# PaNOSC + EOSC + CNRS

**22 Janvier, 2020**

**Presenter: Andy Götz (coordinateur de PaNOSC)**

**Meeting: Journée EOSC au CNRS**

**Venue: Auditorium Marie Curie, Paris**

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

# Questions de Laurent Lellouch

1. Y-a-t-il des leçons (écueils à éviter, difficultés, recettes gagnantes, etc.) que vous pouvez partager avec des communautés moins avancées que vous?

2. Y-a-t-il des outils que vous avez développés qui pourraient être utiles à d'autres communautés?

3. Comment le CNRS pourrait vous être utile dans la mise en place d'un OSC PaN, pas seulement au travers de moyens financiers, mais également en expertise scientifique?

4. Comment pourriez-vous être utiles au CNRS dans la mise en place de son Open Science policy?

# PaNOSC in the EOSC context


Cluster projects
FAIR

Photon + Neutron
ESFRI

ESRF
EBS

# PaNOSC in the EOSC context


Cluster projects — Photon + Neutron — ESRF

FAIR — ESFRI — EBS

# 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

**Co-funded project:** ExPaNDS

**Linked 3rd 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

**Ends:** 1/12/2022

**Duration:** 4 years

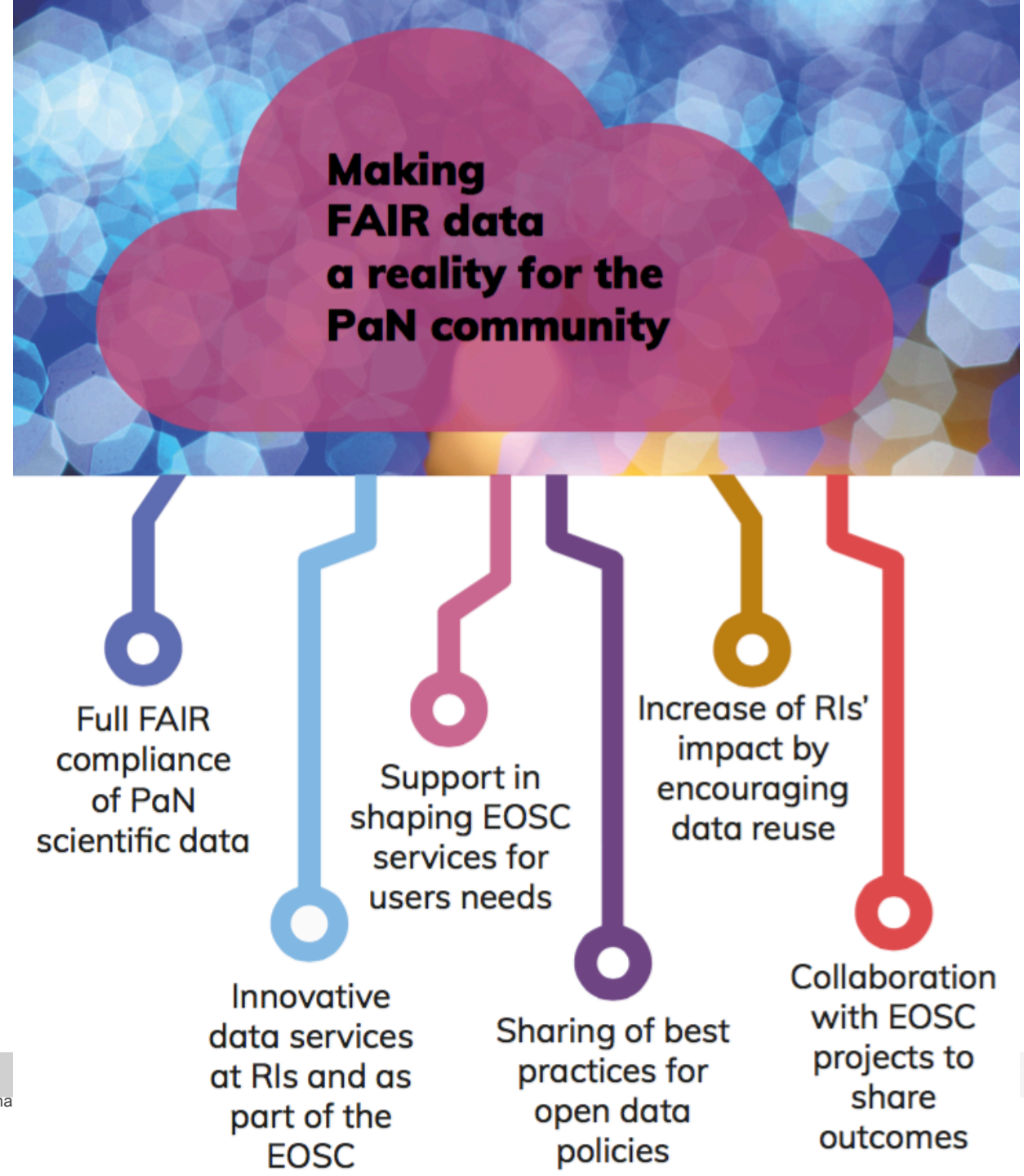

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

ESFRI European Strategy Forum on ResearchInfrastructures

panosc

This project has received funding from the European Union's Horizon 2020 research and innovation programme under

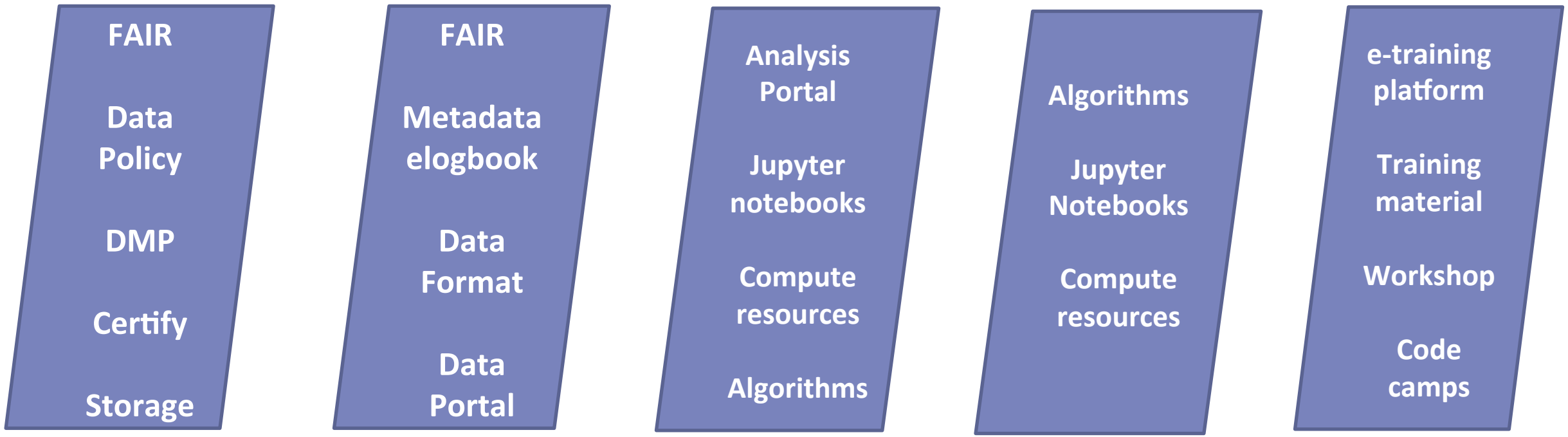
# PaNOSC goals




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.

This project ha

# PaNOSC = data policy + management + analysis + simulation + training




Data Policy | Metadata | Analysis | Simulation | Training

# Data management enables

## Publications

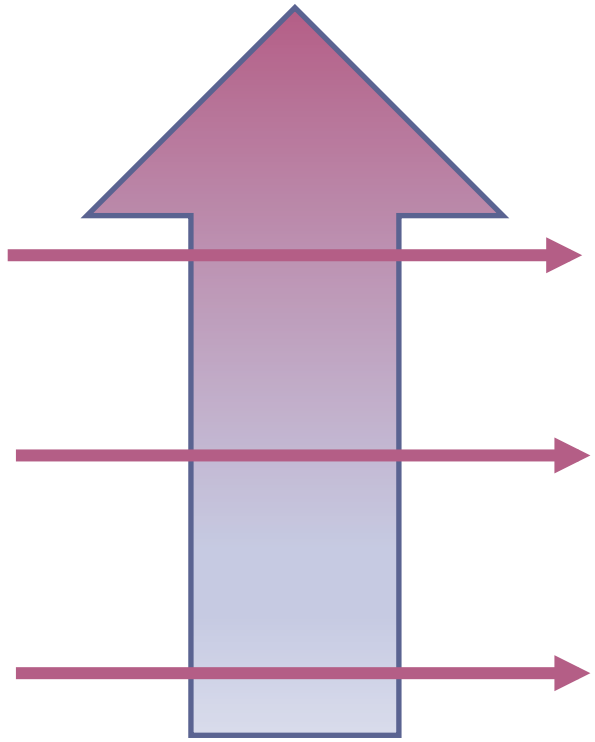
## Data analysis

## Artificial Intelligence

Federated catalogues

On-site data reduction

Data + metadata



Open Data
Analysed Data
Digital Objects

Reduced Data
Archived Data

Raw Data

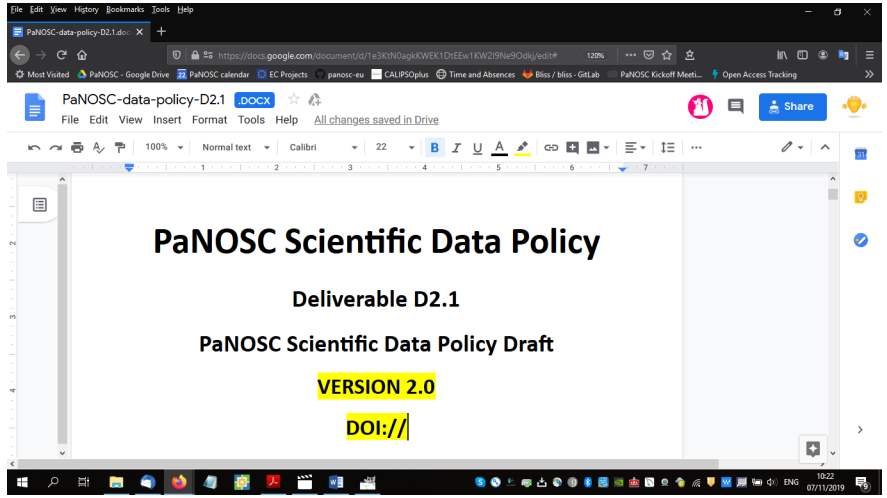
Machine Learning
Algorithms
Data Mining

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

# Data policy WP2

Update the PaNdata data policy

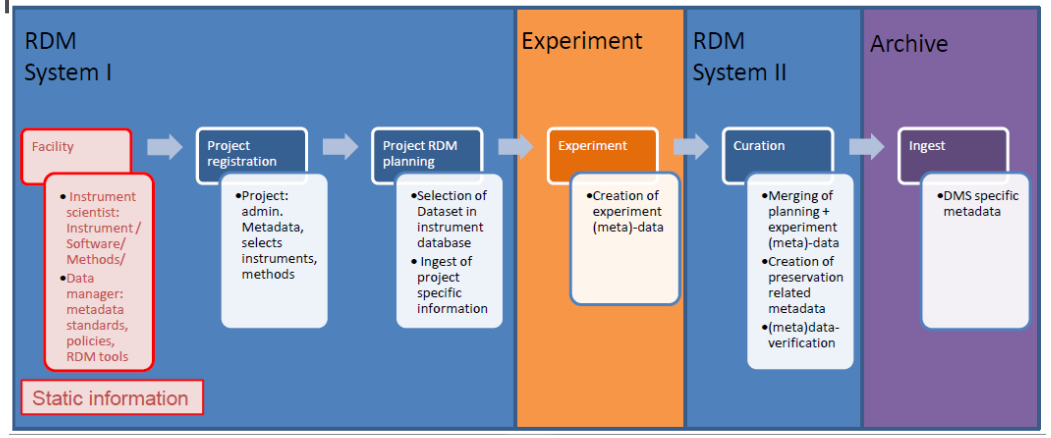
Include the FAIR concepts and make it FAIR compliant (52 criteria) ✓




→ Update existing Data Policies to be compatible with the Data Policy 2.0 framework

→ **Develop a tool for Data Management Plans together with ExPaNDS**

Share outcomes with




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

# Data catalogue WP3

Develop an Application Programmers Interface (API) for searching for FAIR data

Integrate search API into EOSC portal

Use Nexus/HDF5 for metadata

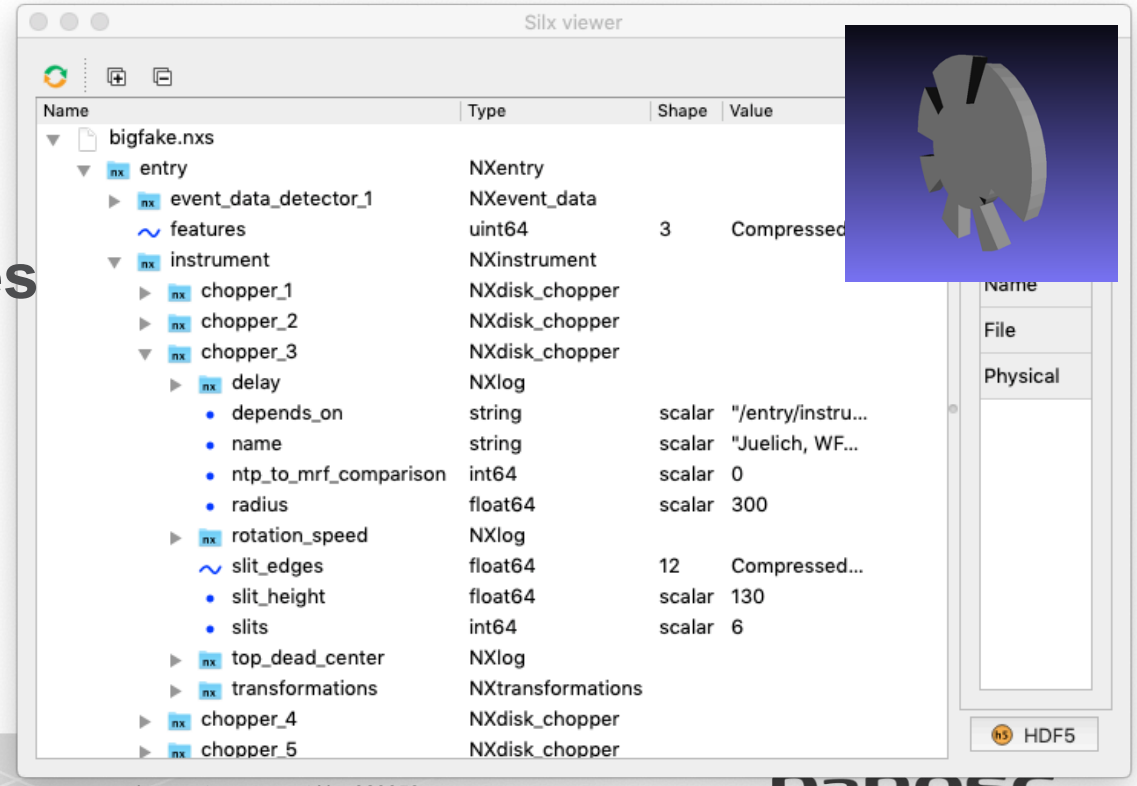
→ Automate metadata collection on beamlines

Use e-logbook to make data FAIRer ✓

→ **Long term storage (100s of Petabytes)** ✓

**Dataset** Show/Hide | List Operations | Expand Operations

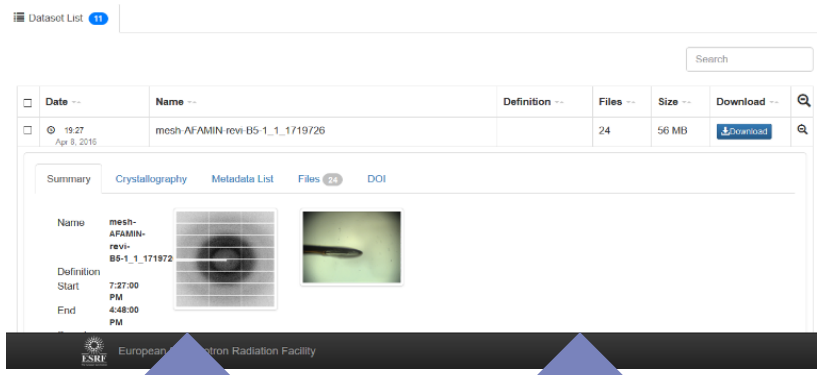
| | | |
|---|---|---|
| PATCH | /Datasets | Patch an existing model instance or insert a new one into the data source. |
| GET | /Datasets | Find all instances of the model matched by filter from the data source. |
| PUT | /Datasets | Replace an existing model instance or insert a new one into the data 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. |

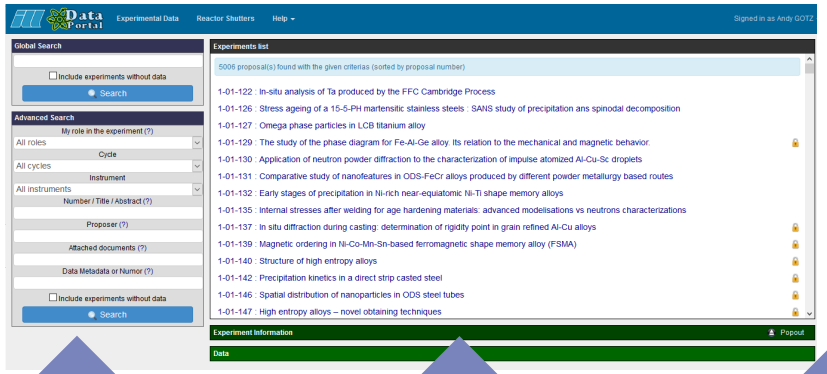


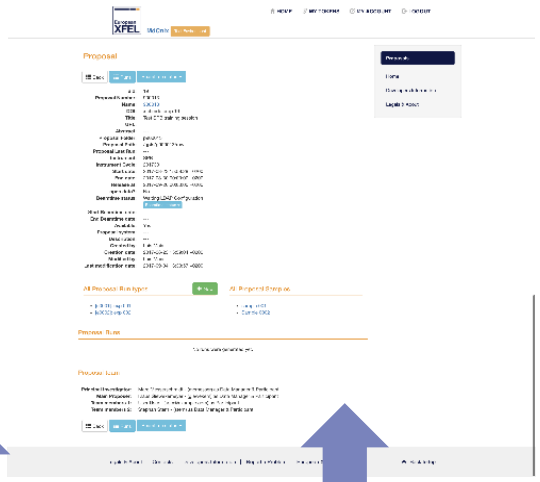

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

# PaNOSC has 6 data catalogues with different APIs + UIs

WP3



50 PB/yr

ESRF
(icat)

15 PB/yr

CERIC
(icat)

<1 PB/yr

ESS
(SciCat)

.6 PB/yr

ILL
(local)

10 PB/yr

ELI
(tbd)

100 PB/yr

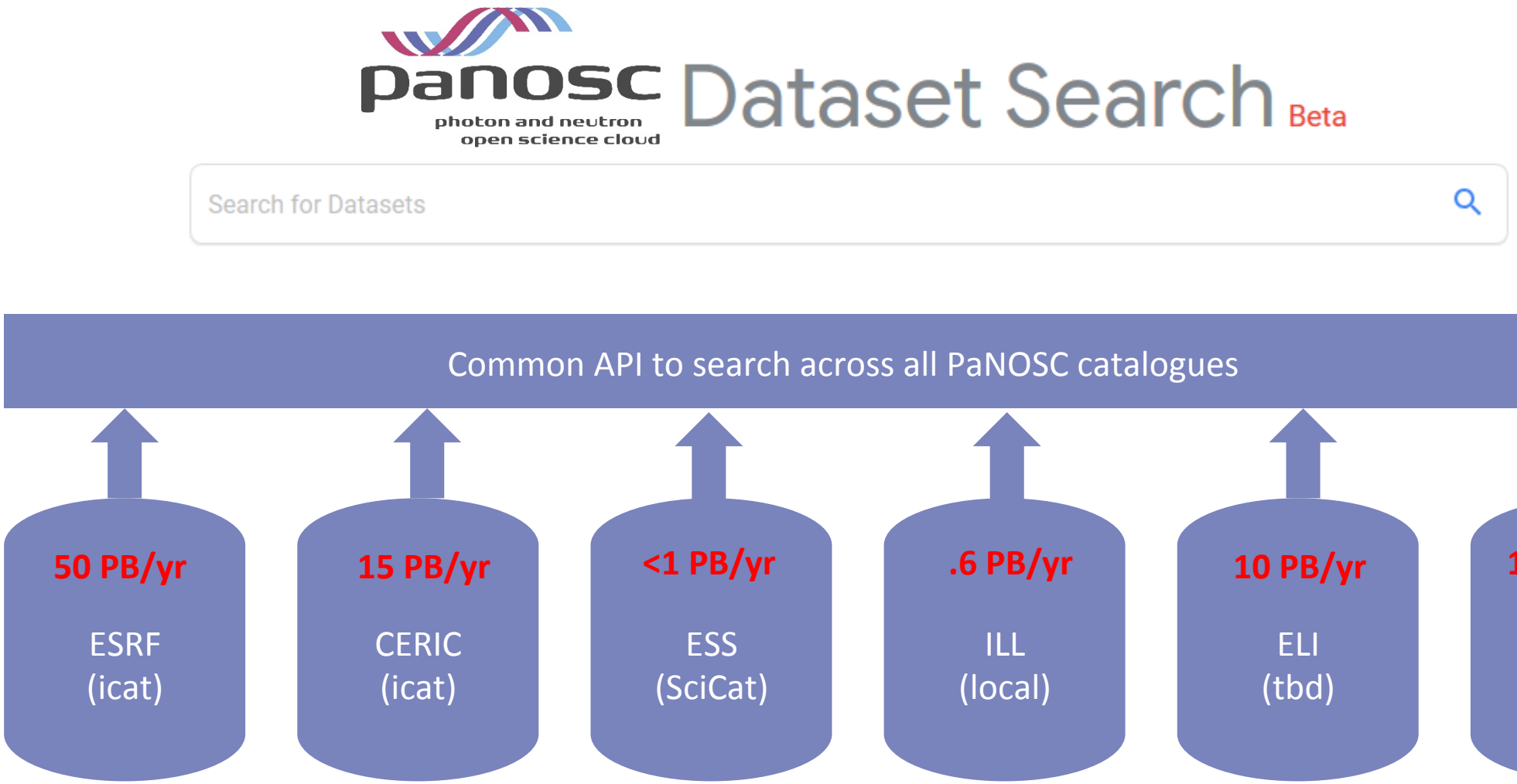
XFEL
(MyMdc)

https://data.esrf.fr

https://data.ill.fr

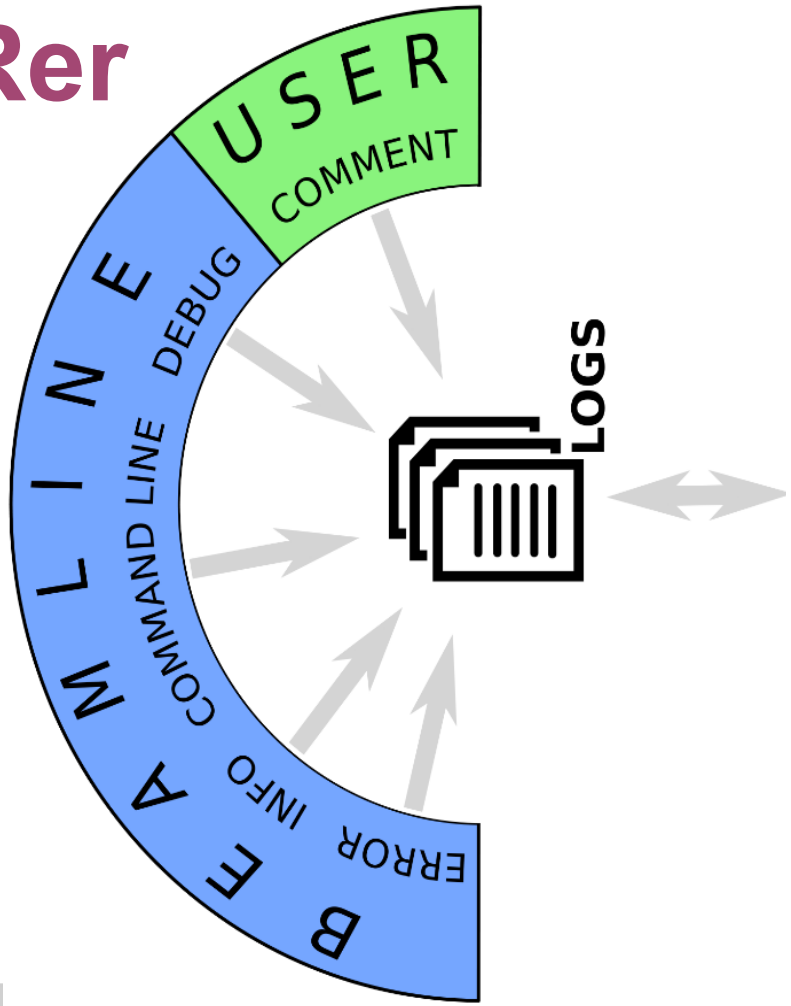
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

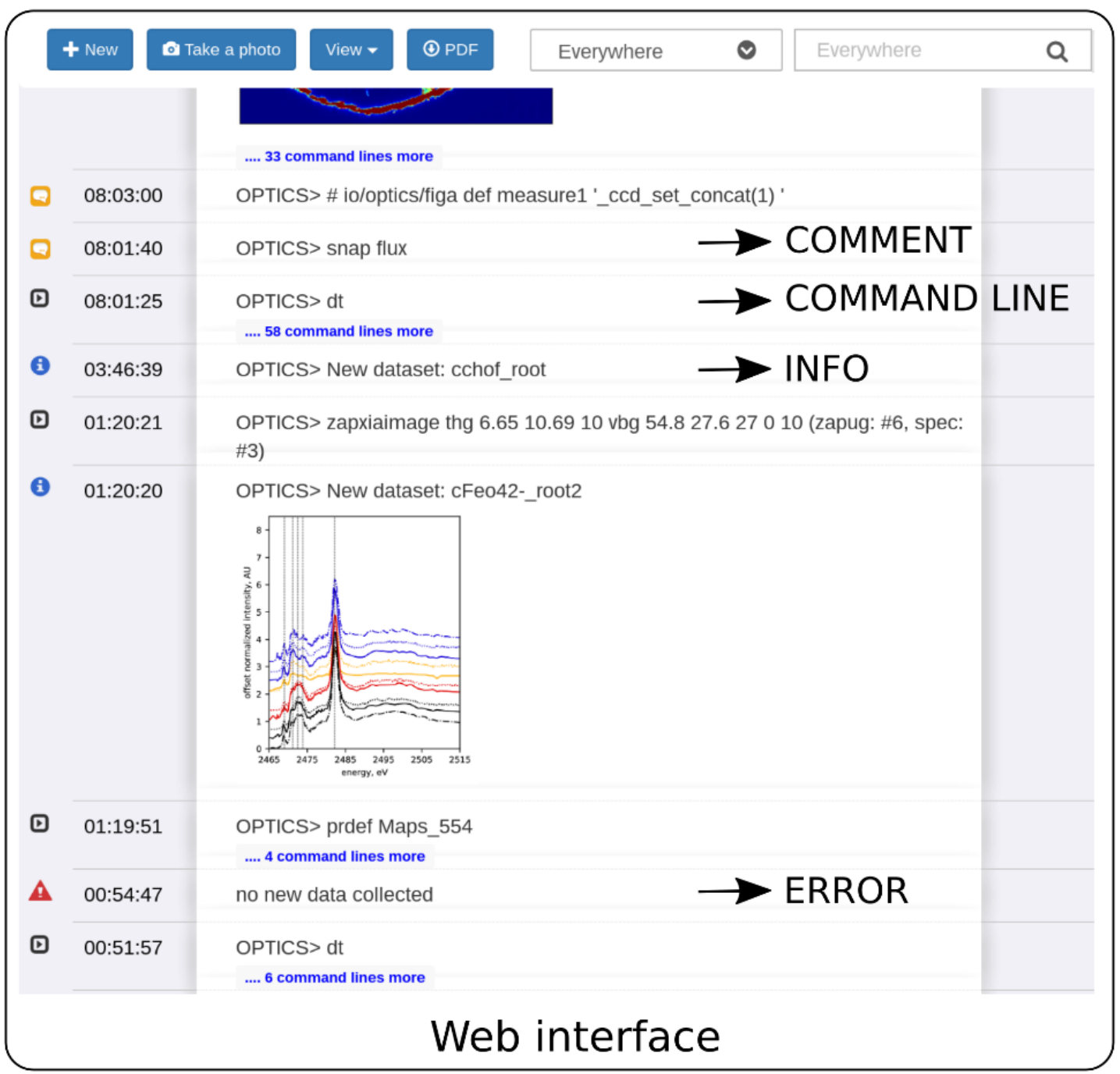
# PaNOSC common API across all sites




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

# E-logbook makes experiments FAIRer ✓



# Data Analysis Services WP4 ✓

- → **Data Analysis portal**
- Jupyter Python Notebooks
- Remote Desktop in Browser
- Remote data analysis portal
- HDF5 and visualisation in notebooks
- Package software for s/w catalogue


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

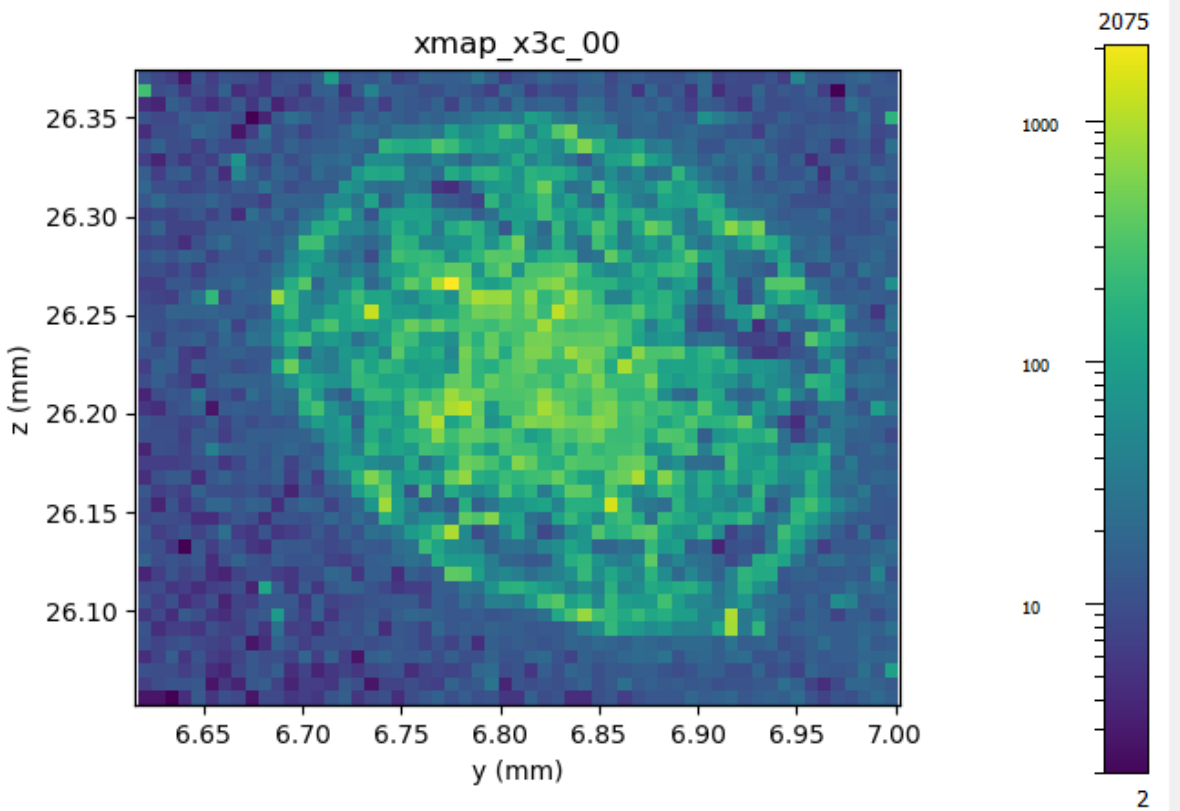
# Example data format (Nexus/HDF5)+ silx viewer

Silx viewer

File Options Views Help

| Name | Description | Type | Shape | Lin |
|---|---|---|---|---|
| id21.h5 | | NXroot | | |
| fe3prior3_sr_coarse.zapimage.#1 | | NXentry | | |
| application:xrf | "NXxrf" | NXsubentry | | |
| definition | "NXxrf" | string | scalar | |
| end_time | "2018-12-15... | string | scalar | |
| i0 | 2D data | float32 | 46 × 57 | So |
| it | 2D data | float32 | 46 × 57 | So |
| mca00 | | | | |
| end_time | "2018-12-15... | string | scalar | |
| instrument | "ESRF ID21: ... | NXinstrument | | |
| mca00 | | NXdetector | | |
| data | 3D data | uint32 | 46 × 57 × 2048 | |
| monochromator | | NXmonochromator | | |
| energy | 7.21974 | float64 | scalar | |
| name | "ESRF ID21: ... | string | scalar | |
| positioners | | NXcollection | | |
| measurement | | NXcollection | | |
| plot:default | | NXdata | | |
| start_time | "2018-12-15... | string | scalar | |

id21.h5::/



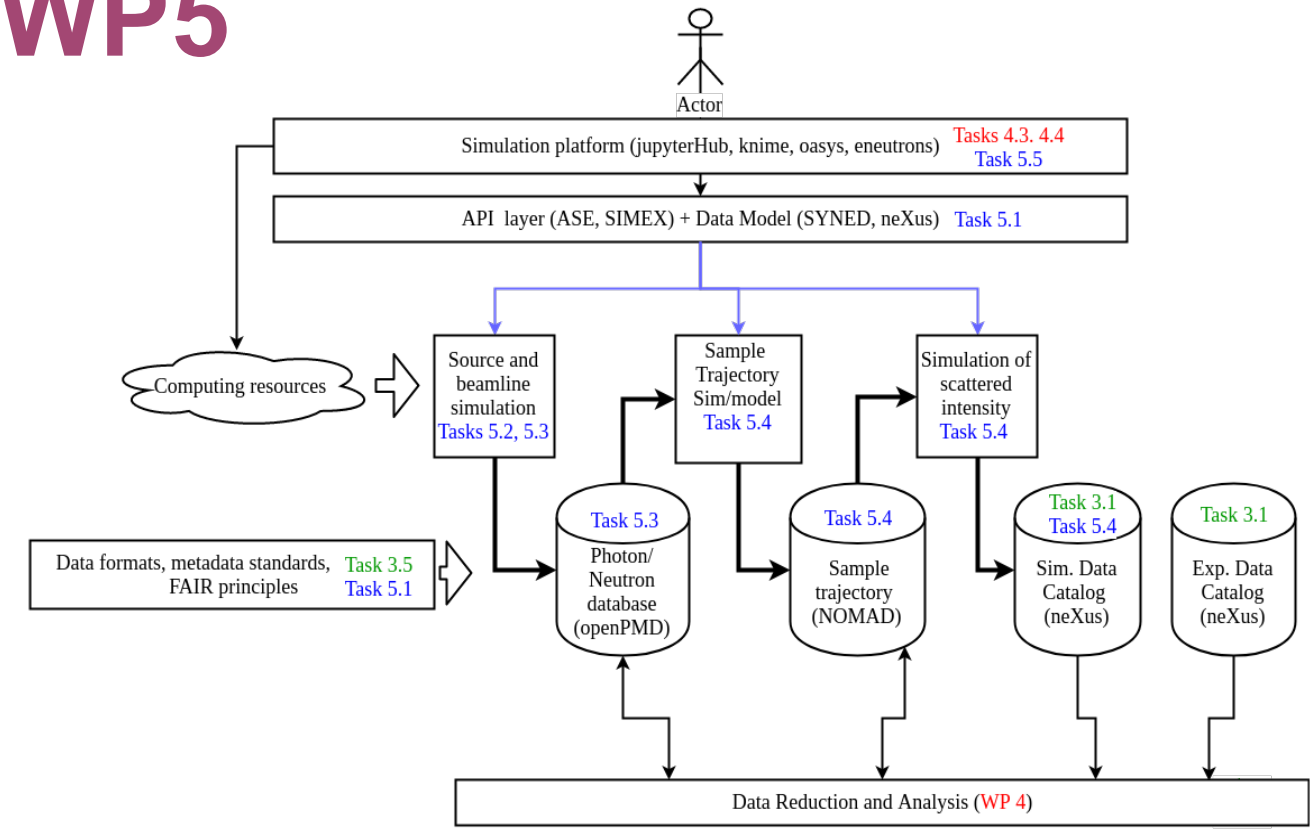

X: 6.614978 Y: 26.28121 Data: - Dims: 57x46

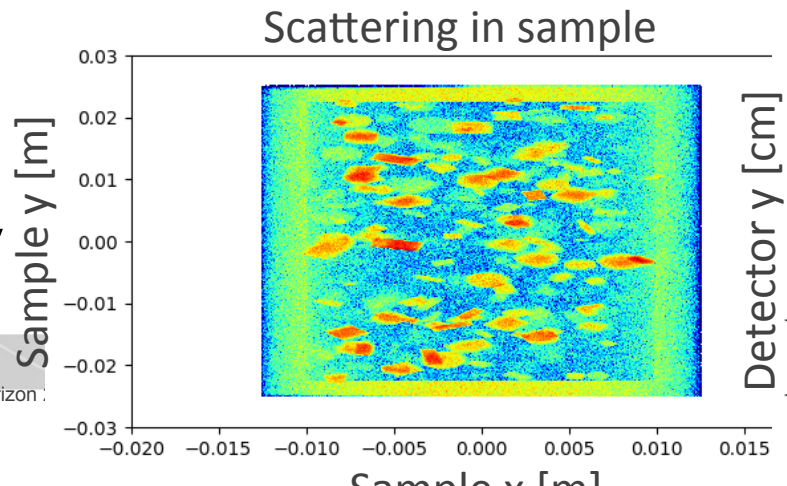
0 limits: 0, 28

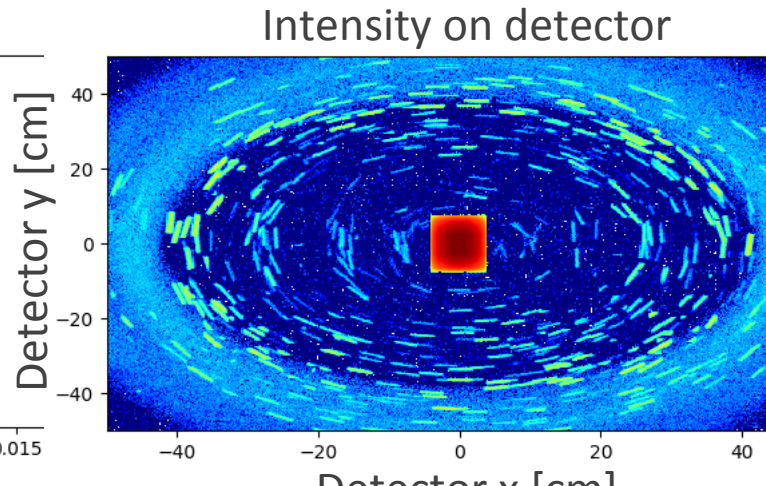
HDF5 NXdata

# Simulation services WP5

- **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




This project has received funding from the European Union's Horizon

# EOSC Integration WP6 ✓

→ • **AAI = Authentication and Authorisation Infrastructure**

  - Critical if we want to identify people and define roles
  - EOSC must solve at least the AAI problem
  - PaNOSC is working with GÈANT to make **UmbrellaId** sustainable

• **Develop a PaNOSC data analysis as a service portal**

→ • **Integrate PaNOSC services into EOSC**

→ • **Data download service (terabytes via internet)**

• **Software catalog service**

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

# 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 with FTS3**)

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**, and simple solution without caching (**webdav**)

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

# Sustainability WP7 ✓

- Develop a model to calculate cost of FAIR data
- Propose a sustainability plan for RIs
- Develop models for different service levels
- Participate and contribute to EOSC Sustainability WG
- Cost-benefit analysis of EOSC for PaN RIs

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

# Training + e-learning WP8 ✓

McStas

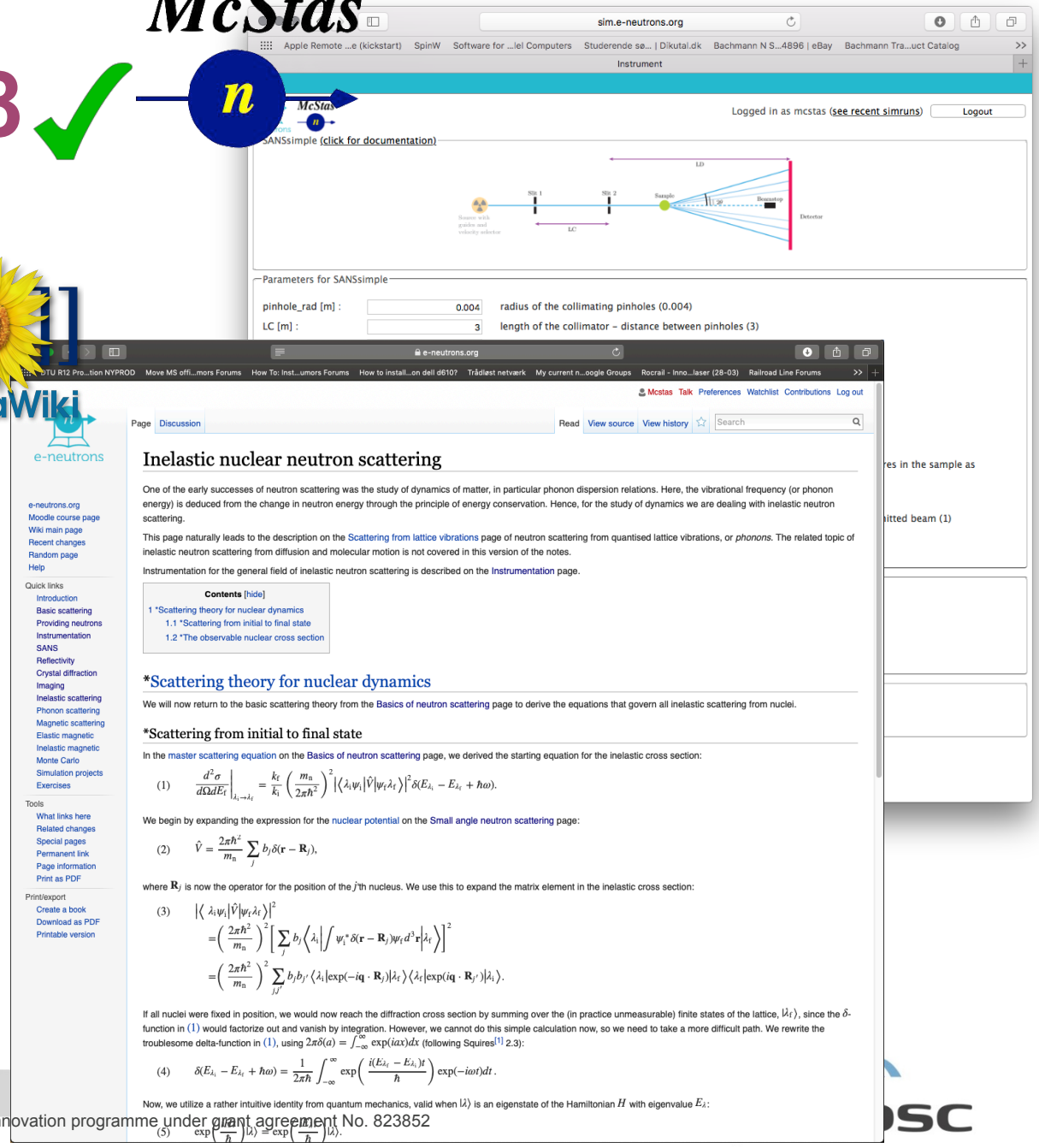
→ **e-neutrons.org**

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

MediaWiki

## Extensions:

- **Support for Jupyter Notebooks**
- **Integrate data analysis services**
- **Integrate simulation services**
- **What is the EOSC training platform?**




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


| | |
|---|---|
| Introduction to Machine Learning and Deep Neural Networks for scattering science | Dr. Bill Triggs<br>14:00 - 14:50 |
| Improving data analysis using artificial intelligence: an ESRF perspective | Dr. Vincent Favre-Nicolin<br>14:50 - 15:15 |
| Machine learning applications for Small Angle X-ray Scattering data collection and analysis at EMBL-Hamburg | Dr. Daniel Franke |
| rnoon 1 · Afternoon 2 · Coffee Break · Wine & Cheese buffet a... | |
| *ESRF entrance hall & mezzanine* | 15:40 - 16:00 |
| DOE's Center for Advanced Mathematics for Energy Research Applications (CAMERA): Artificial Intelligence, Machine Learning, and Experimental Facilities: Present and Future | Dr. Jamie Sethian<br>16:00 - 16:50 |
| Machine Learning at ILL | Dr. Paolo Mutti<br>16:50 - 17:15 |
| Machine le[illegible] | [illegible] Carlos Oscar Sorzano<br>17:15 - 17:40 |
| Machine learning and artificial intelligence in MX | Dr. Melanie Vollmar<br>17:40 - 18:05 |
| Convolutional neural networks for DESY photon science | Dr. Philipp Heuser<br>18:05 - 18:30 |

**AVAILABLE here →** https://www.youtube.com/user/LightforScience/videos



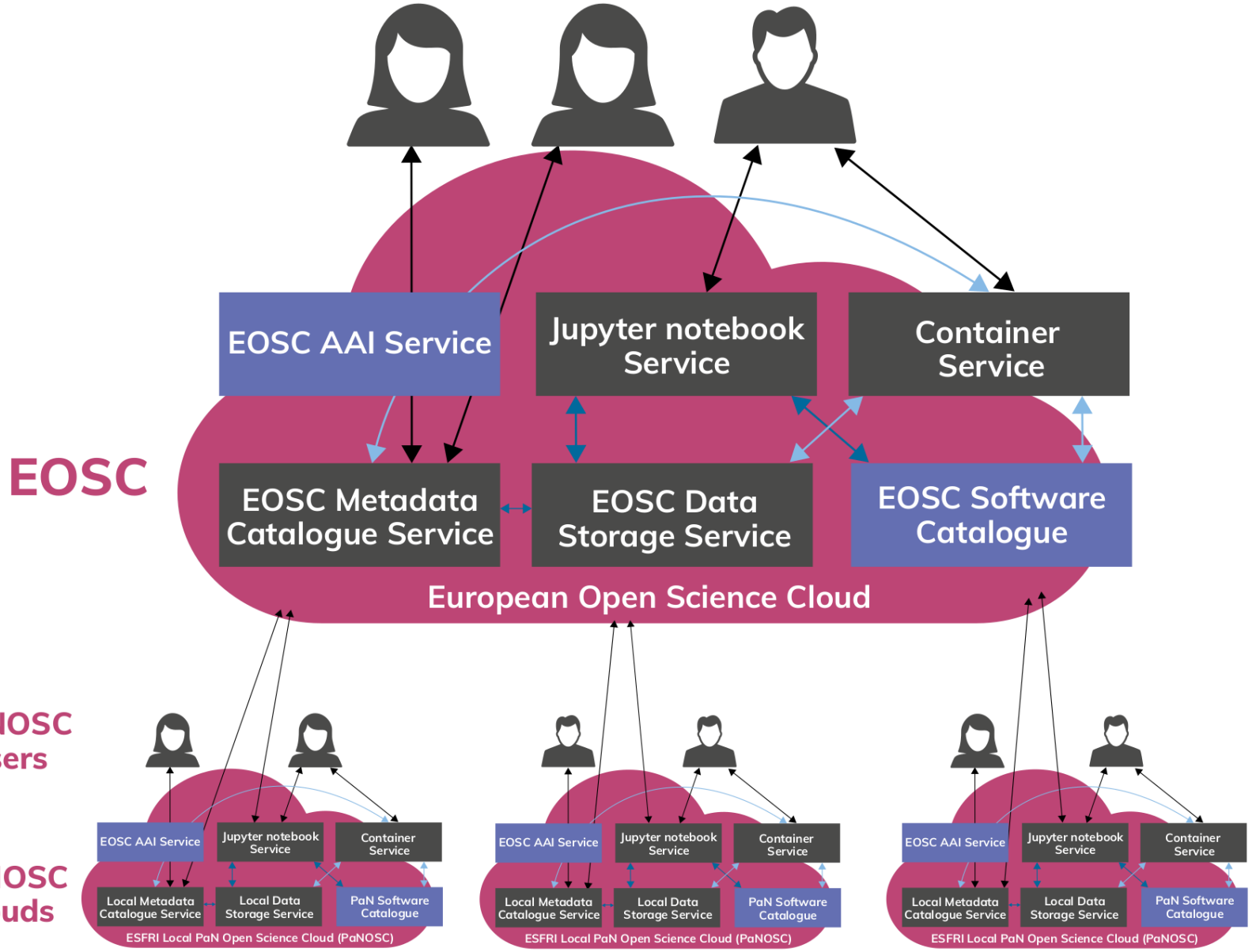
panosc

# HDF5 European Workshop for Science and Industry

17-18 September 2019
ESRF (Grenoble)
Europe/Paris timezone

- Overview
- Timetable
- Call for Abstracts
  - View my Abstracts
  - Submit Abstract
- Registration
  - Modify my Registration
- Contribution List
- Author List
- My Conference
  - My Contributions
- Book of Abstracts
- Practical information

# PaNOSC Vision – a PaN scientific commons




ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

cnrs

# **Minimal Viable EOSC** - *PaNOSC Position Paper (22/11/2019)*

- ✓ 1. A common way of identifying, authenticating, and authorising users (**AAI**) across Europe. The EOSC should operate and sustain AAI as part of the EOSC infrastructure. The EOSC AAI should support the AAI features PaNOSC is implementing on the UmbrellaId AAI [6].
- ✓ 2. A service for **transferring and downloading data efficiently** (distributed and high bandwidth);
- ✓ 3. A solution for **long-term archiving of large quantities of open data** (petabytes) coupled to high-performance storage and compute resources for the (re)analysis of open data;
- ✓ 4. A **federated search capability** for searching and finding scientific data in a wide variety of domains;
- ✓ 5. A set of **services for data simulation and analysis** ranging from generic services like Jupyter notebooks to domain specific applications per scientific application in the PaN software catalogue [5].

# PaNOSC + ExPaNDS contribution to EOSC

1. **Petabytes** of raw and processed data in a wide variety of scientific domains

2. **Meta-data** that will create **FAIR** raw and processed scientific data

3. **Software** for generic and specific data simulation and data analysis

4. **Workflows and expertise** for reducing and analysing data

5. **Reference training material** and **training platform** for understanding photon and neutron science and associated handling of data

6. Interface to large **user communities** of photon and neutron sources and their expectations for services

# Why we make data open ?

# on http://paleo.esrf.fr


## Synchrotron tomography

Porcier S. M., Berruyer C., Pasqali S., Ikram S., Berthet D., Tafforeau P. « **Wild crocodiles hunted to make mummies in Roman Egypt: Evidence from synchrotron imaging** ». Journal of Archaeological Science, 1 octobre 2019. Vol. 110, p. 105009. DOI : https://doi.org/10.1016/j.jas.2019.105009

: agreement No. 823852

panosc

# Questions de Laurent Lellouch

5. Avez-vous des attente vis-à-vis la France comme pays hôte d'ESRF? (e.g. des connections internet à très haut débit, l'accès à des centres de calcul, etc.)
   OUI – connections tres haut debit vers des centres de calcul / cloud
6. Quels sont les coûts financiers et humains associés à la mise en place de PaNOSC au niveau de l'ESRF?
   4-5 CDI + 3 COD + 1 million euros (archivage long terme)
7. Avez-vous une idée des coûts qui seront associés au maintien de la contribution de l'ESRF à PaNOSC?
   3 CDI + 100 mille euros / ans (archivage long terme)
8. Même question pour l'intégration de PaNOSC dans EOSC et l'ouverture à des communautés au delà de celle des utilisateurs d'infrastructures PaN? En particulier, est-ce que ces infrastructures ont actuellement les moyens de proposer les services à des communautés plus grandes?
   NON

# Conclusion

1. PaNOSC vision is to create a Scientific Data Commons for Photon and Neutron sources and make the data available via the EOSC (whatever that may be)

2. PaNOSC will collaborate closely with ExPaNDS to make FAIR data a reality for all PaN RIs

3. PaNOSC has been running for 1 year and is now up to speed however the implementation of EOSC is still not defined

4. PaNOSC et CNRS doit travailler ensemble sur des sujets comme l'archivage long terme, politique des données, API de recherche des donnée, logiciels scientifiques open source, et aider les scientifiques à partager et reutiliser les données



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

# Vision

The EOSC could become the **GitHub of Open Science in Europe**. This means making it a platform for scientists to share their data analysis and workflows and link these back to open data and other workflows – either their own, or that of other scientists. To achieve this, it will be necessary to provide scientists with a personal space where they can create content (data analysis recipes, workflows, publications), store analysed data and share their work with collaborators via a versioning system like git.

**Example of a different approach to building a platform for Open Science à la EOSC → https://cos.io**

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

# PaNOSC Resources

- https://panosc.eu
- https://github.com/panosc-eu/panosc
- "Enabling Open Science for Photon and Neutron sources" presented at ICALEPCS 2019 http://icalepcs2019.vrws.de/papers/tubpl02.pdf
- PaNOSC Position Paper - https://github.com/panosc-eu/panosc/blob/master/Work%20Packages/WP6%20EOSC%20Integration/PaNOSC_EOSC_position_paper.pdf

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

# BACKUP SLIDES



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

panosc

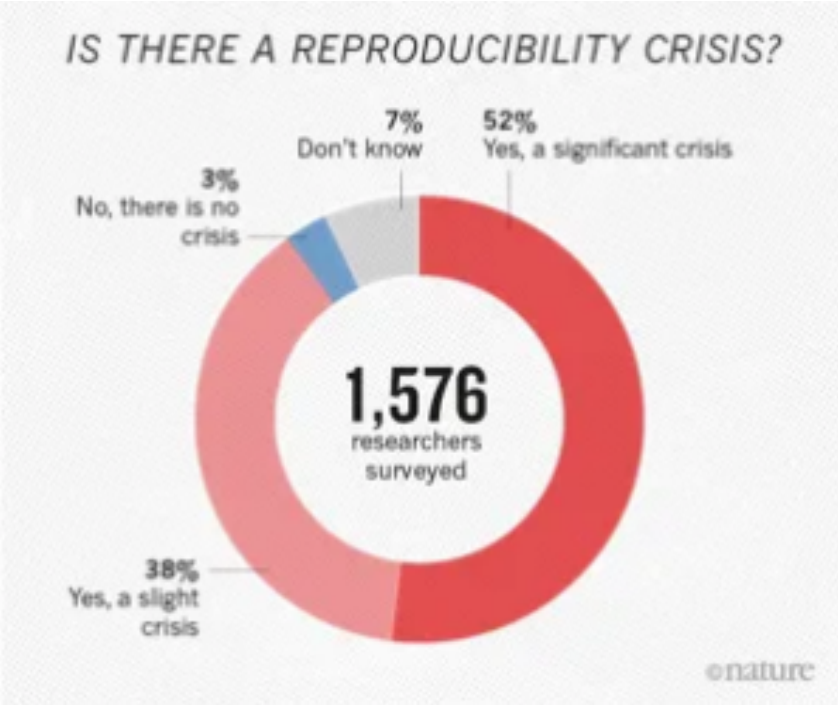
# Reproducibility crisis in science

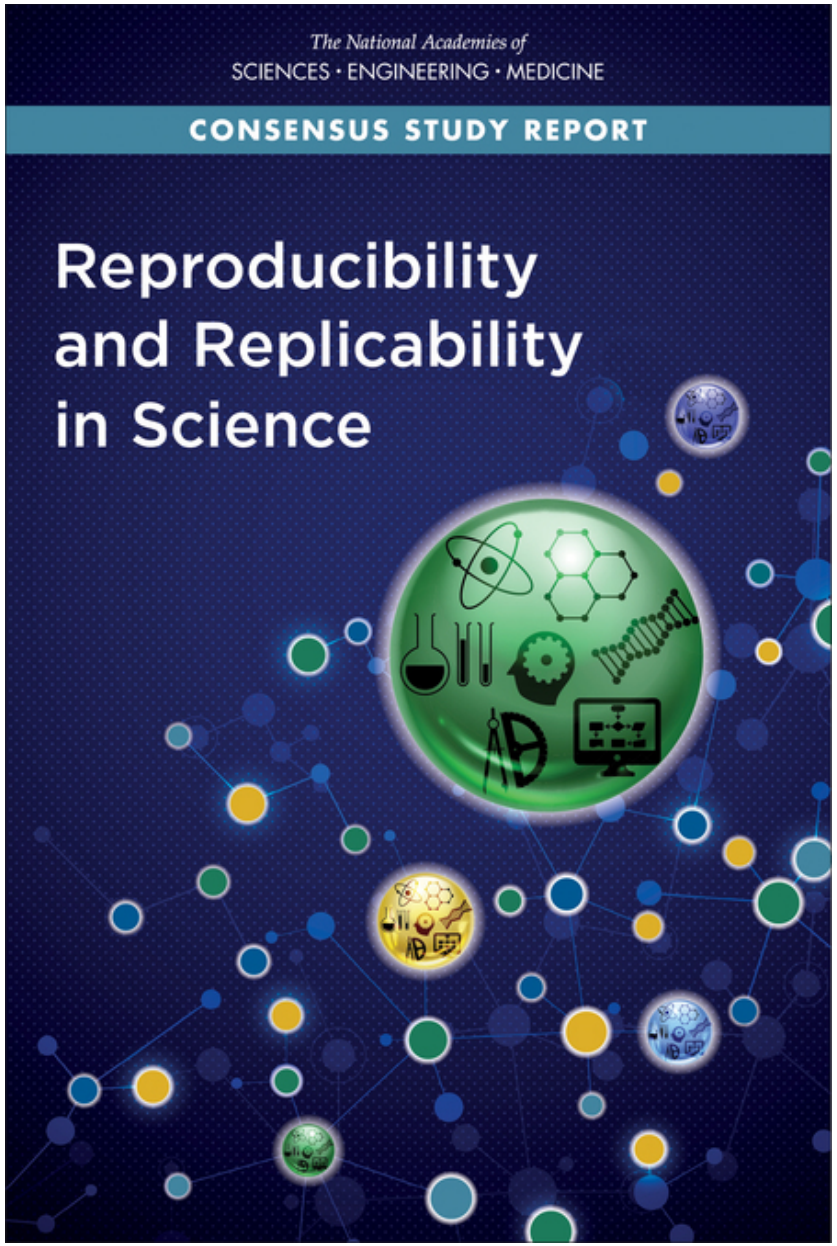
NEWS FEATURE
25 MAY 2016
Nature

## 1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

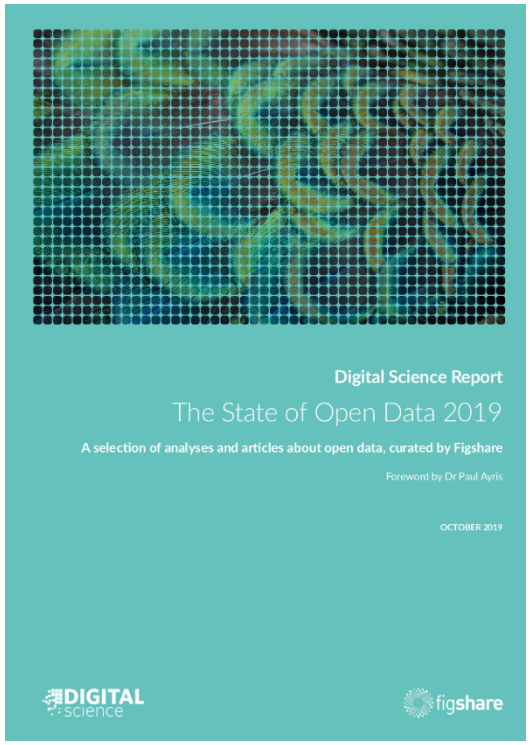
Monya Baker




Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533,** 452–454 (2016) doi:10.1038/533452a

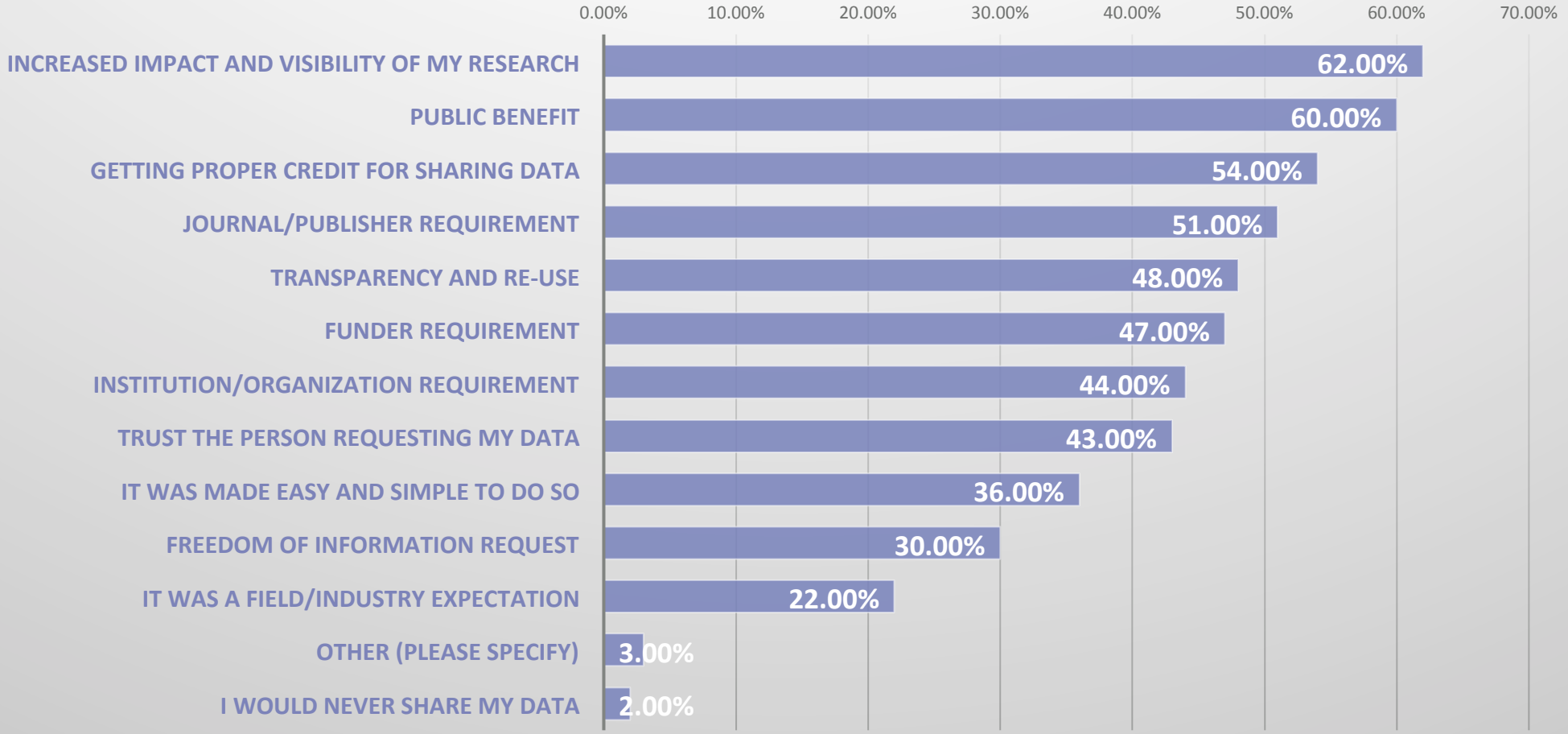
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852




# State of Open Data 2019

## Which circumstances would motivate you to share your data?

*The State of Open Data 2019 report*




https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2019/9980783

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

# What is the European Open Science Cloud ?

**“The main goal of the EOSC initiative was to offer European researchers a virtual environment with free, open, and seamless services for the storage, management, analysis and re-use of research publications, data and software that are linked to their research activities across borders and disciplines.**

The model proposed for realising EOSC was to **federate existing and newly developed research data infrastructures under a common governance structure**, assist the **shared procurement** of additional required capacity from public and private service providers, and **support the development of added-value services** for the exploitation of research data. “

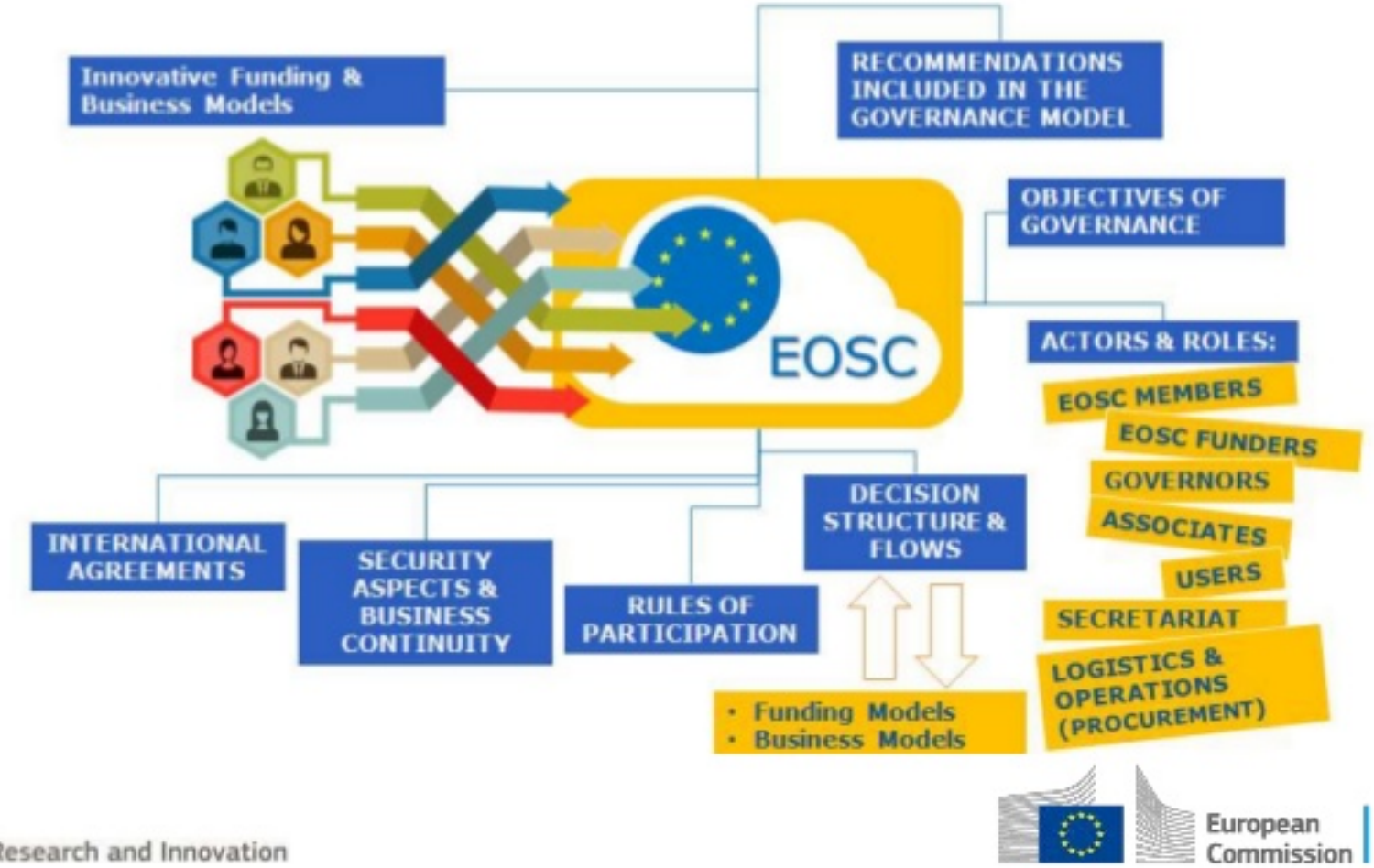
*EOSC Partnership Proposal – 16 December 2019*



panosc

# EOSC

## European Open Science Cloud (EOSC)



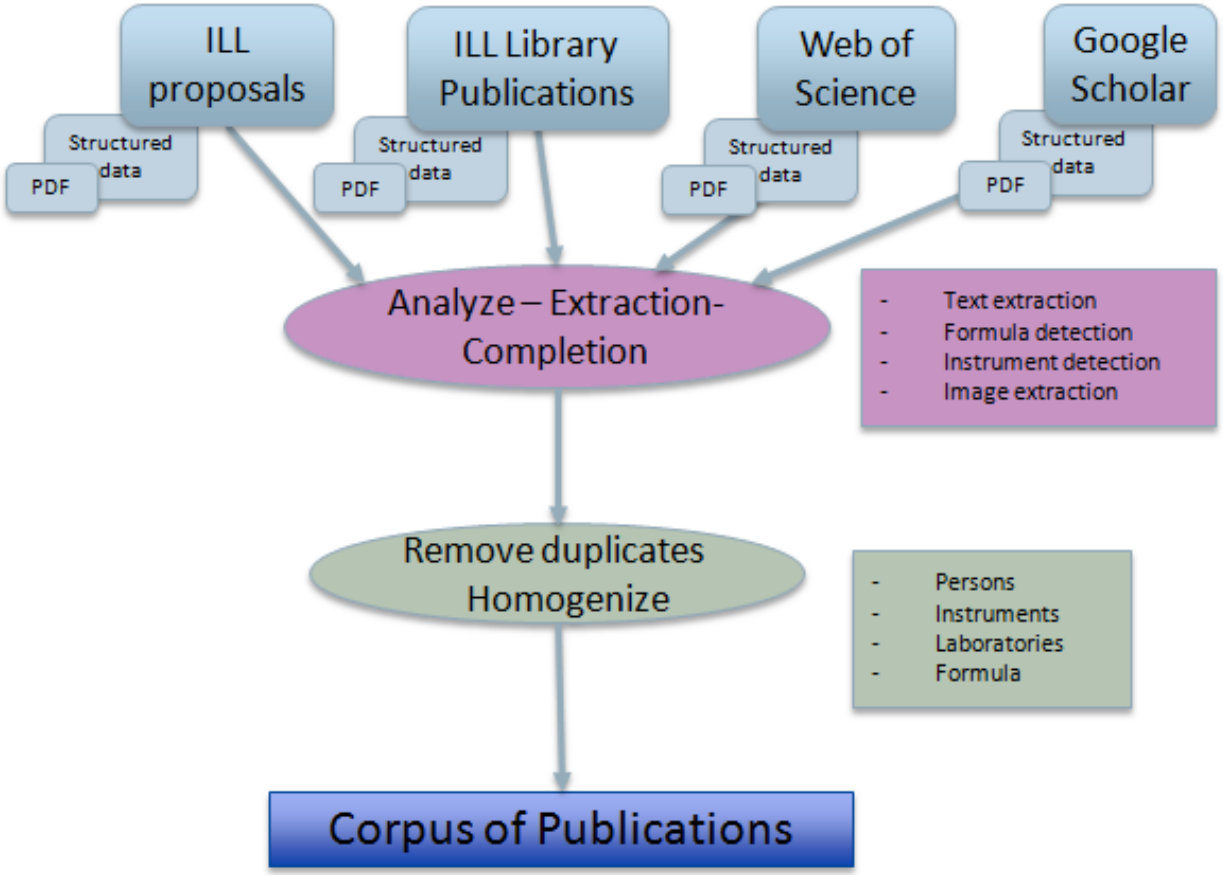

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

# **Feedback to the Executive Board** (PanOSC Position Paper)

1. **Provide** a clear technical definition of **"what is the EOSC"**
2. **Define** common **standards** for **FAIR data** so that the different scientific fields have a common approach and understanding
3. **Provide** **long-term sustainable plan** for how the EOSC will be maintained and financed
4. **Provide** **cloud resources** for running data analysis workflows and simulations, ideally enough to make a significant difference for users needing access to computing resources beyond what can be offered by the PaNOSC partners
5. **Collaborate** with **publishers** to generalise the requirement for citing data in publications and **making open data a publication** in its own right
6. **Provide** **documentation** and training material on the EOSC
7. **Do** a **cost-benefit analysis** of what the EOSC provides

# 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**
- **Integrate in OpenAIRE ?** ✓




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