# PIDs for instruments:  B2INST perspective

Dr. Tibor Kálmán

tibor.kalman@gwdg.de

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

**Persistent Identifiers (PIDs) for Facility Research**

ExPaNDS Workshop

22. October 2021.

SCAN ME

# Agenda

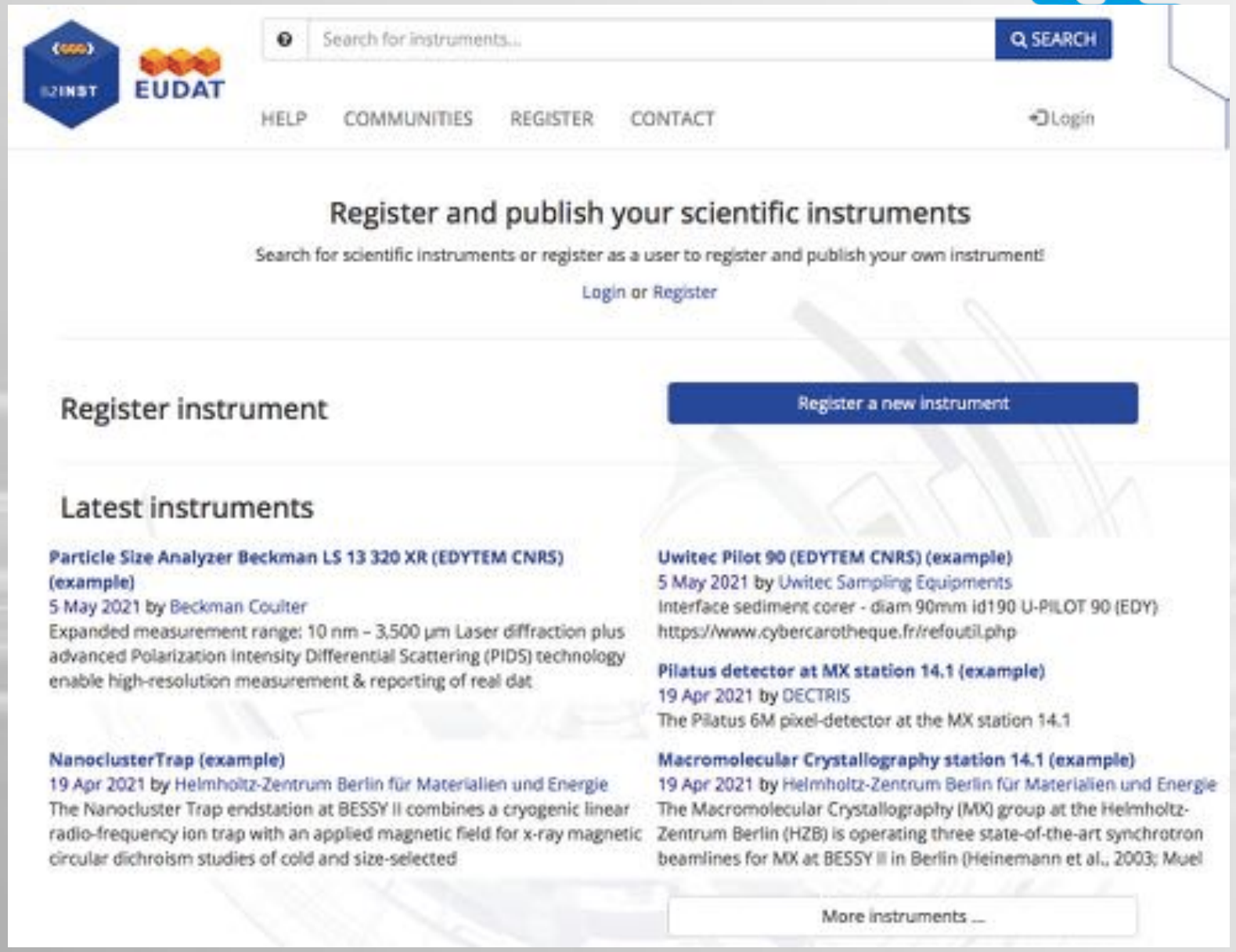
- Motivation for a public service to reference instruments
- Information modeling:
  - adapt a “mature“ schema
  - support community specific requirements
- Sustainability
  - through Persistent Identifiers (PIDs)
  - through organizational structures
  - through technology
  - through data re-use
-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

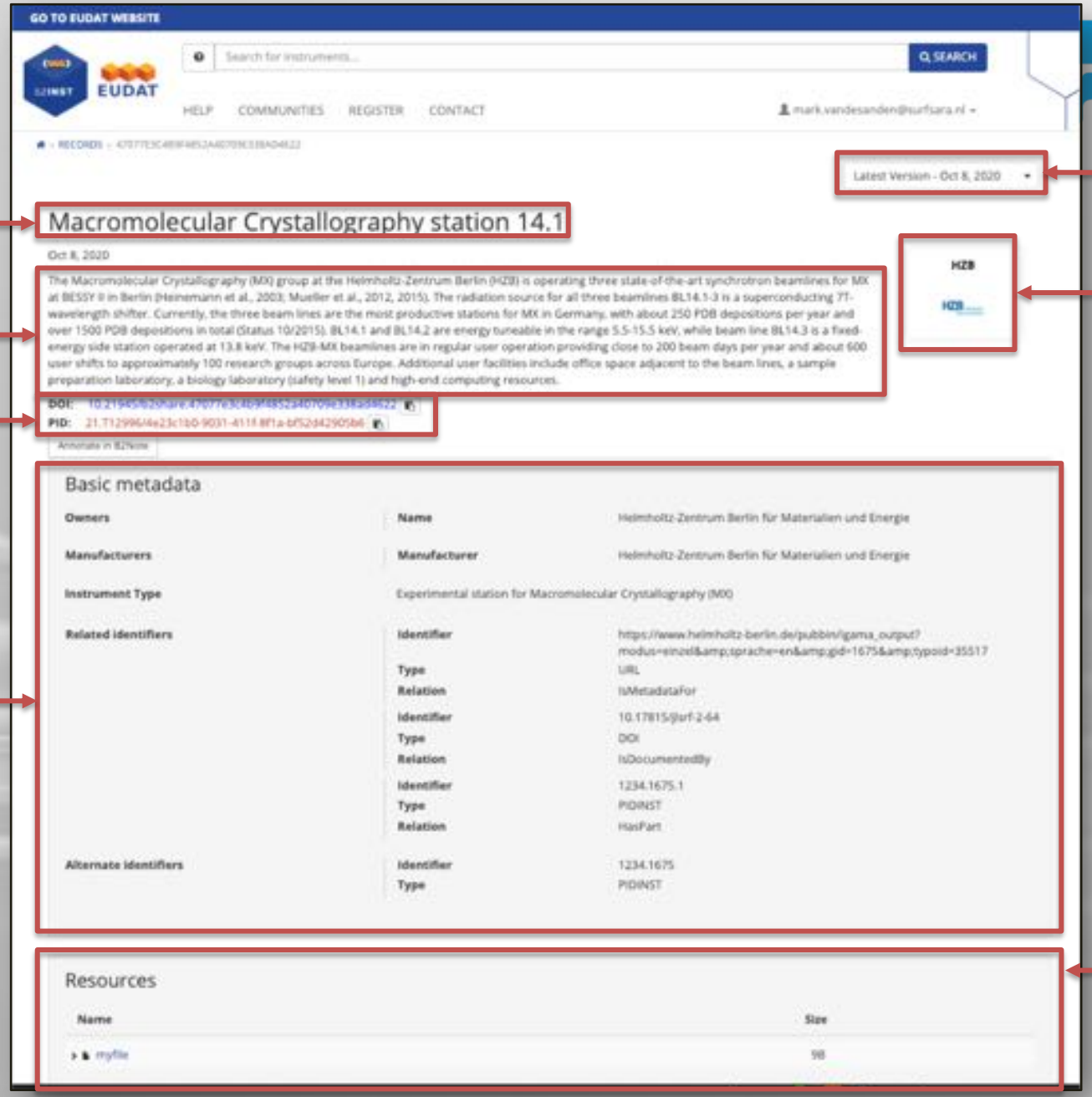
*(Slides also from:*
*Sanden, Piggelen, Fenner, Schwardmann, Kalman: Update on B2INST @RDA #17*
*EUDAT Secreteriat & DICE project)*

# 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**


Search for instruments...
SEARCH
EUDAT
HELP
COMMUNITIES
REGISTER
CONTACT
Login
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 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 µm 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
NanoclusterTrap (example)
19 Apr 2021 by Helmholtz-Zentrum Berlin für Materialien und Energie
The Nanocluster Trap endstation at BESSY II combines a cryogenic linear radio-frequency ion trap with an applied magnetic field for x-ray magnetic circular dichroism studies of cold and size-selected
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 II in Berlin (Heinemann et al., 2003; Muel
More instruments ...




Name

Description

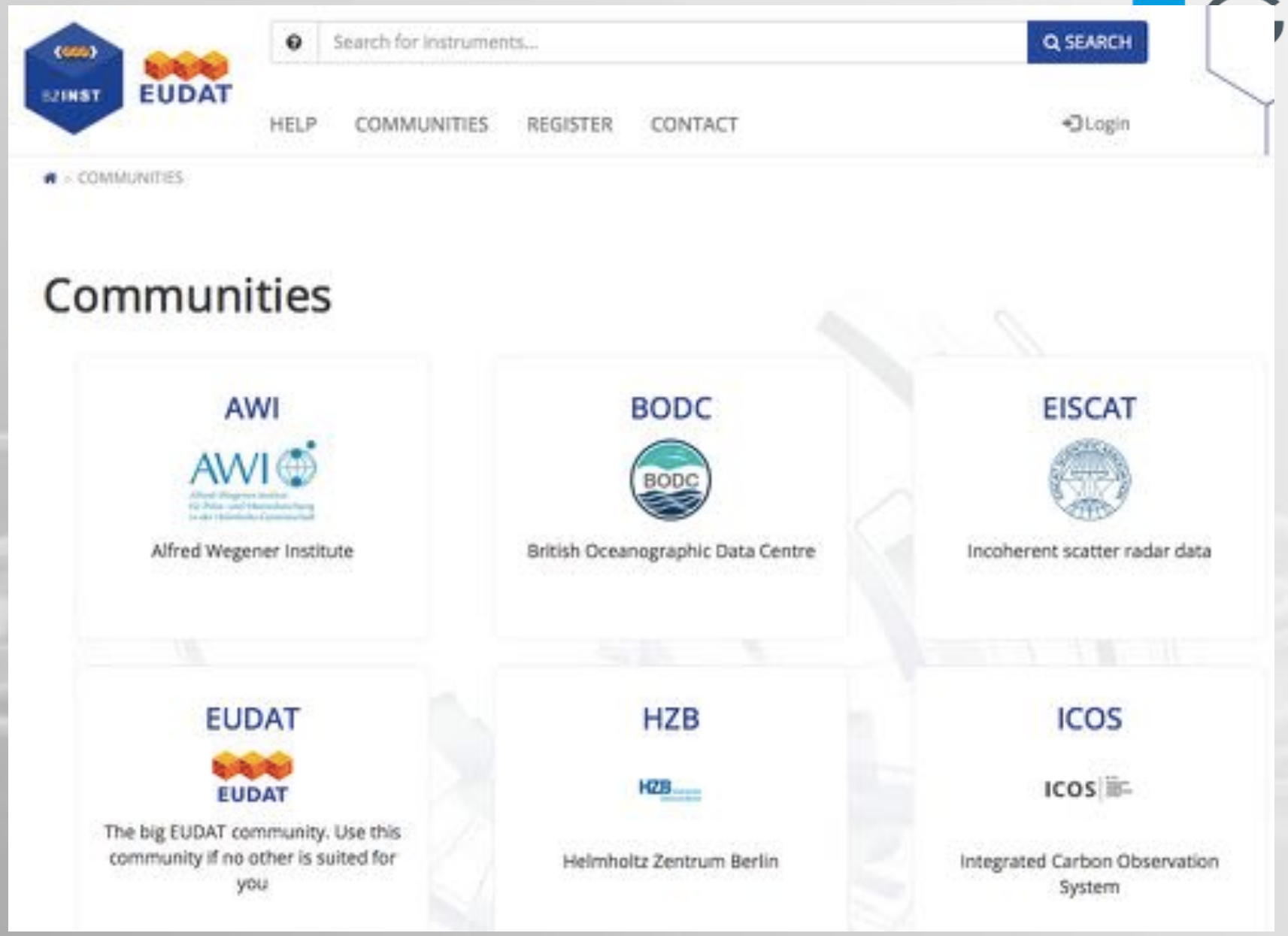
PID

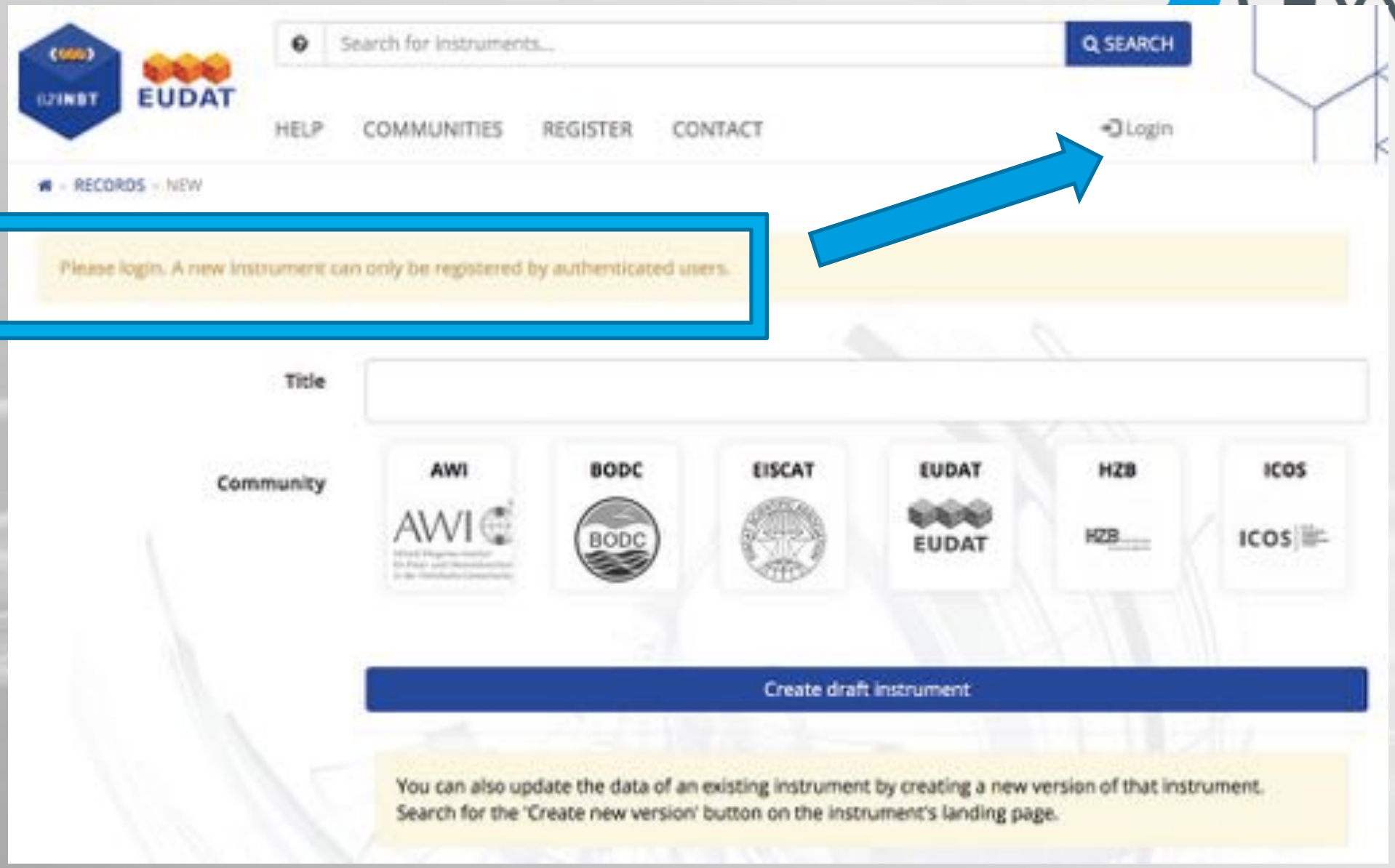
Basic Metadata

Versioning

Community Domain

Optional Files


Search for instruments...
SEARCH
HELP
COMMUNITIES
REGISTER
CONTACT
Login
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


Search for instruments...
SEARCH
EUDAT
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

# 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 loose 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!

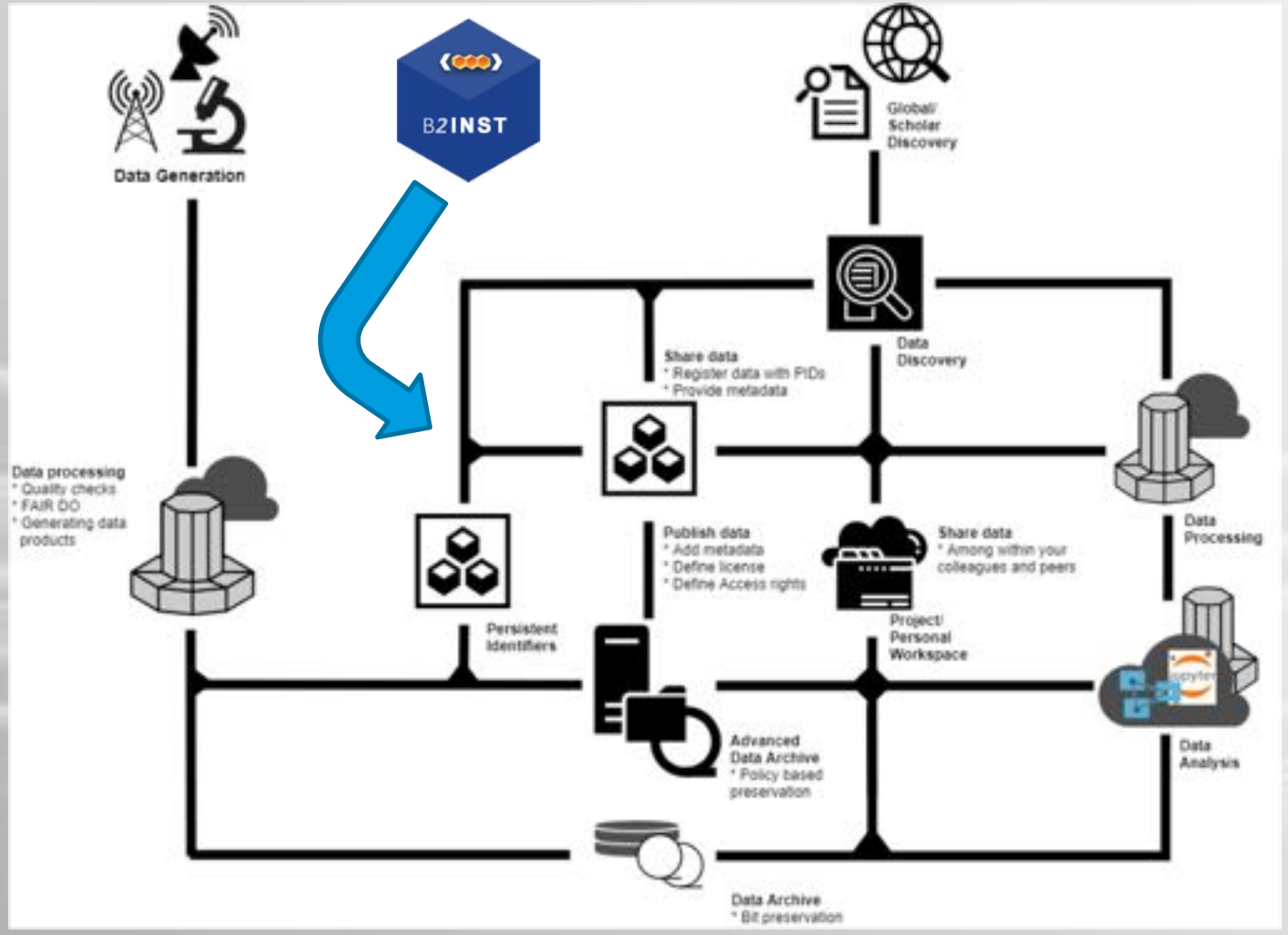
# Site Göttingen, Germany


- 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



EUDAT Collaborative Data Infrastructure

# The EUDAT Service Suite

B2ACCESS
Identity & Authorisation

B2DROP
Synch and Share Research Data

B2FIND
Find Research Data

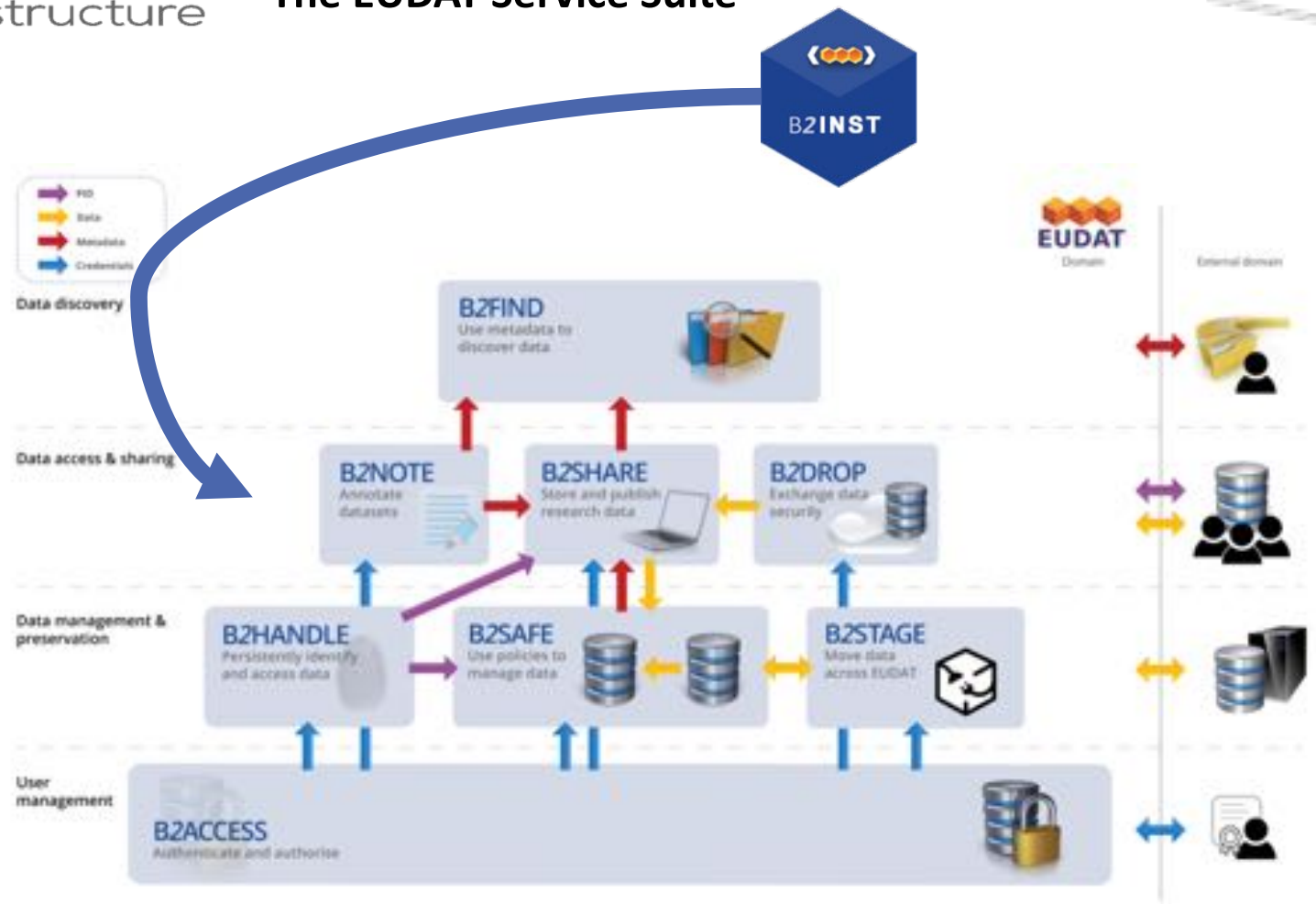
B2SHARE
Store and Publish Research Data

B2SAFE
Replicate Research Data Safely

B2HANDLE
Register your Research Data

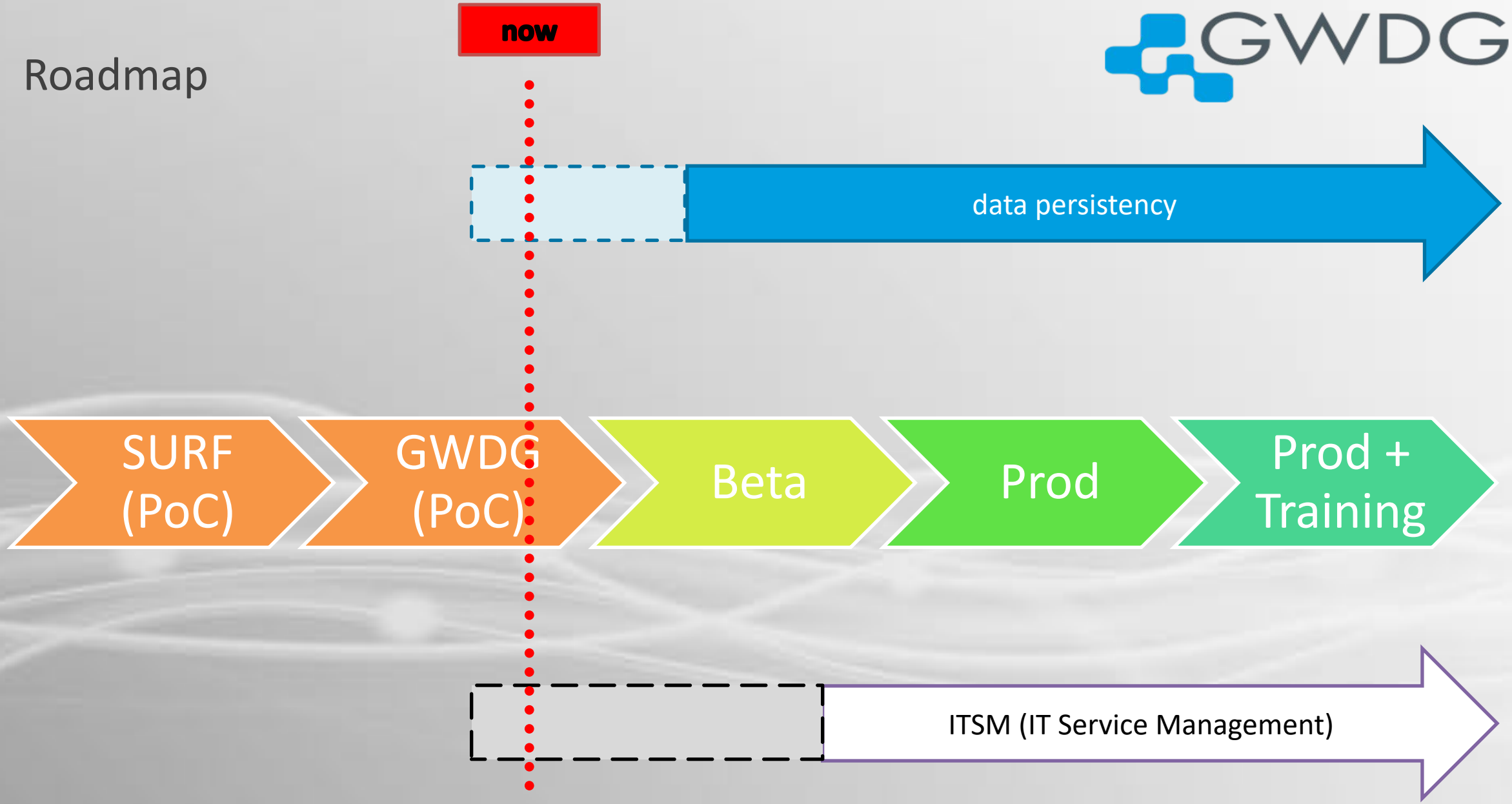
B2STAGE
Get Data to Computation

B2NOTE



# ROADMAP,
# ONGOING ACTIVITIES

# Roadmap


now
data persistency
SURF (PoC)
GWDG (PoC)
Beta
Prod
Prod + Training
ITSM (IT Service Management)

# 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

# DICE WP4 Tasks

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!

PID (Handle):
**21.11101/0000-0007-F40B-A**

Resolvable URL:
https://hdl.handle.net/**21.11101/0000-0007-F40B-A**

Tibor Kálmán
tibor.kalman@gwdg.de

www.gwdg.de

Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen

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

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/**

ePIC
Persistent Identifiers for eResearch

# 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
    - GWDG (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.
  - GWDG: 21.xyz
  - IDF (DOI): 10.xyz
- Web:
  - **https://www.dona.net/**

Common Data Infrastructure
(EUDAT)

# SUSTAINABILITY:
# THE INFRASTRUCTURE

# The EUDAT members

EUDAT Collaborative Data Infrastructure

GWDG

# The EUDAT Service Suite

B2ACCESS
Identity & Authorisation

B2DROP
Synch and Share Research Data

B2FIND
Find Research Data

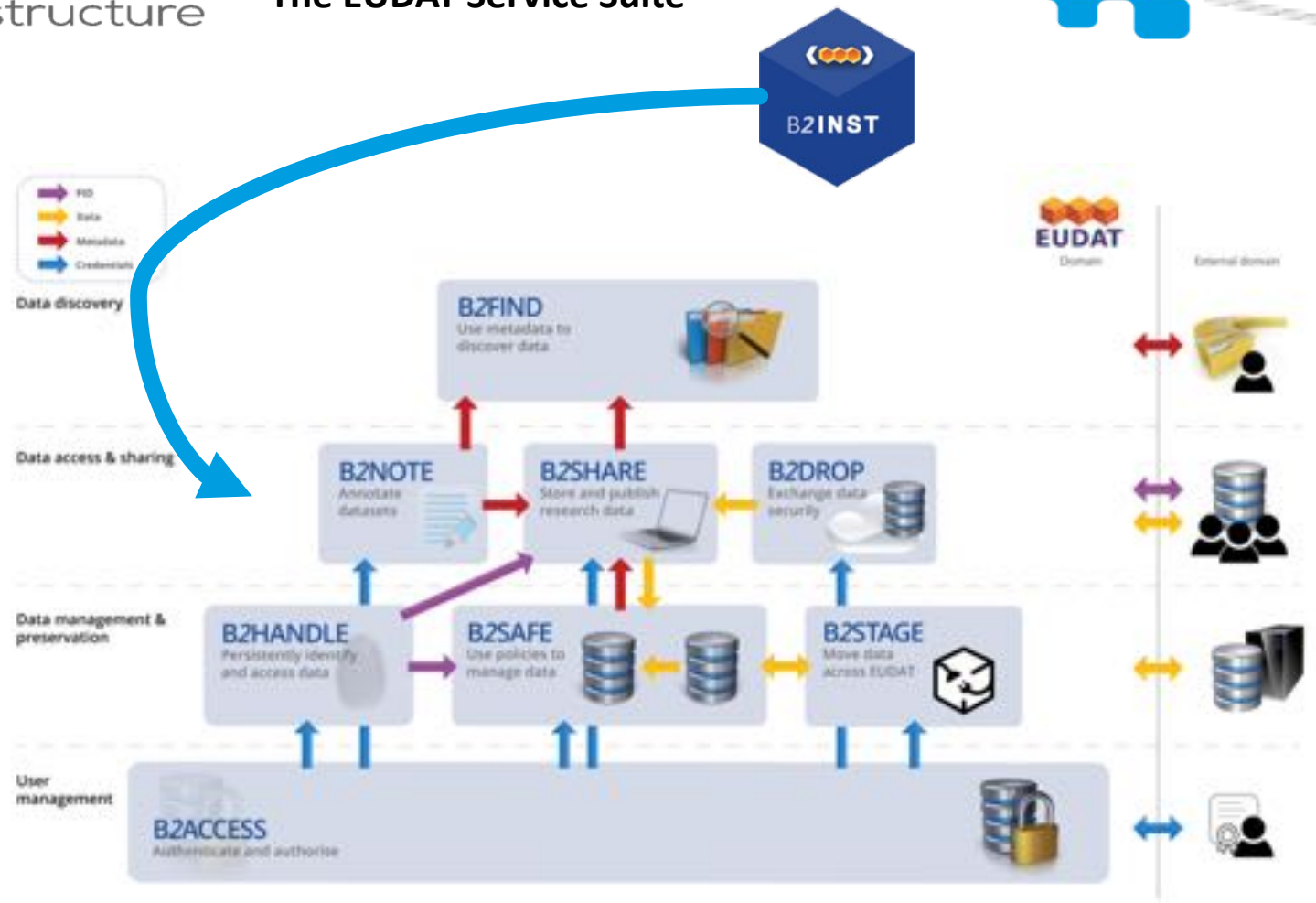
B2SHARE
Store and Publish Research Data

B2SAFE
Replicate Research Data Safely

B2HANDLE
Register your Research Data

B2STAGE
Get Data to Computation

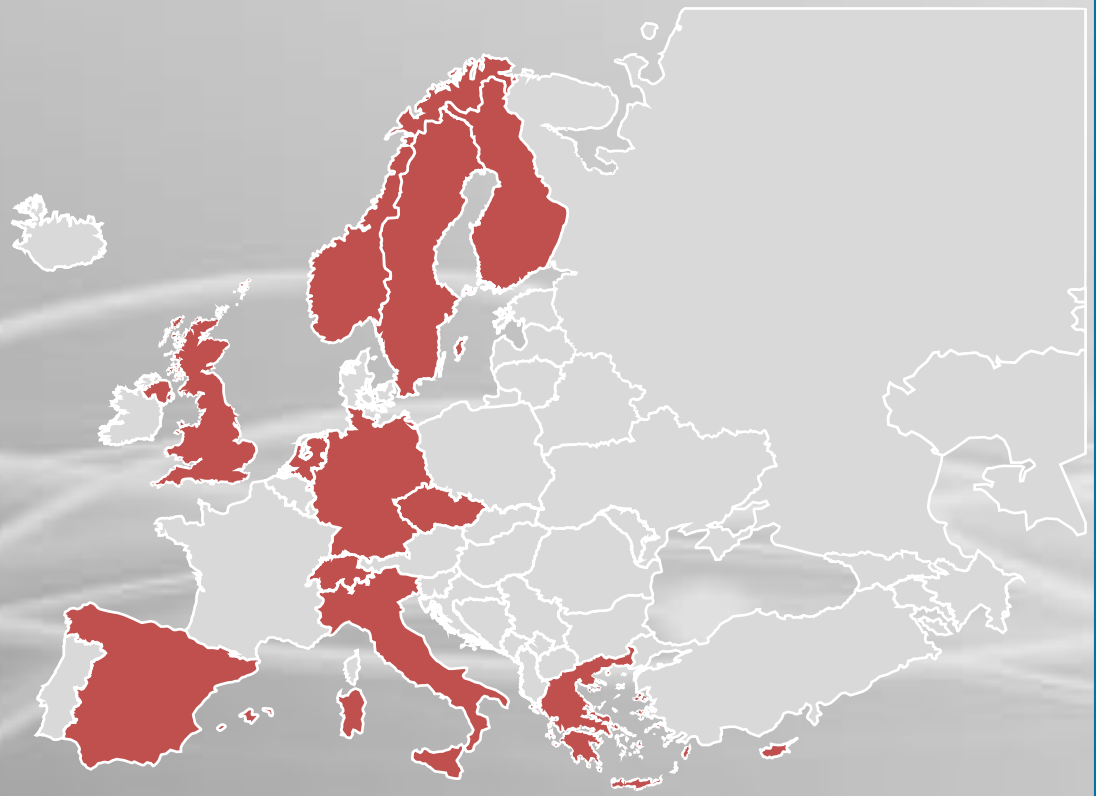
B2NOTE



Data Infrastructure Capacity for EOSC
(DICE)

# THE PROJECT

# Consortium



Service providers

CINECA

JÜLICH Forschungszentrum

CSC

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

MPCDF Max Planck Computing & Data Facility

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

GWDG Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

VŠB Technická univerzita Ostrava | IT4Innovations Národní superpočítačové centrum

cesnet

UNINETT

Sigma2

Users' communities

EGI DG

ASTRON

ICOS Carbon Portal

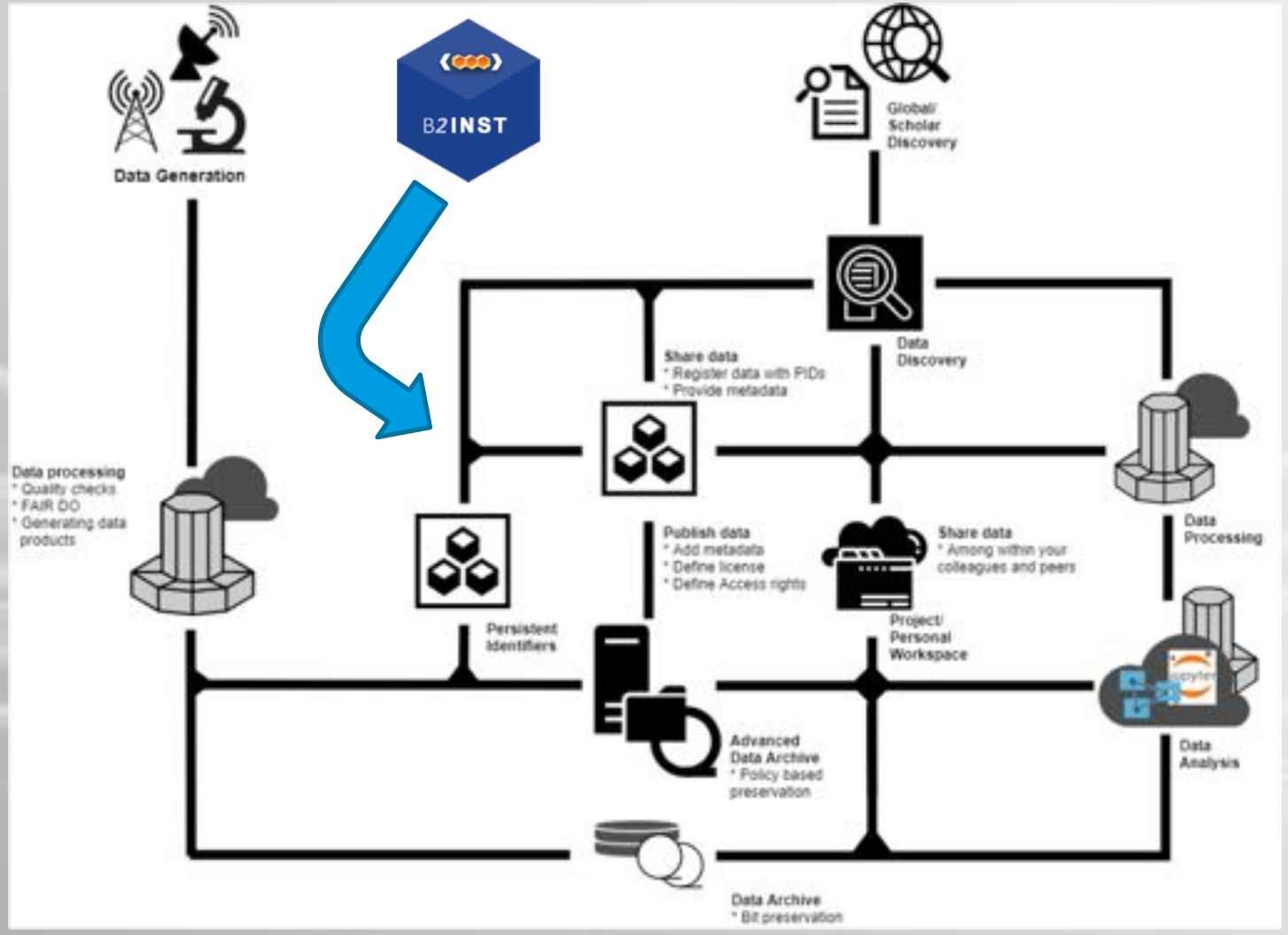
UCL

Outreach & sustainability

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