

PaNOSC project Introduction + Overview

16 June 2020

Andy Götz, on behalf of PaNOSC project



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

Outline

- Motivation → Why
- Overall Concept → What
- Current Status → When
- PaN Commons Vision → Future
- Impact of COVID-19 → Dry Run

PaNOSC & EOSC : built on Boldness

“

Whatever you can do, or
dream you can do, begin it.
Boldness has genius,
power, and magic in it.
BEGIN IT NOW.

- Goethe



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 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

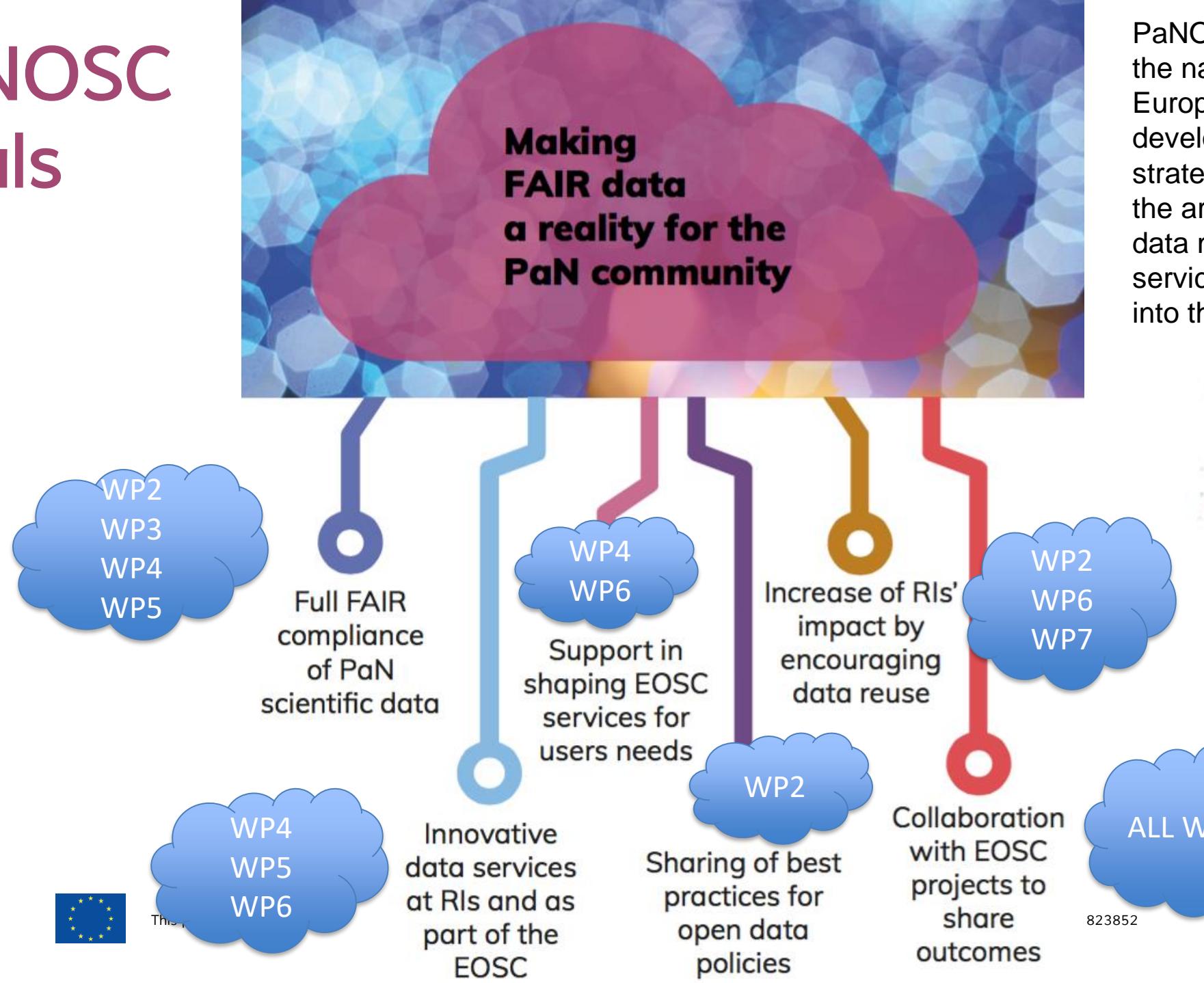
Started: 1/12/2018

Duration: 4 years



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

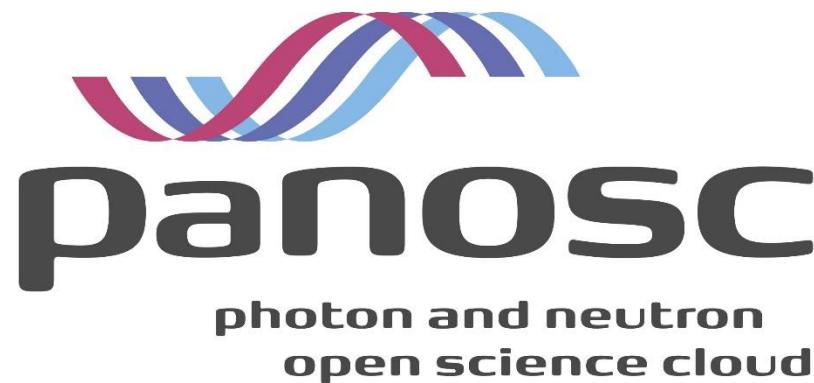
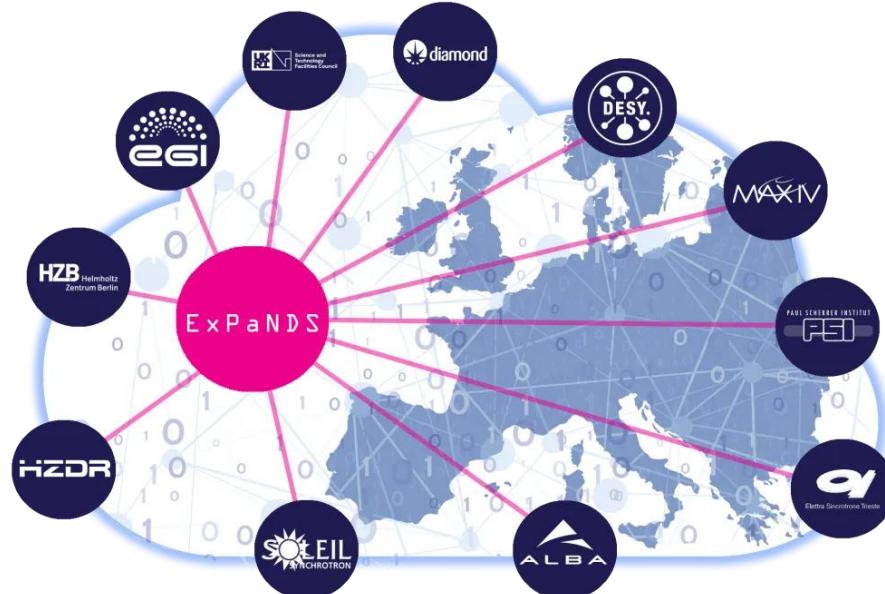
PaNOSC goals



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



ExPaNDS – Extending Photon and Neutron Data Services



- Together PaNOSC+ExPaNDS represent all photon and neutron sources in Europe
- PaNOSC and ExPaNDS are collaborating closely in all common Work Packages i.e. WP1, WP2, WP3, WP4, WP8, WP9



COVID-19 : A Dry Run for PaNOSC + EOSC

Definition of 'dry run'

dry run

Collins COBUILD



Word forms: plural **dry runs**

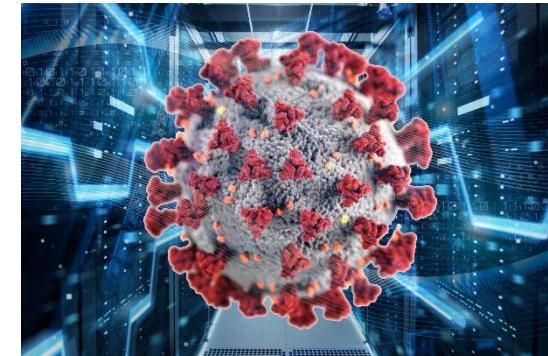
COUNTABLE NOUN

If you have a **dry run**, you practise something to make sure that you are ready to do it properly.

The competition is planned as a dry run for the World Cup finals. [+]

for

COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers



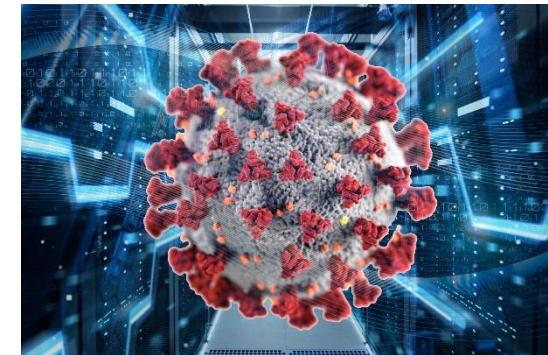
How ready were PaNOSC + EOSC for the COVID-19 pandemic?



COVID-19 : a test for PaNOSC and EOSC

- PaNOSC + EOSC should offer:

1. FAIR data
2. Downloadable Metadata & Raw data
3. Software service to browse and analyse raw data
4. Platform as a service to do computations + simulations
5. Common space to share progress and workflows



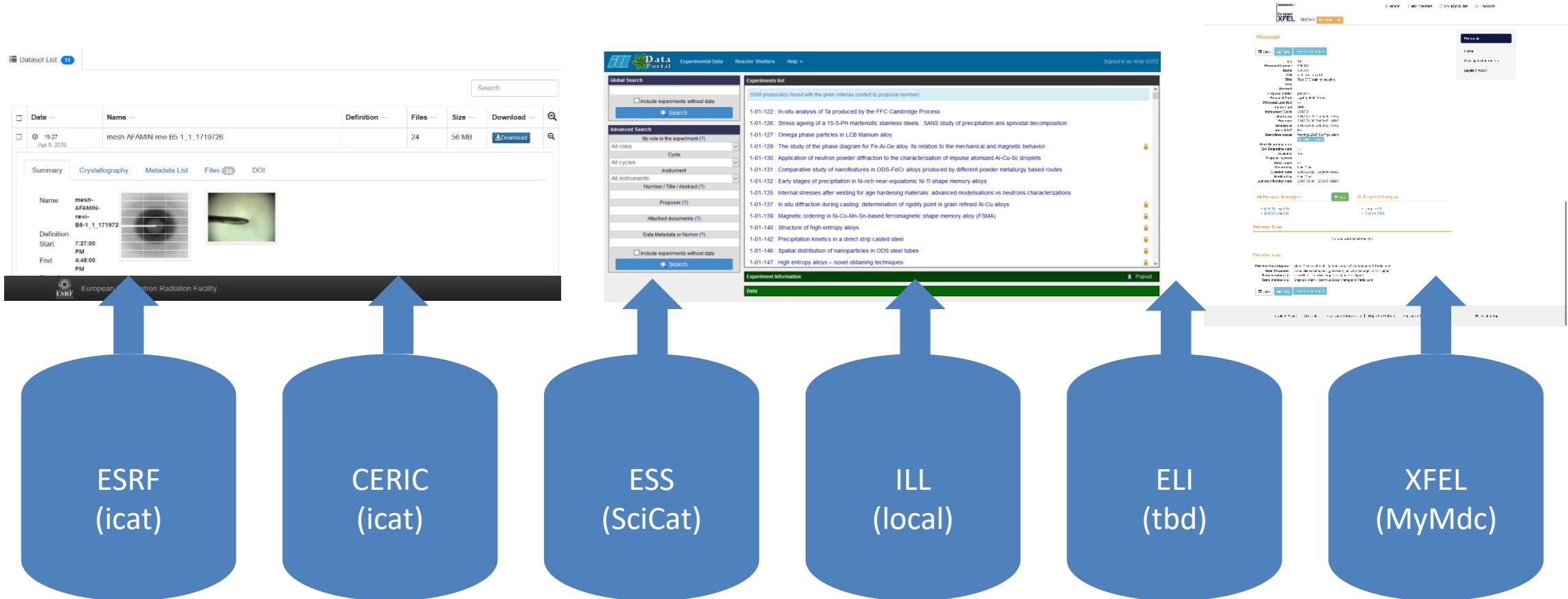
PaNOSC KPIs

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data/year 2018	0.2 PB	8 PB	1 PB	3PB	< 1 PB	0
Data/year 2023	0.6 PB	50 PB	15 PB	100 PB	10 PB	< 1 PB
Data Policy 2018	2011	2010	2014(3/8)	2017	in prog	2017
Data Policy 2023	FAIR	FAIR	FAIR	FAIR	FAIR	FAIR
Metadata catalogue 2018	Local	Icat	Local	myMdC	No	SciCat
Metadata catalogue 2023	Local	Icat	Local	myMdC	[TBD]	SciCat
Metadata definition 2018	Nexus	Nexus	custom	myMdC	?	Nexus
Metadata definition 2023	Nexus	Nexus	Nexus	Nexus	[Nexus]	Nexus
DOI 2018	yes	yes	no	yes	no	yes
DOI 2023	yes	yes	yes	yes	yes	yes



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}



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

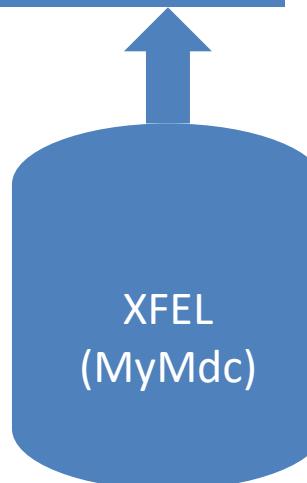
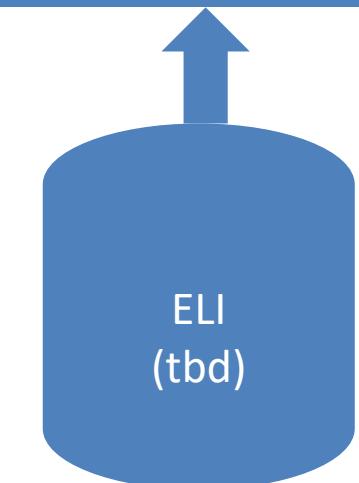
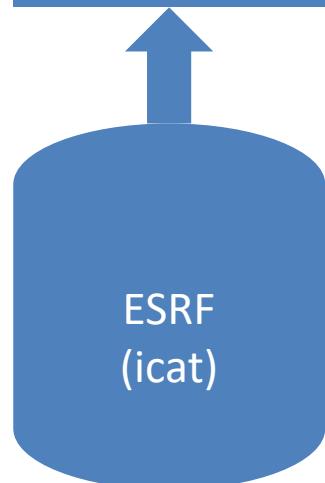
PaNOSC common API across all sites



Search for Datasets



Common API to search across all PaNOSC catalogues



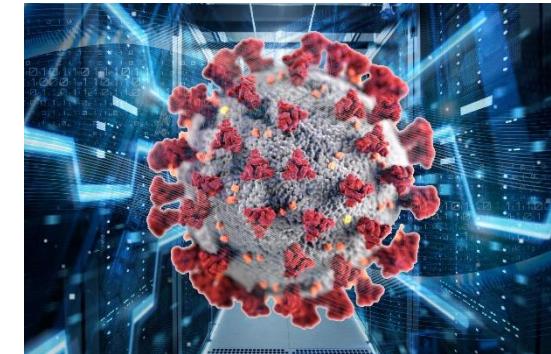
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852



COVID-19 : a test for PaNOSC and EOSC

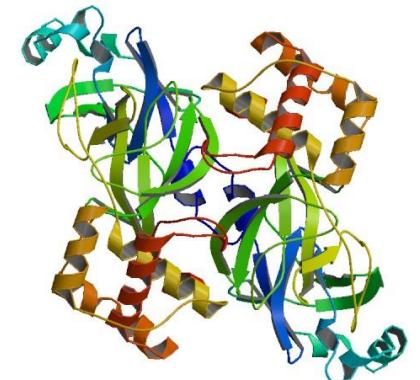
- FAIR data & Downloadable
Metadata & Raw data : In Progress

→ e.g. ESRF has 15 Coronavirus structures in the Protein Data Bank
but no raw data ...



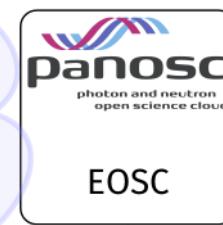
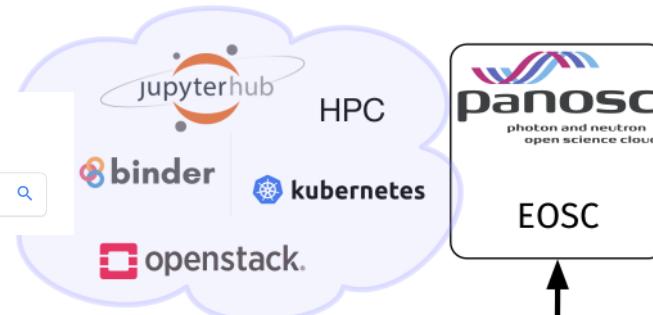
SARS-CoV/2c3s, SARS-CoV/2beq, SARS-CoV/2bez,
SARS-CoV/2h85, bovine-coronavirus/3cl4, bovine-
coronavirus/3cl5, HCoV-OC43/6qfy, MERS-CoV/4ud1,
MERS-CoV/6g13, murine_coronavirus/4c7l,
murine_coronavirus/4c7w, murine_coronavirus/5jif,
murine_coronavirus/5jil, 2cme, SARS-CoV/2xyr, SARS-CoV/2xyv,
SARS-CoV/2fav, SARS-CoV/1qz8, SARS-CoV/1uw7

SARS-CoV/2c3s

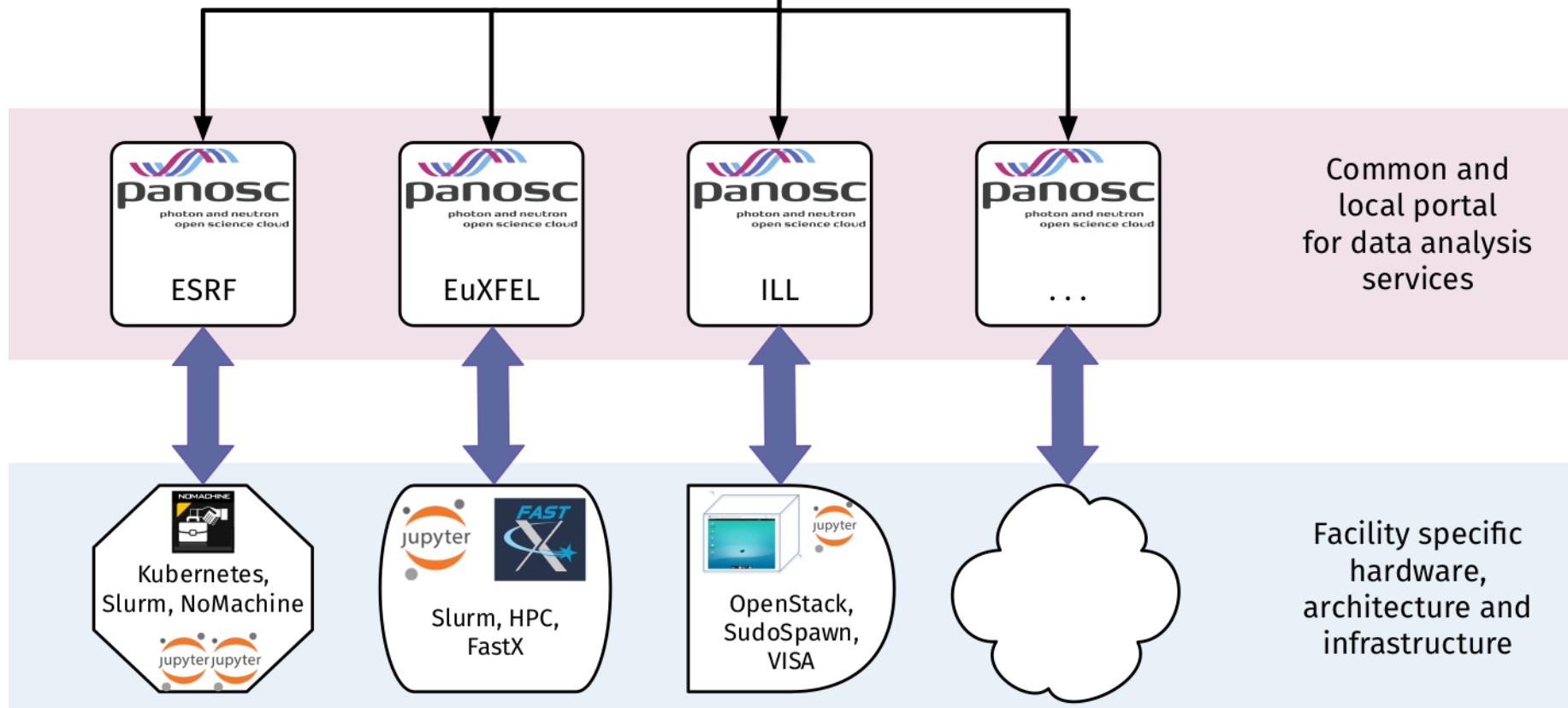




Search for Datasets



Meta portal
for data analysis
services



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



Simulation services - ViNYL

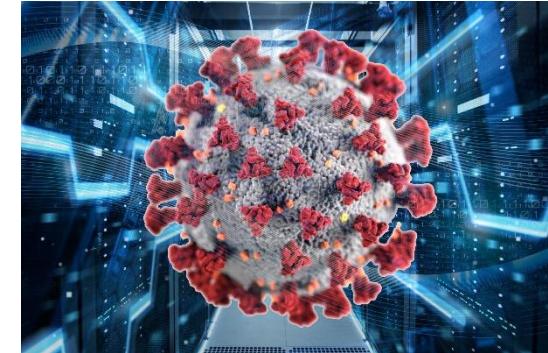
- Virtual Neutron and x-ray Laboratory (**ViNYL**) will:
- Offer services for simulation and modelling of photon and neutron instruments, as well as start-to-end simulations to describe entire experiments at photon and neutron facilities
- Make simulation data services inter-operable among the involved partners, such as **OASYS**, **McSTAS** and **SIMEX**
- Enable RIs to seamlessly link this experiment simulation services to their in-house data reduction, analysis, and visualization infrastructures.
- **SIMEX** is responsible for photon experiment simulations



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

COVID-19 : a test for PaNOSC and EOSC

- Software & Platform service to browse and analyse raw data :
In Production and/or In Test



→ EuXFEL, ILL, ESRF, CERIC-ERIC, ESS, ELI provide remote analysis services :

Jupyter → EuXFEL, ILL, ESRF, CERIC-ERIC, EGI in production,
ESS + ELI in test

Desktop → EuXFEL, ILL, ESRF, CERIC-ERIC, EGI in production
But not scalable to all users

HPC → EuXFEL, ILL, ESRF, EGI in production for local users
But not scalable to all users

Server Options

Simple Advanced

Architecture

Intel (x86_64) IBM Power (ppc64le)
Intel Xeon IBM POWER9

CPUs

Minimum Medium Maximum
1 core(s) 14 core(s) Entire node

Job duration 1 hour

List of available resources:

Current Status		
Partition	# nodes	# avail
nice	41	12
p9gpu	4	2
deb9-fast-io	1	0
deb9-gpu	1	1
id16a	2	1

Start

PaNOSC KPIs

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Open Data 2018	100s	2	0	10s	0	0
Open Data 2023	1000s	1000s	100s	1000s	100s	10s
Data Services 2018	Pilot	In progress	Remote	In progress	?	In progress
Data Services 2023	Desktop Jupyter	Jupyter Desktop	Jupyter Desktop	Jupyter Desktop	Desktop Jupyter	Jupyter Desktop
Common data API 2018	No	No	No	No	No	No
Common data API 2023	Yes	Yes	Yes	Yes	Yes	Yes
User training 2018	No	No	No	No	No	No
User training 2023	Yes	Yes	Yes	Yes	Yes	Yes



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 :**
 - User driven data transfer (e.g. **Globus Online**)
 - Data archiving for RIs (**STFC as the archive center**)
 - Transfer from RIs 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**.
- **Pilots with EGI Support of Use Case 2 &3, evaluating :**
 - Integration with the existing RI IT infra
 - AAI interoperability
 - Limit of the model in terms of data volume

PaNOSC & EOSC – Experience

- **EOSC is under development :**

- 4 Working Groups
- 40+ EOSC-related projects
- Hundreds of potential collaborators
- Many meetings + reports to read / comment on
- EOSC is like the internet, no single view but many views

- **Pros :**

- Clusters are getting together + sharing know-how
- Working with e-infrastructures EGI + GÉANT is beneficial

- **Future : EOSC will improve reproducibility and boost Open Science + Open Data**



**EUROPEAN OPEN
SCIENCE CLOUD**



WP3

WP4

WP5

WP8

Outreach Events & Developments

- **1st PaNOSC OASYS School**, 14-16 May 2019, Grenoble – France
- **HDF5 European Workshop for Science and Industry**, 17-18 September 2019, @ESRF, Grenoble
- **h5py code camp**, 19-20 September 2019, @ESRF, Grenoble
- **Nexus data format** – extend for synchrotron, FELs + lasers, generalize use of Nexus in PaN community
- **hdf5 – h5py** python library for manipulating hdf5 files + **h5web** web viewer of hdf5
- **Jupyter notebooks** – integrate batch schedulers + integrate h5web viewer, provide notebooks for use cases
- **Data portal** – develop a PaN portal for generic services + remote desktop
- **Pan-learning** – integrate jupyter notebooks into moodle

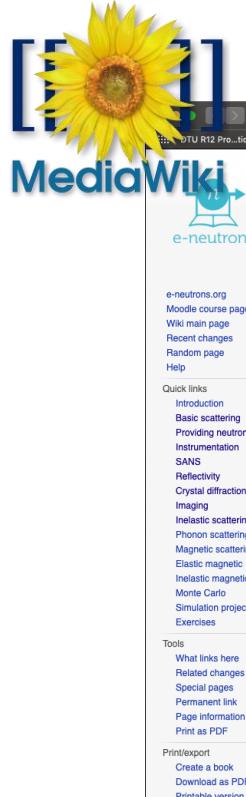


e-learning on Neutrons and Photons

- 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 WP 4 data analysis services
 - Integrate WP 5 simulation services



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





McStas

sim.e-neutrons.org

Instrument

Logged in as mcstas (see recent simruns) Logout

SANSsimple (click for documentation)

Source with guide and velocity selector → Site 1 → Site 2 → Sample → LD → Detector

Parameters for SANSsimple

pinhole_rad [m] : 0.004 radius of the collimating pinholes (0.004)
LC [m] : 3 length of the collimator – distance between pinholes (3)

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.

*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:

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

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:

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

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 δ -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^[1] 2.3):

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

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

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

e-learning on Open Science

WP8

- Train young scientists + PhDs on how to do Open Science and make Data Open and FAIR
- Team up with Eurodoc¹ Hercules school, local universities², FAIRsFAIR, EOSC, etc.

¹ <http://eurodoc.net/>

² <https://www.datacc.org/>

The slide is titled "How to be an Open Scientist at University of Camerino". It features a hand holding a magnifying glass over a green circular graphic, symbolizing research and discovery. The slide is divided into several sections:

- Open Data Management**:
 - Why?**
 - Real-time collaboration
 - Immediate results
 - Increased sharing
 - How?**
 - Uniform data management
 - Open data repositories
 - Intelligent access & interoperability
- Open Source Code**:
 - Why?**
 - Personal benefit
 - Agile development
 - Real-time evaluation
 - How?**
 - Community networking
 - Open source code repositories
- Involving the Public**:
 - Why?**
 - Mutual learning (science <-> public)
 - Civic empowerment
 - Recognition of other perspectives
 - How?**
 - Citizen Science
 - Social Media/Blogs/Apps
 - Public Speaking
- Sharing Results with Colleagues**:
 - Why?**
 - Support wide-spread of information
 - Creating broader networks
 - How?**
 - Open access publishing
 - Pre-print servers
 - University repositories
- Sharing Results with Public**:
 - Why?**
 - Keep shareholders informed
 - Informed public = informed policy
 - Avoid misinformation
 - How?**
 - Traditional/Social media
 - Public speaking
 - Science Cafés, Open Museum nights
- Collaborating with Researchers**:
 - Why?**
 - "Two heads are better than one"
 - How?**
 - Enhanced scope of project
 - Avoid potential bias



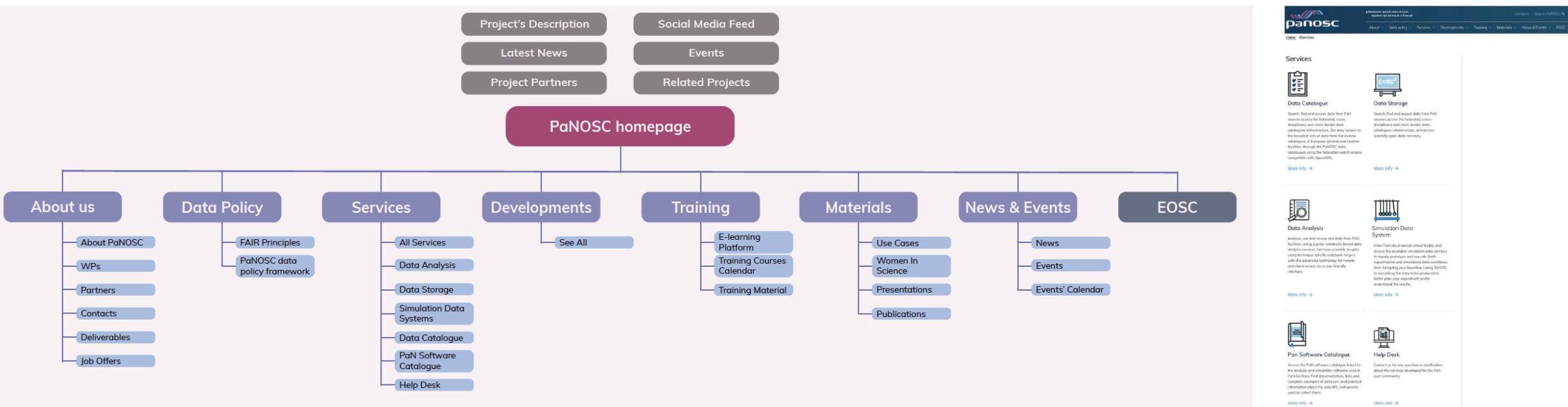
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

<http://eurodoc.net/news/2018/handbook-on-how-to-be-an-open-scientist-for-early-career-researchers>

PanOSC

PaNOSC website – <https://panosc.eu>

WP9



The banner features a dark blue background with a subtle image of a microscopic view of particles. Overlaid text includes the PaNOSC logo, the tagline "photon and neutron open science cloud", and a call-to-action button "Learn more".

The Photon and Neutron Open Science Cloud (PaNOSC)

The Photon and Neutron Open Science Cloud (PaNOSC) is a European project (financed by the INFRAEOSC-04 call) for making FAIR data a reality in 6 European Research Infrastructures (RIs), developing and providing services for scientific data and connecting these to the European Open Science Cloud (EOSC).

Objectives

Tweets by @Panosc_eu

PaNOSC will be at the joint CNECT-RTD project meeting & workshop on 9-10 September in Brussels. The workshop aims to increase the visibility of EOSC-related projects, explore their progress and discuss issues on next steps towards building the #EOSC >>bit.ly/2m1M1hV

The news section header is "News". It displays three news items with images and titles:

- Making FAIR data a reality** (Published on 7 August 2019)
- Reproducible science discussed at the Jupyter for Science workshop** (Published on 18 July 2019)
- Kick-off meeting of PaNOSC WP4 hosted at EuXFEL** (Published on 8 July 2019)

The events section header is "Events". It displays four event details with dates and locations:

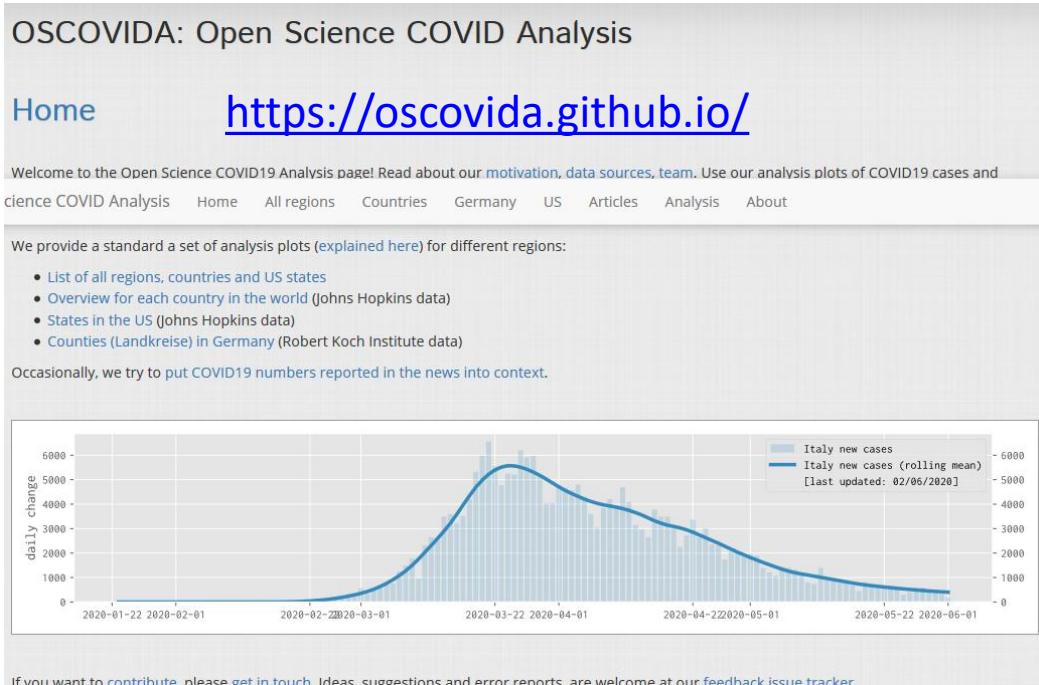
- 04/11/19** PaNOSC 1st Annual Meeting
- 18/09/19** PaNOSC WP3 workshop at ILL
- 17/09/19** Register now to the HDF5 European Workshop for Science and Industry
- 17/09/19** The ESRF, in the frame of PaNOSC and in collaboration with HDF Group, is organizing the HDF5 European Workshop for Science and Industry, taking place at the ESRF headquarters in Grenoble - France, on 17-18 September 2019. The aim of the workshop is to highlight the use of HDF in science and industry, to discuss use [...] More info →



COVID-19 : a test for PaNOSC and EOSC

Common space to share open science workflows : Work in Progress

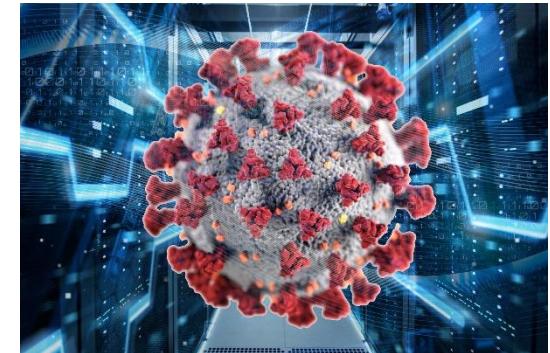
→ e.g. EuXFEL OSCOVIDA example of citizen science Jupyter notebooks



Future → provide a catalogue of notebooks and workflows for analyzing data from different experimental techniques e.g. like the PANGEOT use cases (https://pangeo.io/use_cases/index.html)

We are developing a collection of Jupyter Notebooks that demonstrate real science use cases enabled using Pangeo. If you have a use case that you would like to submit, please raise an issue on the [Pangeo GitHub issue tracker](#) to propose your idea.

- Physical Oceanography
 - Sea Surface Altimetry Data Analysis
 - Initialize Dataset
 - Examine Metadata
 - Create and Connect to Dask Distributed Cluster
 - Visually Examine Some of the Data
 - Timeseries of Global Mean Sea Level
 - Sea Level Variability



PaN Commons

- **Vision** – create a common space for PaNOSC and ExPaNDS facilities where petabytes of PaN data, analysis software, notebooks, analysis software, workflows, and training material can be **Found**, **Accessed** (downloaded and/or executed), **Re-Used** + **Improved** i.e. **FAIR**
- **Remote access** – the PaN commons will be accessible remotely while being executed locally (close to the data) or on the EOSC (data needs to be moved)
- **Remote users** – the PaN commons will enable and encourage remote users and experiments (urgently required in the **post-COVID-19 phase**)

HIGH PRIORITY



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



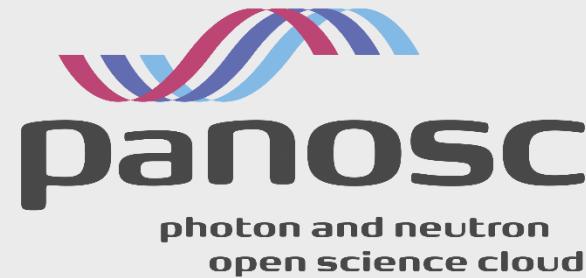
Conclusion

- **Priority** – PaNOSC priority has increased significantly due to increased need for remote access (post-COVID-19). **COVID-19 demonstrated the need for PaNOSC and that we (PaNOSC + EOSC) are not ready for the next pandemic.**
- **EOSC** – interacting with the EOSC remains a challenge but we hope that in 2021 with the legal entity being created the EOSC will stabilise and be easier to interact
- **Collaboration** – with ExPaNDS is essential and is working well; collaboration with INFRAEOSC-04 clusters is increasing
- **Progress** – Objectives and outcomes of PanNOSC are aligned with the needs of all partner sites; all work packages are progressing well



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



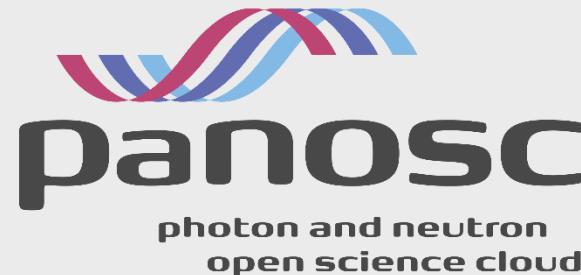


Thank you

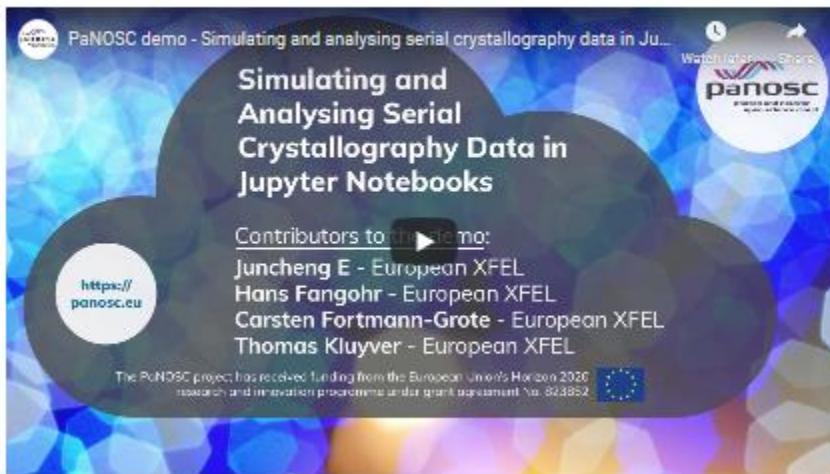
andy.gotz@esrf.fr



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



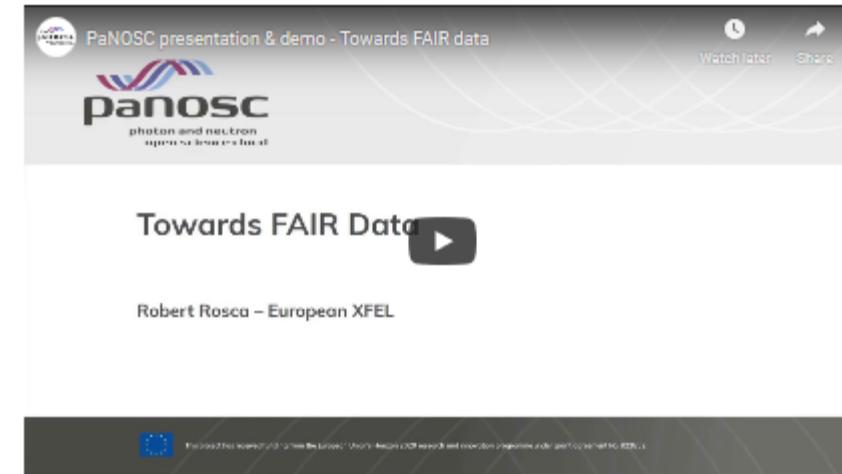
PaNOSC demo – Simulating and analysing serial crystallography data in Jupyter Notebooks



Watch these videos:

<https://www.panosc.eu/video/>

Presentation & Demo – Towards FAIR data



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