

# Enabling Open Science for Photon and Neutron Sources

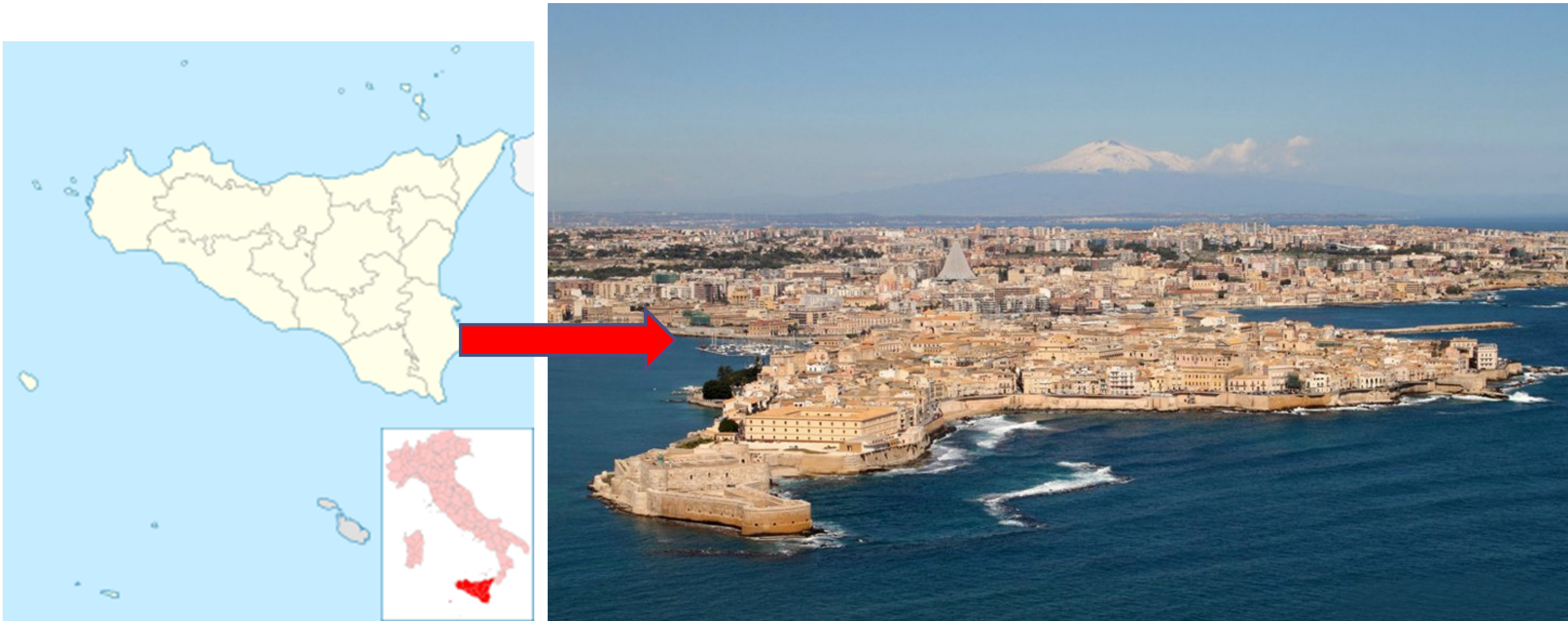
**8 October, 2019**

**Presenter: Andy Götz (on behalf of the PaNOSC collaboration)**

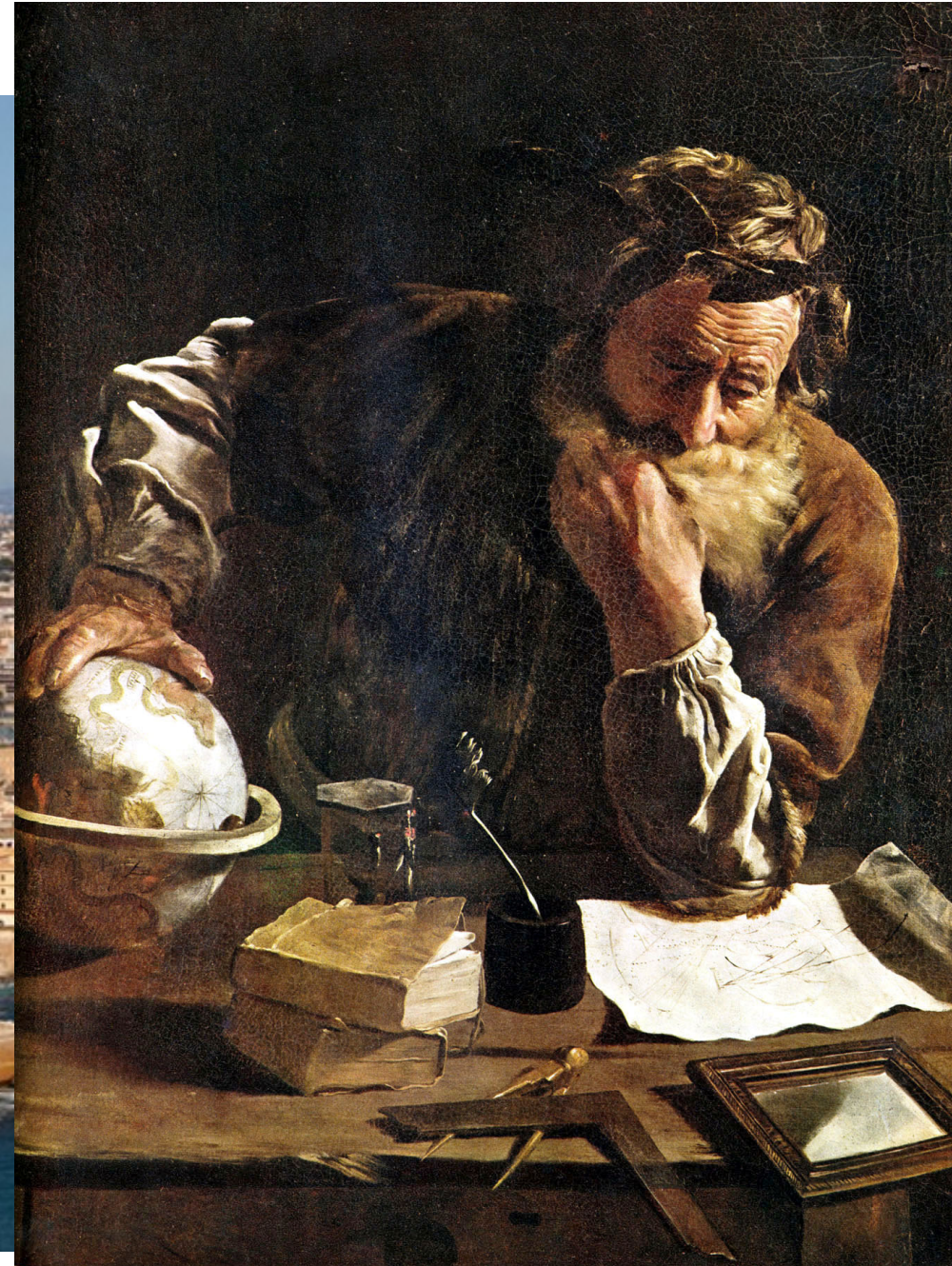
**Venue: ICALEPCS 2019 conference in Brooklyn, NYC, USA**



# Archimedes of Syracuse = a Blueprint for Modern Science



# Archimedes of Syracuse = a Blueprint for Modern Science



# Archimedes Palimpsest *aka* Codex C

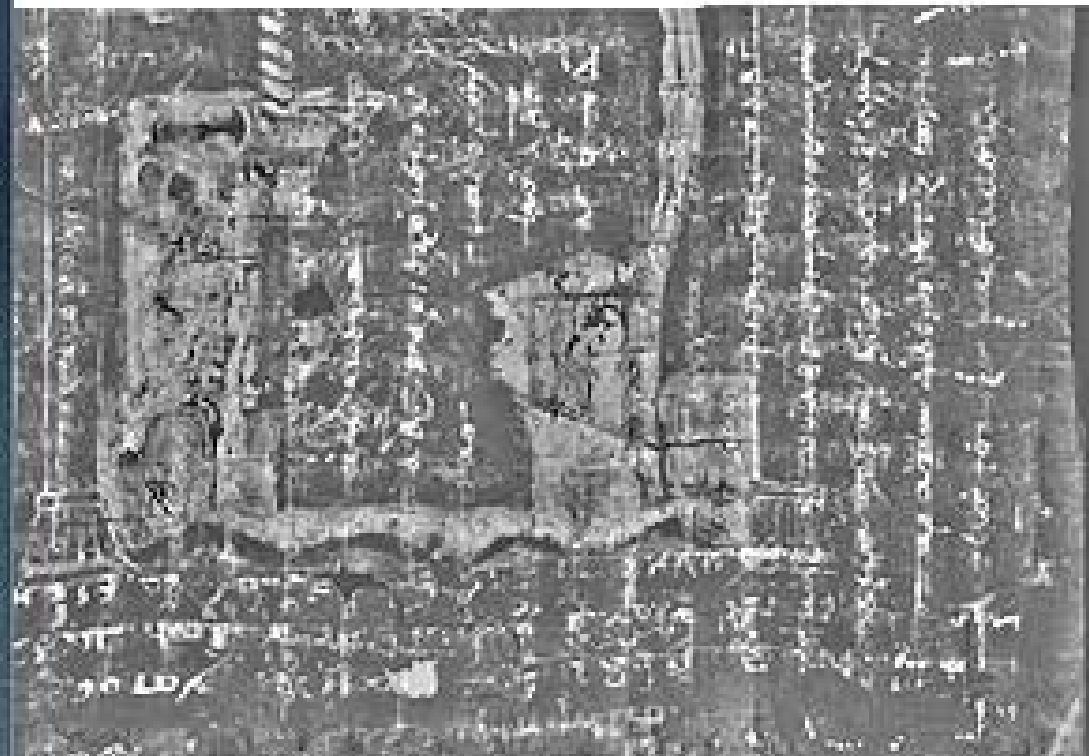
13<sup>th</sup> Century Prayer Book



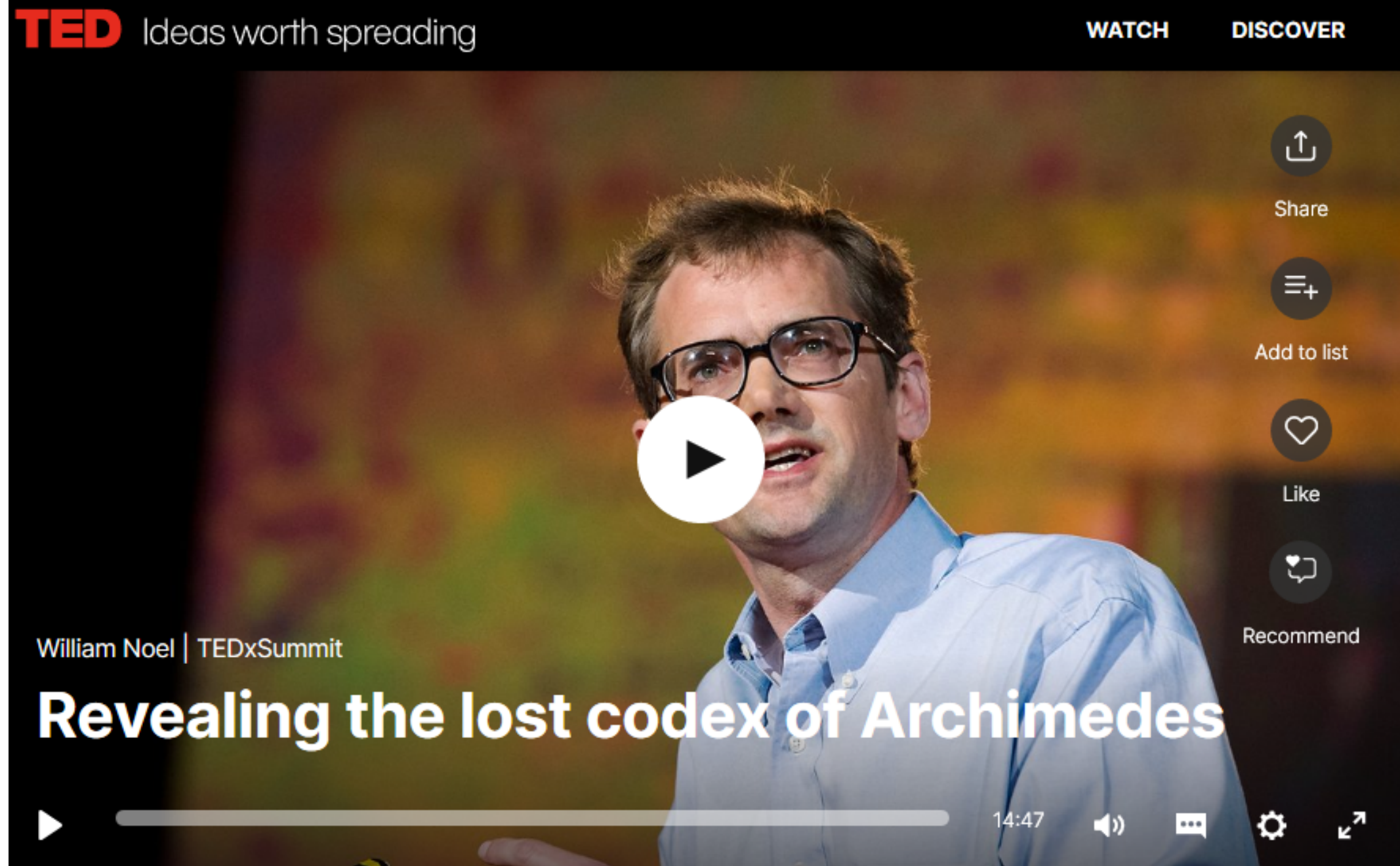
# Archimedes Palimpsest *aka* Codex C



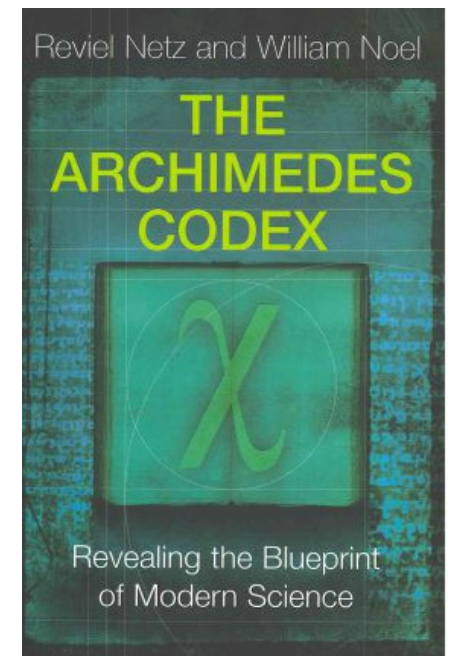
X-ray Fluorescence data taken at Stanford Synchrotron



# Archimedes Palimpsest TED *by* William Noel

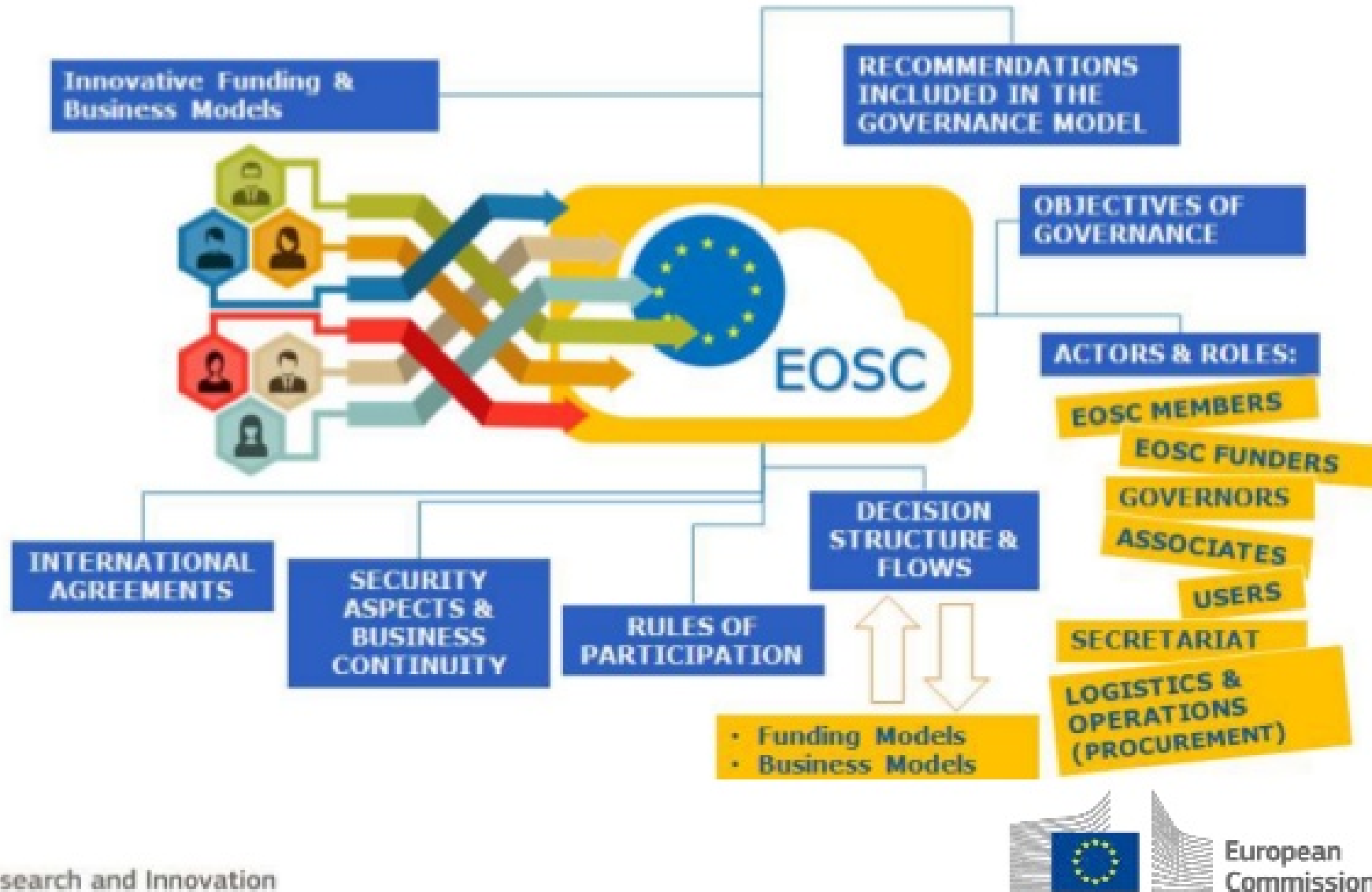


*“Data is going to die if it’s not used, so you want your data to be as easily usable as possible, and a Creative Commons license is an integral part of that.”*



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

# European Open Science Cloud (EOSC)



1. Increase openness, integrity, and reproducibility of research
2. “Making FAIR Data reality”
3. “Github for science in Europe”

Research and Innovation



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

# Open Science in USA

The National Academies of  
SCIENCES · ENGINEERING · MEDICINE

CONSENSUS STUDY REPORT

## Reproducibility and Replicability in Science



**DATA**  
COALITION

ABOUT ▾

ISSUES ▾

[Home](#) > [News](#) > [Press Releases](#) >

### Press Release

## President Signs Government-wide Open Data Bill

Posted on **January 14, 2019** by **Jessica Yabsley**

**Washington, D.C.** – Today, President Trump signed into law the Foundations for Evidence-Based Policymaking (FEBP) Act (H.R. 4174, S. 2046), which includes the [Open, Public, Electronic and Necessary \(OPEN\) Government Data Act](#) (Title II). The package [passed Congress](#) on Monday, December 31, 2018.

The [OPEN Government Data Act](#) requires all non-sensitive government data to be made available in open and machine-readable formats by default. It establishes Chief Data Officers (CDO) at federal agencies, as well as a CDO Council. The law's mission is to improve operational efficiencies and government services, reduce costs, increase public access to government information, and spur innovation and entrepreneurship. This is a win for evidence-based decision-making within the government.

“Today is a historic day for the open data movement.” said **Sarah Iov Hays, Acting Executive Director of**



This project has received funding from the

# PaNOSC factsheet

**Call:** Horizon 2020 InfraEOSC-04

**Partners:** ESRF, ILL, XFEL.EU, ESS, CERIC-ERIC, ELI-DC, EGI

**Description:** cluster of ESFRI Photon and Neutron sources

**Observers/non-funded:** GÉANT, EUDAT, national RIs

**Linked 3<sup>rd</sup> parties via EGI:** DESY, STFC, CESNET

**Status:** Started 1/12/2018

**Github:** <https://github.com/panosc-eu>

**Home page:** <https://panosc.eu>

**Twitter:** @PaNOSC\_eu #PaNOSC

**Budget:** 12 M€

**Coordinator:** ESRF

**Started:** 1/12/2018

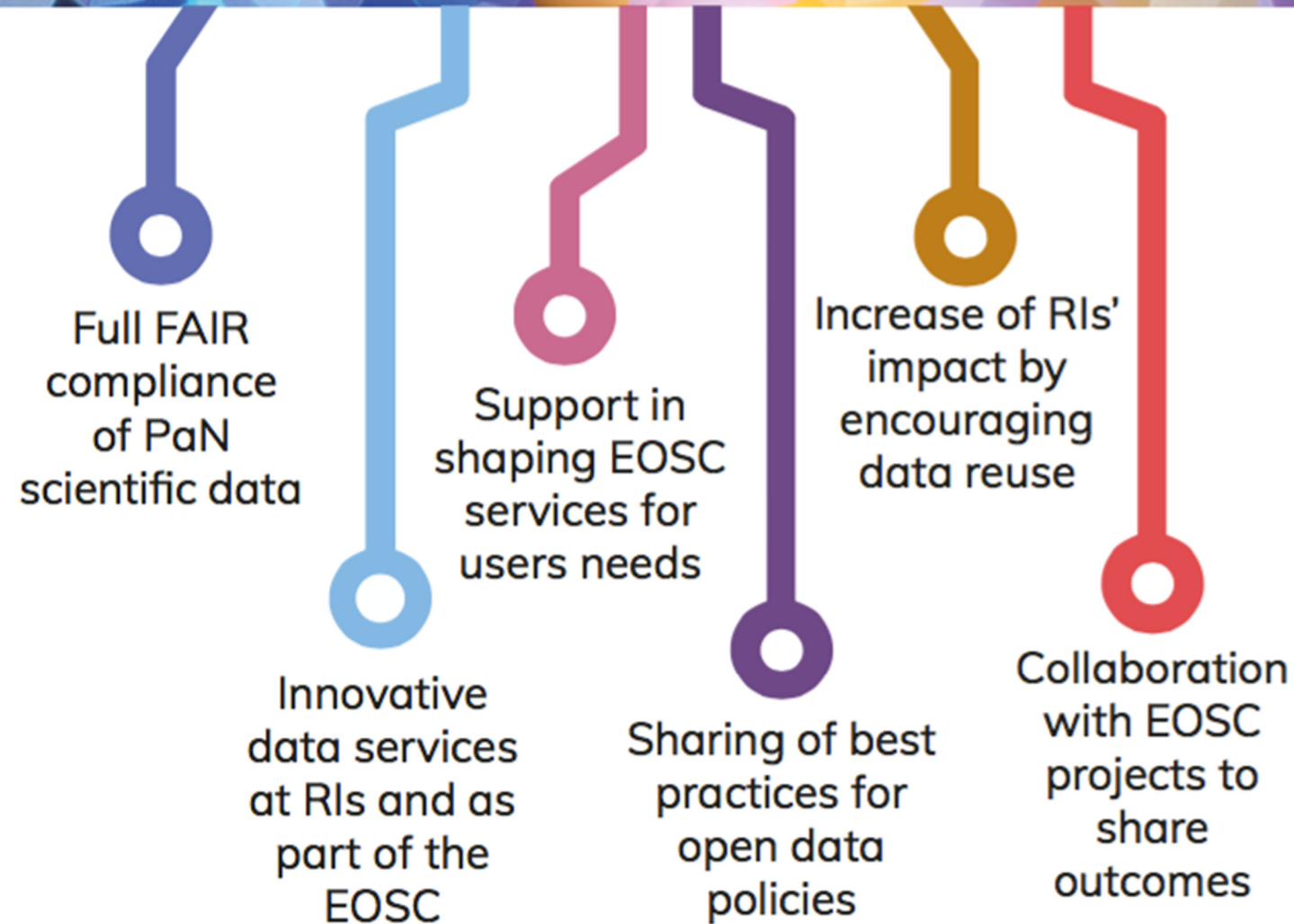
**Duration:** 4 years



# PaNOSC goals



PaNOSC works closely with the PaN sources in Europe to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services, integrating them into the EOSC.



# Data management enables

**Publications**

**On-site  
data analysis**

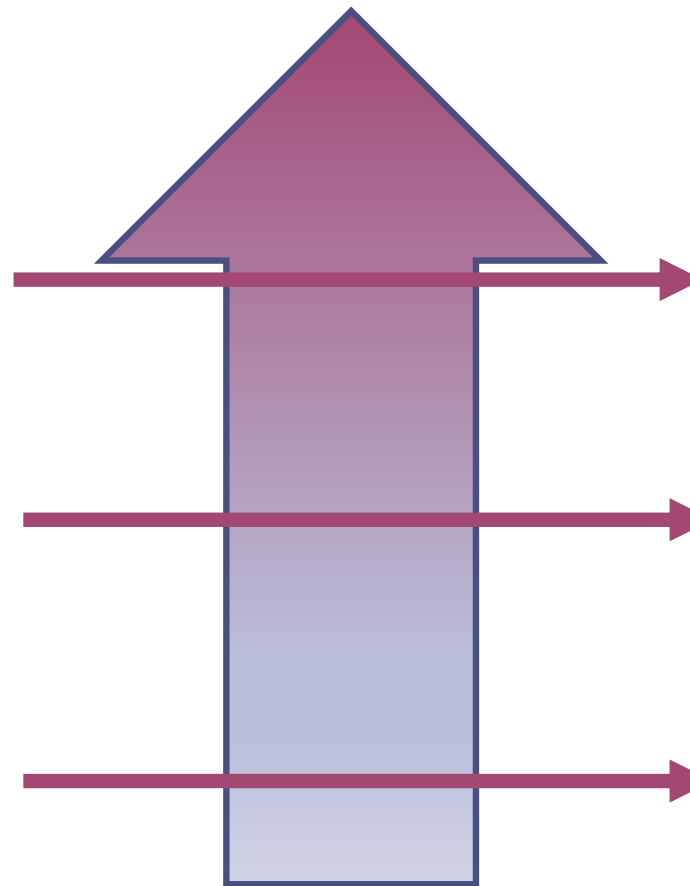
**Data mining**

**AI**

**Federated catalogues**

**On-site data reduction**

**Data + metadata catalogues**



**Open Data  
Analysed Data  
Digital Objects**

**Reduced Data  
Archived Data**

**Raw Data**



**F**indable



**PID**  
**[DOI]**

**A**ccessible



**Open Licence**  
**Downloadable**

**I**nteroperable



**Rich Metadata**

**R**eusable



1

**E-Logbooks**  
**Machine**  
**friendly**

# Data catalogue WP

## Fair Data API development

Identified two sides of EOSC Integration:

- Harvesting Data by EOSC Agents and Repos
- FAIR Data Search to be federated

Existing catalogues and their capabilities as well as third-party options for implementations have been surveyed.

## NeXus Survey for Ontologies

Facility practises and plans around file formats are being shared in sessions with partners. Identified commonalities can be applied everywhere and inform the catalogue search task (see <https://nexusformat.org> )

### Dataset

Method	Endpoint	Description
PATCH	/Datasets	Patch an existing model instance or insert a new one into the data source.
GET	/Datasets	Find all instances of the model matched by filter from the data source.
PUT	/Datasets	Replace an existing model instance or insert a new one into the data source.
POST	/Datasets	Create a new instance of the model and persist it into the data source.
PATCH	/Datasets/{id}	Patch attributes for a model instance and persist it into the data source.
GET	/Datasets/{id}	Find a model instance by {{id}} from the data source.
HEAD	/Datasets/{id}	Check whether a model instance exists in the data source.
PUT	/Datasets/{id}	Replace attributes for a model instance and persist it into the data source.
DELETE	/Datasets/{id}	Delete a model instance by {{id}} from the data source.
GET	/Datasets/{id}/exists	Check whether a model instance exists in the data source.

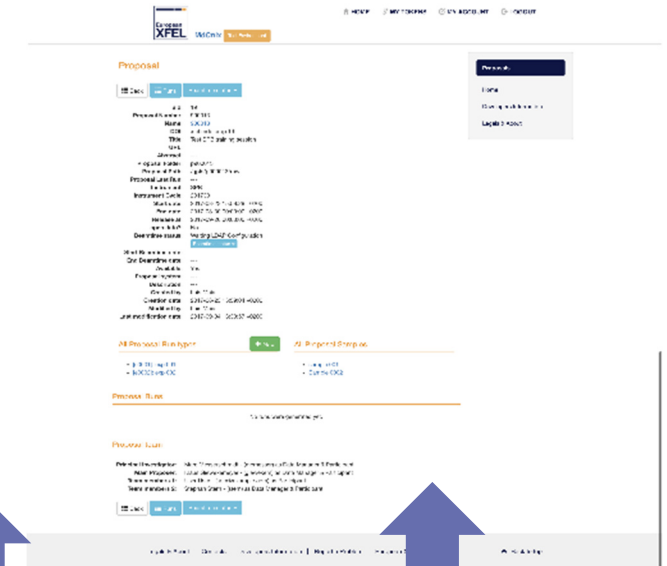
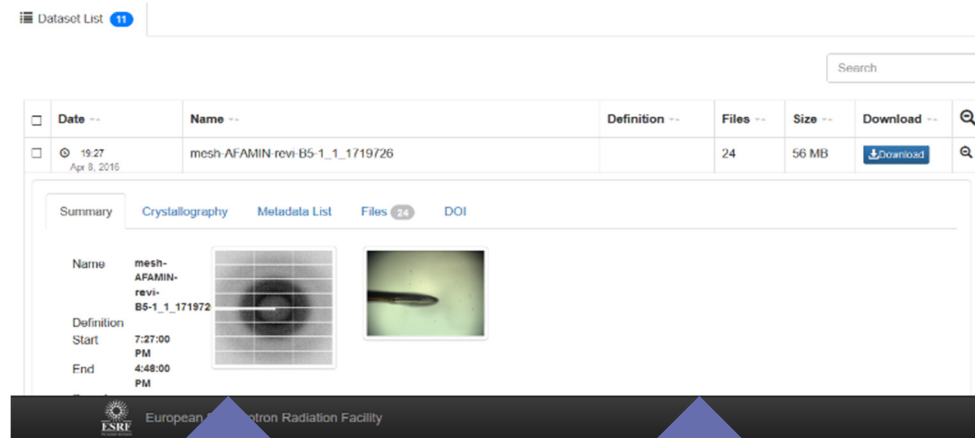
WP3

Silx viewer

Name	Type	Shape	Value
bigfake.nxs			
entry	NXentry		
event_data_detector_1	NXevent_data		
features	uint64	3	Compressed
instrument	NXinstrument		
chopper_1	NXdisk_chopper		
chopper_2	NXdisk_chopper		
chopper_3	NXdisk_chopper		
delay	NXlog		
depends_on	string	scalar	"/entry/instru...
name	string	scalar	"Juelich, WF...
ntp_to_mrf_comparison	int64	scalar	0
radius	float64	scalar	300
rotation_speed	NXlog		
slit_edges	float64	12	Compressed...
slit_height	float64	scalar	130
slits	int64	scalar	6
top_dead_center	NXlog		
transformations	NXtransformations		
chopper_4	NXdisk_chopper		
chopper_5	NXdisk_chopper		



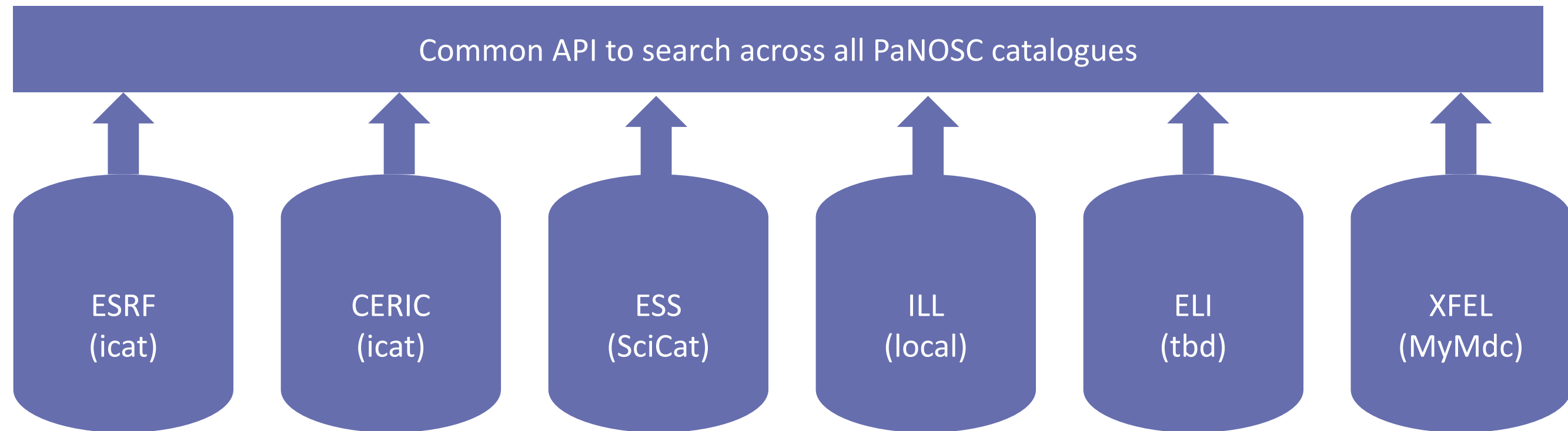
# PaNOSC has 6 data catalogues with different APIs + UIs



# PaNOSC common API across all sites



Search for Datasets 

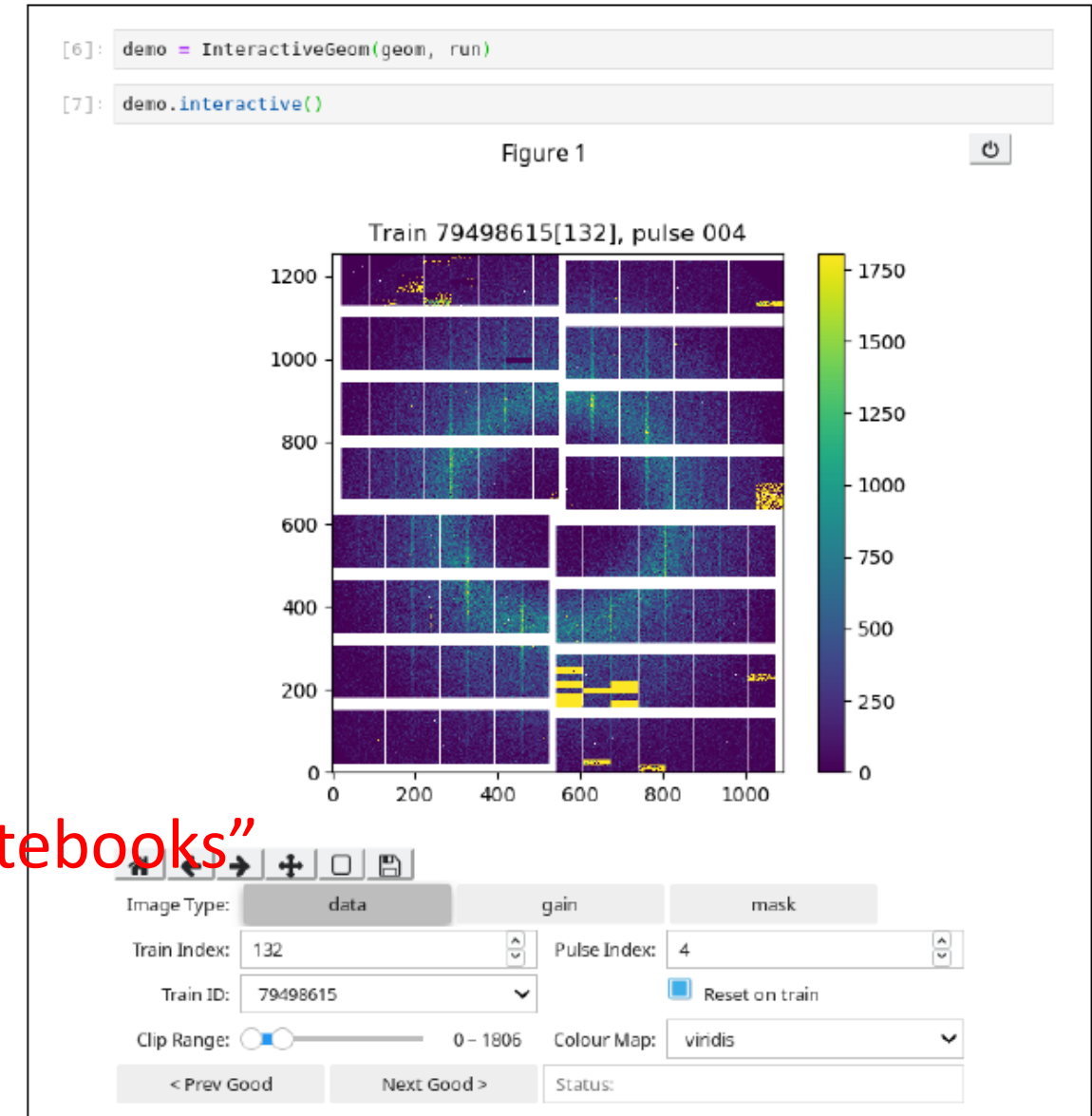


# Data Analysis Services - choices

- Jupyter Notebooks
- Remote Desktop in Browser
- Remote data analysis portal
- HDF5 and visualisation in notebooks

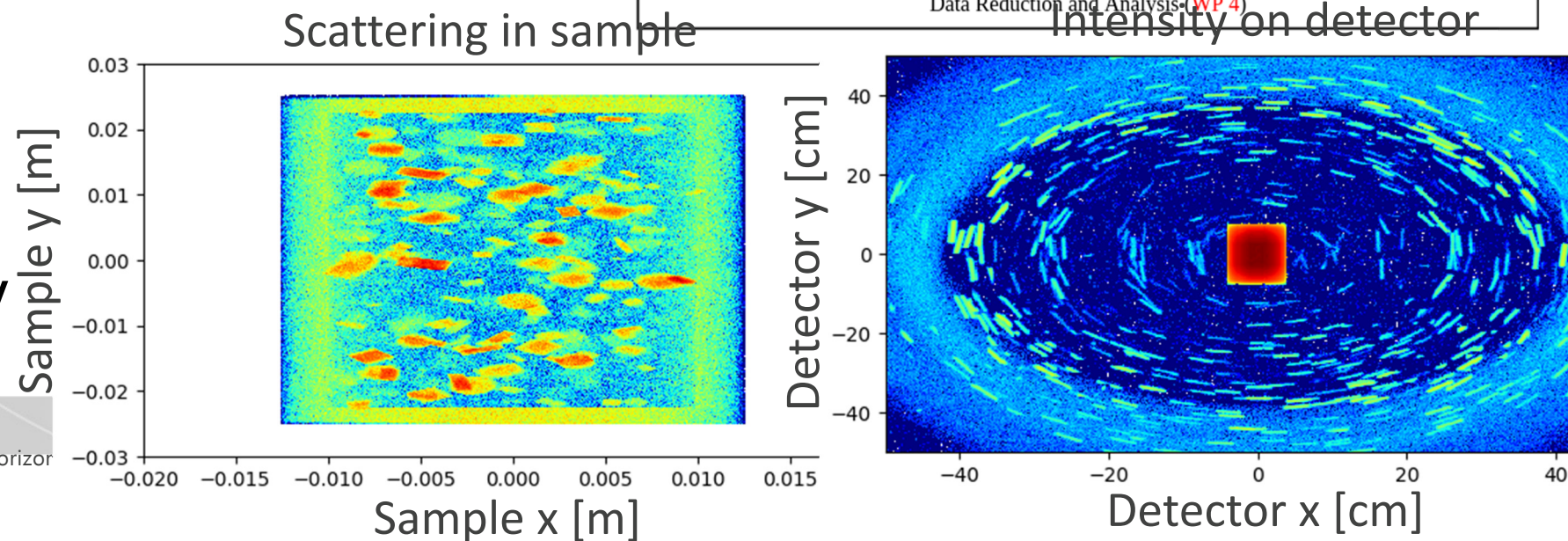
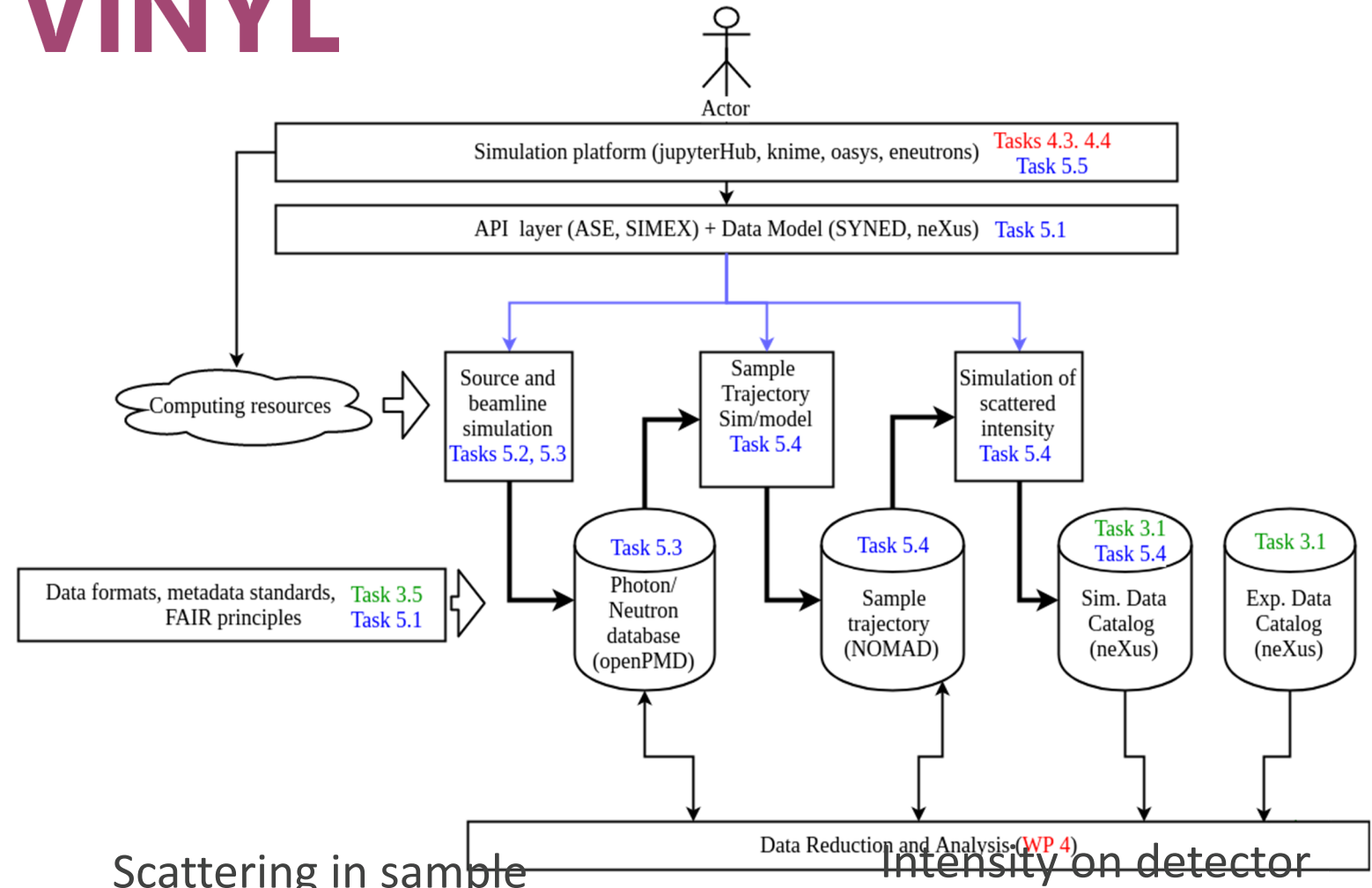
**TUCPR02: "Data Exploration and Analysis with Jupyter Notebooks"**

**WEPHA057: "Building a Data Analysis as a Service Portal"**



# Simulation services - ViNYL

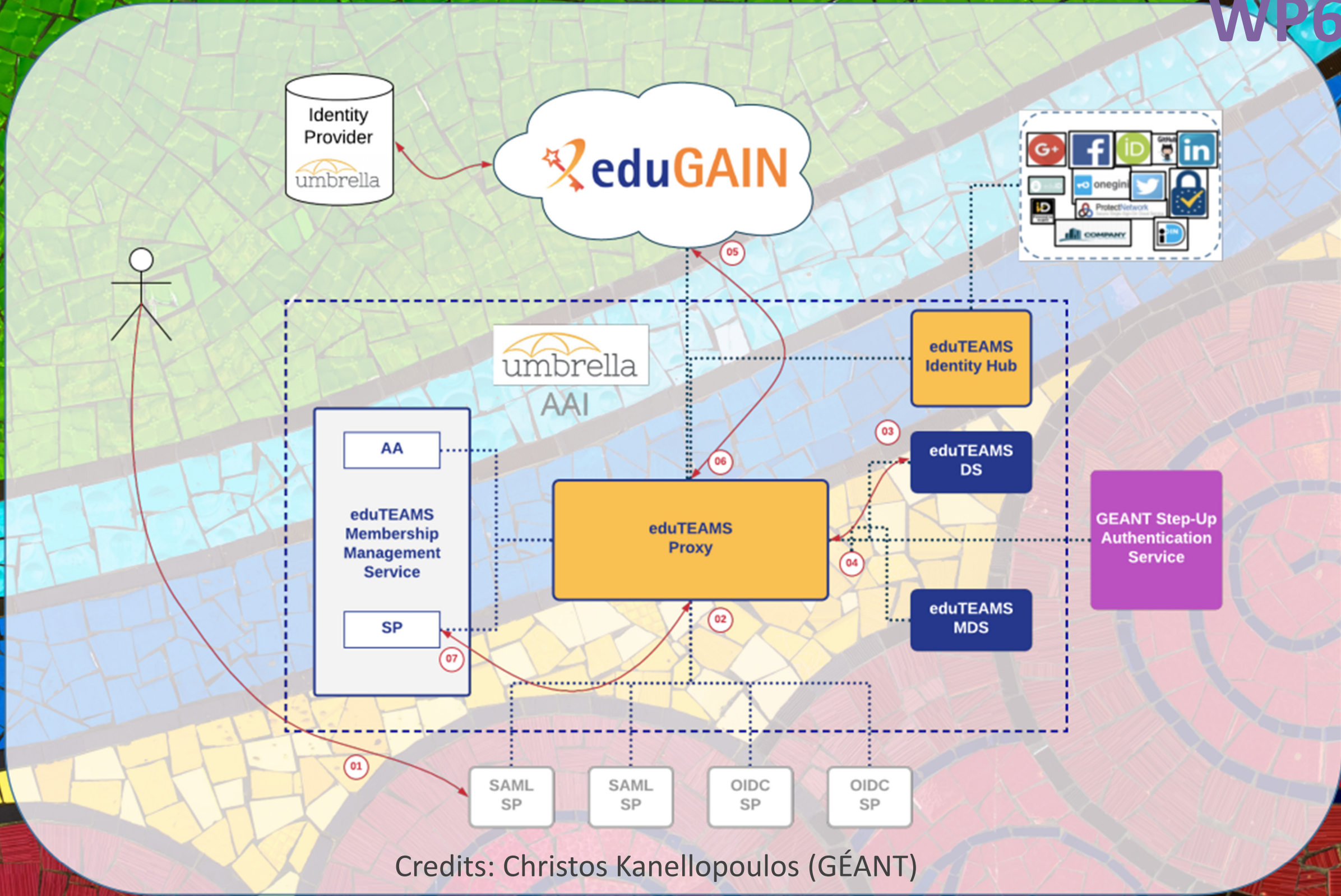
- **SIMEX**: Start-to-end photon experiment simulation library (python)
- **McStas-script**: python API for Neutron ray-tracing with McStas
- **OASYS**: Wavefront propagation for beamline design (WISE)
- **EPOCH** particle in cell + McStas: Simulation of laser driven neutron sources and ray tracing



# EOSC Integration – AAI i.e user identity

- **AAI = Authentication and Authorisation Infrastructure**
- **Critical** if we want to identify people and define roles
- **EOSC** must solve at least the AAI problem





Credits: Christos Kanellopoulos (GÉANT)

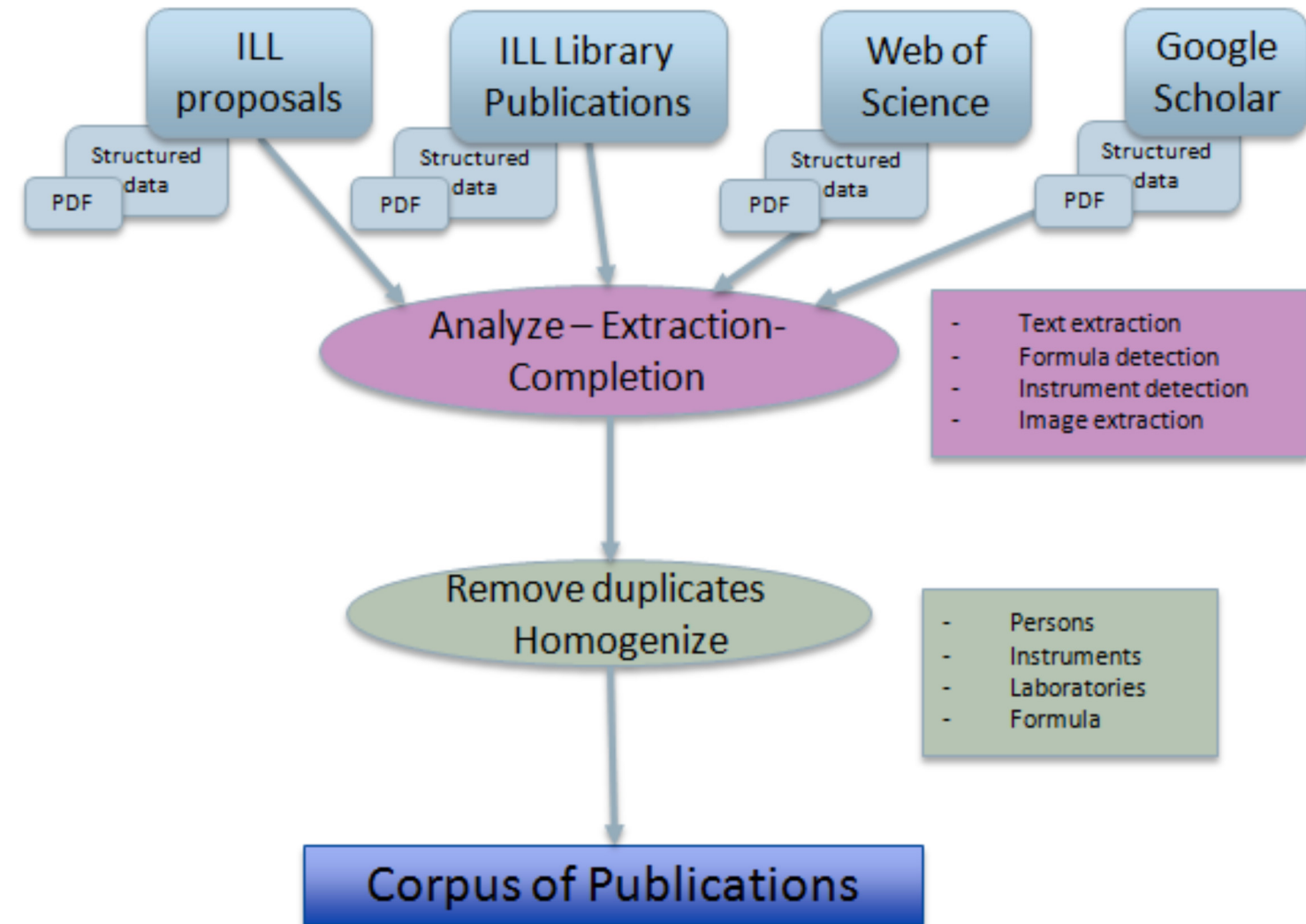
# EOSC Integration – Data transfer

- **3 uses cases :**
  1. User driven data transfer (e.g. **Globus Online**)
  2. Data archiving for RI (**STFC as the archive center**)
  3. Transfer from RI to compute facilities on behalf of users (i.e. based on the scenario where users perform analysis on a different infra than the one of the RI where the data have been produced). Currently exploring **OneData**, **dCache** and simple solution without caching (**webdav**)



# PUMA publication metrics

- Publications and citations per instrument
- **Delay between a proposal and a publication**
- Countries and laboratories which use ILL (ESRF, ...) data
  - Look for potential new users
- **Scientific trends**
- **Impact factor per science facility**



# e-learning

## e-neutrons.org

- Wiki with neutron scattering theory
- Web instrument simulation using McStas
- Quizzes using both theory and simulation

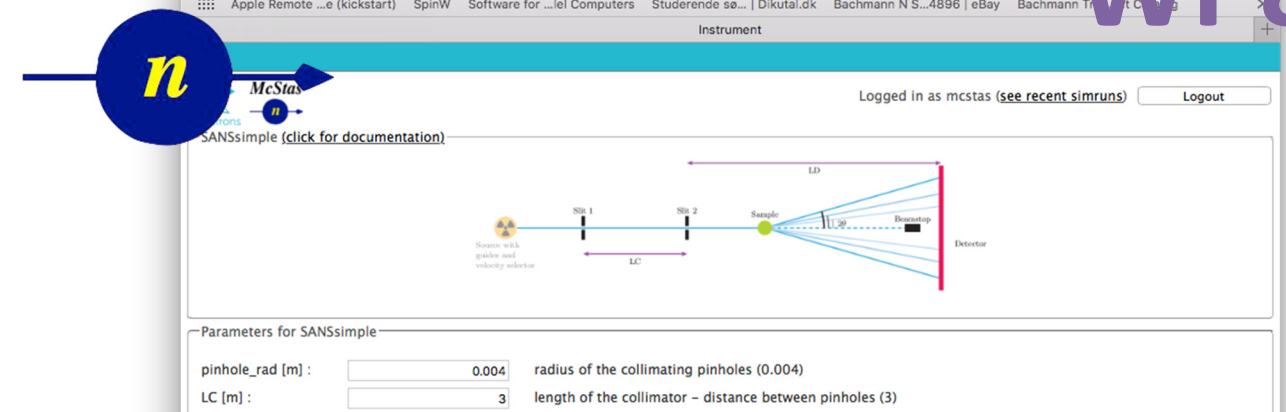
## Migration to ESS servers underway

## Extensions:

- Support for Jupyter Notebooks
- Integrate data analysis services
- Integrate simulation services

# McStas

# WP8



The screenshot shows a MediaWiki page titled 'Inelastic nuclear neutron scattering'. The page content includes an introduction to the topic, a table of contents, and sections on scattering theory for nuclear dynamics and scattering from initial to final state. It contains several mathematical equations, including the differential cross-section equation (1) and the nuclear potential equation (2). The page also features a sidebar with navigation links and a footer with funding information.

**Inelastic nuclear neutron scattering**

One of the early successes of neutron scattering was the study of dynamics of matter, in particular phonon dispersion relations. Here, the vibrational frequency (or phonon energy) is deduced from the change in neutron energy through the principle of energy conservation. Hence, for the study of dynamics we are dealing with inelastic neutron scattering.

This page naturally leads to the description on the [Scattering from lattice vibrations](#) page of neutron scattering from quantised lattice vibrations, or *phonons*. The related topic of inelastic neutron scattering from diffusion and molecular motion is not covered in this version of the notes.

Instrumentation for the general field of inelastic neutron scattering is described on the [Instrumentation](#) page.

**Contents** [hide]

- 1 \*Scattering theory for nuclear dynamics
  - 1.1 \*Scattering from initial to final state
  - 1.2 \*The observable nuclear cross section

**\*Scattering theory for nuclear dynamics**

We will now return to the basic scattering theory from the [Basics of neutron scattering](#) page to derive the equations that govern all inelastic scattering from nuclei.

**\*Scattering from initial to final state**

In the [master scattering equation](#) on the [Basics of neutron scattering](#) page, we derived the starting equation for the inelastic cross section:

$$(1) \quad \frac{d^2\sigma}{d\Omega dE_f} \Big|_{\lambda_i \rightarrow \lambda_f} = \frac{k_f}{k_i} \left( \frac{m_n}{2\pi\hbar^2} \right)^2 |\langle \lambda_i \psi_i | \hat{V} | \psi_f \lambda_f \rangle|^2 \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega).$$

We begin by expanding the expression for the [nuclear potential](#) on the [Small angle neutron scattering](#) page:

$$(2) \quad \hat{V} = \frac{2\pi\hbar^2}{m_n} \sum_j b_j \delta(\mathbf{r} - \mathbf{R}_j),$$

where  $\mathbf{R}_j$  is now the operator for the position of the  $j$ 'th nucleus. We use this to expand the matrix element in the inelastic cross section:

$$(3) \quad \begin{aligned} & |\langle \lambda_i \psi_i | \hat{V} | \psi_f \lambda_f \rangle|^2 \\ &= \left( \frac{2\pi\hbar^2}{m_n} \right)^2 \left[ \sum_j b_j \left\langle \lambda_i \left| \int \psi_i^* \delta(\mathbf{r} - \mathbf{R}_j) \psi_f d^3r \right| \lambda_f \right\rangle \right]^2 \\ &= \left( \frac{2\pi\hbar^2}{m_n} \right)^2 \sum_{j,j'} b_j b_{j'} \langle \lambda_i | \exp(-i\mathbf{q} \cdot \mathbf{R}_j) | \lambda_i \rangle \langle \lambda_i | \exp(i\mathbf{q} \cdot \mathbf{R}_{j'}) | \lambda_i \rangle. \end{aligned}$$

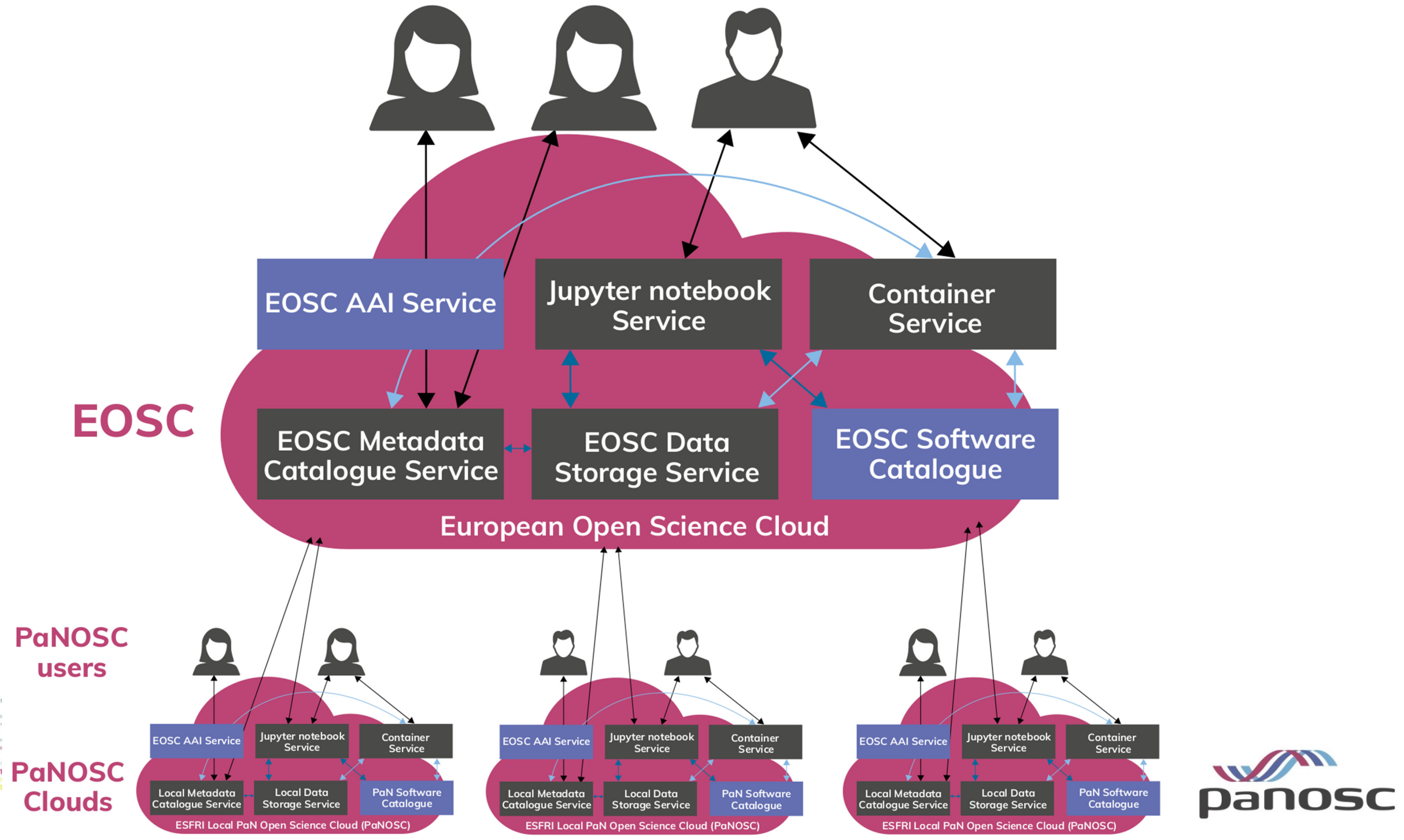
If all nuclei were fixed in position, we would now reach the diffraction cross section by summing over the (in practice unmeasurable) finite states of the lattice,  $|\lambda_i\rangle$ , since the  $\delta$ -function in (1) would factorize out and vanish by integration. However, we cannot do this simple calculation now, so we need to take a more difficult path. We rewrite the troublesome delta-function in (1), using  $2\pi\delta(a) = \int_{-\infty}^{\infty} \exp(iax) dx$  (following Squires<sup>[1] 2.3):</sup>

$$(4) \quad \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \exp\left(\frac{i(E_{\lambda_i} - E_{\lambda_f})t}{\hbar}\right) \exp(-i\omega t) dt.$$

Now, we utilize a rather intuitive identity from quantum mechanics, valid when  $|\lambda\rangle$  is an eigenstate of the Hamiltonian  $H$  with eigenvalue  $E_\lambda$ :

$$(5) \quad \exp\left(\frac{iE_\lambda t}{\hbar}\right) |\lambda\rangle = \exp\left(\frac{iE_\lambda t}{\hbar}\right) |\lambda\rangle.$$


# Our Vision – layers of scientific clouds



# Conclusion

- 1. FAIR data management is an essential part of (Open) Science**
- 2. We need to implement FAIR data management at the source (collect metadata, curate data with PIDs etc.)**
- 3. Data Management is the foundation for Data Analysis as a Service**



# Philosophy

***“Best of all is to win (Data Management of Petabytes for decades). But if you can't win, then you should fight for a noble cause because then you'll be remembered.”***

**Hyperides**

[Archimedes Palimpsest]

