ELI Scientific Data Management System

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

ELI ERIC Statute

ARTICLE 13 DATA POLICY

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 ERIC role, as CUSTODIAN of the Data: *"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."*

ELI Data Policy *has been developed and will soon be submitted to* the International Scientific and Technical Advisory Committee*. Expected to be adopted by the end of the year.*
"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."

For a consistent and efficient implementation of the policies, an integrated Scientific Data Management System is needed!

What Scientific Data Management is? How is this supporting users?

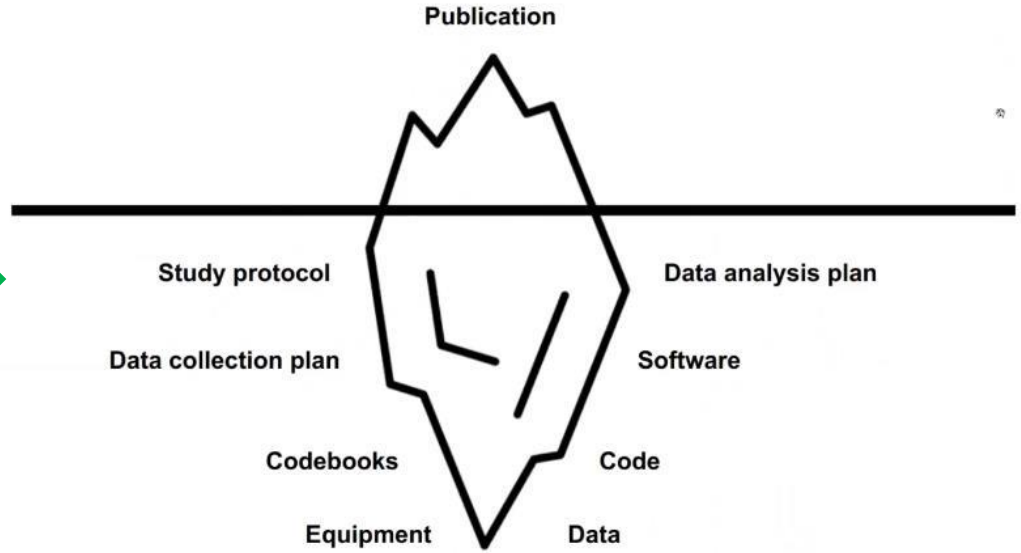
It is ELI's Research Data Management Plan.

It is the **collection of practices** helping us **plan, collect, process, analyse, preserve, share** and make **data re-usable.**

- Data management planning – the **DMP – reflecting the data lifecycle**
- Collecting raw data and metadata – Curation and correlation
- Processing to produce new data
- Analysing data to produce results
- Curating data for the long term
- Sharing data and making it Findable, Accessible, Interoperable and Reusable

The direct impact of FAIR:

- Facilitates new data
- Derived data produced
- Enables new research
- Accelerates science



Common challenges are building communities!

What the users communities are asking:

- Good (meta)data + logbooks
- Performant Download services
- Digital Object Identifiers for Data
- Remote data analysis
- Access to Open Data
- Credit for Data re-use

What Funders are asking:

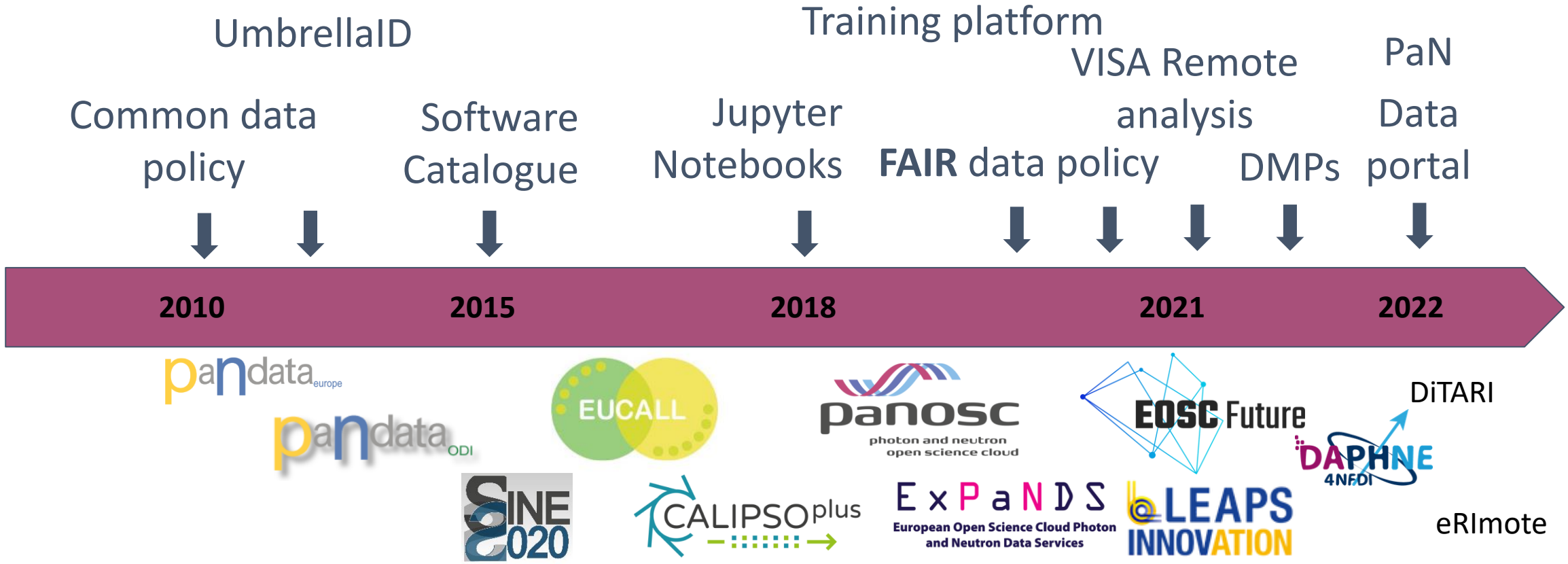
- FAIR Data
- Open Science
- Digital Object Identifiers for Data
- Reproducible Publications
- Participate in the EOSC
- Metrics about Data Re-Use

What is happening pre/after discussing data?

- User: has an idea / need to study a sample
- Proposal: User writes a proposal for one of the PaN facilities
- Review committee: Reviews proposal and rates scientific quality
- Beamline scientist: Review proposal and checks feasibility
- Beamtime allocated: User travels to facility / sends sample
- Experiment: Sample(s) are exposed to beam + data collected
- Analysis:Data is reduced, analysed+ curated (DOI)
- Publication: User publishes results (DOI) in peer review journal

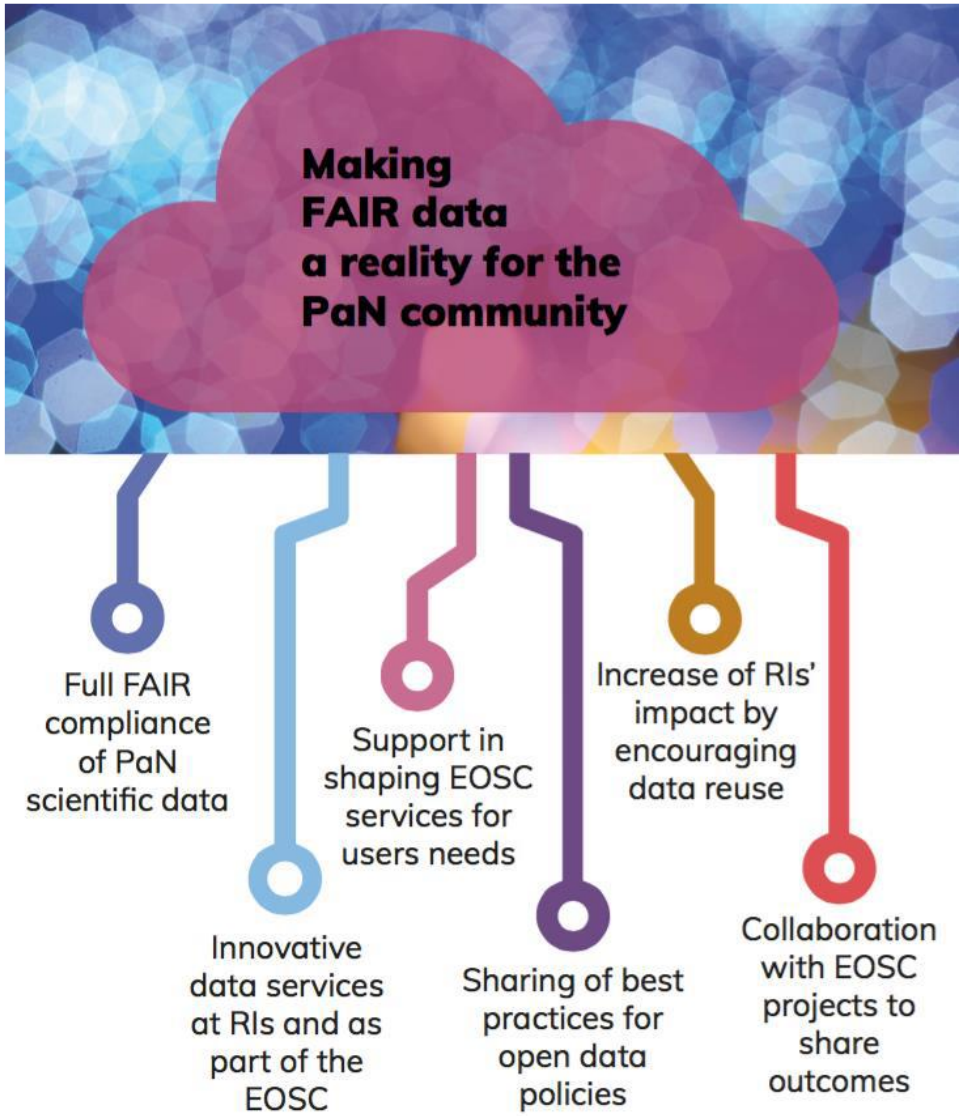
Communities support

PaNOSC community we keep building on



6

PaNOSC is more than tools, is the community sharing the same challenges, same standards and working together to find unique solutions.



What PaNOSC does:

Policies supporting adoption of FAIR policies:

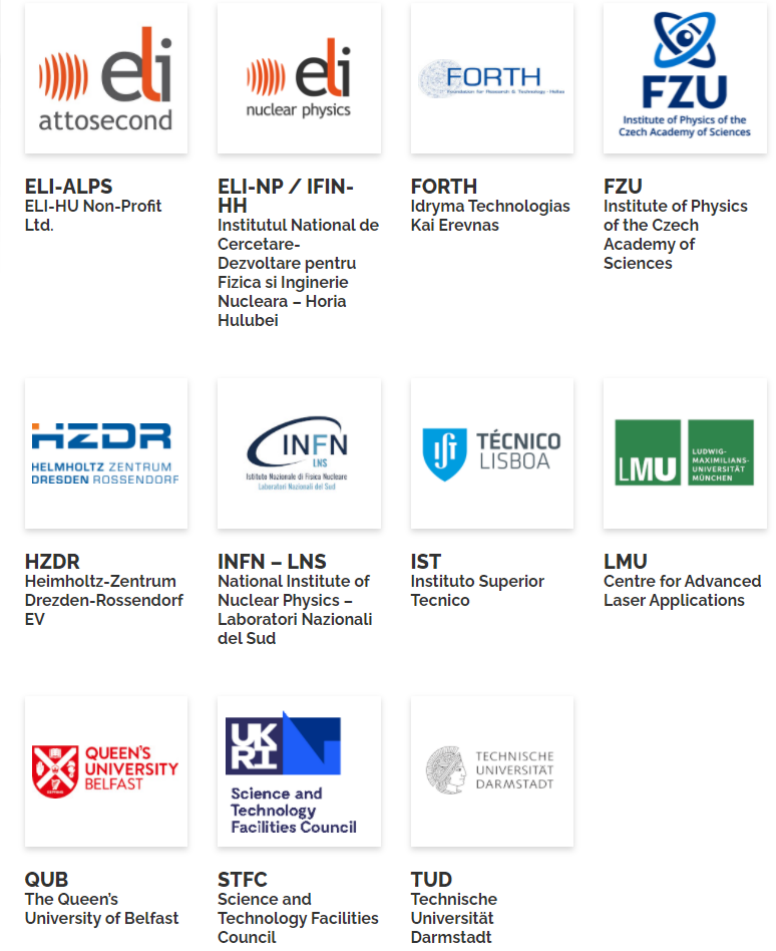
- Data Policy Framework - https://zenodo.org/record/3862701
- Data Policy guidelines - https://zenodo.org/record/4899344

Tools and services:

- AAI
- File Cataloguing solutions and support
- Data tools:
  - Data portal
  - Data transfer tools and solutions for PaN
  - ……

Specific ELI Input for implementing FAIR Principles.

IMPULSE Project Goal: A global platform for high-power laser science and development, uniting the facilities of the Extreme Light Infrastructure together.



ELI-ALPS
ELI-HU Non-Profit Ltd.

ELI-NP / IFIN-HH
Institutul National de Cercetare-Dezvoltare pentru Fizica si Inginerie Nucleara – Horia Hulubei

FORTH
Idryma Technologias Kai Erevnas

FZU
Institute of Physics of the Czech Academy of Sciences

HZDR
Heimholtz-Zentrum Drezden-Rossendorf EV

INFN – LNS
National Institute of Nuclear Physics – Laboratori Nazionali del Sud

IST
Instituto Superior Tecnico

LMU
Centre for Advanced Laser Applications

QUB
The Queen's University of Belfast

STFC
Science and Technology Facilities Council

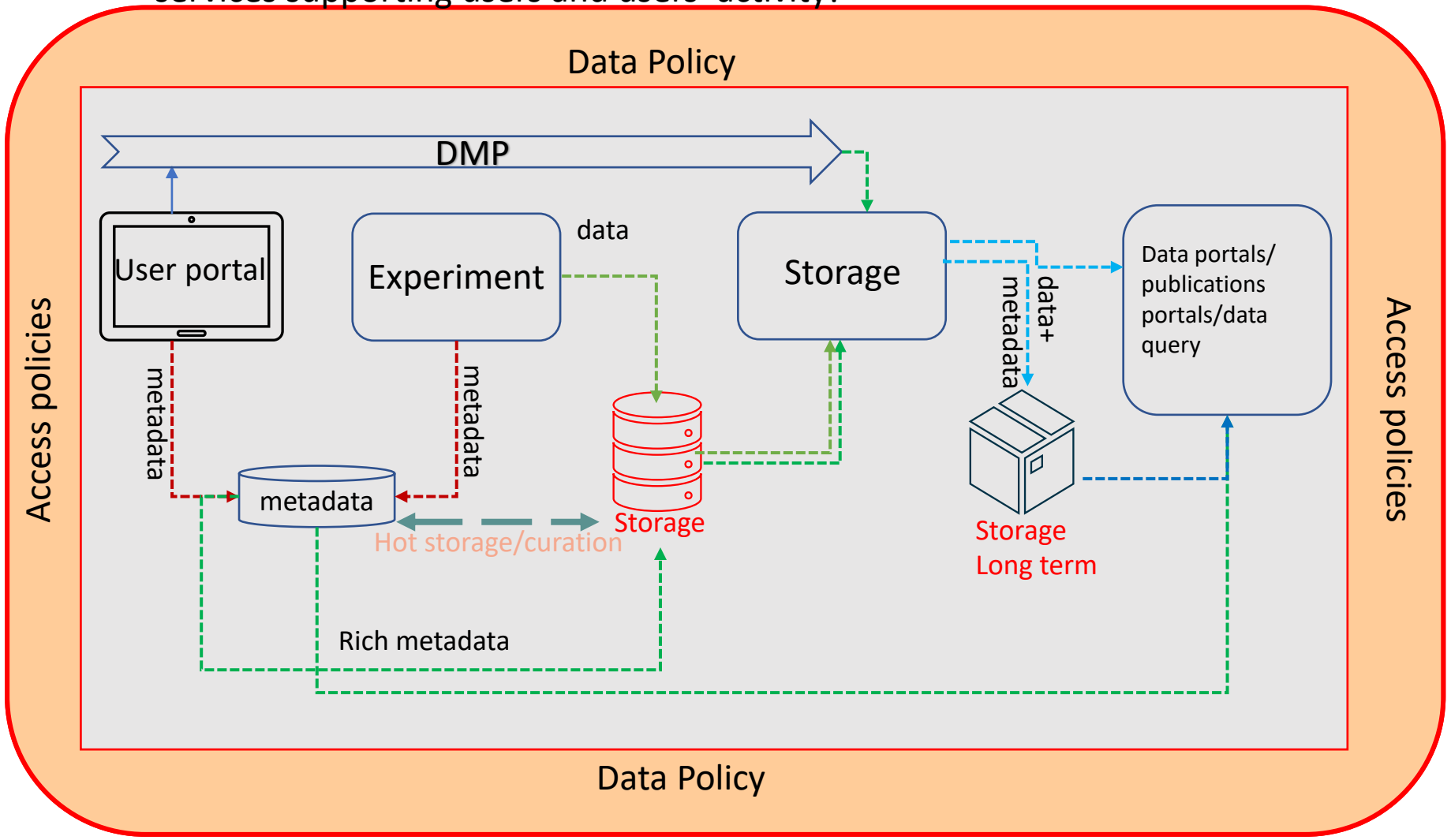
TUD
Technische Universität Darmstadt

Provides the necessary support for having the FAIR principles and all tools and services implemented based on ELI Specific requirements.
Major outcomes that are already used in the design:

- Users office workflow and user portal processes - supporting the implementation of the **DMP;**
- Simulation software expected to improve operations – supporting the data analysis and simulation services for users;
- CS teams are joining efforts – accelerating the development of data tagging, data correlation and data curation processes;
- …….
- Most of the activities are boosting the design and implementation of the Data Policies and data services.

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#.YW-xLxpByUk, https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/)

FAIR (Meta)DATA matters and makes a difference!

Unified Data and Computing Experience

WHO?

Identity Management System?

Federated ID(providers):

Reuse existing:

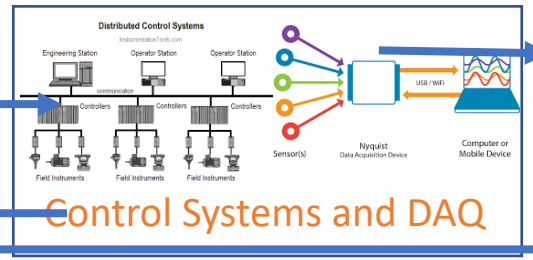
- AD
- LDAP

Add :

- UmbrellaID
- ORCID

.....

WHERE FROM?



Control Systems and DAQ

- Control/Operate
- Acquire & Generate (metadata)
- Archive Big

Online storage

ELI ALPS/Beamlines/NP

HOW?

User portal

Proposal submission:

- Technical feasibility
- Scientific peer review
- Approval
- Scheduling

HOW TO LOOK FOR

Metadata

ICAT

SciCat

INVENIO

(META)DATA?

Ingest Metadata
Make it searchable

??offline and Long Term Storage??

WHAT?

ELI internal (embargoed data)

panosc
photon and neutron open science cloud

ELI external portal for open data

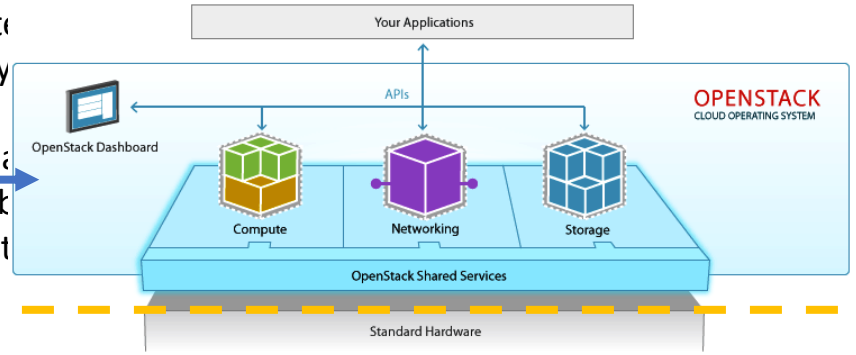
Internal portal:

- Access for Authenticated users(AI)
- Dedicate
- capacity

PaN Portal:

- Open Da
- Access k
- Connect

WHERE TO ANALYSE?



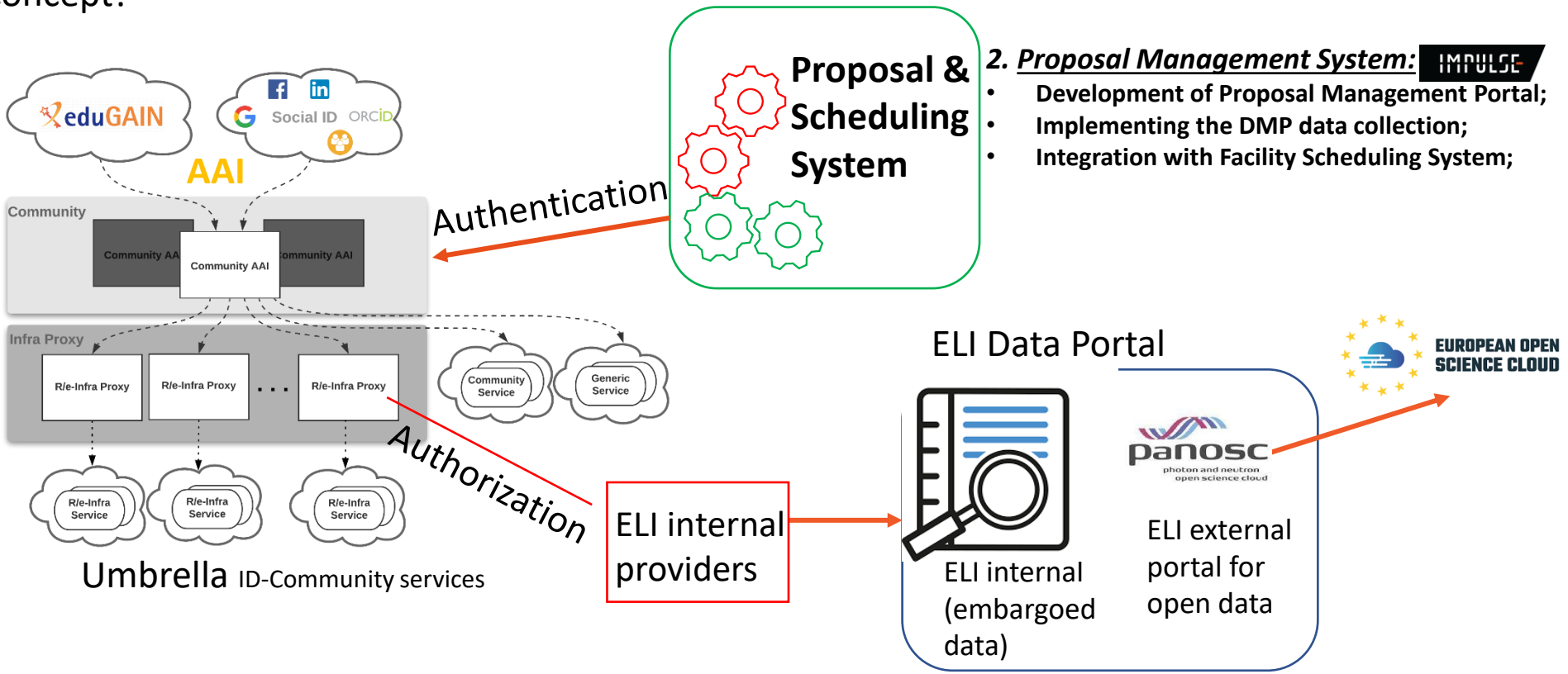
HPC

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

ELI Private Computing Cloud

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

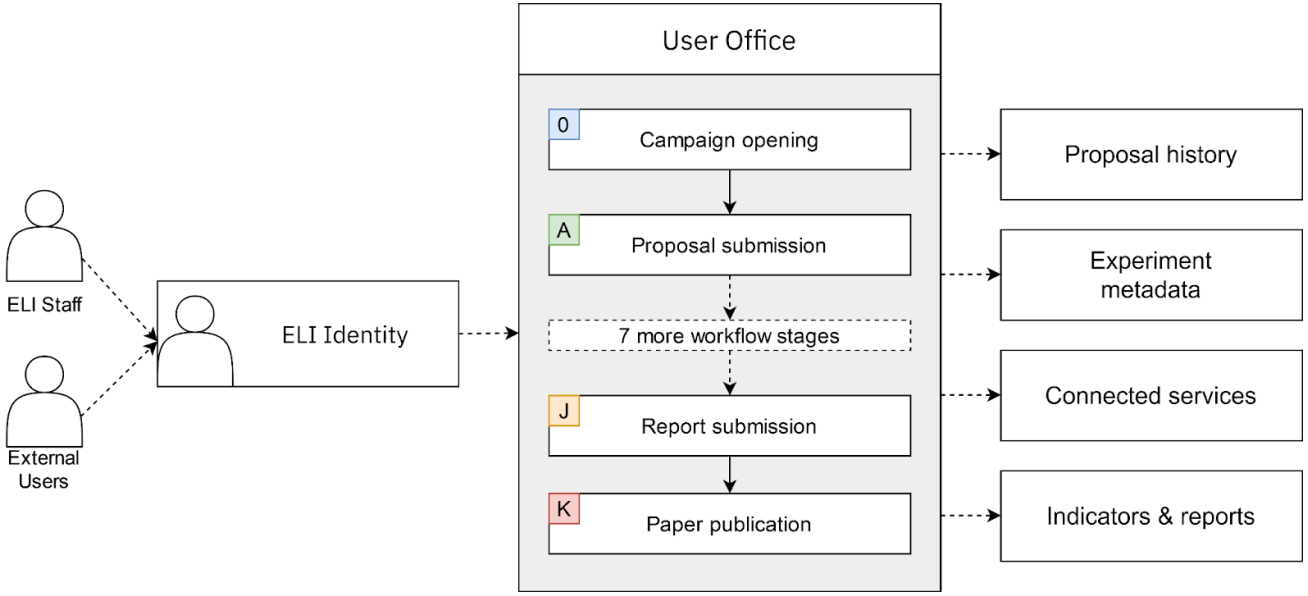
Concept!

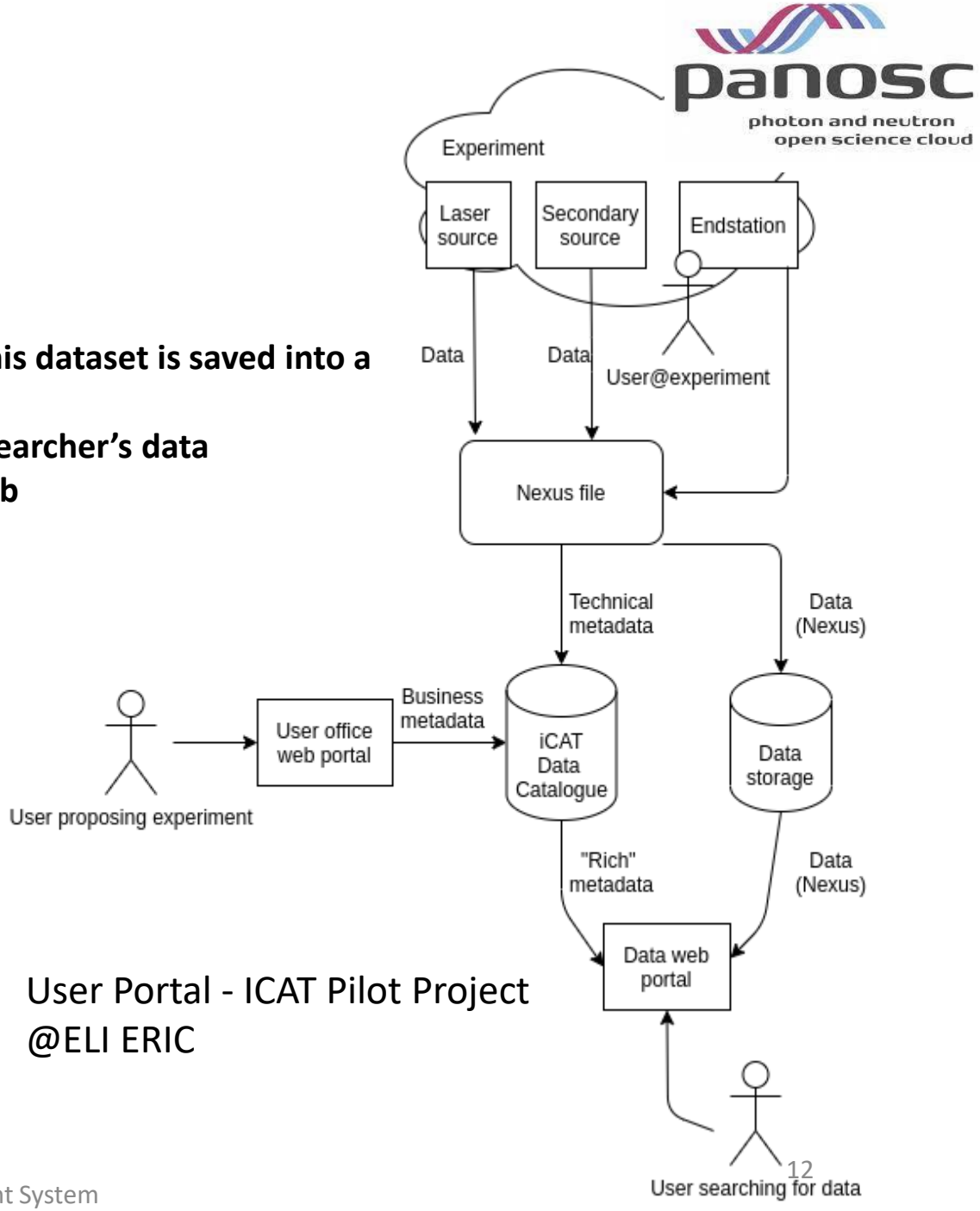


Pilots:

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

NEXT Steps

ELI ALPS and Beamlines CS and IT teams are actively engaging users and identifying requirements/challenges/needs.

Data matters and users' input adds value to data!

Based on the details we have collected:

- Electronic logbook solution, allowing collaborative work has been identified – pilot project to be started with ELI Beamlines CS Team, a similar approach will be prepared with ELI Alps.
- Data Curation and computing back-end strategy for supporting our users.
- ACCELERATE the implementation of the Data Policy becomes a priority!

THANK YOU

ELI ERIC Computing Team
Teodor Ivanoaica
Martin Dostal
Jiri Bartos