Integrated Carbon Observation System

The screenshot shows the EUDAT website interface. At the top left, there are logos for B2INST and EUDAT. A search bar with the placeholder text "Search for Instruments..." and a "Q SEARCH" button is located at the top right. Below the search bar, there are navigation links for "HELP", "COMMUNITIES", "REGISTER", and "CONTACT", along with a "Login" button. A yellow message box with a blue border contains the text: "Please login. A new instrument can only be registered by authenticated users." A blue arrow points from this message box towards the "Login" button. Below the message box, there is a form with a "Title" input field and a "Community" selection area. The "Community" area includes buttons for AWI, BODC, EISCAT, EUDAT, HZB, and ICOS, each with its respective logo. A blue button labeled "Create draft instrument" is positioned below the community selection. At the bottom, another yellow message box states: "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

## Information modeling

- RDA PIDINST
  - Research Data Alliance - Persistent Identification of Instruments Working Group
- PIDINST aim:
  - *“to explore a community-driven solution for globally unique identification of measuring instruments operated in the sciences”*
- Idea:
  - Create persistent identifiers (PIDs) for instruments.
- Outputs:
  1. Persistent Identification of Instruments, doi:10.5334/dsj-2020-018
  2. PIDINST White Paper (RDA)



# Information modeling & Data modeling



1. Identify use cases
  2. Gather requirements
  3. Design an information model
  4. Define a data model
  5. Implement a solution
1. Collected real[\*] use cases (PIDs for instruments)  
[\*] we deal with researchers!
  2. What information is needed to describe instruments?
  3. Define a schema and the necessary attributes to store the information about the instruments
  4. Explored potential PID service providers & their PIDs. Identified two candidates:
    1. ePIC (Handles)
    2. DataCite (DOIs)

# **SUSTAINABILITY THROUGH ORGANIZATIONAL STRUCTURES**

# Sustainability through Organizational Structures



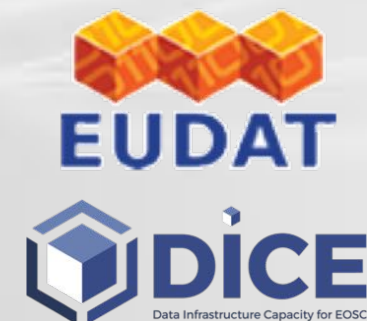
- Technologies are constantly changing
  - Outdated formats and interfaces
  - The requirements grow with the growing volume of data
- **Technologies alone are not sustainable!**
- Projects are limited:
  - Usually, services cannot be operated beyond the project timeline
  - Even partners may lose their funding or change their focus
- Services should be maintained at a professional level. It's not just ITSM:
  - Choose (when possible) proven and easily accessible technologies
  - Openness (open source, open standards, transparent governance) is important
  - Diversity might help to sustain services
  - Providing resources require long-term and stable funding

Components of a FAIR data ecosystem (eg PID service) can be operated sustainably when a sustainable organizational structure exists!

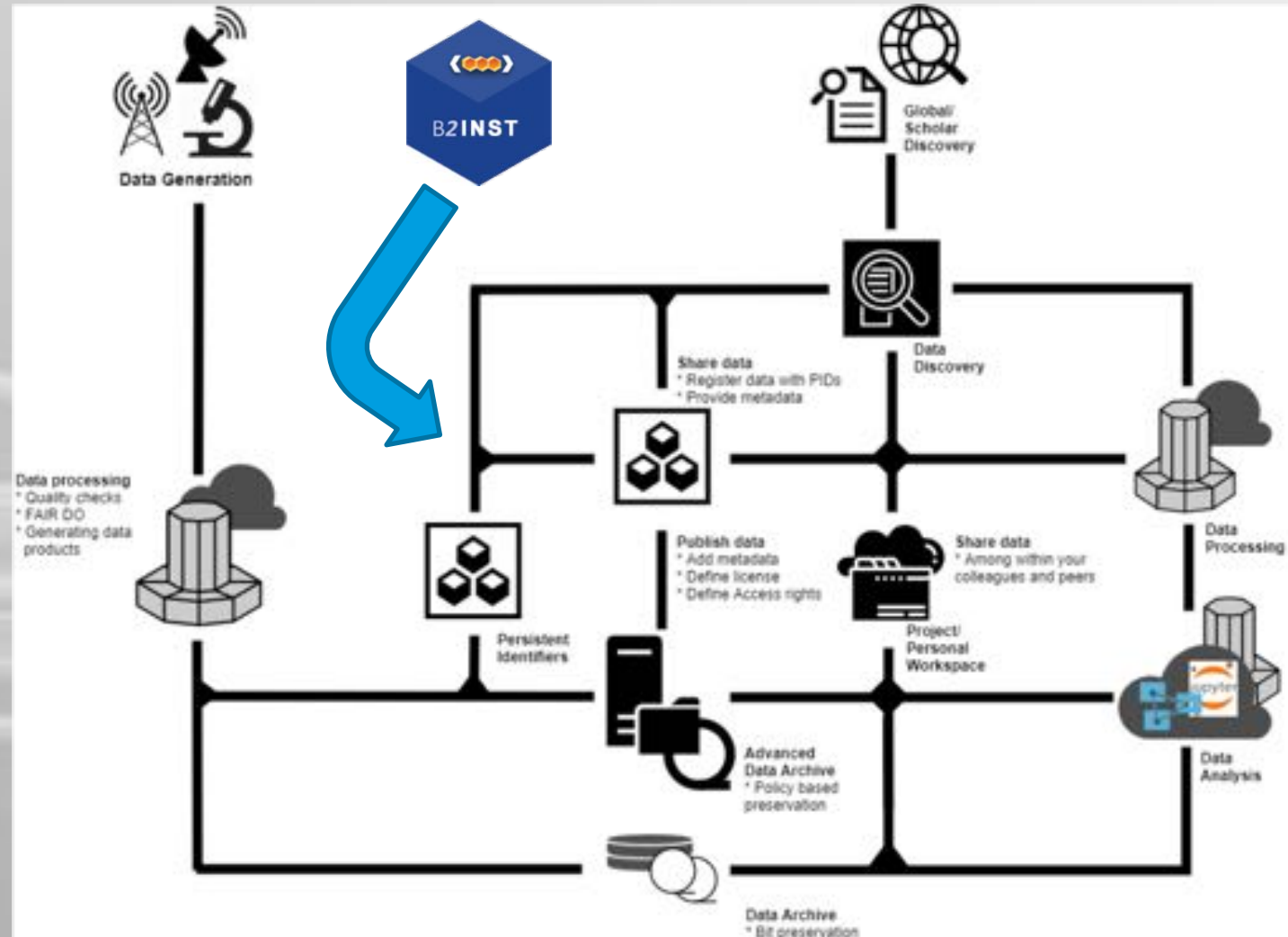
## Site Göttingen, Germany

- Regarding PIDs
  - Sustainable organizational structures already exist
- Göttingen participates in such organizational structures
  - Example of Campus Göttingen
    - ePIC
    - DataCite
    - DONA
    - [EUDAT CDI / DICE project – more than just PIDs]

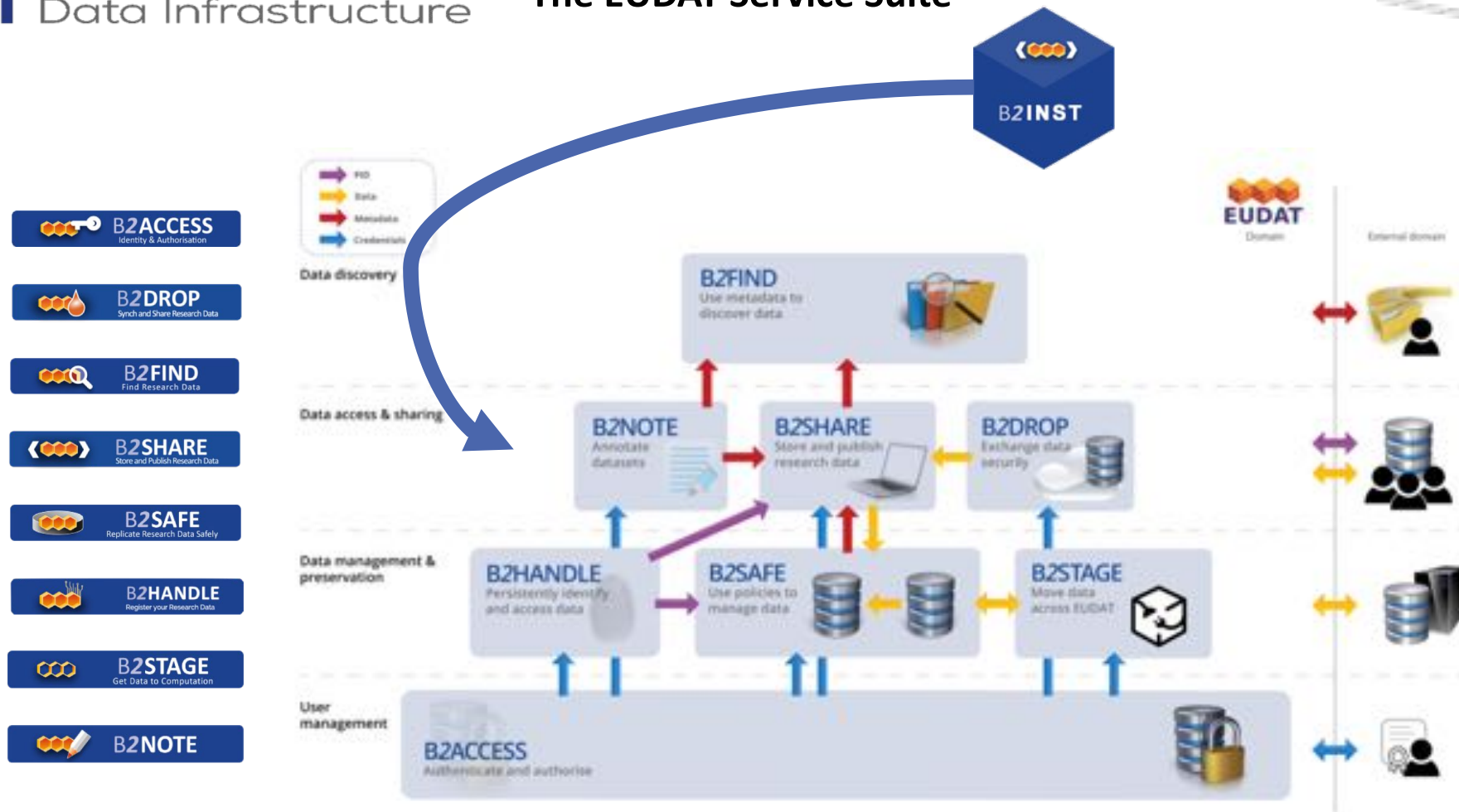
*(check the Backup Slides at the end for more details)*



# Research data workflow



# The EUDAT Service Suite



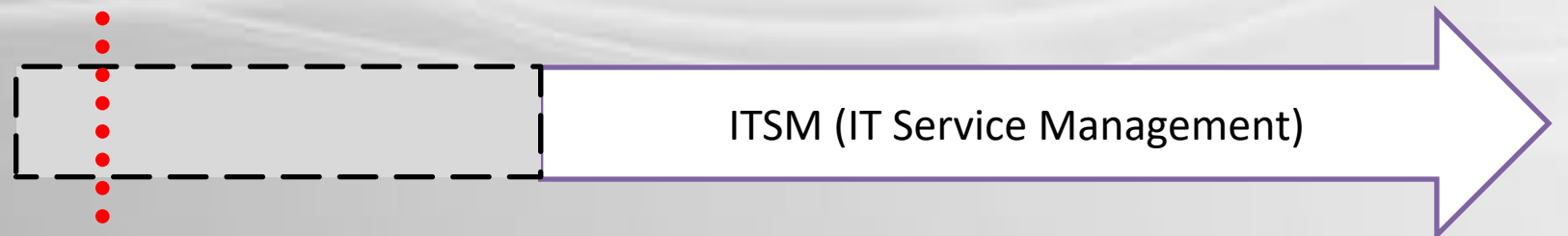
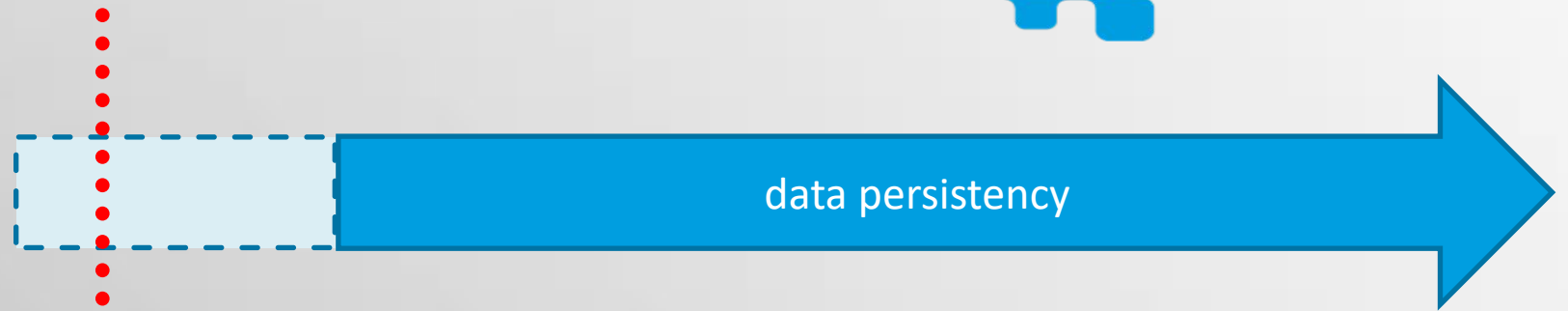


# ROADMAP, ONGOING ACTIVITIES

# Roadmap



**now**



## Ongoing activities

- Service provider partner:
  - SURF → GWDG / EUDAT
- ITSM
  - Towards a Beta and a Production service
  - Still waiting for B2SHARE v3
  - Providing a Production and a Demo (Training) instance
  - Refine Service Level & agree on Terms of Use
- Schema
  - Fully comply with the PIDINST metadata schema
  - Allow community extentions
- Keeping data persistent from Beta onwards
- PIDs
  - DOI prefix provided by DataCite
  - PID prefix provided by ePIC
- Discovery service
- Types & Type registries

## ➤ T4.2 Discovery and Referencing

Lead: GWDG (8PM)      Participants: DKRZ (8PM), DataCite (4PM)

- Main concern of the Task:
  - Integration, Interoperability, Scalability of PID systems

### Referencing

- 'PID for Instruments' (PID4INST)
  - Metadata schema: adaptation for instruments (based on RDA)
  - Types exist in the type registry
  - Defined types can be used directly
- [...]
- Integrity check
  - tool for PID and Type integrity check

### Discovery

- [...]
- B2FIND enhancement
  - metadata schema development to support enhanced interoperability (e.g. EUDAT Core)
- [...]

## Summary

- B2INST: public service to reference instruments
- Information modeling:
  - adapt a “mature” schema
  - support community specific requirements
- Sustainability
  - through Persistent Identifiers (PIDs)
  - through organizational structures
  - through technology
  - through data re-use & added-value services (PID Graph, Types)
  
- Direct involvement of users to co-design & co-create the service
  - ➔ Q & A.

Some final thoughts...

... for those involved in projects or research collaborations:

**Free lunch here!**





Thank you!

Tibor Kálmán  
tibor.kalman@gwdg.de

www.gwdg.de

Tel.: +49 (0)551 201-1545/ -1510  
Fax: +49 (0)551 201-2150

PID (Handle):

**21.11101/0000-0007-F40B-A**

Resolvable URL:

**<https://hdl.handle.net/21.11101/0000-0007-F40B-A>**

Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen

Am Fassberg 11  
37077 Göttingen

# BACKUP SLIDES



# SUSTAINABILITY THROUGH ORGANIZATIONAL STRUCTURES

## Site Göttingen

- Regarding PIDs
  - Sustainable organizational structures already exist
- Participation in such organizational structures
  - Example of Campus Göttingen
    - ePIC
    - DataCite
    - DONA
    - [EUDAT CDI / DICE project – more than just PIDs]

# Sustainability through Organizational Structures

## Example: ePIC



- Not every institute can or wants to run own services (for managing PIDs)
  - A consortium was formed to provide these services to researchers:
  - **Persistent Identifier Consortium for eResearch** (ePIC)
  - Main focus: the research landscape and cultural institutions
- Founders were data and computing centers, which
  - are sustainably financed
  - have long experience in the operation of stable and highly available services
  - have the possibility to offer SLAs
  - are involved in various eScience projects
- Some extra services:
  - PID information types and templates: Data Type Registry, etc (for PID information types)
  - PID Graph
- <http://www.pidconsortium.net/>



# Sustainability through Organizational Structures

## Example: DataCite



- DataCite
  - DOI „Registration Agency“
  - The focus is on the assignment of DOIs for published data sets ( & co.)
- Registry offices distributed worldwide
  - Example: Germany
    - Germany is represented by several DataCite partners
    - Partners are responsible for specific disciplines
    - Responsibility for "Arts and Humanities"
      - State- and University Library Göttingen
- Web:
  - <https://www.datacite.org/>



# Sustainability through Organizational Structures

## Example: DONA



- Digital Object Naming Authority (DONA)
  - DONA is a Swiss foundation hosting an international consortium
  - It governs the Handle structure at the top level in close collaboration with ITU-T
  - DONA was founded 2012 in Geneva
    - GWWDG (on behalf of ePIC) was one of the first members
    - the consortium will moderately grow in future
- Prefix registration:
  - A Multi-Primary Administrator (MPA) of the Global Handle Registry (GHR), is authorized by the DONA Foundation to allot prefixes to users of the Handle System.
  - GWWDG: 21.xyz
  - IDF (DOI): 10.xyz
- Web:
  - <https://www.dona.net/>



Common Data Infrastructure  
(EUDAT)

# **SUSTAINABILITY: THE INFRASTRUCTURE**

**The CDI members** (October 2020)

**Generic, Integrated Service Providers**

- CERN - Centre for Software (CERN) - FR
- CNRS - Centre National de la Recherche Scientifique (CNRS) - FR
- Barcelona Supercomputing Center (BSC) - ES
- Centre for Technology Enabled Science (CTES) - UK
- DLR - German Aerospace Establishment (DLR) - DE
- European Institute of Technology (EIT) - IT
- Max Planck Computing and Data Facility (MPDF) - DE
- CNRS - National Computing Center for Higher Education (CCIN) - FR
- Swiss Research and Learning Network S.4 (SRN4) - CH
- Jülich Research Centre (JRC) - DE
- Institut für Experimentelle Physik (IPEP) - DE
- The Open University - UK

**Generic, Interoperable Service Providers**

- United Kingdom e-Science Centre (UK e-SC) - UK
- Leibniz Computing and Information Services (LWI) - DE
- JRC - Jülich Research Centre - DE
- French National Research Institute for Computing (IRIT) - FR
- University of Edinburgh - UK
- Central and Eastern European Gateway (CEEG) - PL
- The National Scientific Computing Centre for Transdisciplinary Science and Technology (CNRS) - FR
- Institute of Information Technology (IIT) - IN
- Karlsruhe Institute of Technology (KIT) - DE
- Fermilab - US
- Fermi National Accelerator Laboratory (FNAL) - US

**Thematic, Integrated Service Providers**

- German Research Computing Centre (GRCC) - DE
- Research Center for Information Technology (RCIT) - CH

**Thematic, Interoperable Service Providers**

- European Organization for Nuclear Research (CERN) - CH
- Culham Centre for Fusion Energy (CCFE) - UK
- Max Planck Computing & Data Facility (MPDF) - DE
- Curious OTC - UK
- European Centre for Research and Advanced Training in Science Computing (ECCAS) - FR

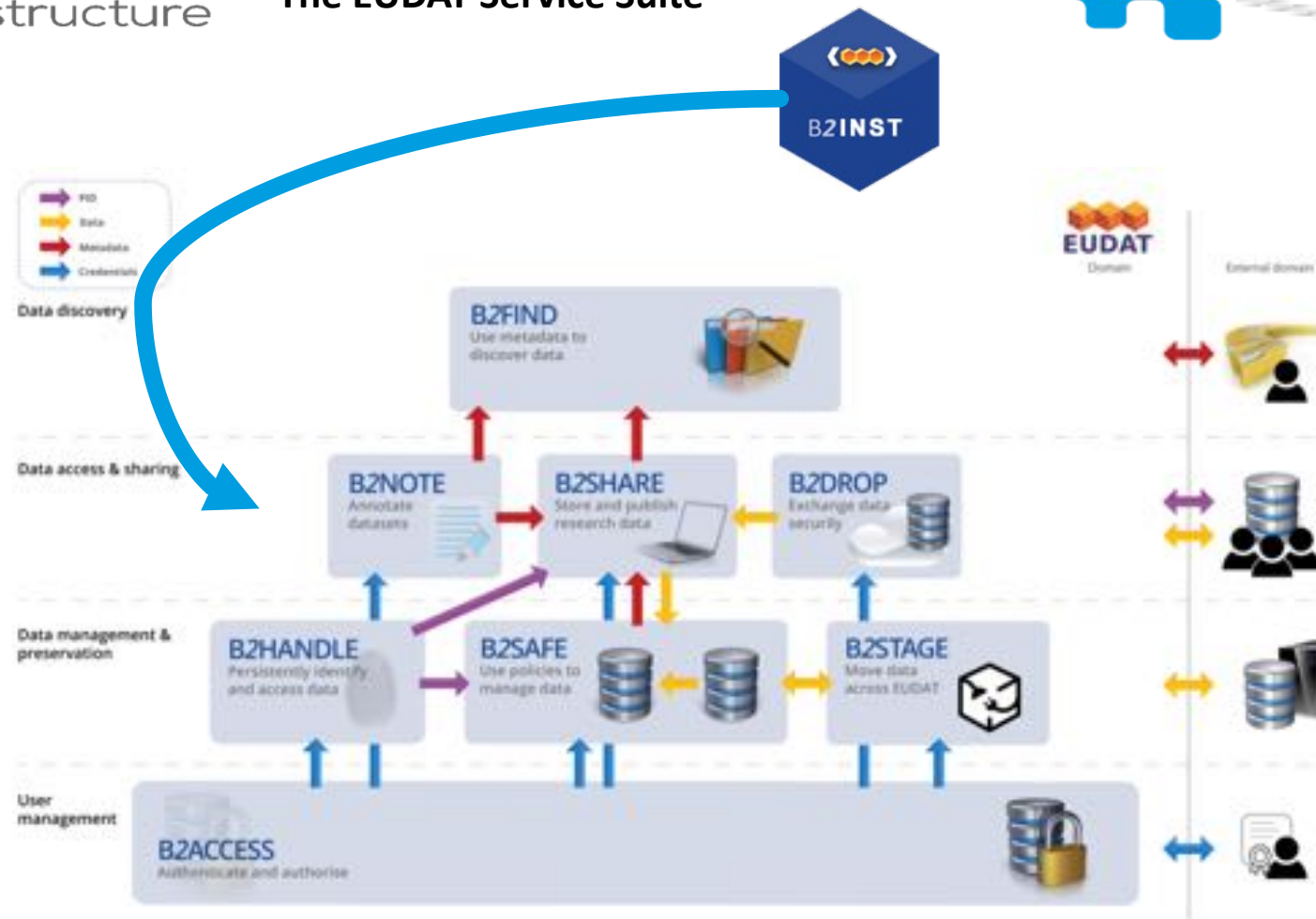


A growing network of 29 European research organisations, data and computing centres from 16 countries

The EUDAT Service Suite



- B2ACCESS**  
Identity & Authorisation
- B2DROP**  
Sync and Share Research Data
- B2FIND**  
Find Research Data
- B2SHARE**  
Store and Publish Research Data
- B2SAFE**  
Replicate Research Data Safely
- B2HANDLE**  
Register your Research Data
- B2STAGE**  
Get Data to Computation
- B2NOTE**

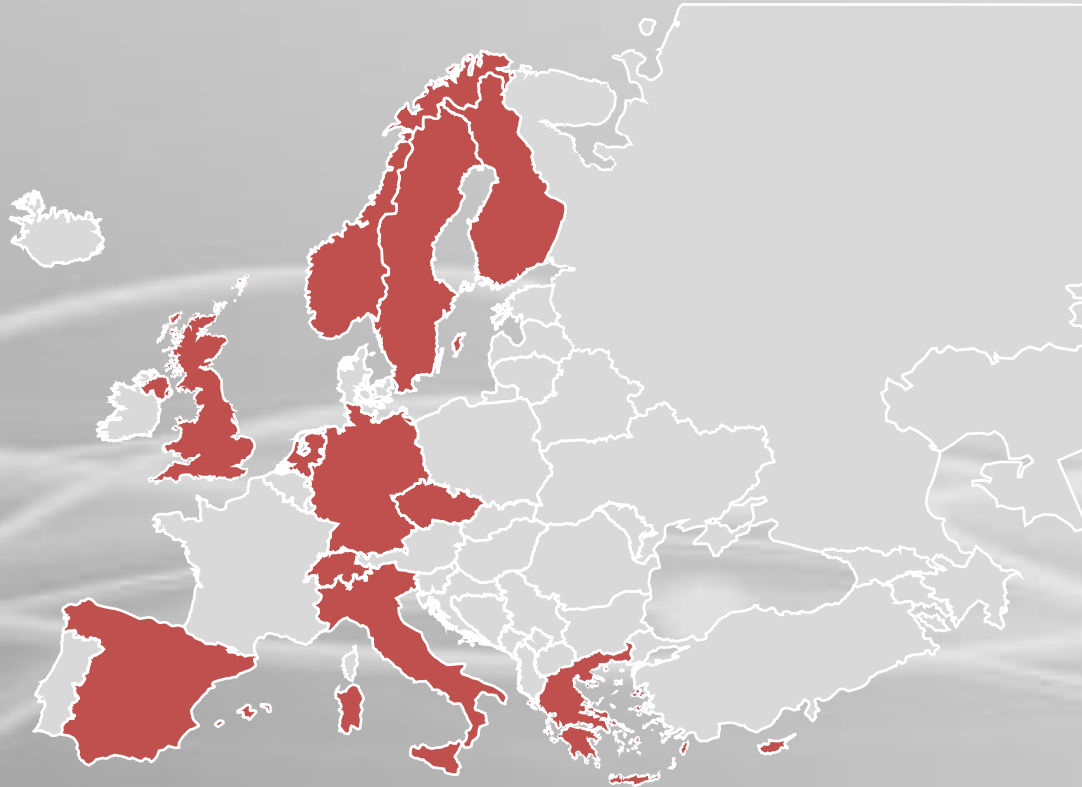




Data Infrastructure Capacity for EOSC  
(DICE)

# THE PROJECT

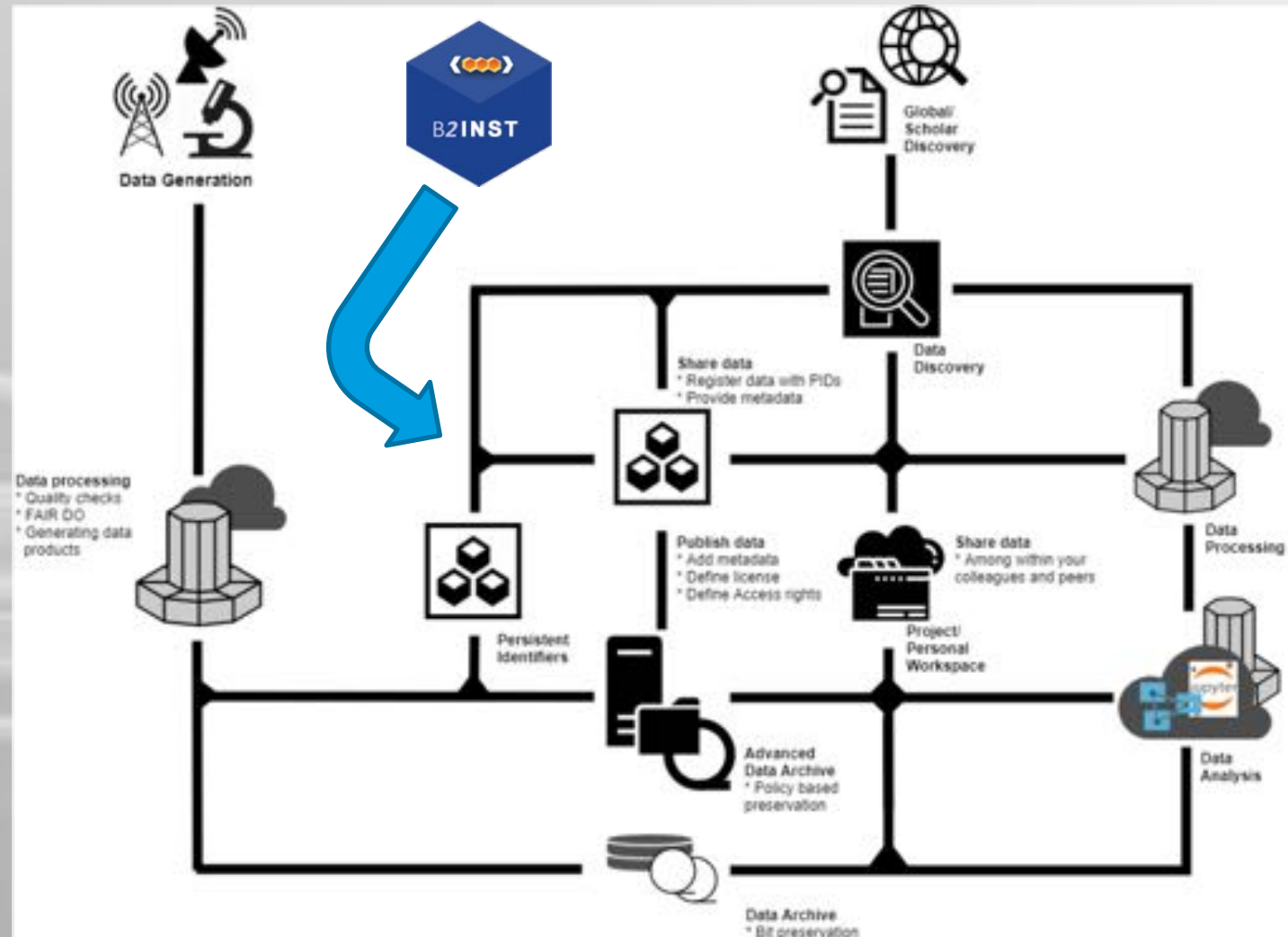
# Consortium



The consortium members and partners are represented by a grid of logos:

- Service providers:** A logo featuring a cloud with a circuit board inside.
- Users' communities:** A logo featuring three stylized human figures.
- CINECA**
- JÜLICH Forschungszentrum**
- ASTRON**
- CSC**
- BSC Barcelona Supercomputing Center** (Centro Nacional de Supercomputación)
- MPCDF MAX PLANCK COMPUTING & DATA FACILITY**
- ICOS Carbon Portal**
- grnet**
- KIT Karlsruhe Institute of Technology**
- DKRZ DEUTSCHES KLIMARECHENZENTRUM**
- ETH zürich CSCS** (Centro Svizzero di Calcolo Scientifico / Swiss National Supercomputing Centre)
- SURF**
- INFN Istituto Nazionale di Fisica Nucleare**
- KNAW**
- DataCite** (FIND, ACCESS, AND REUSE DATA)
- SNIC**
- THE CYPRUS INSTITUTE** (RESEARCH • TECHNOLOGY • INNOVATION)
- GWWDG** (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen)
- VŠB TECHNICKÁ UNIVERZITA OSTRAVA**
- IT4INNOVATIONS NÁRODNÍ SUPERPOČÍTAČOVÉ CENTRUM**
- cesnet**
- UNI-NETT**
- figma2**
- Outreach & sustainability** (represented by a circular arrow icon)
- Trust-IT Services** (Communicating ICT to markets)
- EUDAT LTD**

# Research data workflow



## Task 4.2: Planned Work

### Referencing

- Integration and interoperability of B2HANDLE and the DOI (Handle) service
- ‘PID for Instruments’ (PID4INST)
  - Metadata schema: adaptation for instruments (based on RDA)
  - Types exist in the type registry
  - Defined types can be used directly
- ‘B2TYPE’
  - Type registry / registries
- Integrity check
  - tool for PID and Type integrity check

### Discovery

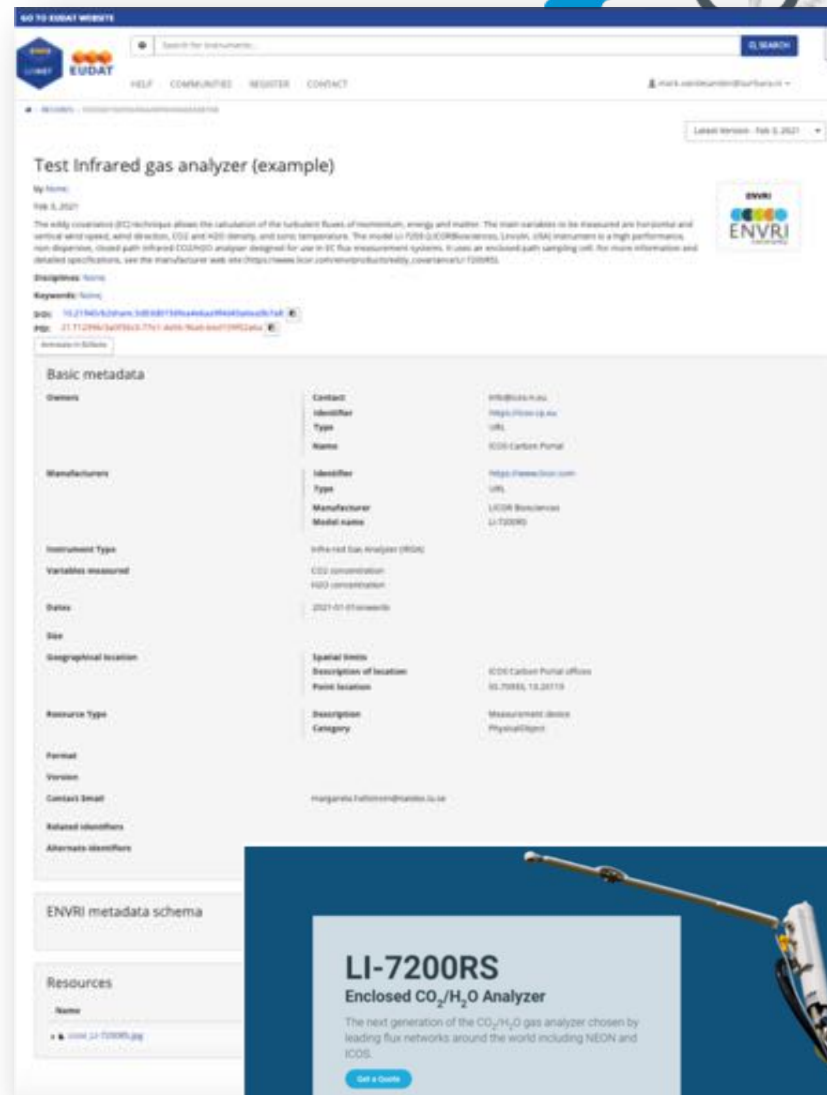
- PID Graph technology:
  - available for the DOI service as well as for Handles with information types
  - Integrate it for the DOI service with Handle info types
- Metadata:
  - datasets with handles + metadata aligned with DataCite metadata schema:
    - integrate into PID graph
- PID graph resources
  - identifiers offered by Datacite for e.g. people (ORCID), Institutions (ROR), Funders (ROR) in B2FIND
    - will be made available to B2FIND
- B2FIND enhancement
  - metadata schema development to support enhanced interoperability (e.g. EUDAT Core)

Data Infrastructure Capacity for EOSC  
(DICE)

# **USER INVOLVEMENT**

## General feedback

- ◆ Very good starting point
  - ◆ 50/50% quite useful/could be useful
  - ◆ Tips can be useful to understand which values should be registered
  - ◆ Date field not in free text
  - ◆ Size field not clear
  - ◆ Resource Type not necessary!
  - ◆ User registration procedure can be optimised
  - ◆ Less suitable for institutes with 1000's of instruments, but probably great for smaller research groups/institutions
- 
- ◆ Thanks to Maggie and ENVRI-FAIR



**Test Infrared gas analyzer (example)**

By Name: Feb 3, 2021

The eddy covariance (EC) technique allows the calculation of the turbulent fluxes of momentum, energy and water. The main variables to be measured are horizontal and vertical wind speed, wind direction, CO<sub>2</sub> and H<sub>2</sub>O density, and sonic temperature. The model LI-7200 (Licor Biosensors, Lincoln, USA) instrument is a high performance, non-dispersive, closed path infrared (CO<sub>2</sub>/H<sub>2</sub>O) analyzer designed for use in EC flux measurement systems. It uses an enclosed path sampling cell for more information and detailed specifications, see the manufacturer web site <https://www.licor.com/resources/technology/eddycovar/7200s/>.

Keywords: None

URI: <https://doi.org/10.21203/rs.3.rs-10000000/v1>

Page: 1/1

**Basic metadata**

<b>Owners</b>	<b>Contact</b>	ENVRI-FAIR
<b>Manufacturers</b>	<b>Identifier</b>	https://www.licor.com
<b>Instrument Type</b>	<b>Type</b>	URL
<b>Variables measured</b>	<b>Name</b>	CO <sub>2</sub> Carbon Portal
<b>Dates</b>	<b>Identifier</b>	https://www.licor.com
<b>Size</b>	<b>Type</b>	URL
<b>Geographical location</b>	<b>Manufacturer</b>	LICOR Biosensors
<b>Resource Type</b>	<b>Model name</b>	LI-7200RS
<b>Format</b>	<b>Instrument Type</b>	Infrared Gas Analyzer (IRGA)
<b>Version</b>	<b>Variables measured</b>	CO <sub>2</sub> concentration H <sub>2</sub> O concentration
<b>Contact Email</b>	<b>Dates</b>	2021-01-01 onwards
<b>Related identifiers</b>	<b>Size</b>	
<b>Alternate identifiers</b>	<b>Geographical location</b>	<b>Spatial Entity</b>
	<b>Description of location</b>	CO <sub>2</sub> Carbon Portal offices
	<b>Point location</b>	55.75855, 12.28119
	<b>Description</b>	Measurement device
	<b>Category</b>	Physical Object
		<a href="https://www.licor.com">https://www.licor.com</a>

**ENVRI metadata schema**

**Resources**

Name

- LI-7200RS.jpg



**LI-7200RS**  
Enclosed CO<sub>2</sub>/H<sub>2</sub>O Analyzer

The next generation of the CO<sub>2</sub>/H<sub>2</sub>O gas analyzer chosen by leading flux networks around the world including NEON and ICOS.

[Get a Quote](#)