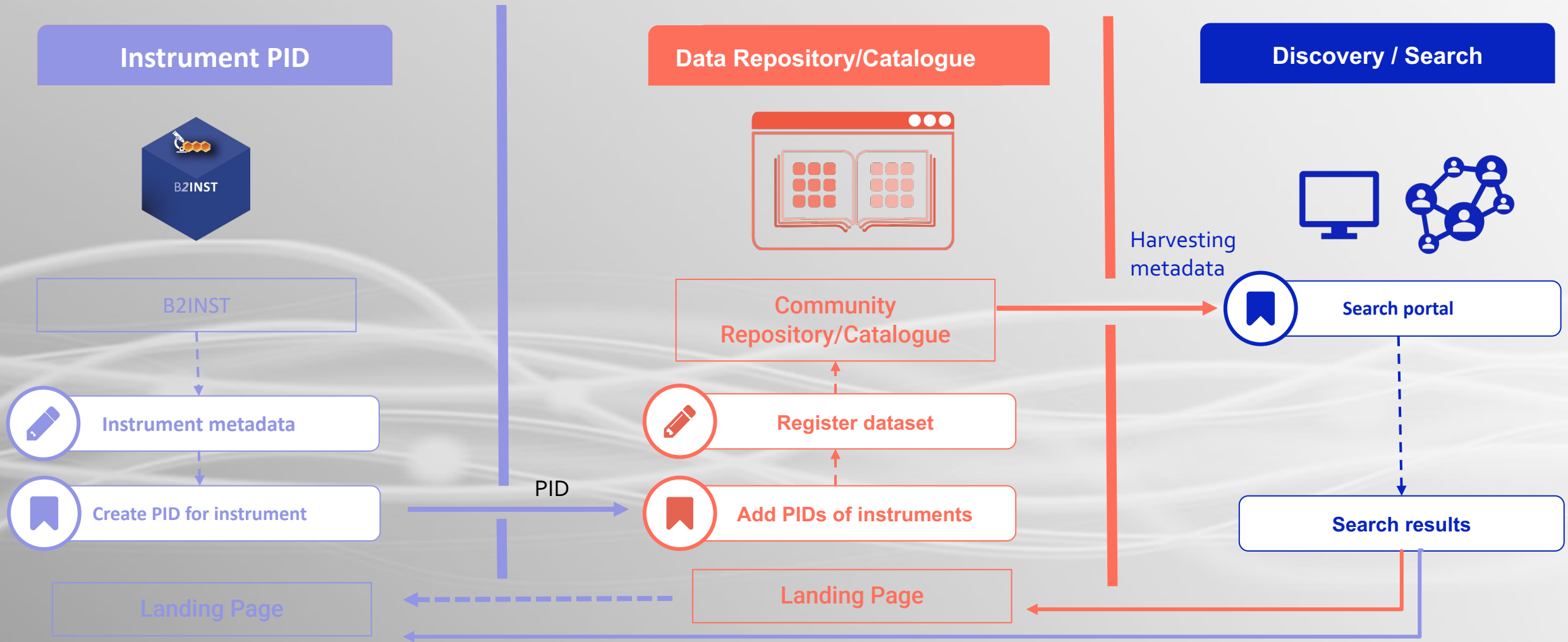
B2INST: Registration and Identification of Instruments

Dr. Tibor Kálmán
tibor.kalman@gwdg.de
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Online Seminar - PIDs for Instruments
PID Network Deutschland
07. May 2024

MOTIVATION

Integration of Instrument PIDs



Motivation

Trends identified:

- Dedicated (community) registries for sensors
  - Metadata is heterogeneous
- Registries started to assign PIDs for their instruments
  - Instrument PID: an emerging PID type

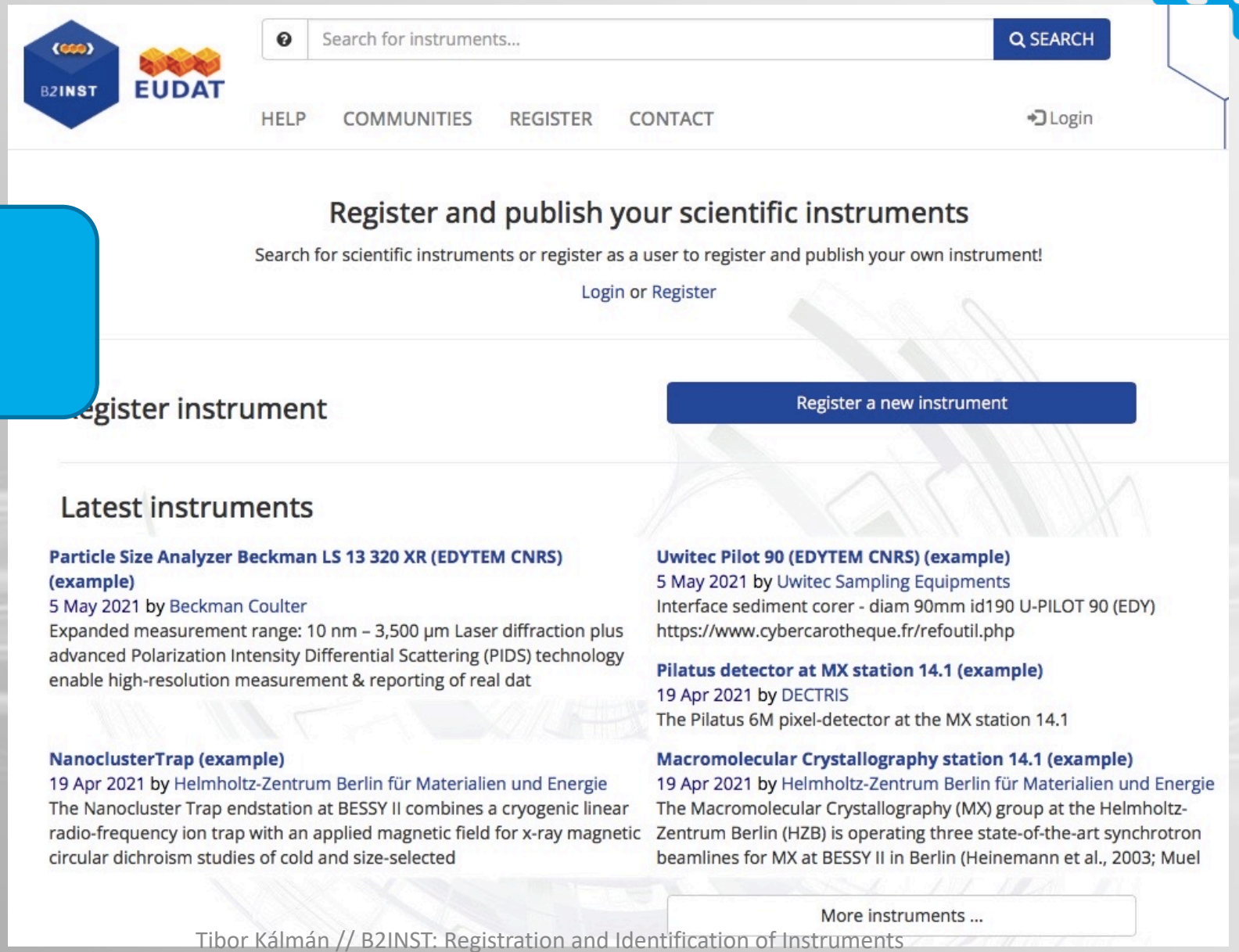
Create a public service to describe, register, and reference instruments

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**

B2INST

The B2INST service



B2INST: Some design principles

- **REST** backend + graphical/**web-UI**-based tool; based on B2SHARE (Zenodo)
- Designed to deal with the **complex metadata** requirements of various communities.
- Enables **global** and **unique identification** of instruments.
- Support various instrument types:
  - measuring instruments, such as sensors used in environmental science,
  - DNA sequencers used in life sciences,
  - microscopes used in medical domains
  - and many more.
- Each instrument has its own landing page (public), which:
  - displays the **basic** data of the instrument record (according to RDA PIDINST WG),
  - but can include **community specific** metadata elements,
  - possible to **add files** (like manuals, pictures, etc)
- **Authenticated users** (SSO) can create new instrument records or maintain the metadata
- **„Publication“ workflow**
  - A record is initially created as a **draft** and then can be published.
  - **Versioning**
- **Search** functionality
  - UI + API

Instrument Identifiers: the 'B2INST' service

B2INST - service to reference, describe and register (scientific) instruments

Instrument Name

Instrument Description

Persistent Identifier

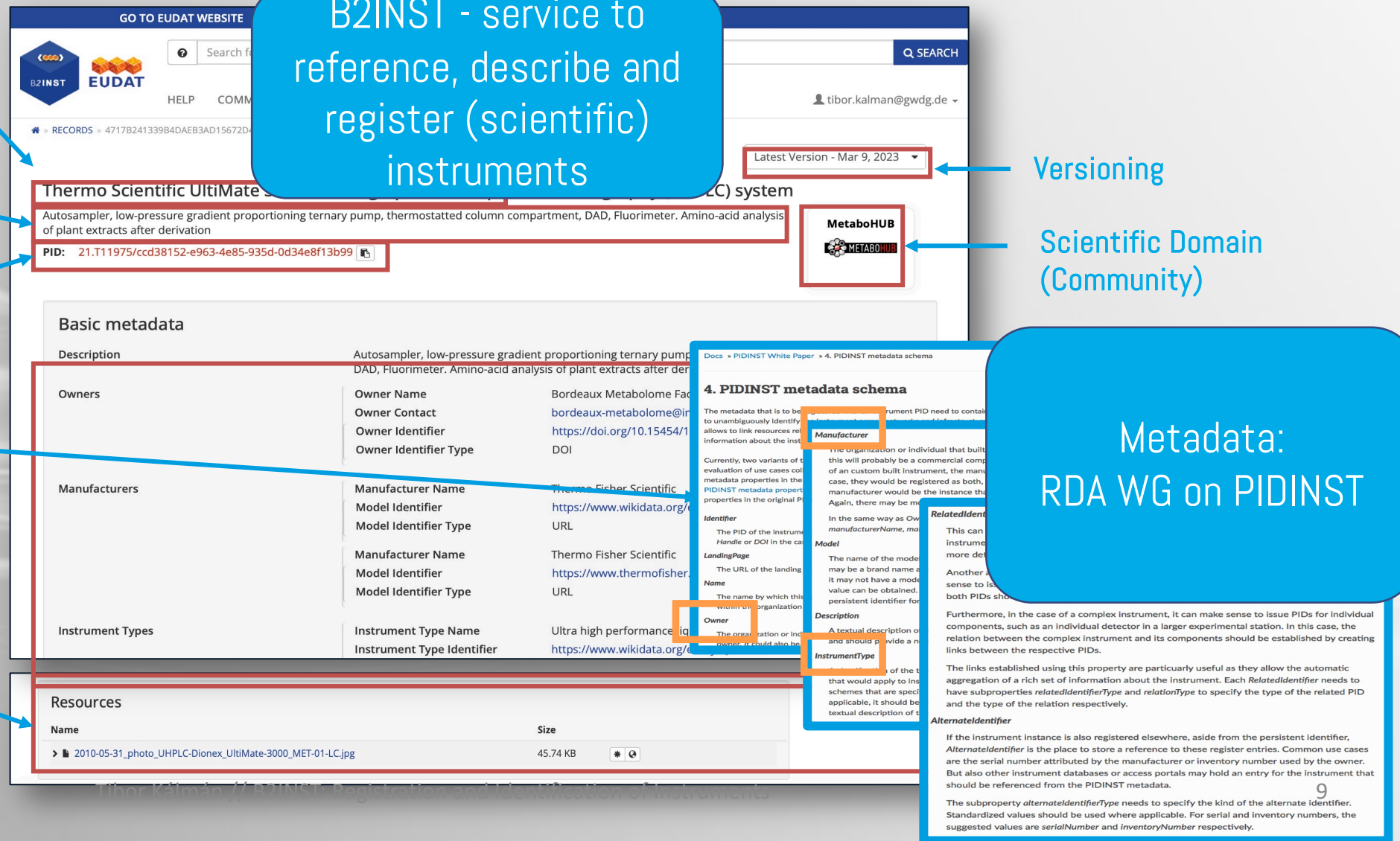
Basic Metadata

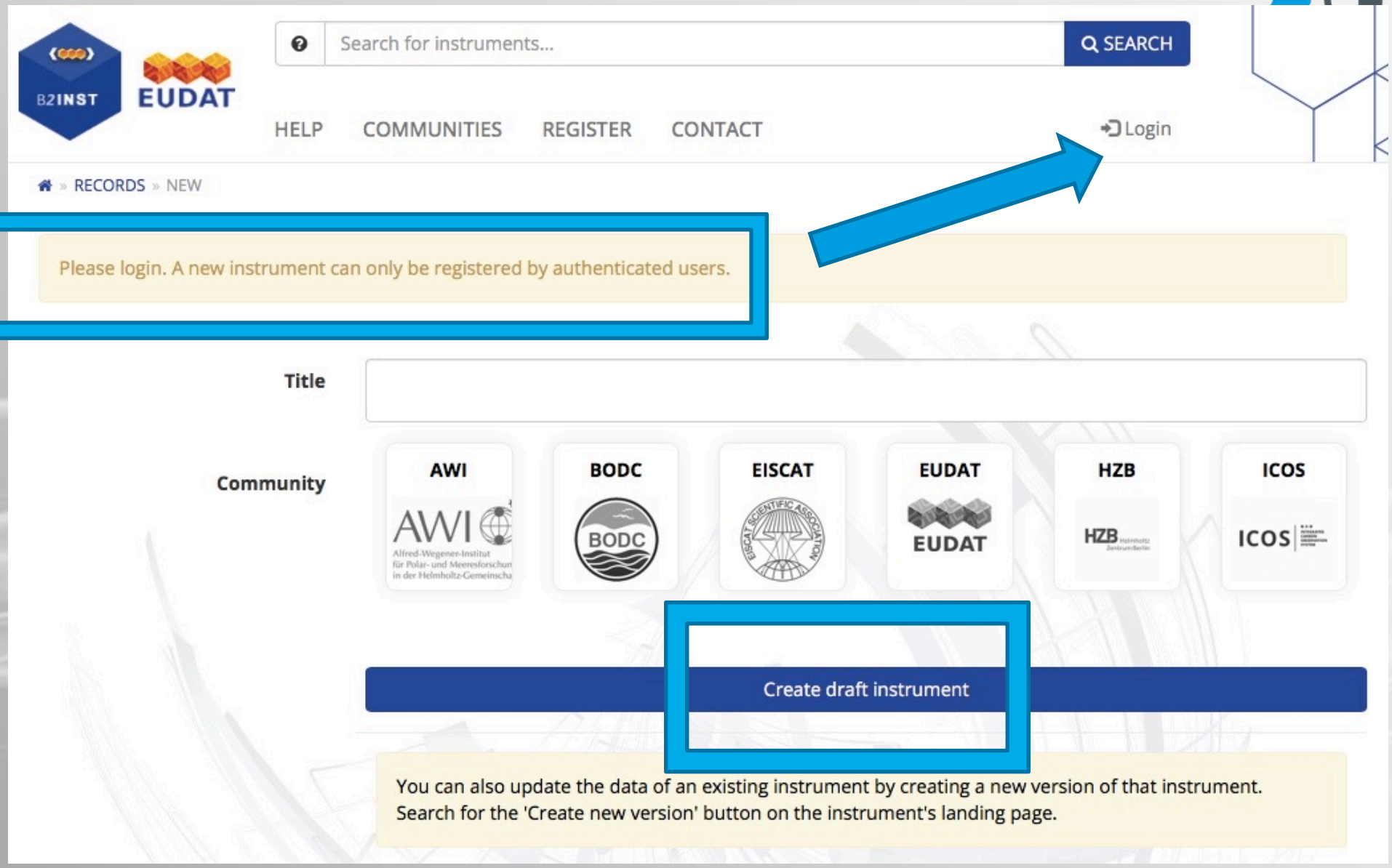
Optional Files (manuals, pictures of the instrument, etc)

Versioning

Scientific Domain (Community)

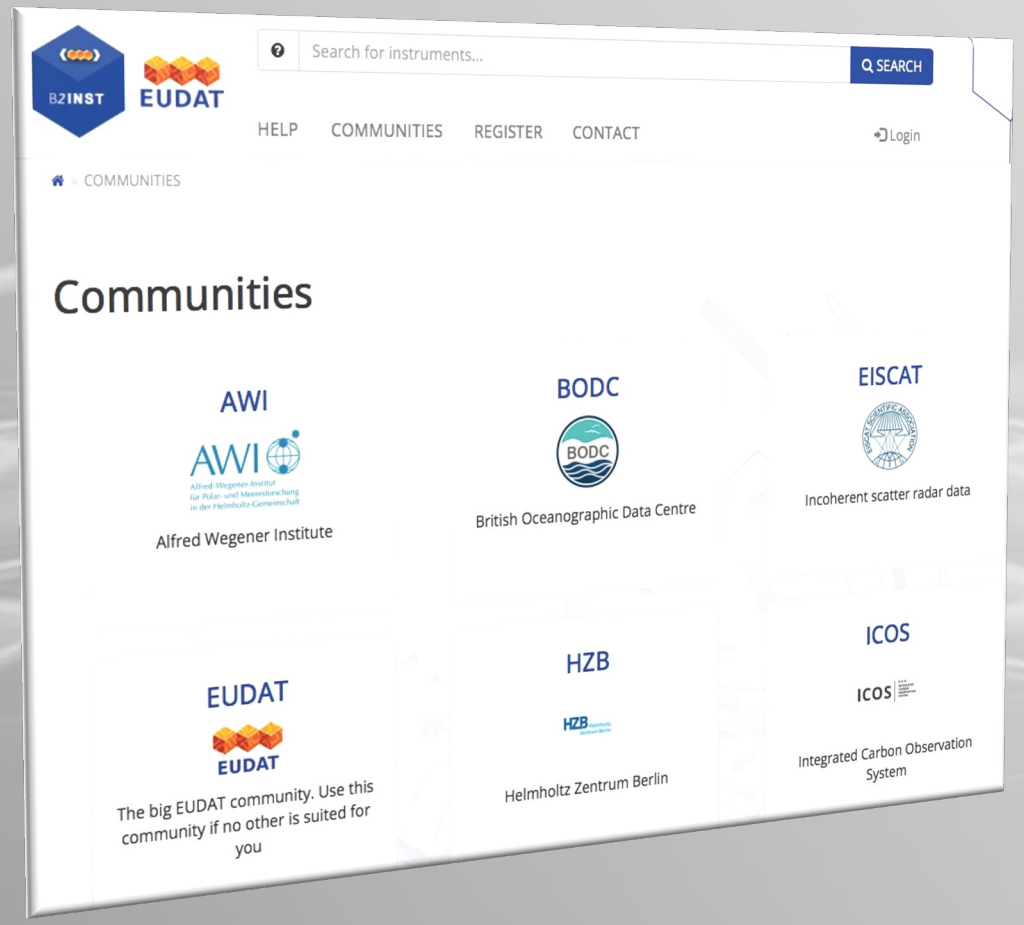
Metadata: RDA WG on PIDINST



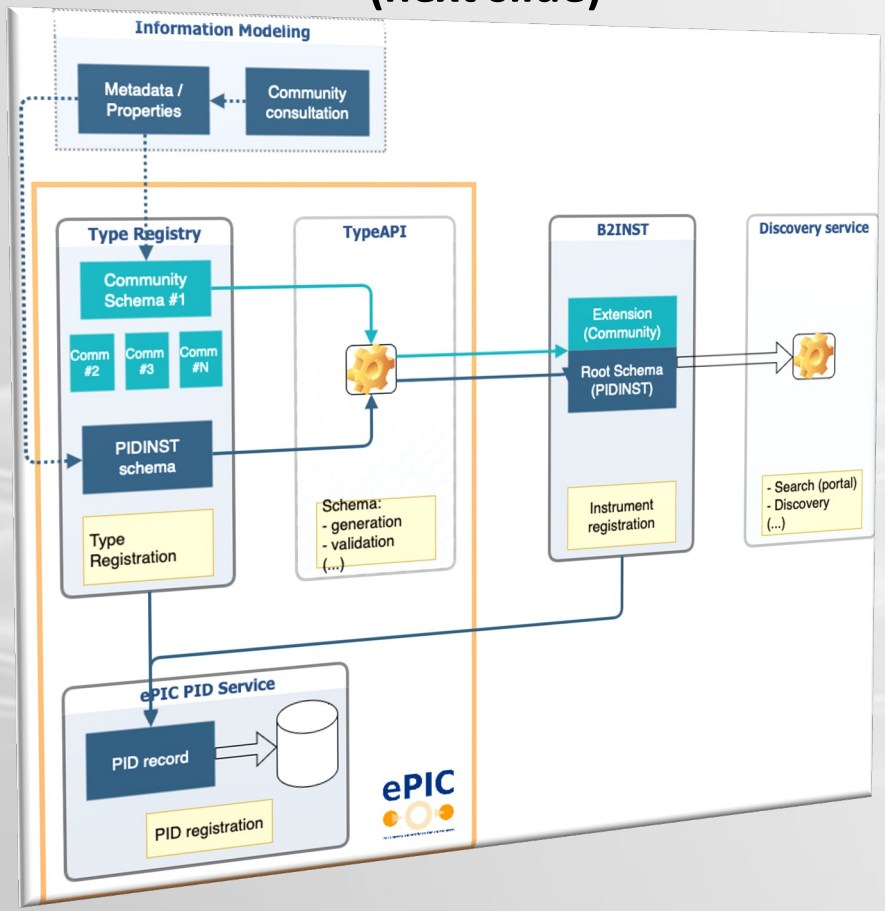

B2INST
EUDAT
Search for instruments...
SEARCH
HELP
COMMUNITIES
REGISTER
CONTACT
Login
» RECORDS » NEW
Please login. A new instrument can only be registered by authenticated users.
Title
Community
AWI
BODC
EISCAT
EUDAT
HZB
ICOS
Create draft instrument
You can also update the data of an existing instrument by creating a new version of that instrument.
Search for the 'Create new version' button on the instrument's landing page.

Supporting Communities

Schema Customization for Communities



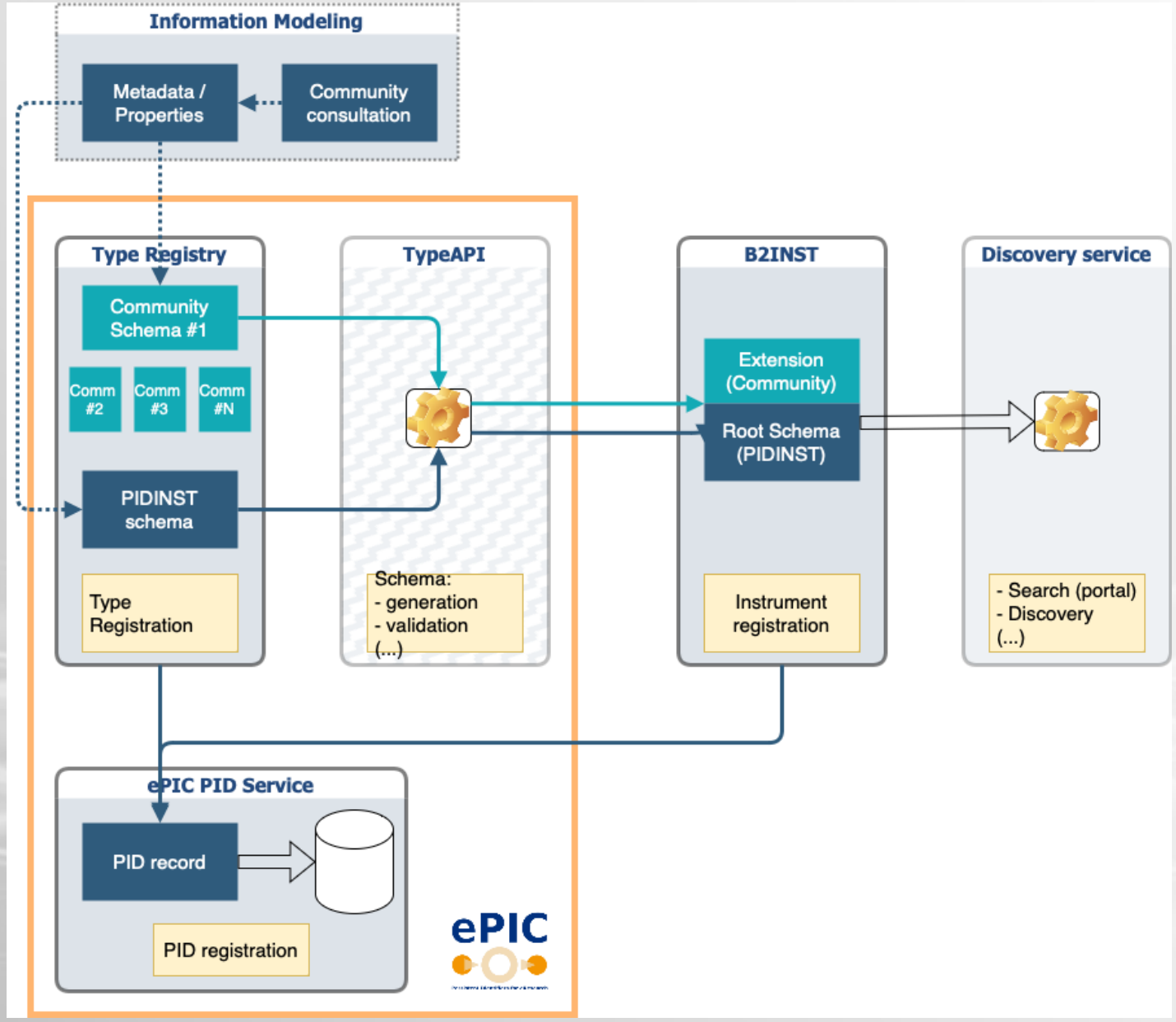
Schema Definition & Management (next slide)



FAIRify Instruments...

- FAIRify information on instrument by
  - registering its metadata and
  - assigning the instrument a PID
- Instruments are non-digital objects.
- The challenge:
  - how to FAIRify non-digital objects
  - and make these discoverable

➔B2INST offers this information as digital representations making the information discoverable when PIDs are assigned.



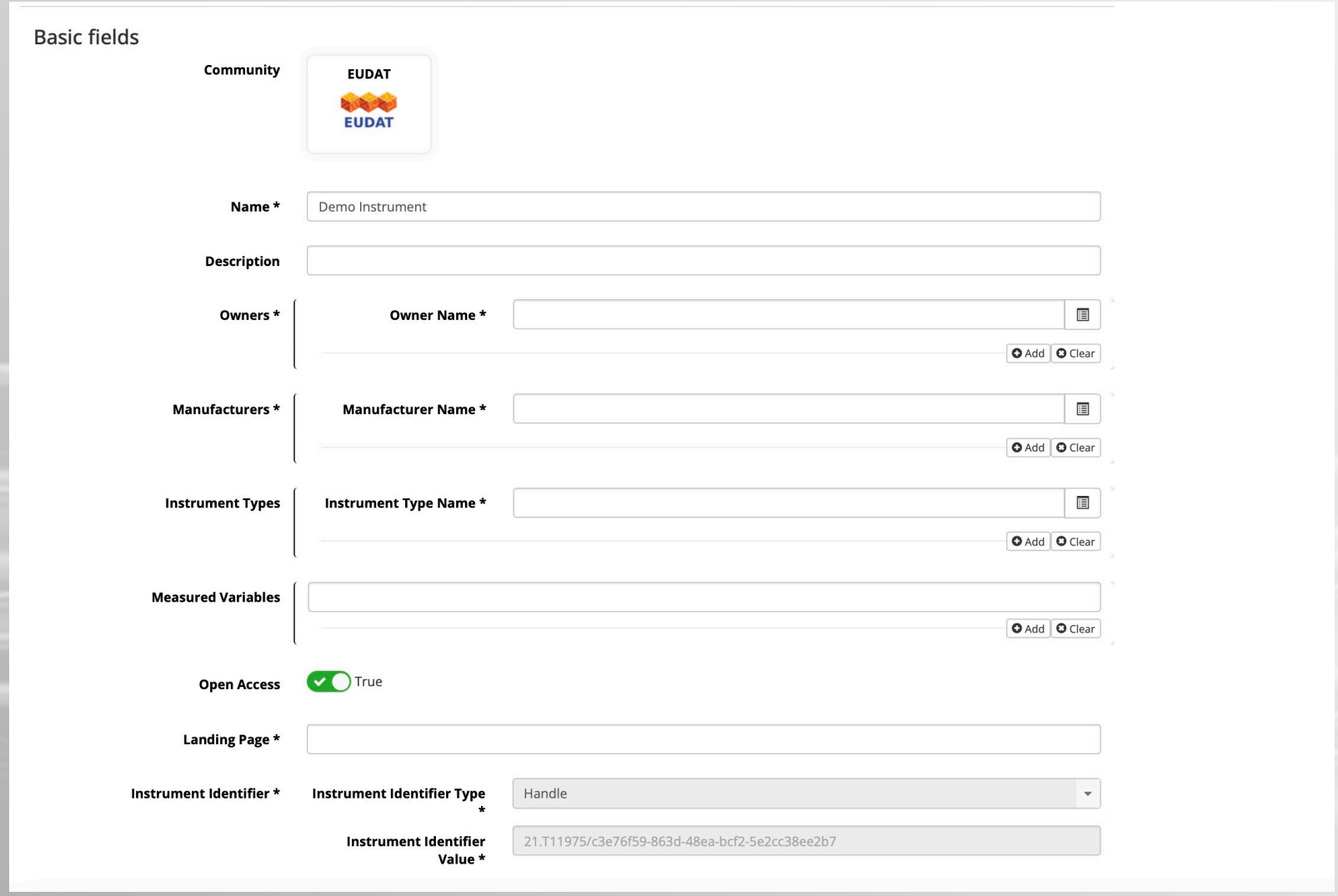
Thank you!

Dr. Tibor Kálmán
tibor.kalman@gwdg.de
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Online Seminar - PIDs for Instruments
PID Network Deutschland
07. May 2024

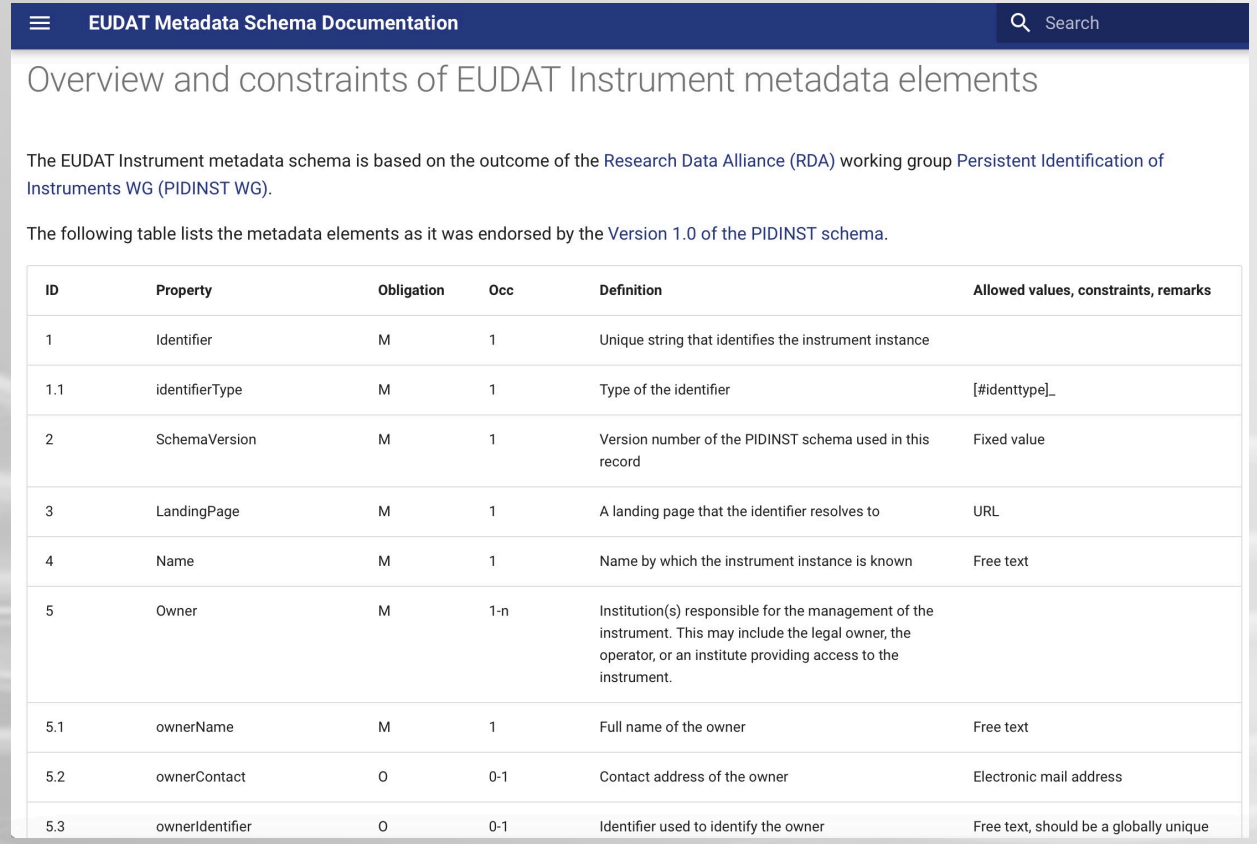
BACKUP SLIDES

Register a new instrument



Additional elements: Community Metadata + Files

Metadata Schema



EUDAT Metadata Schema Documentation

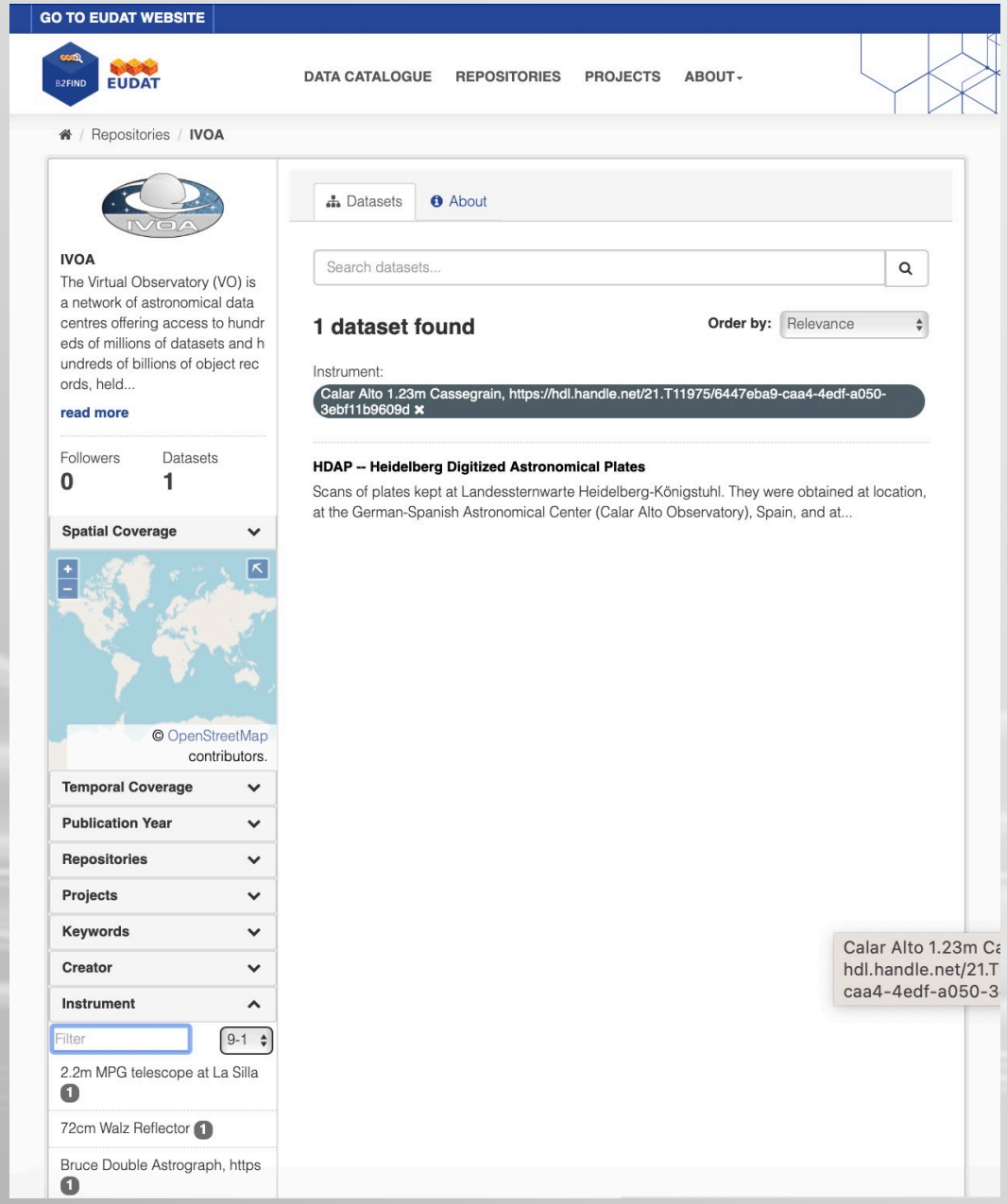
Search

Overview and constraints of EUDAT Instrument metadata elements

The EUDAT Instrument metadata schema is based on the outcome of the Research Data Alliance (RDA) working group Persistent Identification of Instruments WG (PIDINST WG).

The following table lists the metadata elements as it was endorsed by the Version 1.0 of the PIDINST schema.

ID	Property	Obligation	Occ	Definition	Allowed values, constraints, remarks
1	Identifier	M	1	Unique string that identifies the instrument instance	
1.1	identifierType	M	1	Type of the identifier	[#identtype]_
2	SchemaVersion	M	1	Version number of the PIDINST schema used in this record	Fixed value
3	LandingPage	M	1	A landing page that the identifier resolves to	URL
4	Name	M	1	Name by which the instrument instance is known	Free text
5	Owner	M	1-n	Institution(s) responsible for the management of the instrument. This may include the legal owner, the operator, or an institute providing access to the instrument.	
5.1	ownerName	M	1	Full name of the owner	Free text
5.2	ownerContact	O	0-1	Contact address of the owner	Electronic mail address
5.3	ownerIdentifier	O	0-1	Identifier used to identify the owner	Free text, should be a globally unique


GO TO EUDAT WEBSITE
B2FIND
EUDAT
DATA CATALOGUE
REPOSITORIES
PROJECTS
ABOUT
Repositories / IVOA
IVOA
The Virtual Observatory (VO) is a network of astronomical data centres offering access to hundr eds of millions of datasets and h undreds of billions of object rec ords, held...
read more
Followers
0
Datasets
1
Spatial Coverage
© OpenStreetMap contributors.
Temporal Coverage
Publication Year
Repositories
Projects
Keywords
Creator
Instrument
Filter
9-1
2.2m MPG telescope at La Silla 1
72cm Walz Reflector 1
Bruce Double Astrograph, https 1
Datasets
About
Search datasets...
1 dataset found
Order by:
Relevance
Instrument:
Calar Alto 1.23m Cassegrain, https://hdl.handle.net/21.T11975/6447eba9-caa4-4edf-a050-3ebf11b9609d
HDAP -- Heidelberg Digitized Astronomical Plates
Scans of plates kept at Landessternwarte Heidelberg-Königstuhl. They were obtained at location, at the German-Spanish Astronomical Center (Calar Alto Observatory), Spain, and at...
Calar Alto 1.23m Ca
hdl.handle.net/21.T
caa4-4edf-a050-3

GWDG B2FIND

B2INST and its harvesting API

- B2INST and its harvesting API
  - Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) API has been used in B2INST
  - The instrument metadata is exposed to other services in a structured manner
    - so that the other services can make requests to the OAI-PMH API to harvest all the metadata related to the instruments.
  - One of the key advantages of the OAI-PMH is its simplicity. The protocol is based on the HTTP protocol and uses a small set of verbs and nouns to define the metadata harvesting operations.

- We adapted the record serializers in B2INST to support the following metadata formats via the OAI-PMH API:
  - Dublin Core (OAI-DC)
  - EUDAT Core
  - EUDAT Extended
  - DataCite
- Our record serializers were implemented in a way that they are able to expose the following metadata fields for the registered instruments (as shown on Figure XXX):
  - Title, Description and Date
  - Resource Identifier (as Handle PID),
  - Resource Identifier (as URL to the backend API),
  - b2rec internal format
  - Rights management

OAI-PMH record (ListRecord) in B2INST

OAI Record: oai:b2inst-test.gwdg.de:b2rec/4717b241339b4daeb3ad15672d4fc28a

OAI Record Header

OAI Identifier	oai:b2inst-test.gwdg.de:b2rec/4717b241339b4daeb3ad15672d4fc28a **oai_dc** **formats**
Datestamp	2023-03-09T13:31:36Z
setSpec	69ea278d-457c-48b4-8904-74019037672b **Identifiers** **Records**

Dublin Core Metadata (oai_dc)

Date	2023-03-09T13:31:36.631015+00:00
Description	Autosampler, low-pressure gradient proportioning ternary pump, thermostatted column compartment, DAD, Fluorimeter. Amino-acid analysis of plant extracts after derivation
Resource Identifier	http://hdl.handle.net/21.T11975/ccd38152-e963-4e85-935d-0d34e8f13b99
Resource Identifier	https://b2inst-test.gwdg.de/api/records/4717b241339b4daeb3ad15672d4fc28a
Resource Identifier	oai:b2inst-test.gwdg.de:b2rec/4717b241339b4daeb3ad15672d4fc28a
Rights Management	info:eu-repo/semantics/openAccess
Title	Thermo Scientific UltiMate 3000 ultra-high-pressure liquid chromatography (UHPLC) system

Common Data Infrastructure
(EUDAT)

SUSTAINABILITY: THE INFRASTRUCTURE

The EUDAT members

The CDI members (December 2020)

Generic, Integrated Service Providers

1. CSC - IT Center for Science (CSC) - FI
2. CINECA, Consortium of universities (CINECA) – IT
3. Barcelona Supercomputing Center (BSC) - ES
4. Science and Technology Facilities Council (STFC) – UK
5. SURFsara Bv (SURFsara) – NL
6. Karlsruhe Institute of Technology (KIT) - DE
7. Max Planck Computing and Data Facility (MPCDF) – DE
8. Cines - National Computing Center for Higher Education (CINES) – FR
9. Greek Research and Technology Network S.A. (GRNET) – GR
10. Jülich Research Centre (JÜLICH) - DE
11. Institute of Bioorganic Chemistry Polish Academy of Sciences (IBCH PAS) - PL
12. The Cyprus Institute - GR

Generic, Interoperable Service Providers

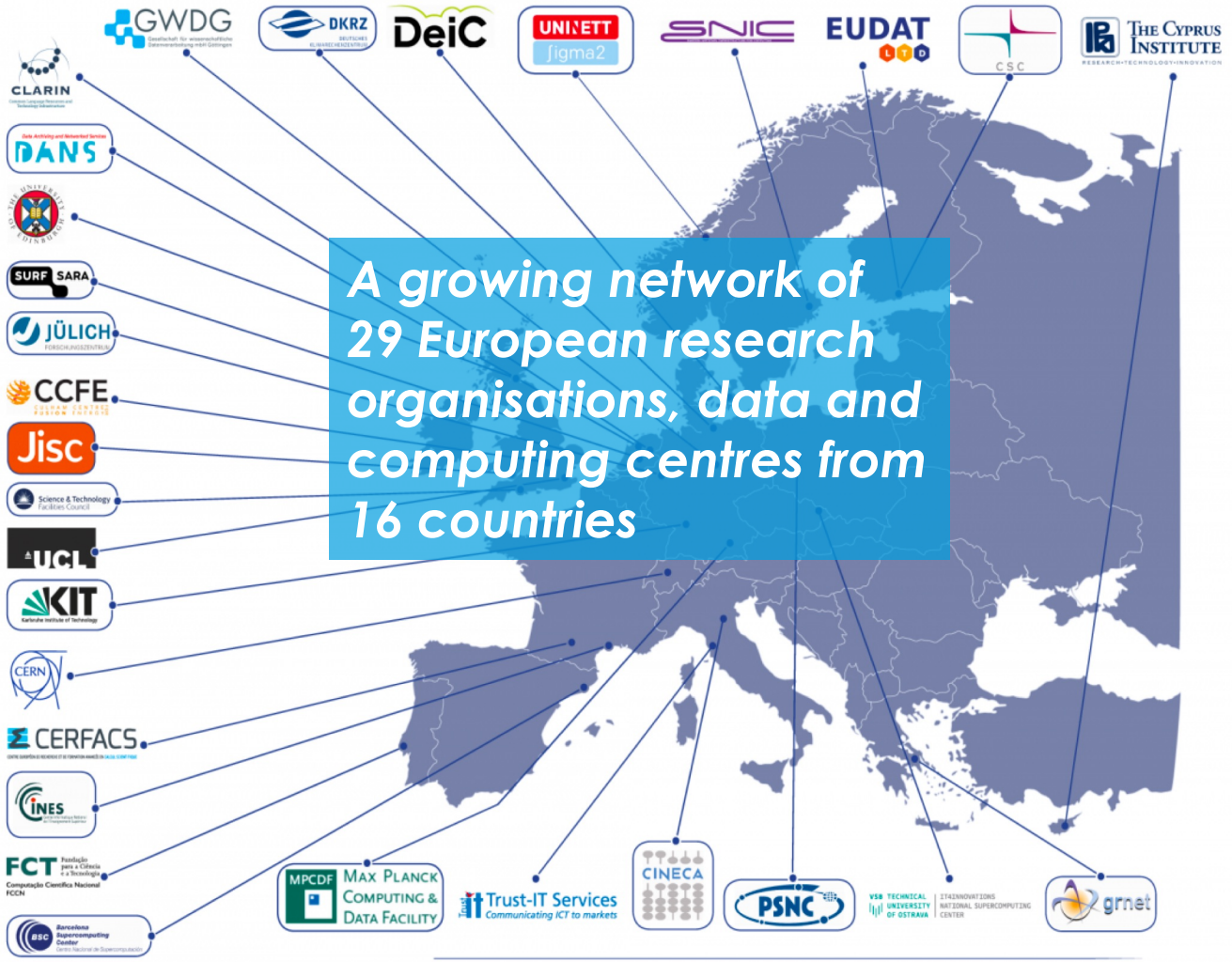
1. Uninett Sigma 2 As (SIGMA) - NO
2. Data Archiving and Networked Services (DANS) - NL
3. JISC LBG (JISC) – UK
4. Swedish National Infrastructure for Computing (SNIC) - SE
5. University of Edinburgh – UK
6. Danish eInfrastructure Cooperation (DeIC) - DK
7. The National Scientific Computing Unit of the Foundation for Science and Technology (FCT|FCCN) - PT
8. Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) - DE
9. IT4Innovations - CZ
10. Trust-IT Services - IT

Thematic, Integrated Service Providers

1. German Climate Computing Centre (DKRZ) - DE
2. University College of London (UCL) - UK

Thematic, Interoperable Service Providers

1. European Organization for Nuclear Research (CERN) - INT
2. Culham Centre for Fusion Energy (CCFE) – UK
3. Meertens Institute (MEERTENS) - NL
4. CLARIN-ERIC - NL
5. European Center in Research and Advanced Training on Scientific Computing (CERFACS) - FR



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

The EUDAT Service Suite