

**IT Services** 

### ETH ORD M2: API interoperability projects

Caterina Barillari Scientific IT Services (SIS), ETH Zurich

PSI, 03.07.2024

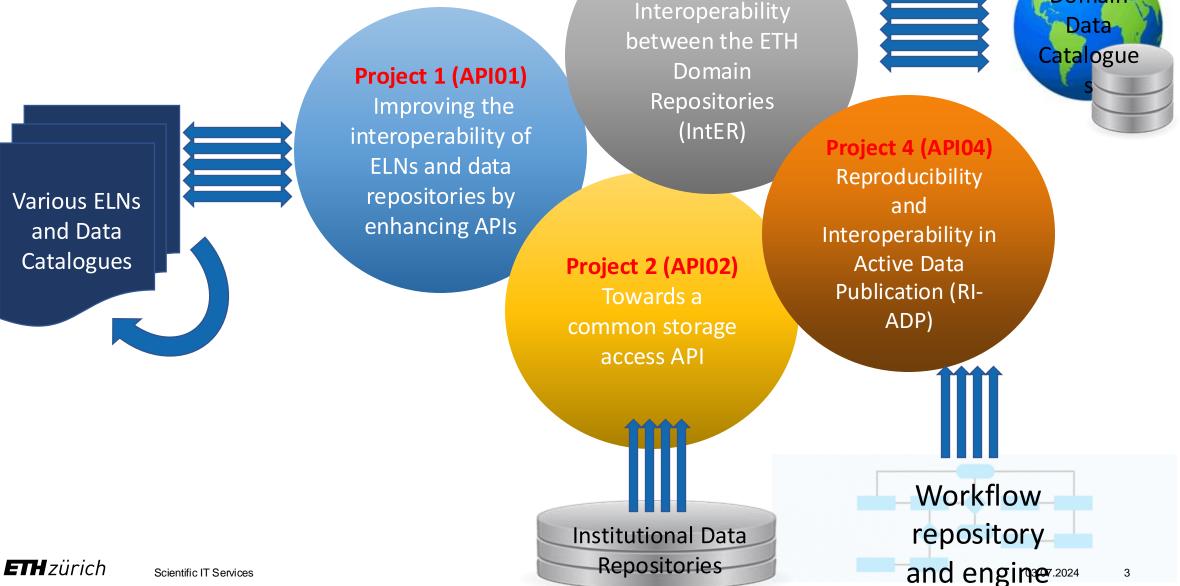


### ETH ORD Measure 2 (M2): API interoperability calls

- Several tools in use by research labs in the ETH Domain:
  - Electronic Lab Notebooks (ELN) / Laboratory Information Management Systems (LIMS)
  - $\circ$  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



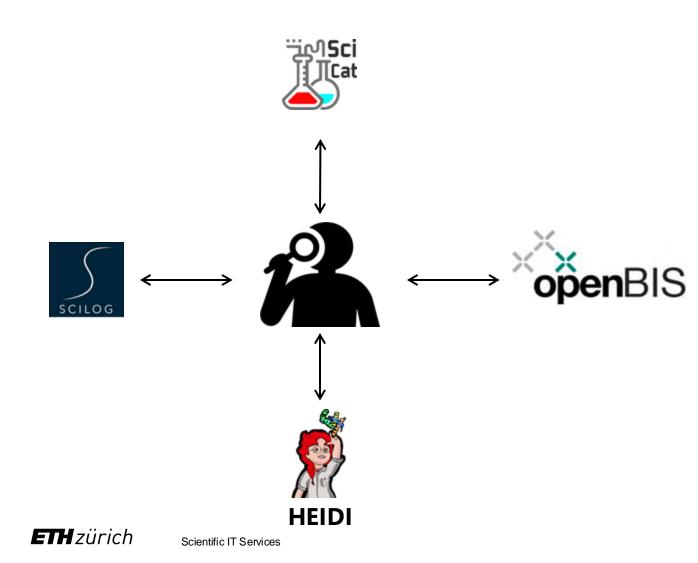
**Project 3 (API03)** 

Scientific IT Services

3

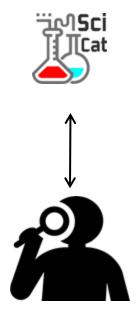
Domain

# 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

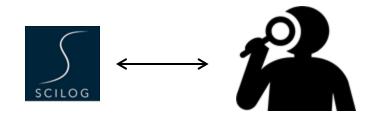


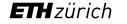
- Open-source software originally developed by PSI and the European Spallation Source in collaboration with the Swedish MAXIV synchrotron
- Forms the basis for the Data Catalogue which is the central repository for storing research data at PSI
- Each dataset is uniquely identified with a globally unique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

**ETH** zürich

### SciLog

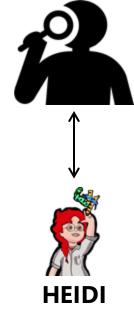
• ELN developed and operated at PSI, which organises information captured during an experiment, both by humans and machines, similar to a logbook.





### HEIDI

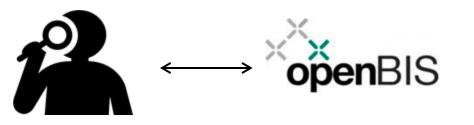
- Open-source ELN developed and operated at PSI for the molecular crystallography (MX) community.
- Collects live info from MX experiments, displays the live acquisition and runs automated pipelines to score the accuracy of the measurement

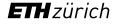




### openBIS

- 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

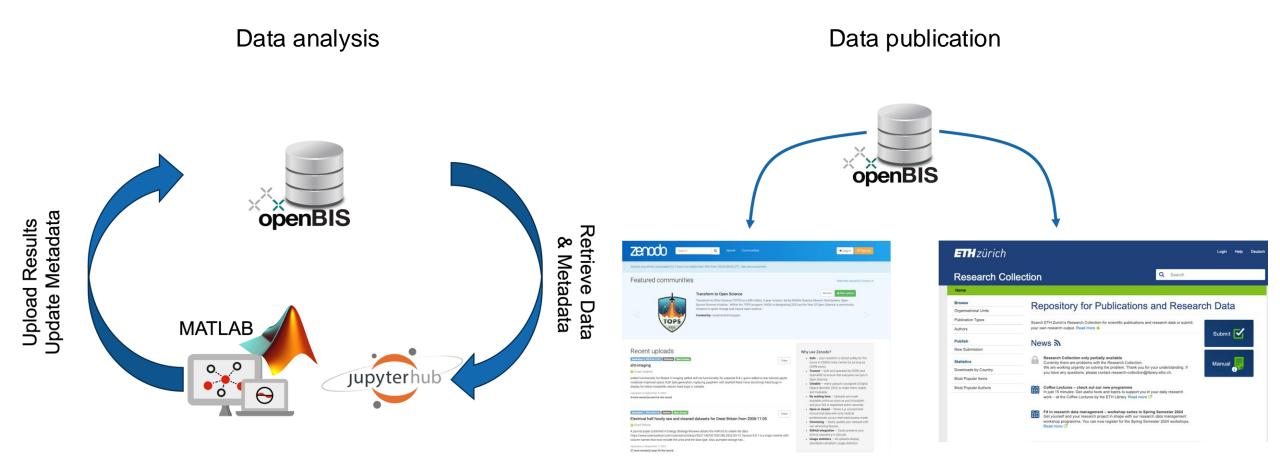
#### Developed at ETHZ since 2007

Collection: Chemical collection					More ✓ General Name:	Q Global Search     W      Filter     Wesselson or wre sensorphone research     Voljects	
C C 1-30 of 30 > > Rows per page: 100 -	COLUMNS	LTERS EXPORTS			Detection of LexA-ER-B42 induction by flow cytometry	✓ ▲ Detection of LexA-ER-B11	✓ General
□ Name ↑	Supplier	Art. Number	Storage conditions	Storage	Owner: Díana Ottoz	✓ Data Sets	Name: Detection of LexA-ER-B42 induction by flow cytometry
lperator					Experimental goals:	🛢 gels	Owner: Diana Ottoz
ND - Filter					Analyze the induction of LexA-ER-B42 in a concentration series of beta-estradiol using a fluorescence readout	E proview	Experimental goals: Analyze the induction of LexA-ER-B42 in a concentration series of beta-estradiol using a fluorescence readout
Acrylamide 4K -Solution (30%)-mix 37.5:1	AppliChem	A1672	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos		✓ ▲ Detection of LexA-ER-B4;	Pring/de une induction for searching and a concentration denies of desired addressing a moneous centrebularity
Albumin bovine fraction V, pH=7.0 (BSA)	Serva	11930	+4 degrees	FRIDGE1_A1 [ 1 , 1 ] NoBox - N	Experimental results:	✓ Data Sets	
Ammonium sulfate ((NH4)2SO4)	Sigma-Aldrich	31119	room temperature	RT_A3[1,1]NoBox - NoPos	Experimental results: The Lax-RF-R42 induction can be measured by using a target gene encoding a fluorescence protein. Lax-A-ER-B42 induction is different at different concentrations of induce:	Analysis results	Experimental results:
Ammoniumperoxdisulfate (APS)	Merck	1.01201.0500	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos		E scripts Flow citometry file:	The LexA-ER-B42 induction can be measured by using a target gene encoding a fluorescence protein. LexA-ER-B42 induction is different at different concentric
						Flow citometry ties > A Detection of LexA-ER-B42	
Beta-estradiol	Sigma	E8875	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos	MITTER 1 Madeine	Philase (disabled)	Micro Di Landone
Beta-mercaptoethanol	Merck	8.0574	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos			
Cycloheximide	Sigma	C7698	+4 degrees	FRIDGE1_A1 [ 1 , 1 ] NoBox - N		🖀 💂 ANALYSIS_SCRIPTS : scripts	
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D5879	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos		FC_LEXA-ER-B42-script.R (2.9 kb)	
Ethanol (EtOH)	Fluka	02860-2.5L	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos		C FRY418t24hCitrine.jpg (68.6 kb)	
Glycerol	Applichem	A2957	room temperature	RT_A3[1,1]NoBox - NoPos		FRY418t24hCitrine.pdf (5.2 kb)	Stantin M
Glycine (Gly)	Applichem	A3707	room temperature	RT_A3 [ 1 , 1 ] NoBox - NoPos		FRY418t24hmKate2.pdf (5.3 kb)	
						을 룬 RAW_DATA : Flow citometry files 이 술 를 룬 DIA4-FRY418-24h	Parents
Lithium acetate dihydrate (LIAc)	Applichem	A3478	room temperature	RT_A3[1,1]NoBox - NoPos	Parents	<ul> <li>1.fcs (510.4 kb)</li> <li>10.fcs (510.4 kb)</li> </ul>	COLUMNS FILTERS EXPORTS
Milk powder	Applichem	A0830	room temperature	RT_A3[1,1]NoBox - NoPos		11.fcs (510.4 kb)	
Polyethylene glycol (PEG 3015-3685)	Aldrich	P3640		RT_A3[1,1]NoBox - NoPos	IC COLUMNS FILTERS EXPORTS	2.fcs (510.4 kb)	Permid / Default Storage
Potassi				08		3.fcs (5 4.fcs (5	dib National Anti-Anti-Anti-Anti-Anti-Anti-Anti-Anti-
Potassk				08	Code	5.fcs (5)	
Inventory					ELN	ELN Data manageme	

#### https://openbis.ch

### Integration of openBIS with 3rd-party applications



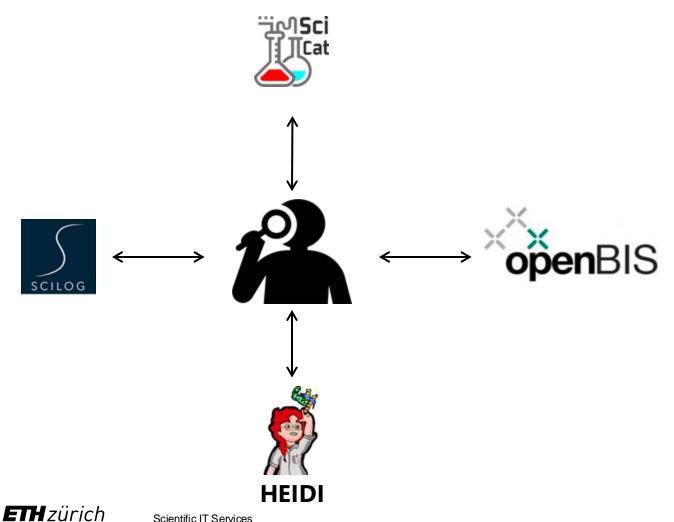


### Mock-up openBIS-SciCat integration

•••	0	🔒 openbis-sis-ci-sprint.ethz.c	h ®ago	ů + C
♥ Q Global Search	SciCat Builde			
✓ ▲ My afternoon t	Export Selected			
✓ Children		of the accessible openBIS structure to export:		
> 👗 EXP3		d do not expand it, everything below this node will be arts of a tree, open the nodes and select what to exp		
> Others (disabled)				
✓ Sinventory	○ □ 壹 / ○ □ 童 Default			
> Materials	C C E En Settings			
	O 🗆 💼 Storage			
> Methods	C 🗋 Default Lab	Notebook		
> Publications	C C m Methods			
✓ 〒 Stock	C     Materials     C     C     Materials     Stock Catal			
	C C Stock Order			
> Stock Catalog	C D m Publications			
> Stock Orders				
✓ ≁ Utilities	- SciCat Metodata			
1 User Profile	Owner u	iser_id		
Object Browser	Authors			
Vocabulary Browser	Description p	erm_ld_a, perm_ld_b, perm_ld_c		
Q Advanced Search	Derived			
O Dropbox Monitor	Scientific Metadata			
✓ ≛ Exports	Ke	у	Value	
Export to ZIP	🕀 🖸			ן נ
Export to Zenodo				
Storage Manager				
🖀 User Manager				
Trashcan				
Settings				
✗ Other Tools				
About				

**ETH** zürich

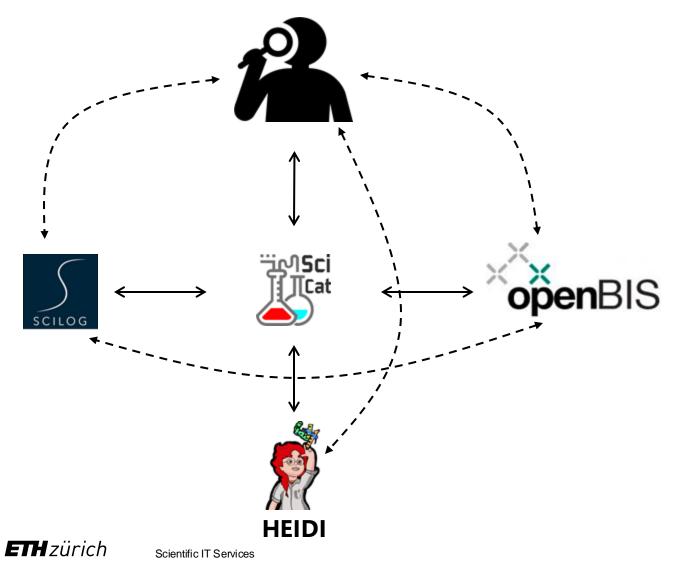
### Why do we need interoperability between these systems?



#### 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

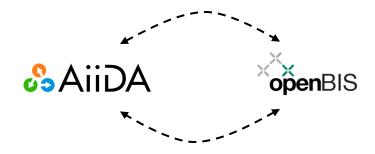


- Integrate these 4 applications used at PSI, ETHZ, Empa to have data/metadata exchange between platforms.
- Connect from one application to the other via UI.
- This can be done on the API level. This means 1-1 integration between each platform.
- Based on current work from other ETH ORD projects, we 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": "<u>https://json-ld.org/contexts/person.jsonld</u>",
"@id": "<u>http://dbpedia.org/resource/John\_Lennon</u>",
"name": "John Lennon", "born": "1940-10-09",
"spouse": "<u>http://dbpedia.org/resource/Cynthia\_Lennon</u>"
}



### 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.