ETH zürich

IT Services

ETH ORD M2: API interoperability projects

Caterina Barillari
Scientific IT Services (SIS), ETH Zurich

PSI, 03.07.2024

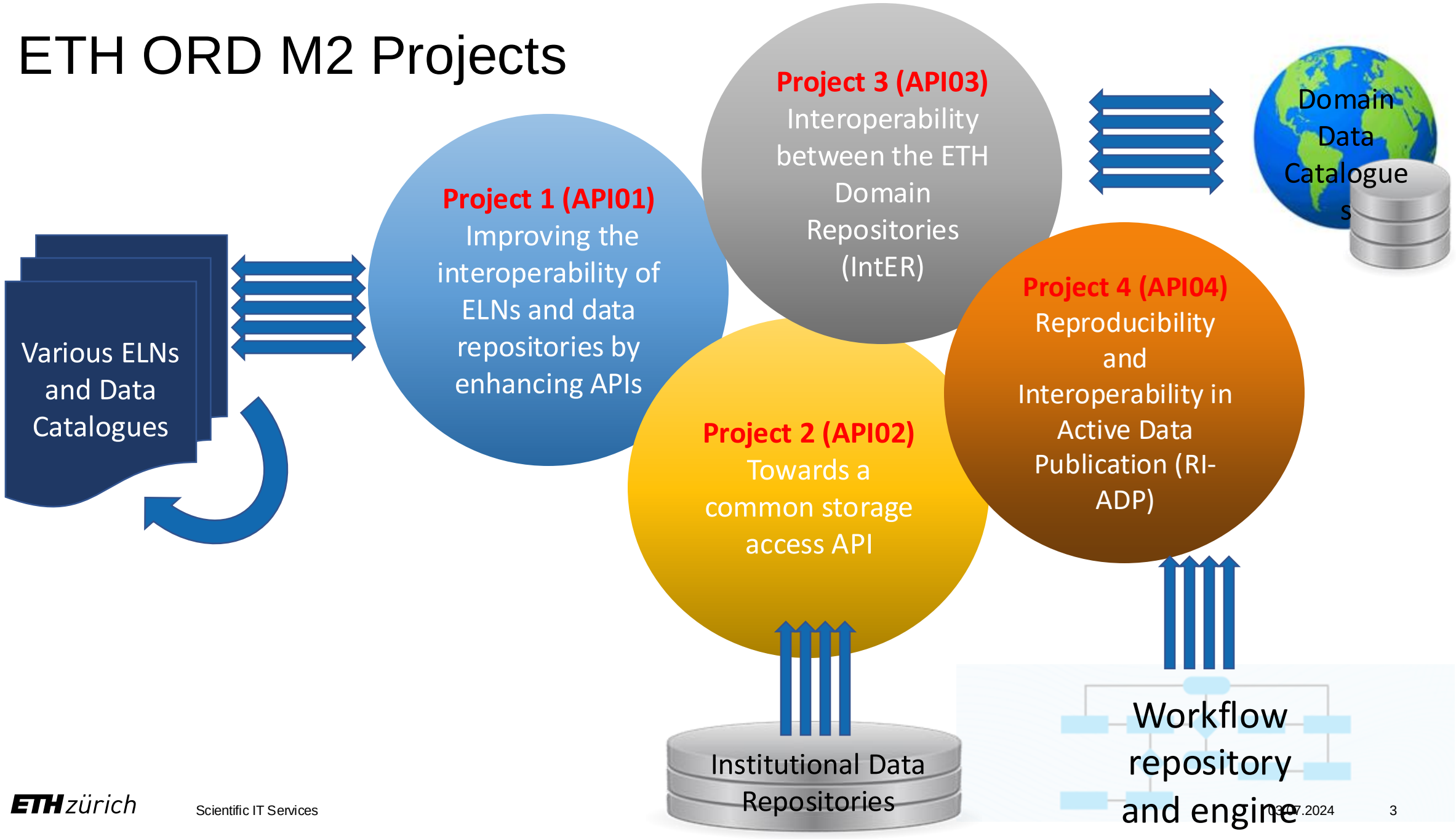
ETH ORD Measure 2 (M2): API interoperability calls

- Several tools in use by research labs in the ETH Domain:
  - Electronic Lab Notebooks (ELN) / Laboratory Information Management Systems (LIMS)
  - Data repositories
  - Data catalogues
  - Storage solutions

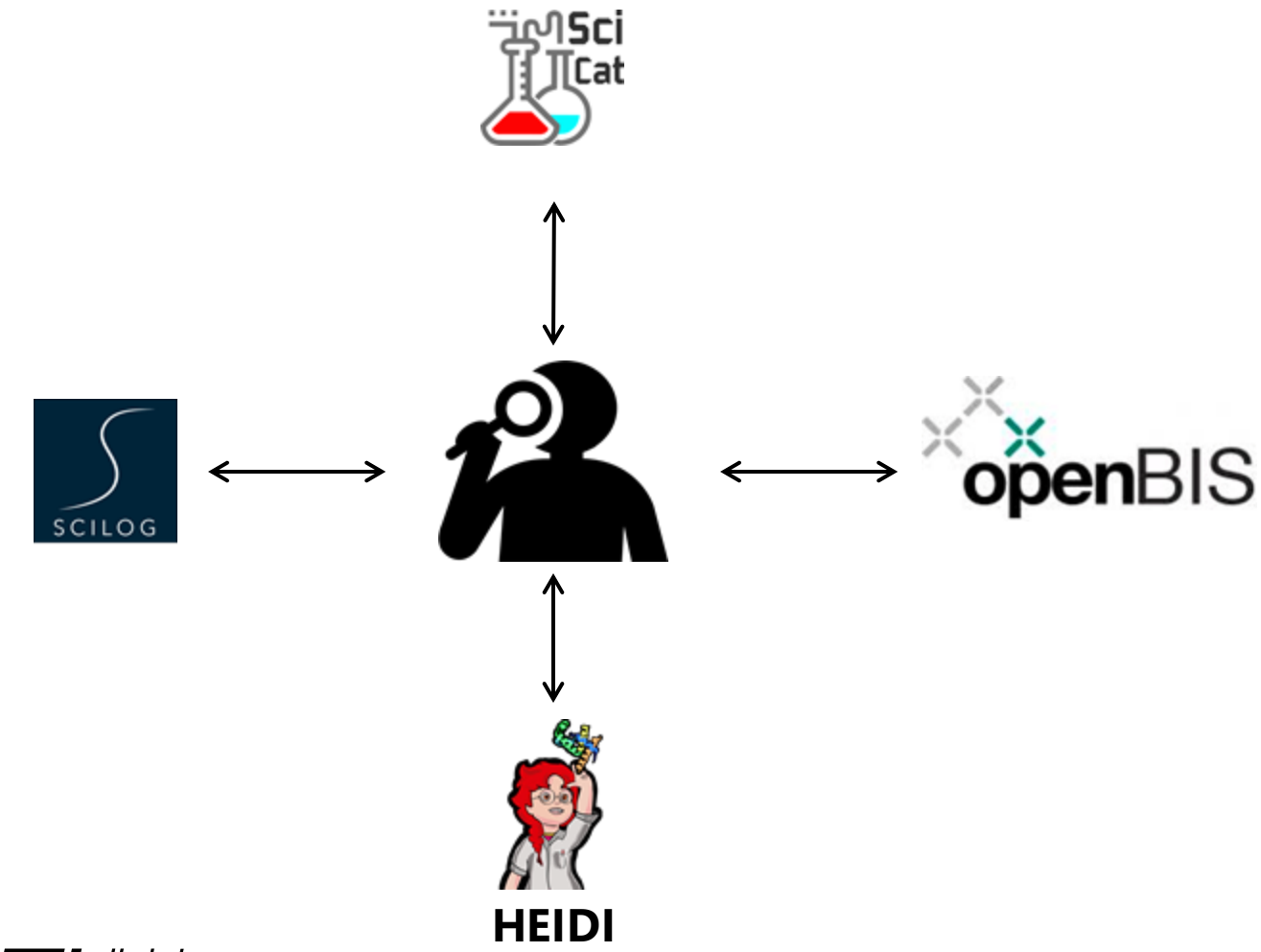
- Scientists often need to use 2 or more of these tools, which are independent from one another → information spread in different places and/or information duplication.

- There is a need for a seamless integration of these systems.

ETH ORD M2 Projects

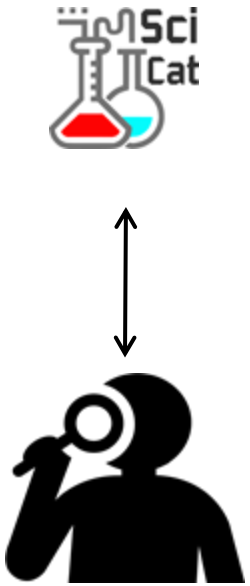


API01: Improving the interoperability of ELNs and data repositories by enhancing APIs



- Different ELNs and 1 data catalogue used in some institutions of the ETH Domain.
- Project lead: PSI.
- Project partners: ETHZ, Empa.
- Project timeline: 2.5 years, starting 07.2024.

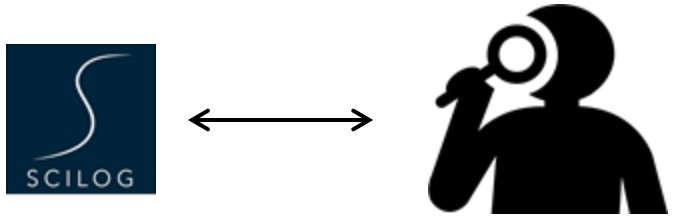
SciCat



- Open-source software originally developed by PSI and the European Spallation Source in collaboration with the Swedish MAXIV synchrotron
- Forms the basis for the Data Catalogue which is the central repository for storing research data at PSI
- Each dataset is uniquely identified with a globally unique persistent identifier, with published datasets assigned a DOI
- Data are tagged with searchable metadata and can be archived in a Petabyte Archive System.

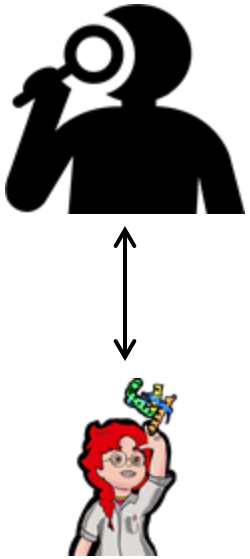
SciLog

- ELN developed and operated at PSI, which organises information captured during an experiment, both by humans and machines, similar to a logbook.

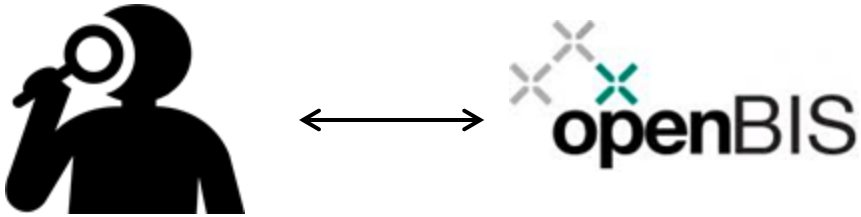


HEIDI

- Open-source ELN developed and operated at PSI for the molecular crystallography (MX) community.
- Collects live info from MX experiments, displays the live acquisition and runs automated pipelines to score the accuracy of the measurement



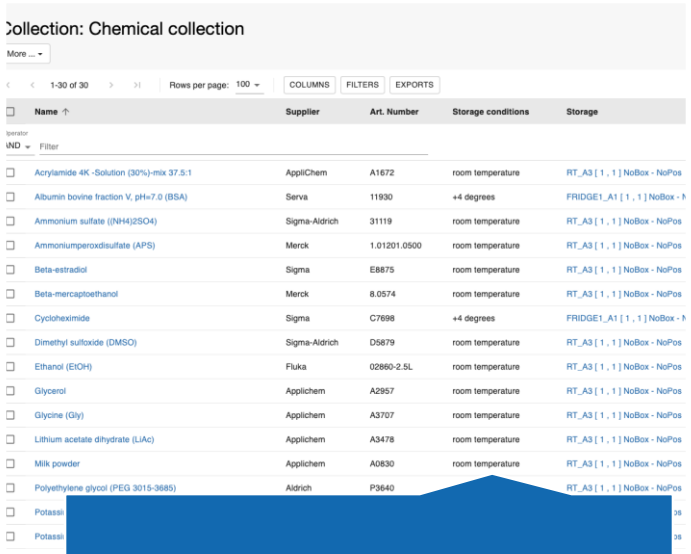
openBIS

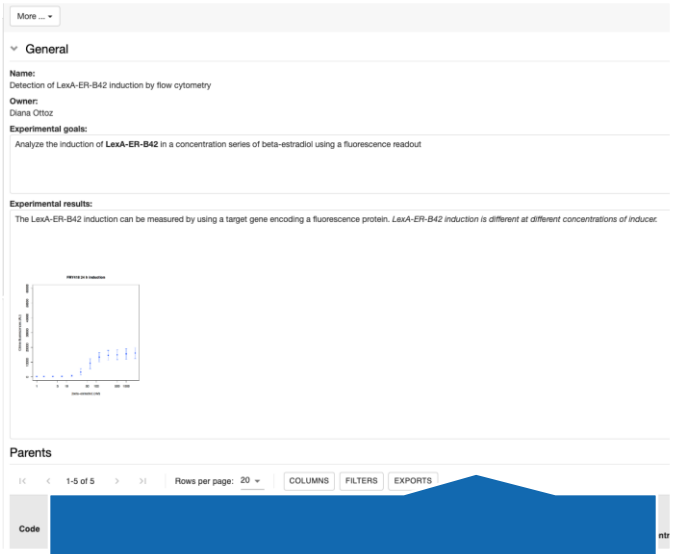


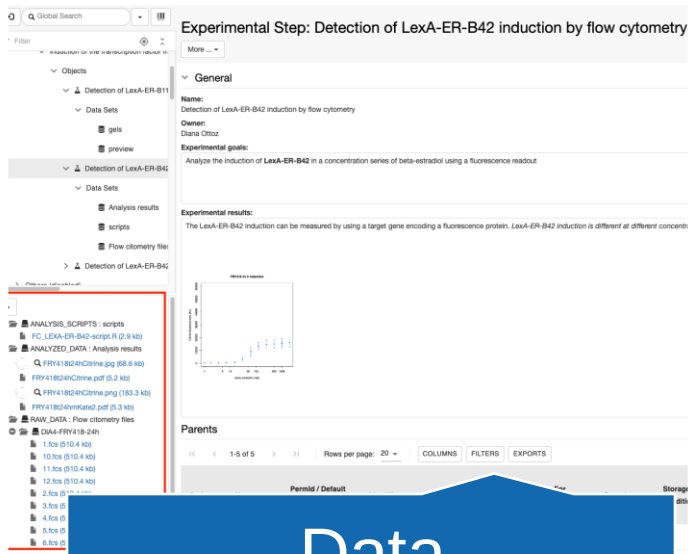
- Open-source ELN, LIMS & RDM solution developed and supported by Scientific IT Services at ETH Zurich.
- Used in many scientific disciplines (life sciences, material sciences, engineering, etc)

openBIS: a complete solution towards FAIR data management

- Developed at ETHZ since 2007



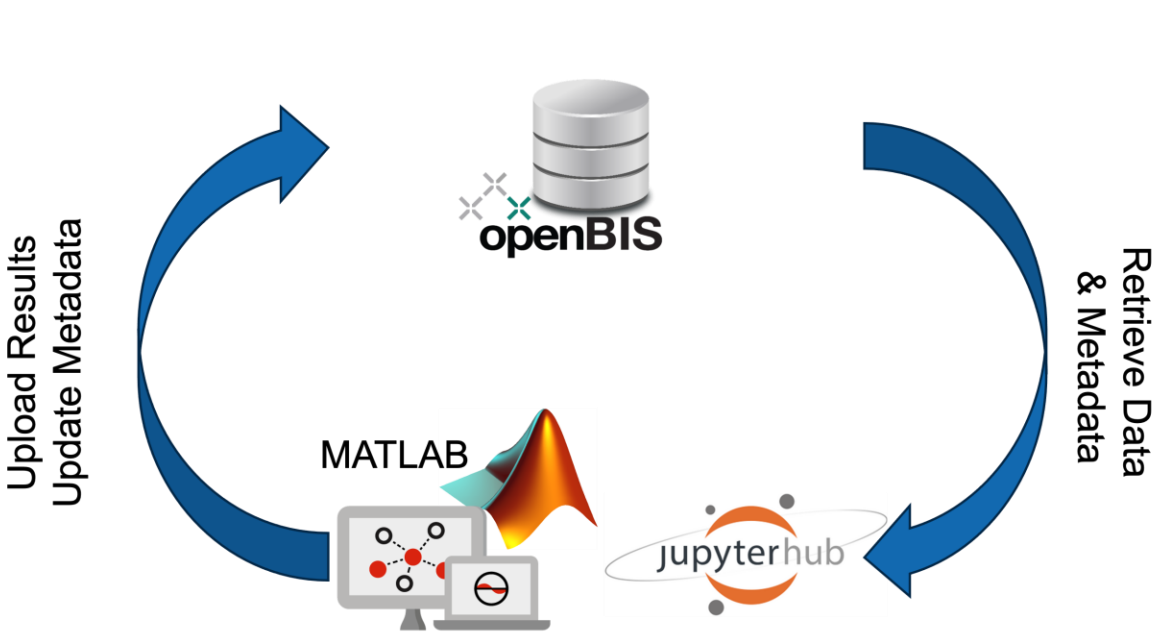




https://openbis.ch

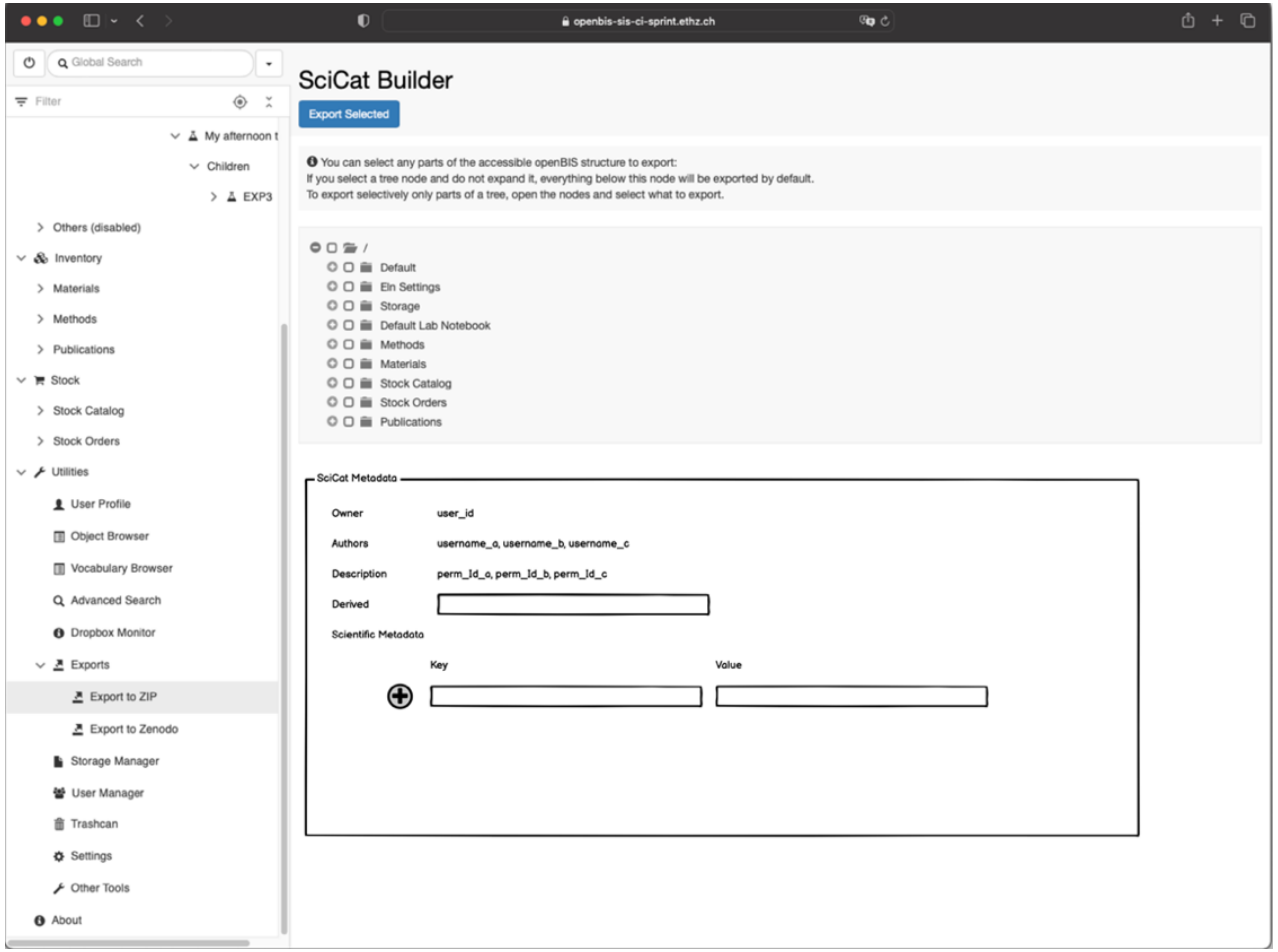
Integration of openBIS with 3rd-party applications

Data analysis

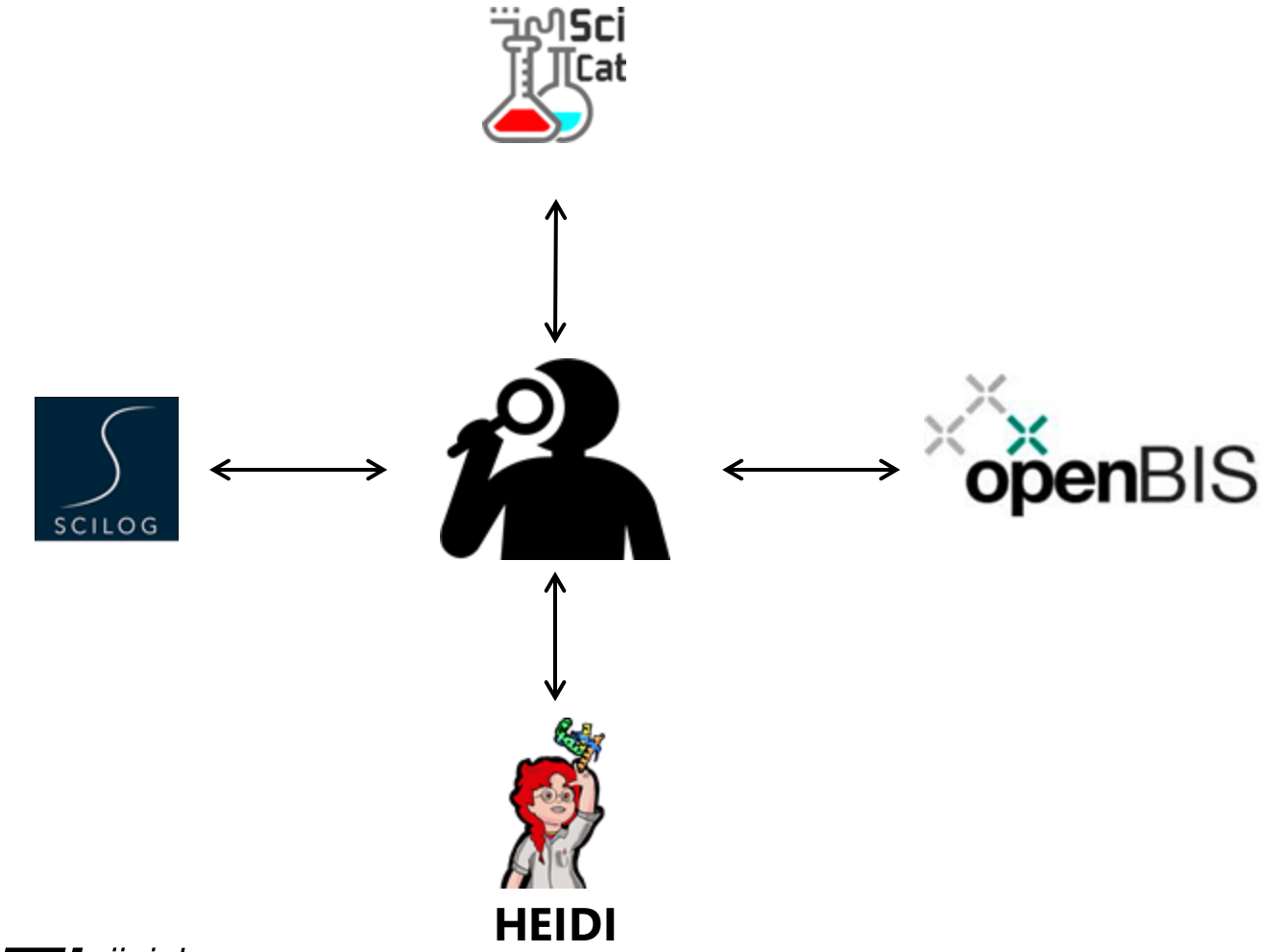


Data publication

Mock-up openBIS-SciCat integration



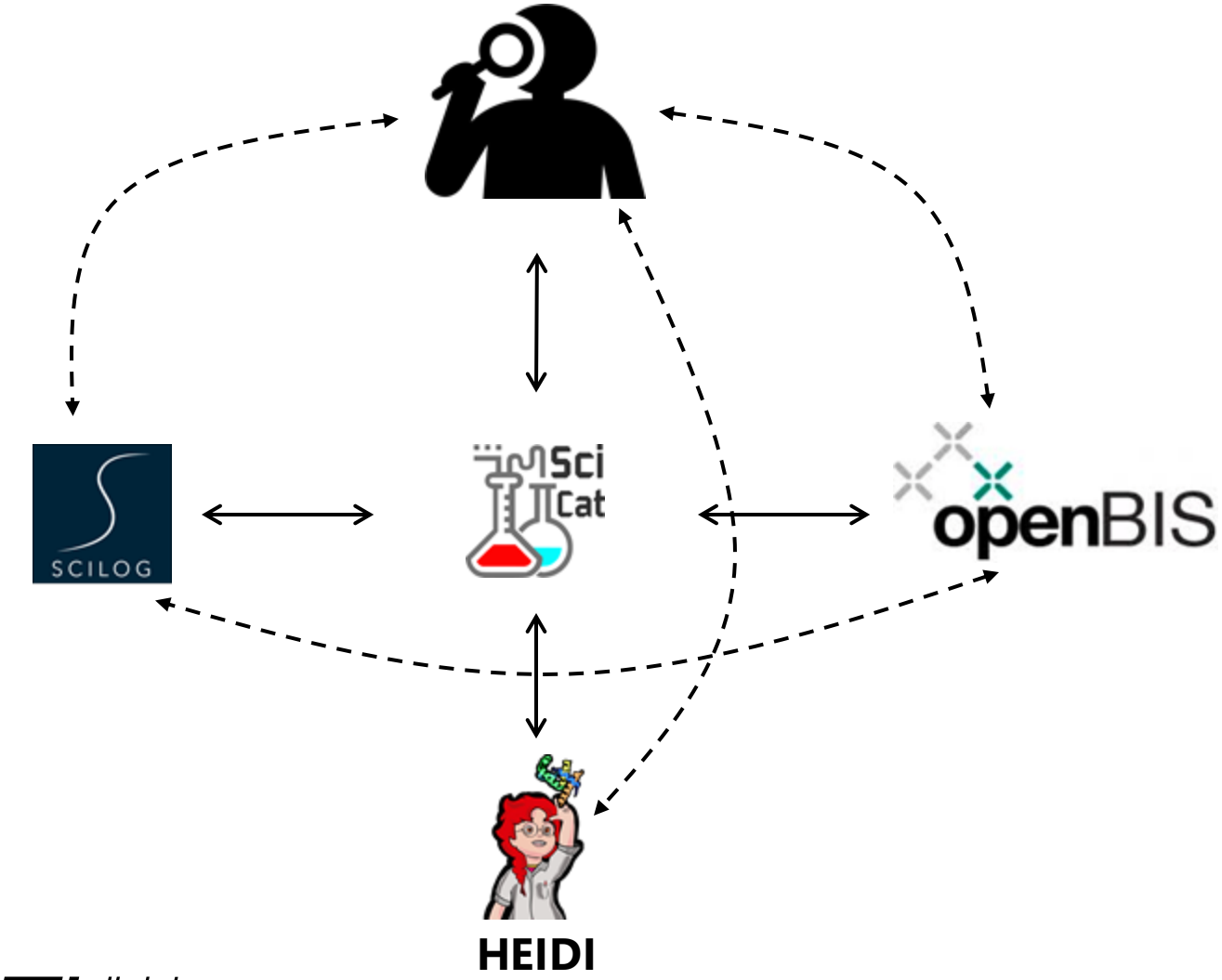
Why do we need interoperability between these systems?



Examples:

- environmental scientists using openBIS (ELN/LIMS) need to manually transfer information/data to SciCat (data catalogue)
- Heidi (MX LIMS) users, need to manually transfer information/data to SciCat and SciLog (ELN)
- SciLog users do not have a way to persist logbooks, which contain valuable metadata, in particular using SciCat

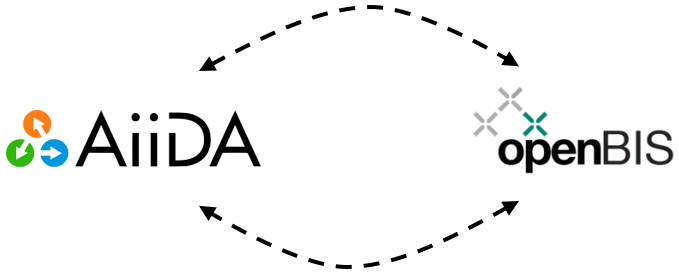
Goals of the API01 project



- Integrate these 4 applications used at PSI, ETHZ, Empa to have data/metadata exchange between platforms.
- Connect from one application to the other via UI.
- This can be done on the API level. This means 1-1 integration between each platform.
- Based on current work from other ETH ORD projects, we will investigate platform-agnostic integration making use of Ontologies and RO-crates.

ETH ORD M1: PREMISE project

- Collaboration between PSI, ETHZ, Empa
- Focus: integration of ELNs and Workflow Management Systems (WFMs)
- Use cases for project: material sciences (batteries production, enzymatic catalysis, microscopy)



- Currently looking into RO-crates to have a platform-agnostic integration of ELNs and WFMs
- Same approach can be extended to integration of other platforms → sync between different ETH ORD projects is crucial to leverage on each other's work

Research Object Crate (RO-crate)

- Community effort to package data and metadata
- An RO-crate is formed by a collection of data (crate) and a ro-crate-metadata.json file that describes the data
- ro-crate-metadata.json is a JSON-LD file
- JSON-LD is a lightweight Linked Data format.

```
{
"@context": "https://json-ld.org/contexts/person.jsonld",
"@id": "http://dbpedia.org/resource/John_Lennon",
"name": "John Lennon", "born": "1940-10-09",
"spouse": "http://dbpedia.org/resource/Cynthia_Lennon"
}
```

Current limitations and open points for further investigations

- Each platform has their own metadata schema. Use of standards and ontologies we can help in the translation of terms.
- JSON-LD does not contain the metadata schema, which is necessary for interoperability between platforms.
- A separate JSON schema file would have to be used.
- Different file formats (e.g. YAML) can be investigated.
- These questions will be investigated in API01 in collaboration with the other M2 API projects, with PREMISE and potentially other projects looking at the topic of interoperability.