PAUL SCHERRER INSTITUT

PSI

Giovanni Pizzi :: Group Leader "Materials Software and Data" :: Paul Scherrer Institute

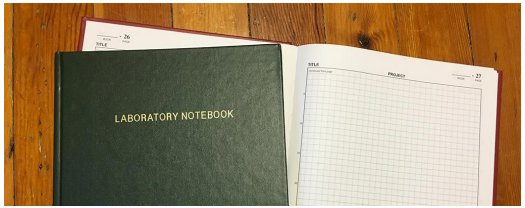
ORD-R Establish "PREMISE" project: "Open and Reproducible Materials Science Research"

ORD meeting, PSI, 4 May 2023

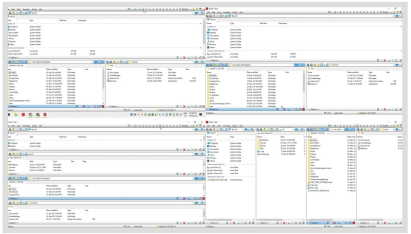
State of the art and challenges

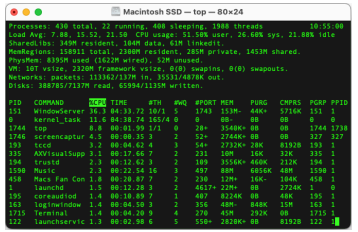
Typical scenario in Materials Science

Experiments



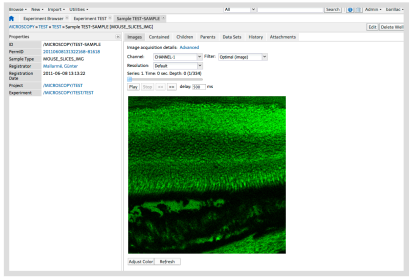
Simulations

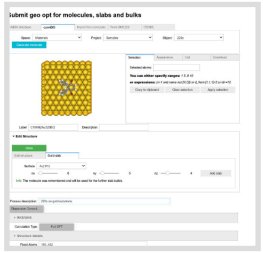


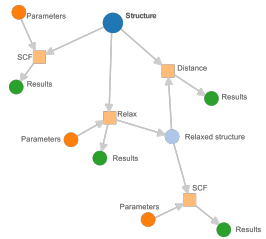


Better scenario

ELN/LIMS







WF manager + GUI

But how to interchange data between experiments and simulations using common metadata formats?

PREMISE goals

- Establish, promote and facilitate the **adoption of FAIR ORD practices** in Materials Science.
- Provide missing critical components to **enable open and reproducible research** (accessible, shareable)
- Address **interoperability between data from simulations and experiments** (currently: no established RDM practices)
- Key enabler of **emerging AI/ML-driven autonomous laboratories**, with native support for RDM and ORD practices

Partners and technology

- **+ @PSI: Dr. Edan Bainglass post-doc,** from Apr 2023

- Partners
  - **PSI:** Pizzi (lead)
  - **ETHZ:** Rinn, Barillari, Lütcke
  - **Empa:** Pignedoli (microscopy); Battaglia (robotic battery experiments)
- **Start date: 1 April 2023**

- Exploit existing ORD platforms in the ETH domain

openBIS

ELN + LIMS (ETHZ)

Reproducible experiments

AiiDA AiiDAlab

WF Manager + GUI (PSI, and also Empa/EPFL)

Reproducible simulations

Both focused on tracking the whole provenance and ensuring reproducibility!

PAUL SCHERRER INSTITUT

PSI

AiiDA

COMPUTATIONAL SCIENCE INFRASTRUCTURE FOR HIGH THROUGHPUT WORKFLOWS WITH FULL DATA PROVENANCE

Language: implemented and API in python

License: MIT open source **http://www.aiida.net/**

Source: https://github.com/aiidateam/aiida-core

Scalable workflow engine: *robustness*

Built-in support for HPC: *performance*

Automated full data provenance: *reproducibility*

Flexible plugin system: *interoperability*

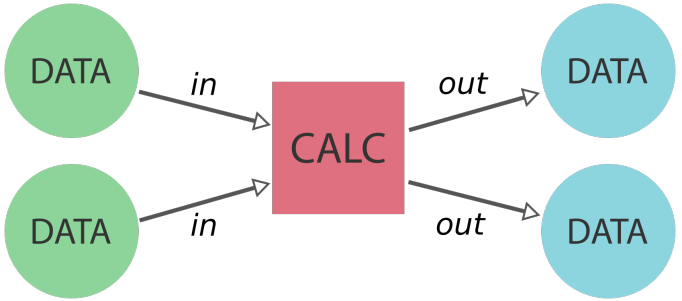
G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)

S.P. Huber et al., Scientific Data 7, 300 (2020)

Data provenance

Simple recipe

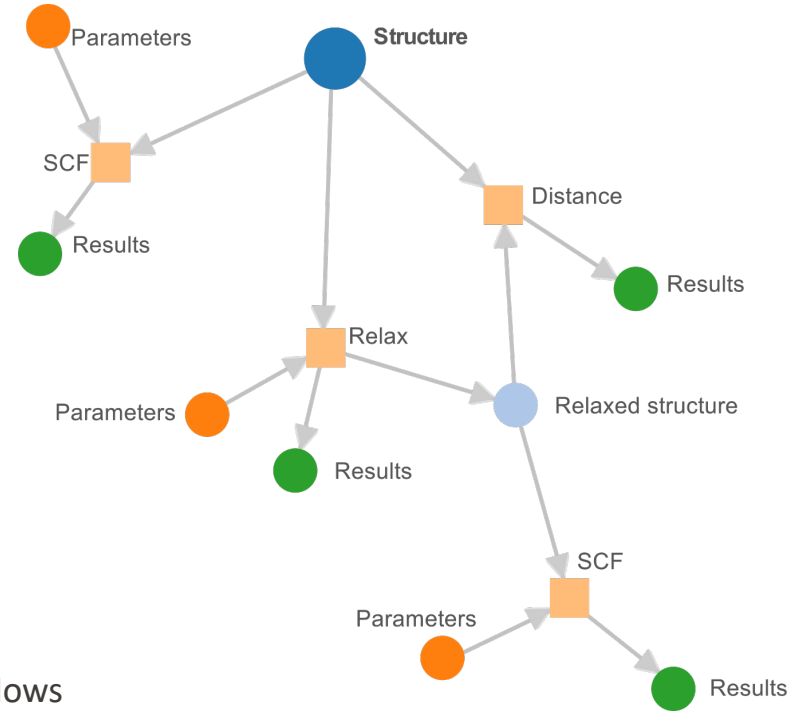
- Store data transformations or '**calculations**'
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**



Data provenance

Simple recipe

- Store data transformations or '**calculations**'
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**



Provenance graphs

- When data gets reused, a directed graph is created
- That quickly grow in complexity even for "simple" workflows

Data provenance

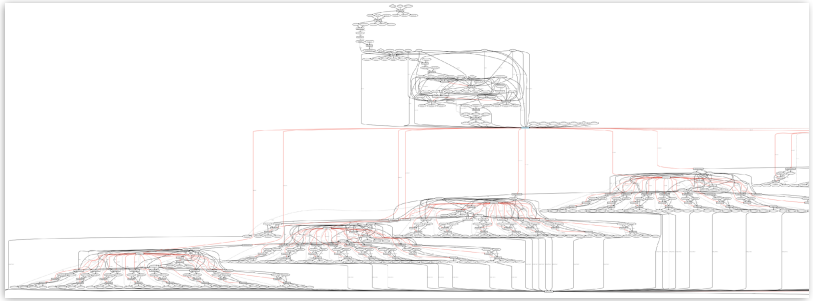
Simple recipe

- Store data transformations or '**calculations**'
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**

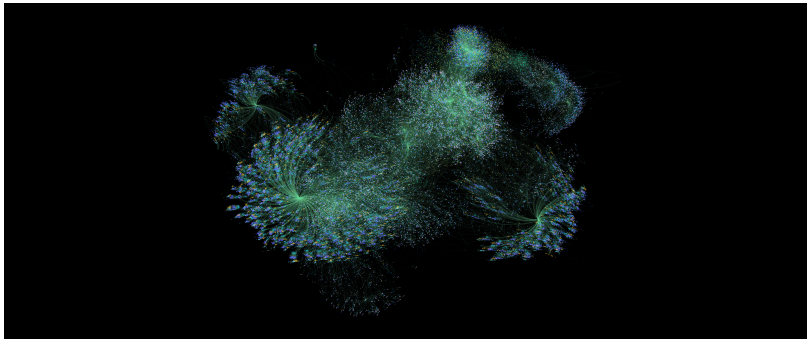
Graph requirements

- Needs to be automated
- Needs to be stored *as data is created*

Complexity grows quickly even for simple workflows and is impossible to reconstruct *a posteriori*



Molecular dynamics study of Lithium in a solid electrolyte



Graphical representation of actual AiiDA database

Data provenance

Simple recipe

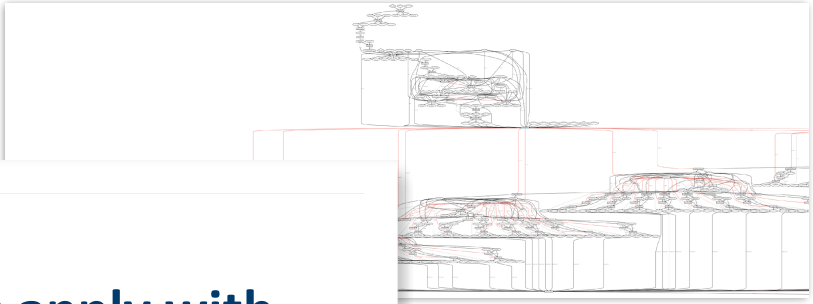
- Store data transforma
- Store its **inputs** and t
- Store its **outputs** and
- Most **crucially** store t

Similar concepts apply with openBIS (for experiments)

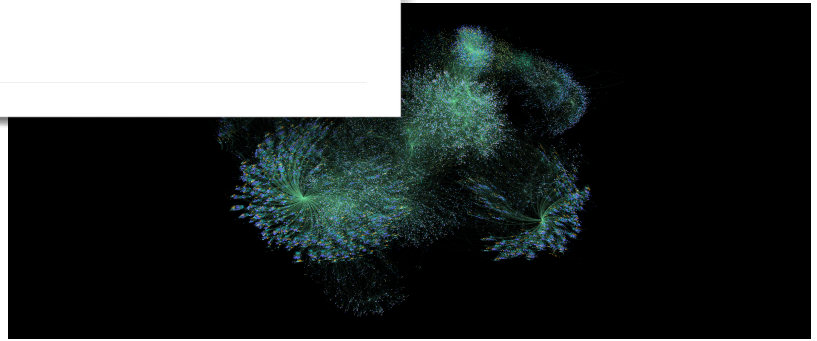
Graph requireme

- Needs to be automated
- Needs to be stored *as data is created*

Complexity grows quickly even for simple workflows and is impossible to reconstruct *a posteriori*

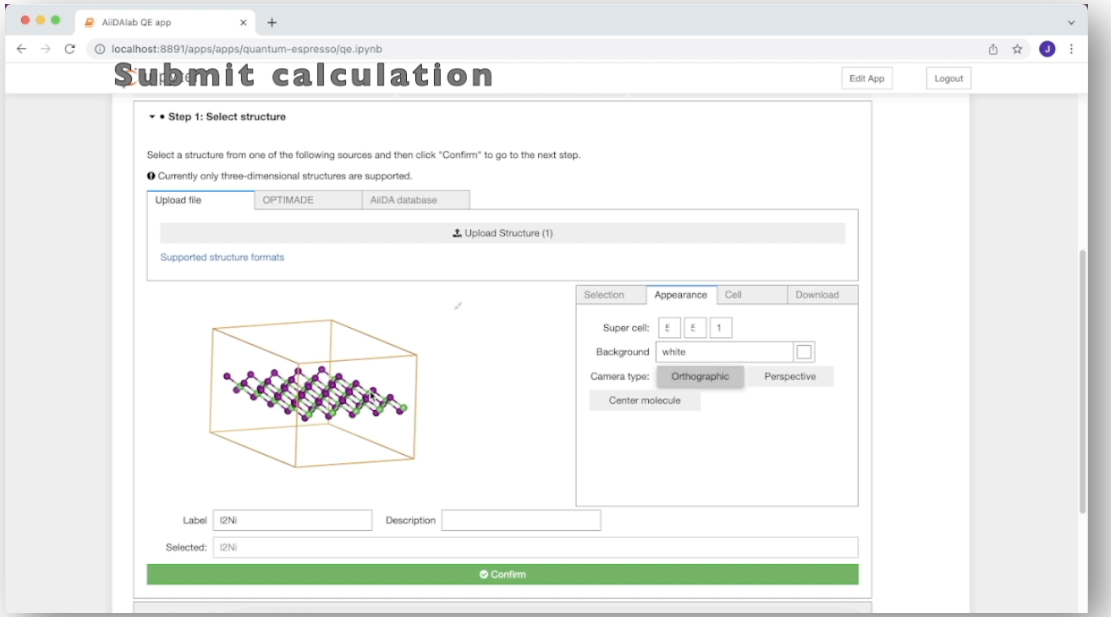


of Lithium in a solid electrolyte



Graphical representation of actual AiiDA database

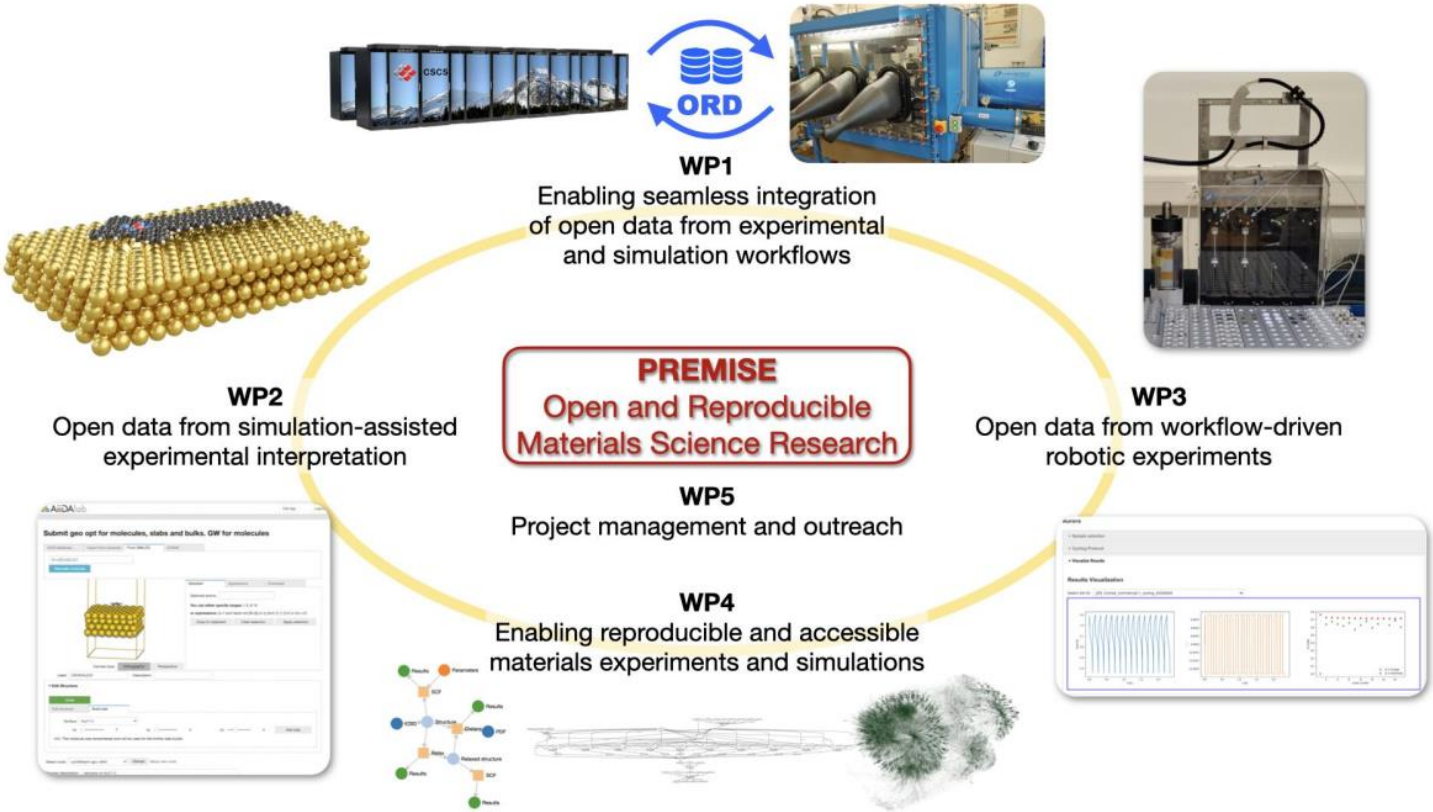
Access to advanced simulations for everybody



https://www.materialscloud.org/work/aiidalab

- GUI exposing workflows also to non experts
- Example: **AiiDAlab app** (GUI) to run Quantum ESPRESSO simulations (DFT simulations: relaxation, bands, density of states, ...)
- We are working to make more apps to predict experiments run at PSI (e.g. XPS, XAS, ...)

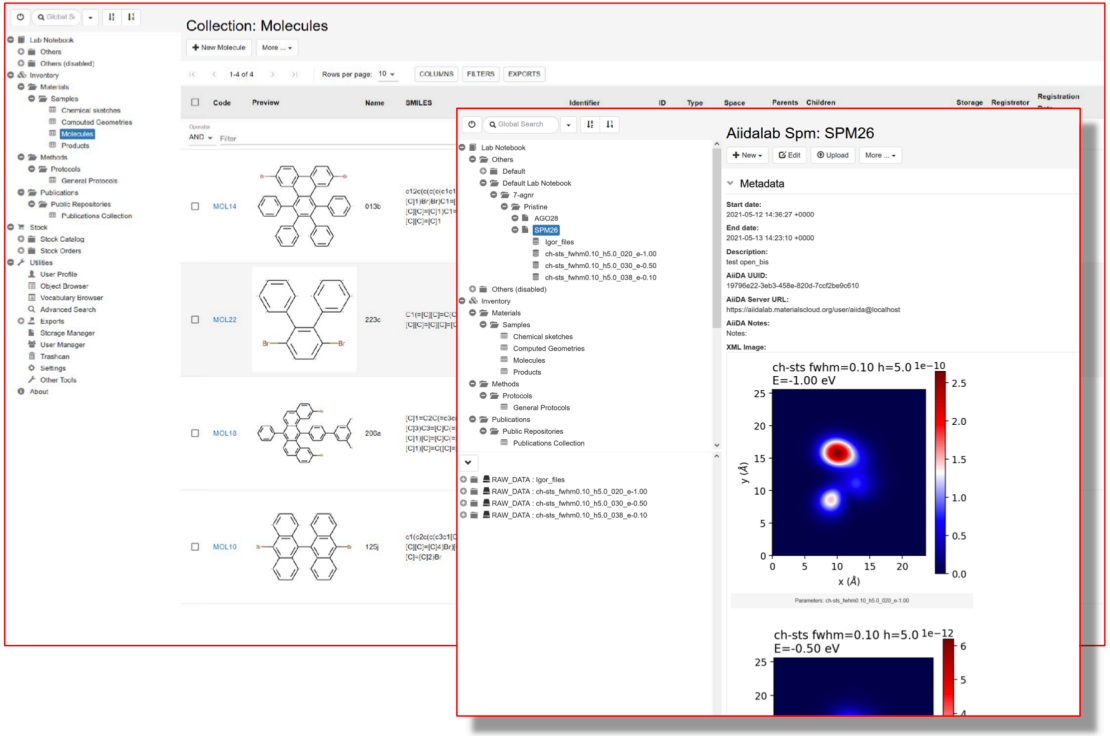
PREMISE structure: structure and workpackages



Usecase 1: Combining the two worlds

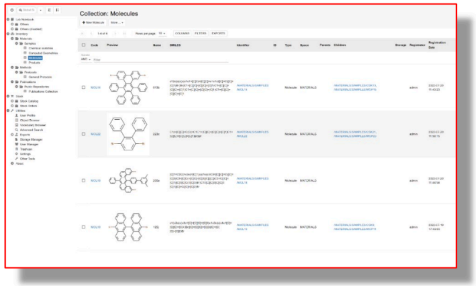
(in collaboration with Pignedoli @ Empa)

- *Task:* identify the adsorption configuation of a molecule on a surface.
- Often this is done by comparing experimental and simulated STM images.
- To enable comparison we need to ensure a seemless data transfer from an ELN to AiiDAlab and back.

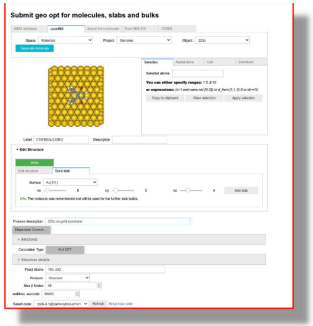


Usecase 1: Combining the two worlds

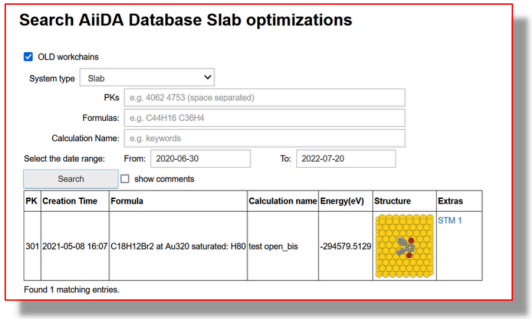
(in collaboration with Pignedoli @ Empa)



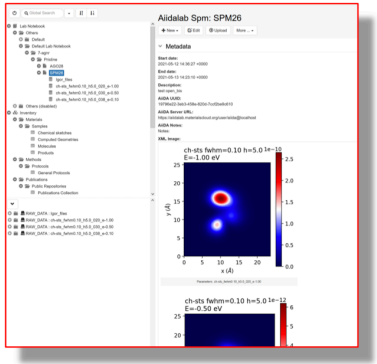
openBIS ELN: list of molecules



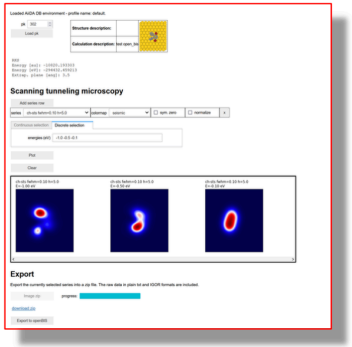
AiiDAlab: imported molecule



AiiDAlab: list of completed simulations



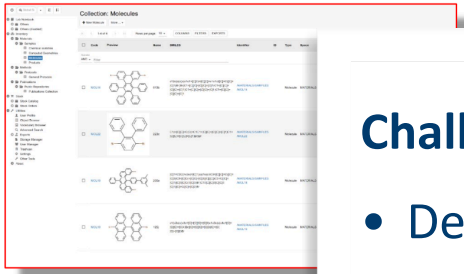
openBIS ELN: imported simulated STM image



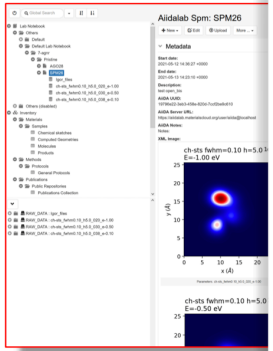
AiiDAlab: simulated STM image

Usecase 1: Combining the two worlds

(in collaboration with Pignedoli @ Empa)



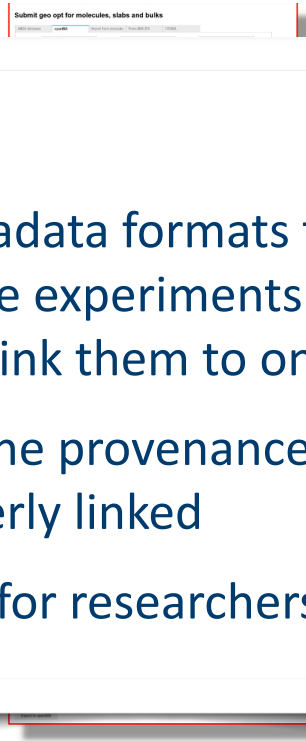
openBIS ELN: list of mole



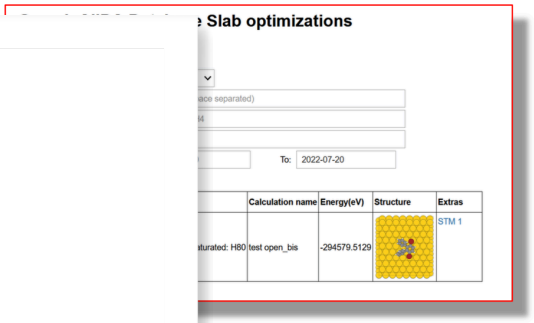
openBIS ELN: imported simulated STM image

Challenges

- Develop metadata formats that can accommodate experiments and simulations, link them to ontologies
- Ensure that the provenance on the two sides is properly linked
- Make it easy for researchers



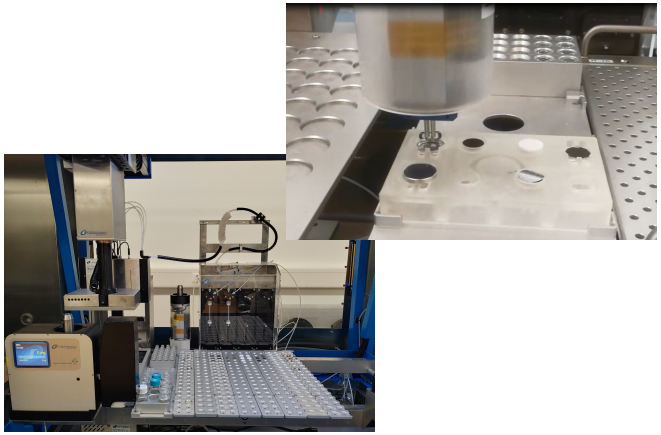
AiiDAlab: simulated STM image

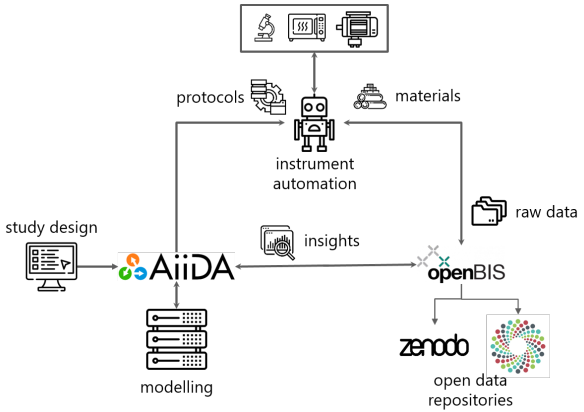


completed simulations

Usecase 2: Robotic experiments toward autonomous labs

- Empa (Prof. Corsin Battaglia): **robotic setups** to
  - assemble coin cells (batteries) changing formulations (32+ channels)
  - perform cyclic testing (128 channels, soon 256)
- Collaboration started within the Battery2030+ BIG-MAP project, now continuing in PREMISE
- **Goal**:
  - Both robotic experiments and simulations driven by our workflow engine *AiiDA*
  - Experimental+simulation data stored in the same ELN: *OpenBis (ETHZ)*
- Designed from the ground to be an open and reproducible materials-science research platform enabling future autonomous labs



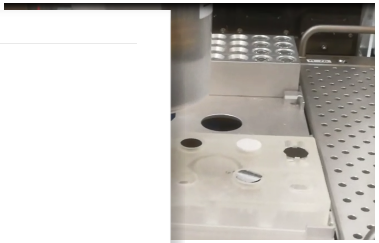


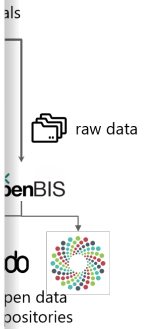
Usecase 2: Robotic experiments toward autonomous labs

- Empa
  - ass
    for
  - pe
- Colla
  MAP
- **Goal**
  - Bo
    by
  - Exp
    ELI
- Desig
  repro
  enab

Challenges

- Automate experiments
- Take care of security requirements at the institutions to access hardware
- Create appropriate digital twins:
  - Simulations can be run many times on same input, but experiments change the history of a sample
  - Many-to-many relation between simulation inputs (e.g. gold crystal structure) and samples (e.g. gold samples)





However: focus on generality

- Build on strength and maturity of AiiDA and openBIS
  - *both developed in ETH domain*
  - *both focused on ORD+reproducibility*
- Combine them and use the two scientific use cases @Empa to demonstrate and stress-test the concepts
- **However, generality of the project**: ensure extensibility to other software and other research projects

PREMISE structure: summary