

## *ELI Data Standards and Data Strategy*

**Teodor Ivănoaica**

Sr Coordinator for Scientific Computing and Data Management



# AGENDA

- Data @ELI – Definitions, roles and (support) policies
- ELI User Journey – The data perspective
- SDMS - Design Principles and scientific community standards
- ELI ERIC Scientific Data Management System Concept
- Next Steps

**ARTICLE 13 ELI STATUTE:**

13(1) ‘Data’ refers to all information collected by USERS and the staff while performing scientific experiments under the ACCESS FOR USERS Policy and performing operations of the ELI FACILITIES.

13(2) Open Access to FAIR data sets and metadata stored in Open Access repositories shall be favoured for data collected as a result of the use of the ELI FACILITIES and, to the extent possible in case of software and computer programmes created by the ELI ERIC and the ELI FACILITIES; open source principles shall be considered.

**ELI Data Policy:**

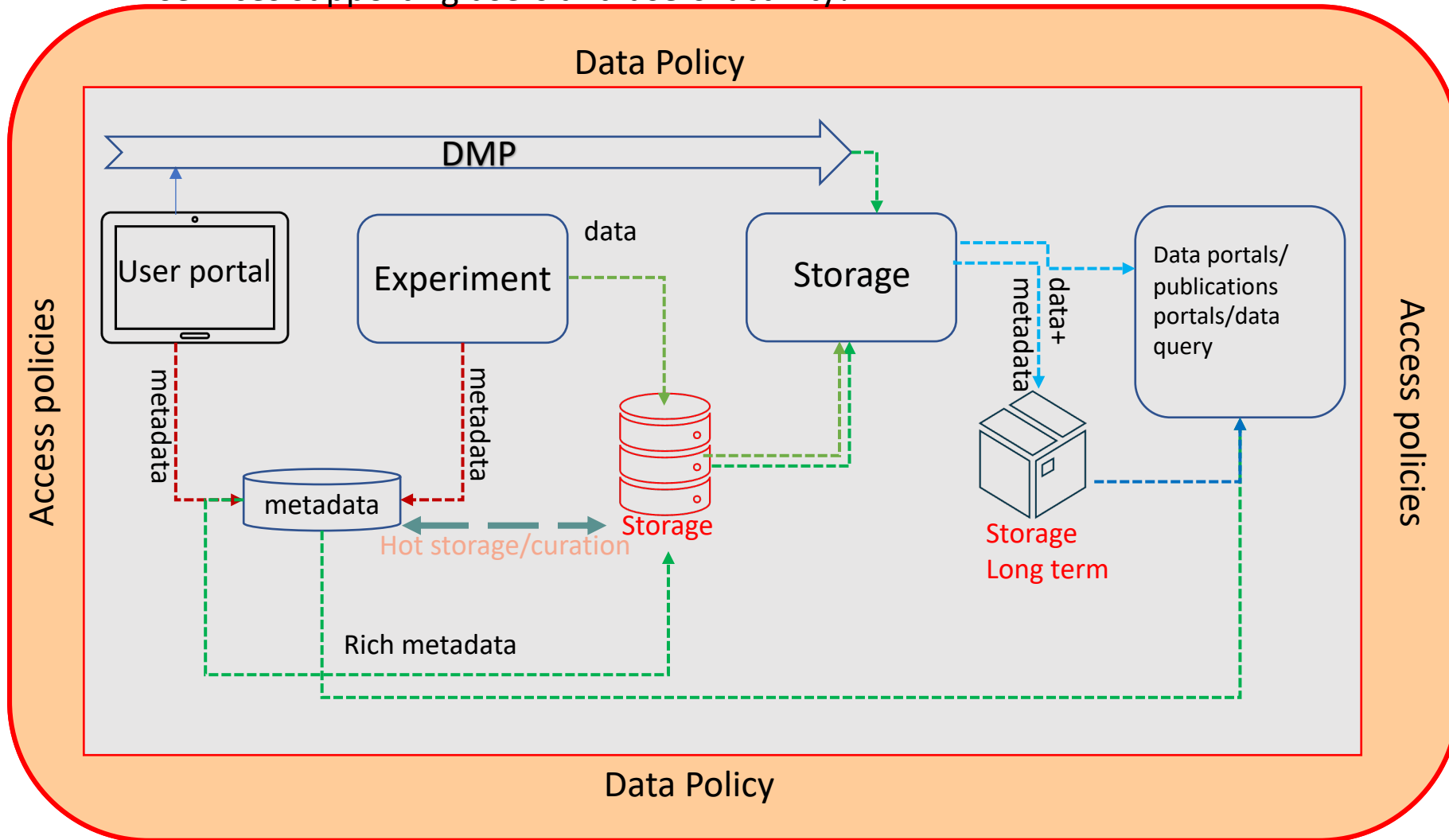
*“Data Policy governs the management of and access to data relevant to perform and calibrate experiments as well as from experiments performed at the Extreme Light Infrastructure ERIC (ELI ERIC). It pertains to the curation, storage and access to data and metadata collected from the operation and scientific usage of the ELI Facilities.”*

*“ELI ERIC shall be the custodian of and steward for the Data, with the responsibility to collect, secure, archive and provide access to the Data. ELI ERIC shall aim at managing Data according to the ‘FAIR’ principles, meaning that Data shall be Findable, Accessible, Interoperable and organised in Reusable datasets.”*

“For experiments performed at ELI facilities a Data Management Plan (DMP) will be provided by the users applying for experiments. The DMP will be uploaded during the entire duration of the experiment and is the core document used to define the experimental data lifecycle.

# ELI Users' Journey, the Data Perspective!

The Data Policy provides the necessary support to address the above challenges and implement tools and services supporting users and users' activity!



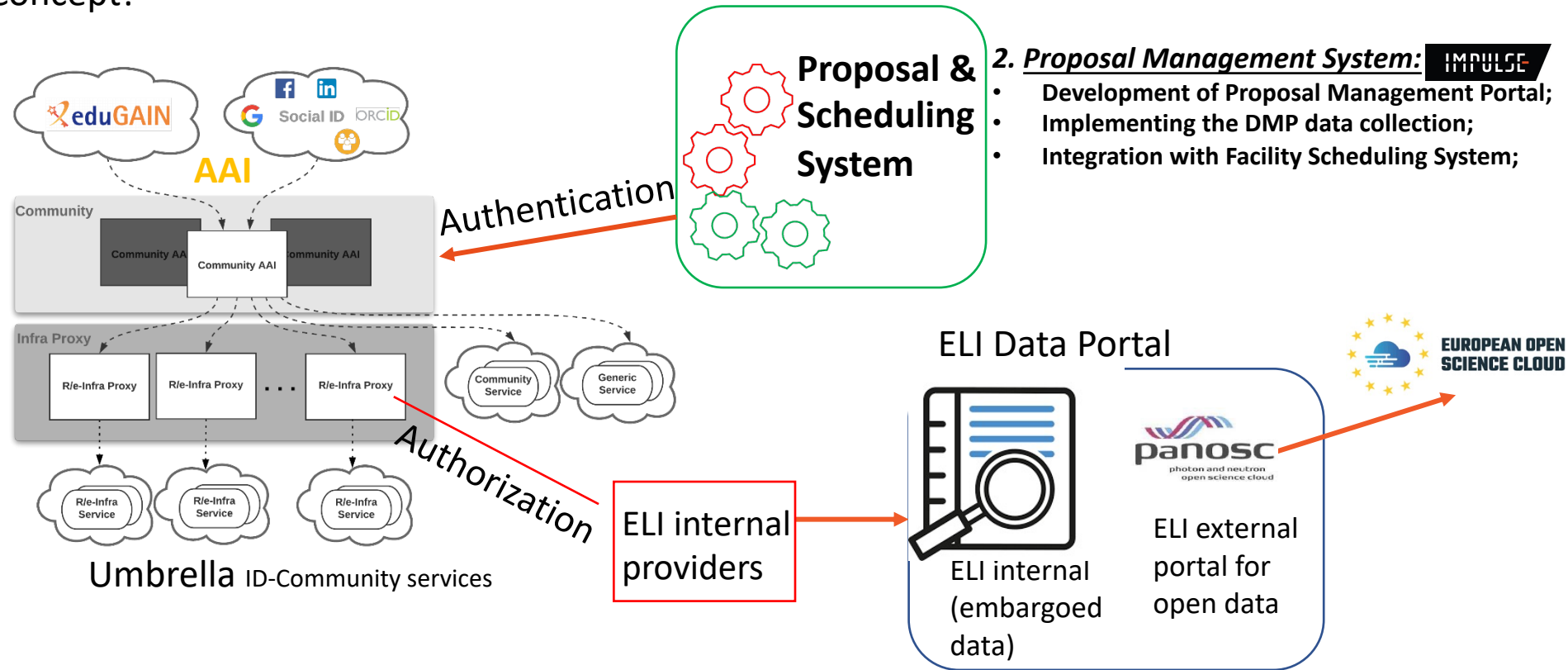
We have the metadata, CS is actively engaging with users to identify/collect/secure:

\***metadata** is the key for searchable/reproducible data, it does not contain the data but it contains enough information to reproduce a data set/conditions that could be used to reproduce that dataset. (<https://en.wikipedia.org/wiki/Metadata>)

\***rich metadata**  
<https://zenodo.org/record/3862701#.YWXLxpByUk>, <https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/>)

# Supporting the users by providing a fully integrated Scientific Data Management System!

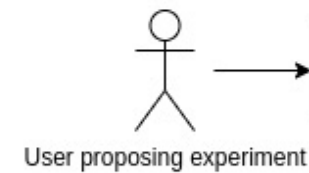
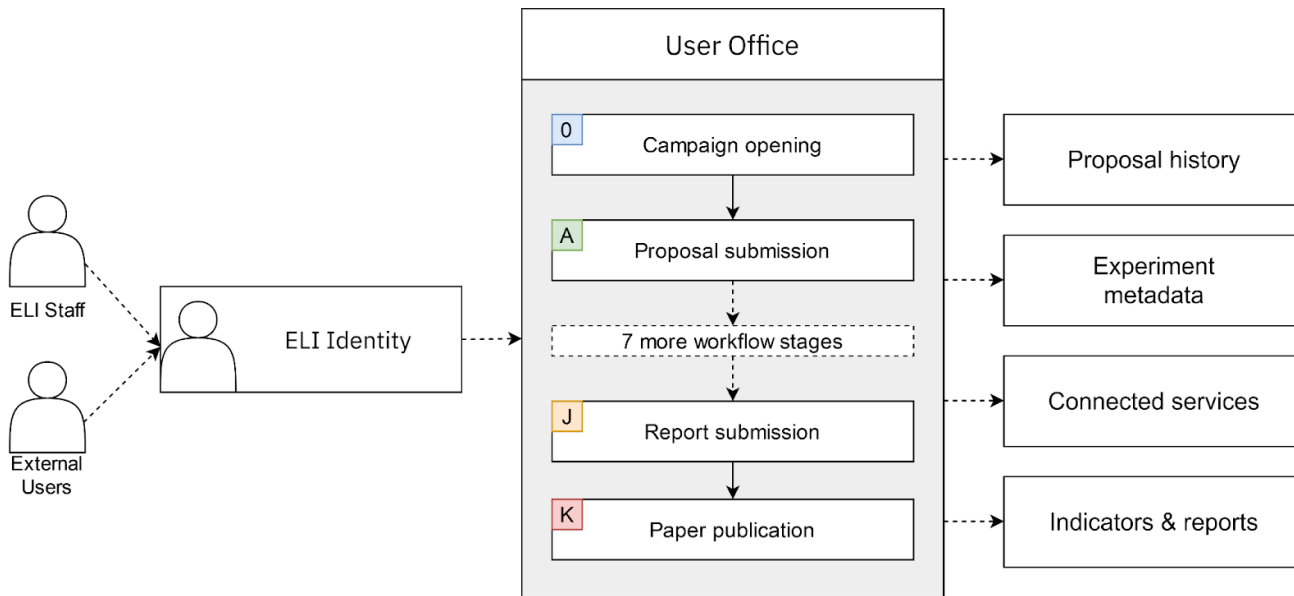
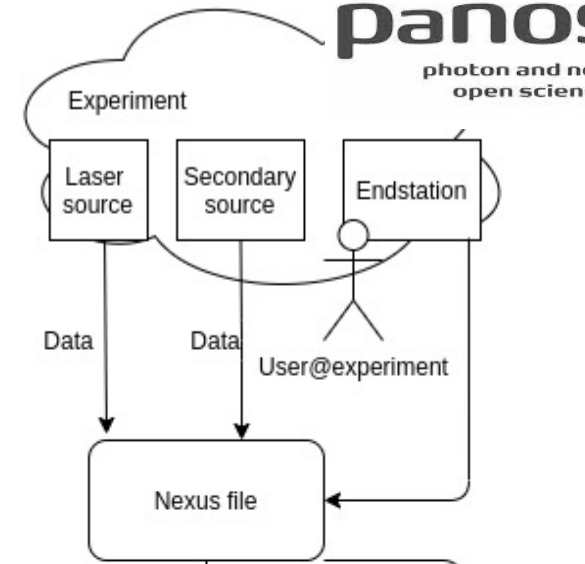
Concept!



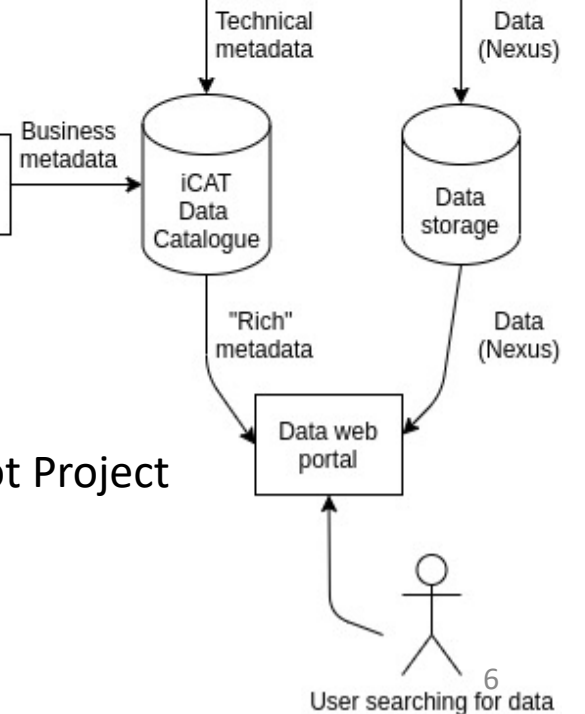
Pilots:

Catalogue integrations:

- User creates proposal via User Office web portal
  - Defines what he wants to do, when and how
- User does the experiment, the full data adhering to his dataset is saved into a single Nexus file
- User wants to search for his old data/some other researcher's data
  - Searches and downloads via the Data Portal web



User Portal - ICAT Pilot Project @ELI ERIC

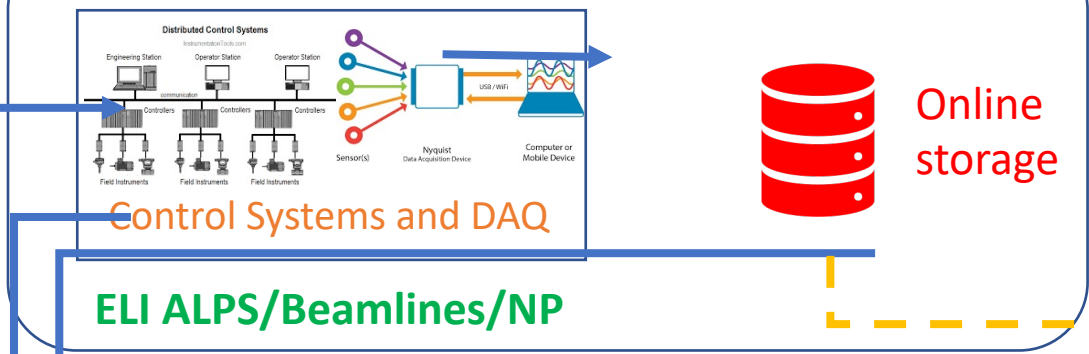


# Unified Data and Computing Experience

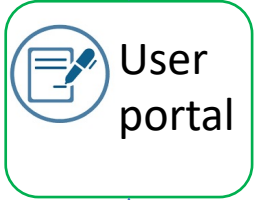
**WHO?**



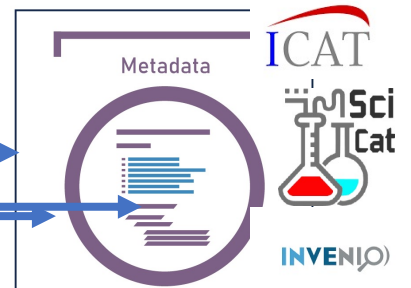
**WHERE FROM?**



**HOW?**



**HOW TO LOOK FOR**

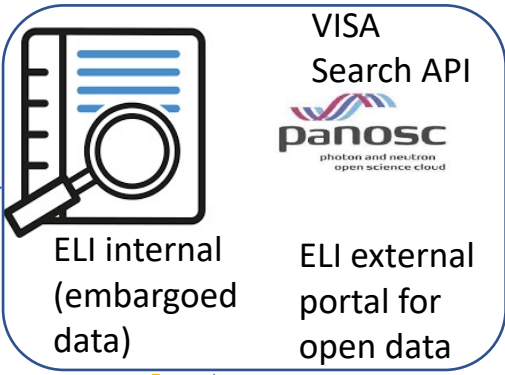


**(META)DATA?**

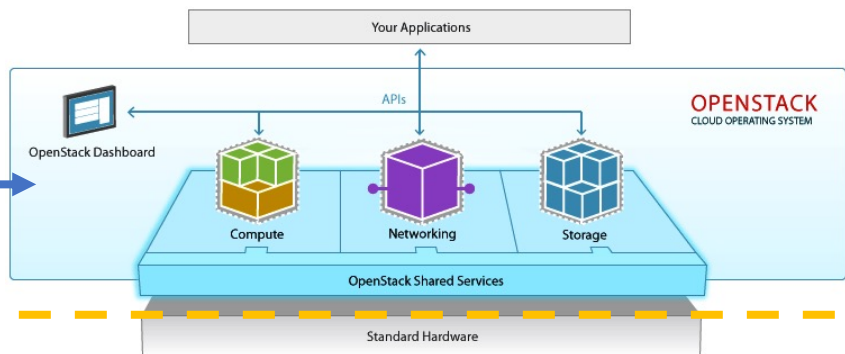
**ELI ALPS/Beamlines/NP**



**WHAT?**



**WHERE TO ANALYSE?**



HPC

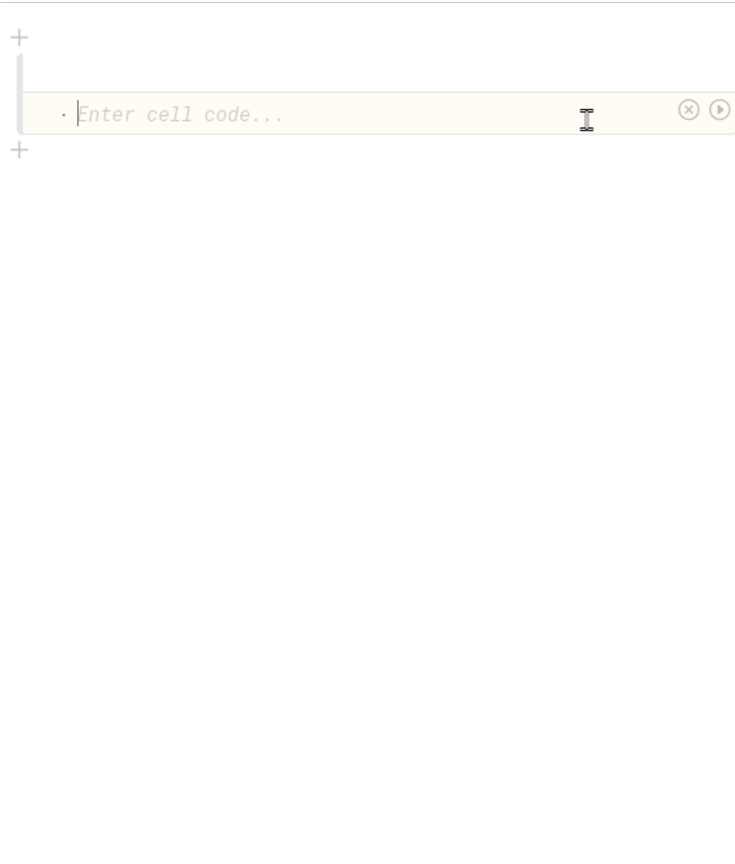
- Highly Scalable
- Easy to manage
- Vendor agnostic
- Automated provisioning
- Software can be automatically distributed

ELI Private Computing Cloud

**ELI ERIC** Electronic Logbook project, started with one PhD student is ready to be tested:

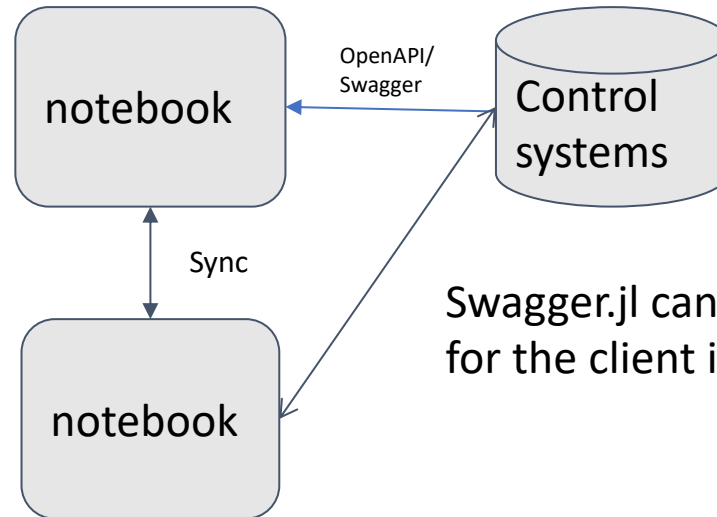
- A pilot is discussed with ELI Beamlines CS
- Another pilot to be started also @ELI ALPS
  - Coherent notebook state for everyone
  - Simple, easy to share
  - Based on Julia

### Distributed lookbooks technology



### Pluto notebooks

- Coherent notebook state for everyone
- Simple, easy to share
- Use julia



Swagger.jl can be used to generate a Julia package for the client interface from a .json spec.



We should be on-time with the implementation, we have the support of the Data Policy and DMP shall be implemented soon:

- Adopt standards for ELI (PID, DOI, Strategy)
- UmbrellaID – running, tests with ICAT
- VISA in place – with Proxmox (OpenStack analysed, in case of a delay we go with Proxmox as temporary solution)
- Proposal Management System
- ICAT POC – (using real data) + new front-end
- Data Portal
- New IT Infrastructure needed (storage and computing back-end)



THANK YOU

ELI ERIC Computing Team

T. Ivanoaica, M. Dostál, J. Bartoš, M. Schneider

L. Schrettner, B. Bago, J. Majer