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

PaNOSC & EOSC : built on Boldness

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

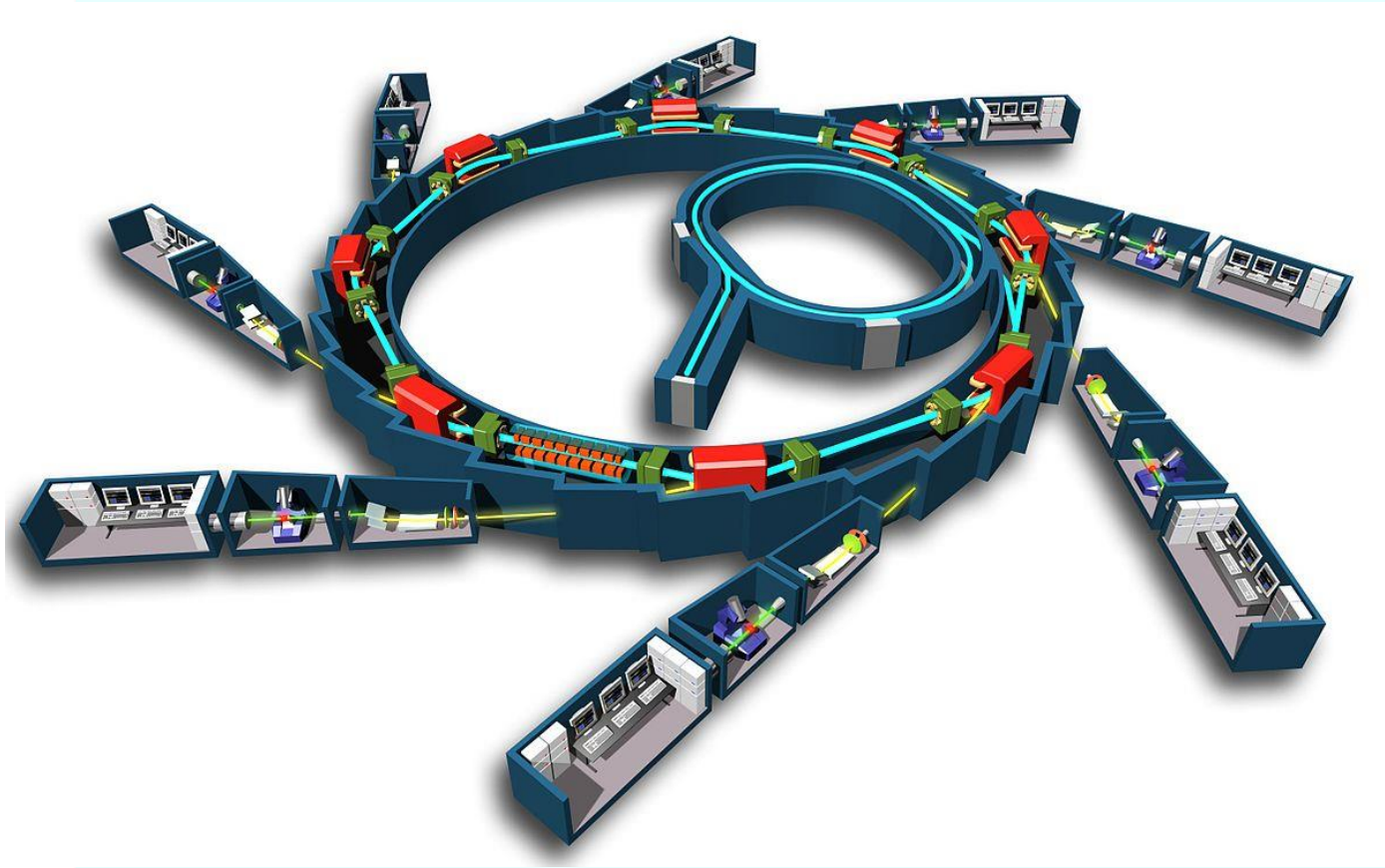
Duration: 4 years

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

panosc

How do PaN facilities work?

1. **User**: **has an idea / need to study a sample**



ality

l

8. **Publication**: **User publishes results (DOI) in peer review journal**

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

panosc

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.

WP2
WP3
WP4
WP5

Full FAIR compliance of PaN scientific data

WP4
WP6

Support in shaping EOSC services for users needs

Increase of RIs' impact by encouraging data reuse

WP2
WP6
WP7

WP4
WP5
WP6

Innovative data services at RIs and as part of the EOSC

WP2

Sharing of best practices for open data policies

Collaboration with EOSC projects to share outcomes

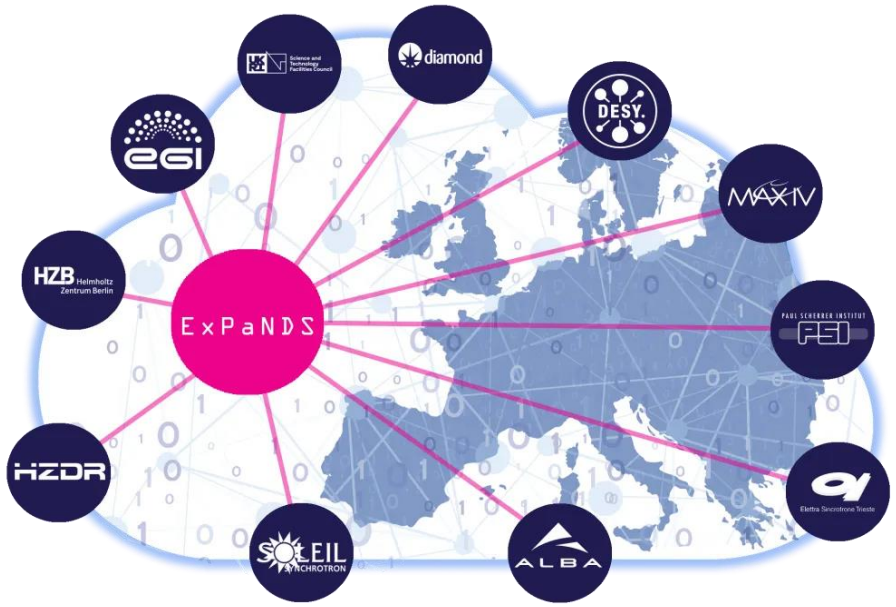
ALL WPs

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

This
823852

panosc

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

panosc
photon and neutron
open science cloud

panosc

COVID-19 : A Dry Run for PaNOSC + EOSC

Definition of 'dry run'

dry run

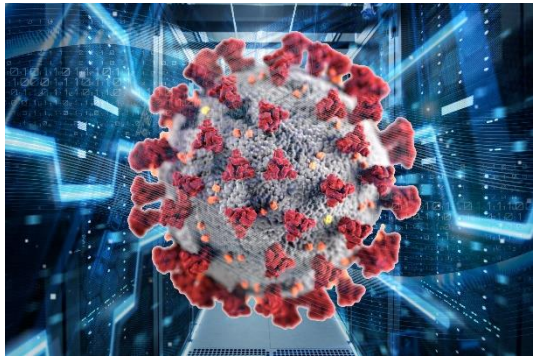
Collins COBUILD

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]

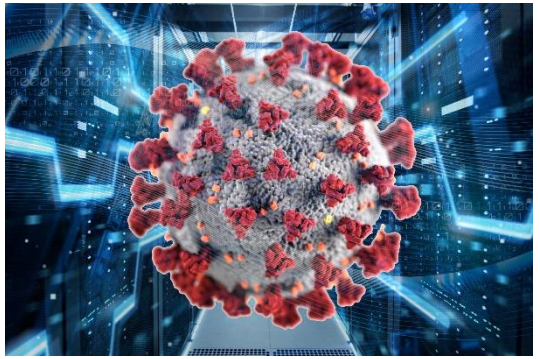


How ready are PaNOSC + EOSC for the COVID-19 pandemic?

panosc

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	2016	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 is making Data Policies FAIR

WP2

zenodo Search Upload Communities

February 11, 2011

Project deliverable | Open Access

Common policy framework on scientific data

Dimper, Rudolf

The increasingly highprofile of issues surrounding preservation and access of research data makes it important to work towards harmonisation of data policesacross the research base. PaN-data brings together a significant number of major world class European research infrastructures to lay the foundation for a fully integrated, pan-European, information infrastructure supporting the complete scientific cycle from experiment definition to publicatione. The participating facilities are used by researchers in universities, publicly funded research entities, and industry.This document describes the common framework for scientific data management at photon and neutron facilities. This document and any future revisions is also available at pan-data.eu .

Preview

Files (352.4 kB)

nion's Horizon 2020 research and innovatio

zenodo Search Upload Communities

May 20, 2020

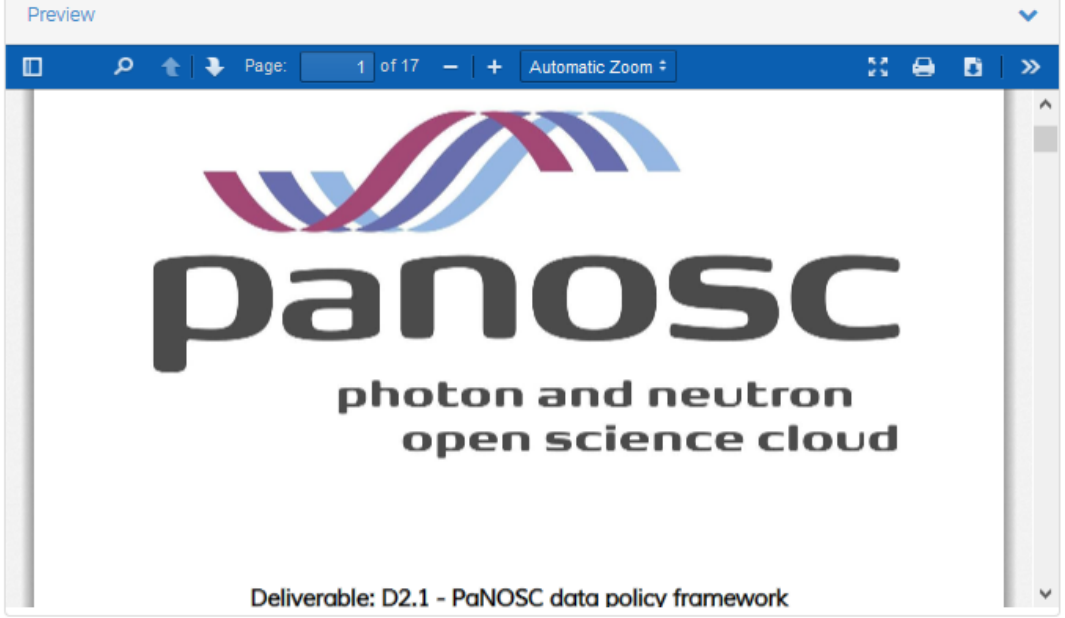
Project deliverable | Open Access

PaNOSC FAIR Research Data Policy framework

Gotz, Andy; Perrin, Jean-Francois; Fangohr, Hans; Salvat, Daniel; Gliksohn, Florian; Markvardsen, Anders; McBirnie, Abigail; Gonzalez-Beltran, Alejandra; Taylor,Jonathan; Matthews, Brian

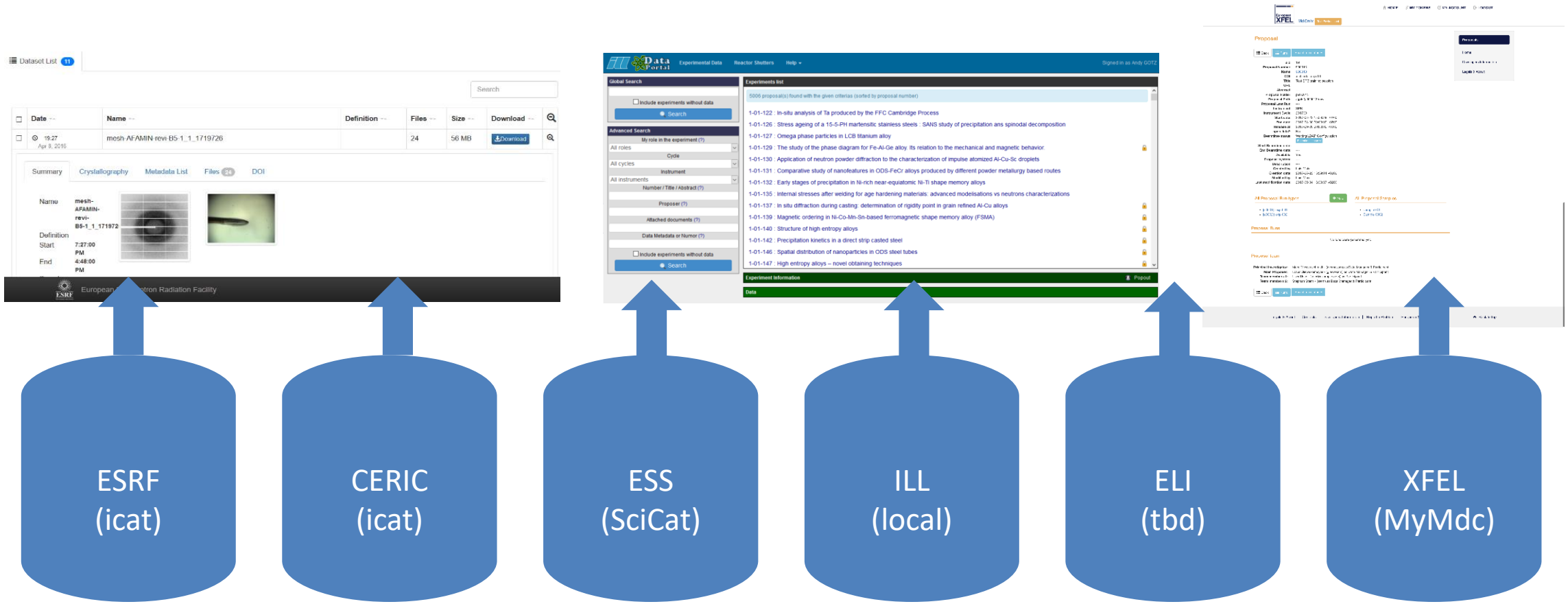
This paper presents the new photon and neutron research data policy framework based on the previous PaNData policy (https://doi.org/10.5281/zenodo.3738497) applicable to all photon and neutron facilities and scientific research data in general. The data policy framework is strongly aligned with the FAIR principles. The aim of the policy is to ensure that the FAIR principles are applied in research data policies. This deliverable has been prepared by the EOSC projects PaNOSC (https://panosc.eu) and ExPaNDS (https://expands.eu) together to ensure harmonisation of the updated data policies for the photon and neutron communities.

Preview



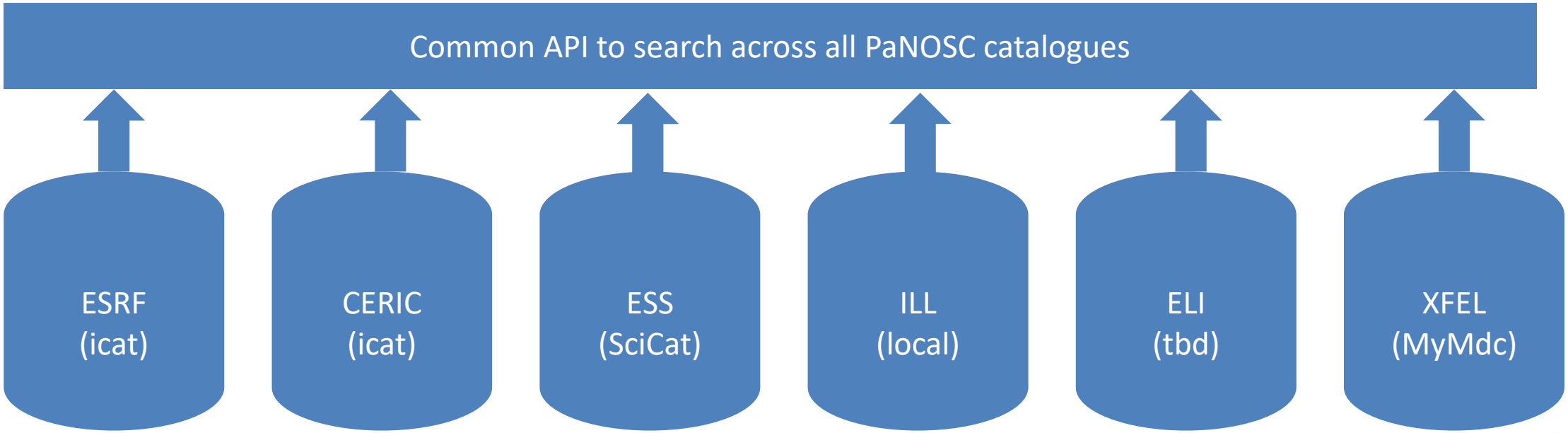
PaNOSC has 6 data catalogues with different APIs + UIs

WP3



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

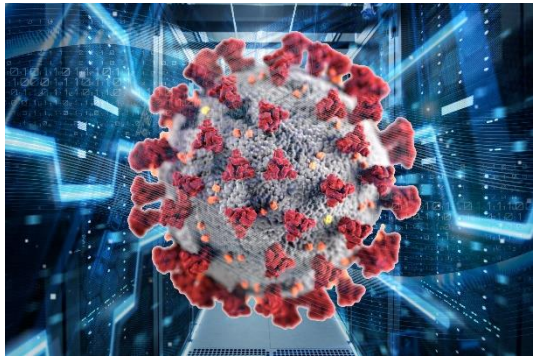
PaNOSC common API across all sites



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

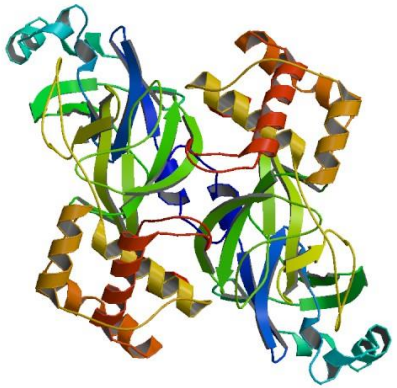
COVID-19 : a test for PaNOSC and EOSC

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



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

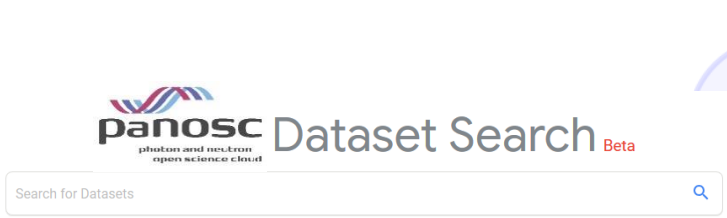
SARS-CoV/2c3s

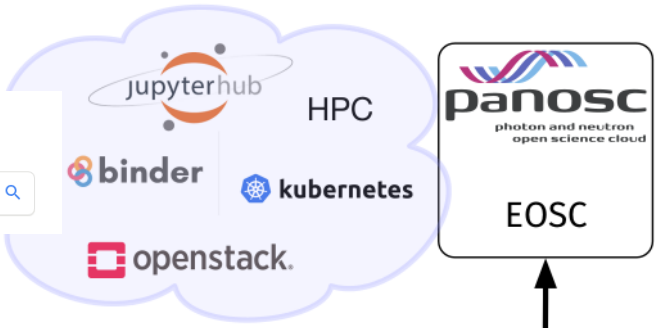


SARS-CoV/2c3s, SARS-CoV/2beq, SARS-CoV/2bez, SARS-CoV/2h85, bovine-coronavirus/3cl4, bovine-coronavirus/3cl5, HCoV-OC43/6qfy, MERS-CoV/4ud1, MERS-CoV/6g13, murine_coronavirus/4c7l, murine_coronavirus/4c7w, murine_coronavirus/5jif, murine_coronavirus/5jil, 2cme, SARS-CoV/2xyr, SARS-CoV/2xyv, SARS-CoV/2fav,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

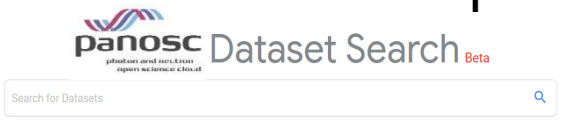
panosc

WP4

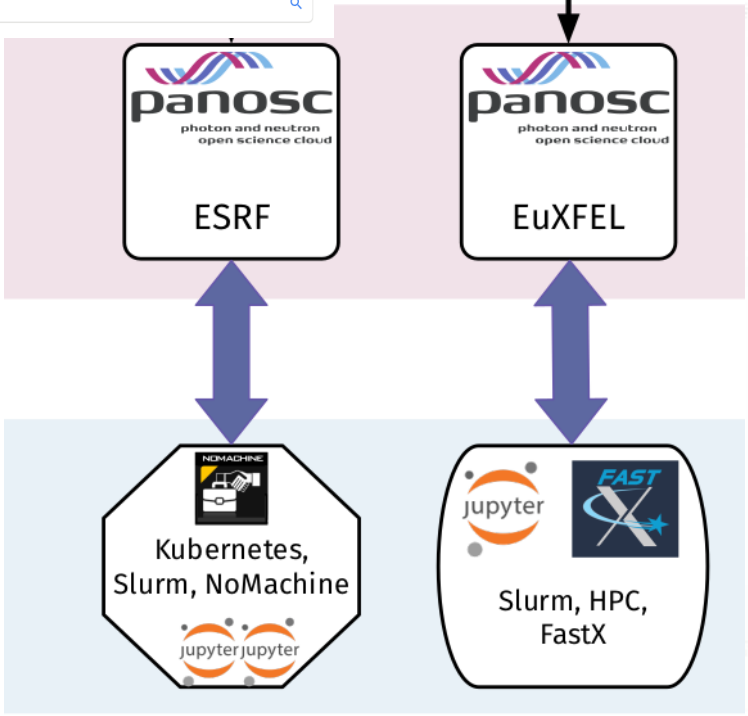




Meta portal
for data analysis
services



Presentation & Demo – Towards Reproducible Publications with FAIR Data



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**

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

- 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

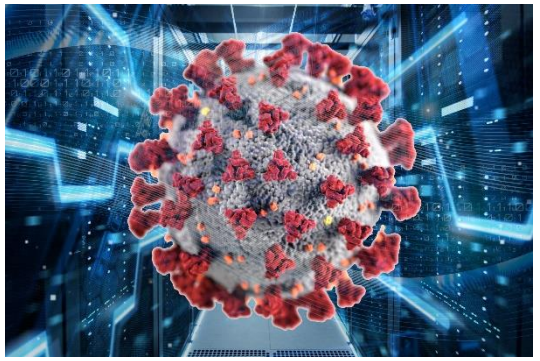
PaNOSC KPIs

	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

COVID-19 : a test for PaNOSC and EOSC

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

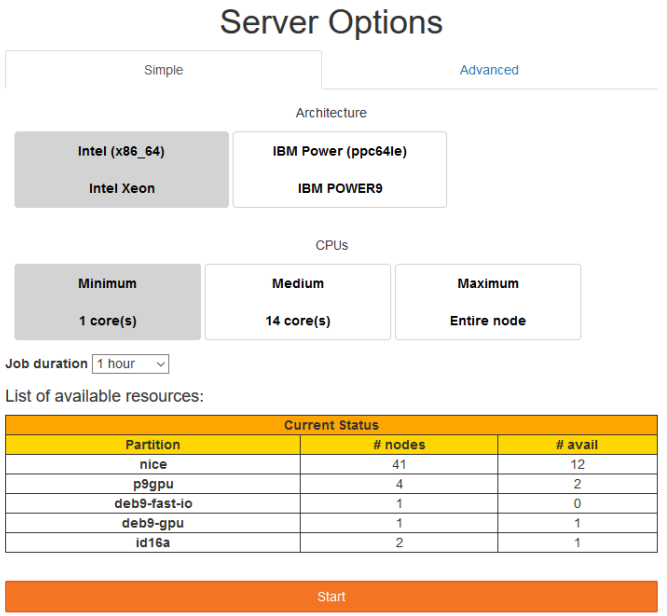


→ **EuXFEL, ILL, ESRF, CERIC-ERIC, ESS,ELI provide remote analysis services :**

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



EOSC Integration – Data transfer

- **3 uses cases :**
  - User driven data transfer of GB to TBs (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

WP6

- **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 – there are 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

Progress on Open Science: Towards a Shared Research Knowledge System

Final Report of the Open Science Policy Platform

Outreach Events & Developments

WP3
WP4
WP5
WP8

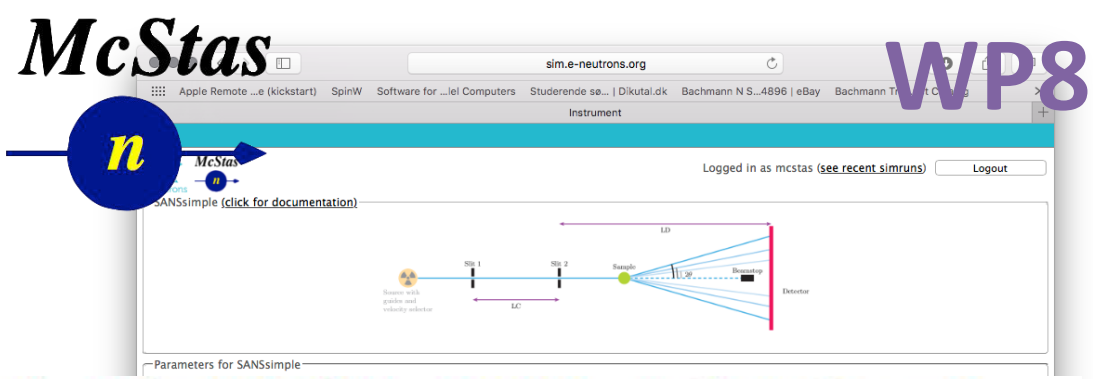
- **1st PaNOSC OASYS School**, 14-16 May 2019, Grenoble – France
- **HDF5 European Workshop for Science and Industry**, 17-18 September 2019, @ESRF, Grenoble
- **h5py code camp**, 19-20 September 2019, @ESRF, Grenoble

- **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

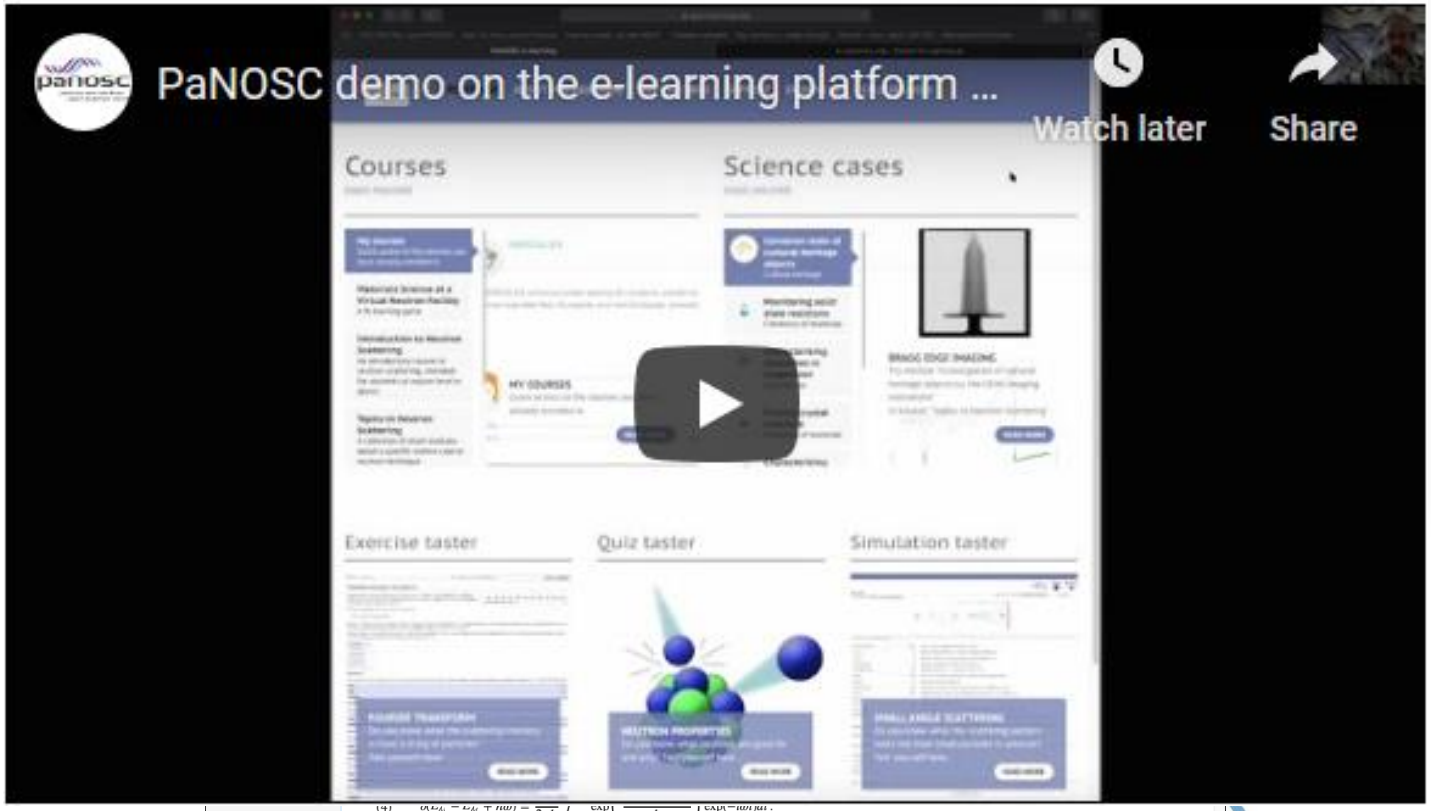
panosc

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

Demo on the e-learning platform www.pan-learning.org



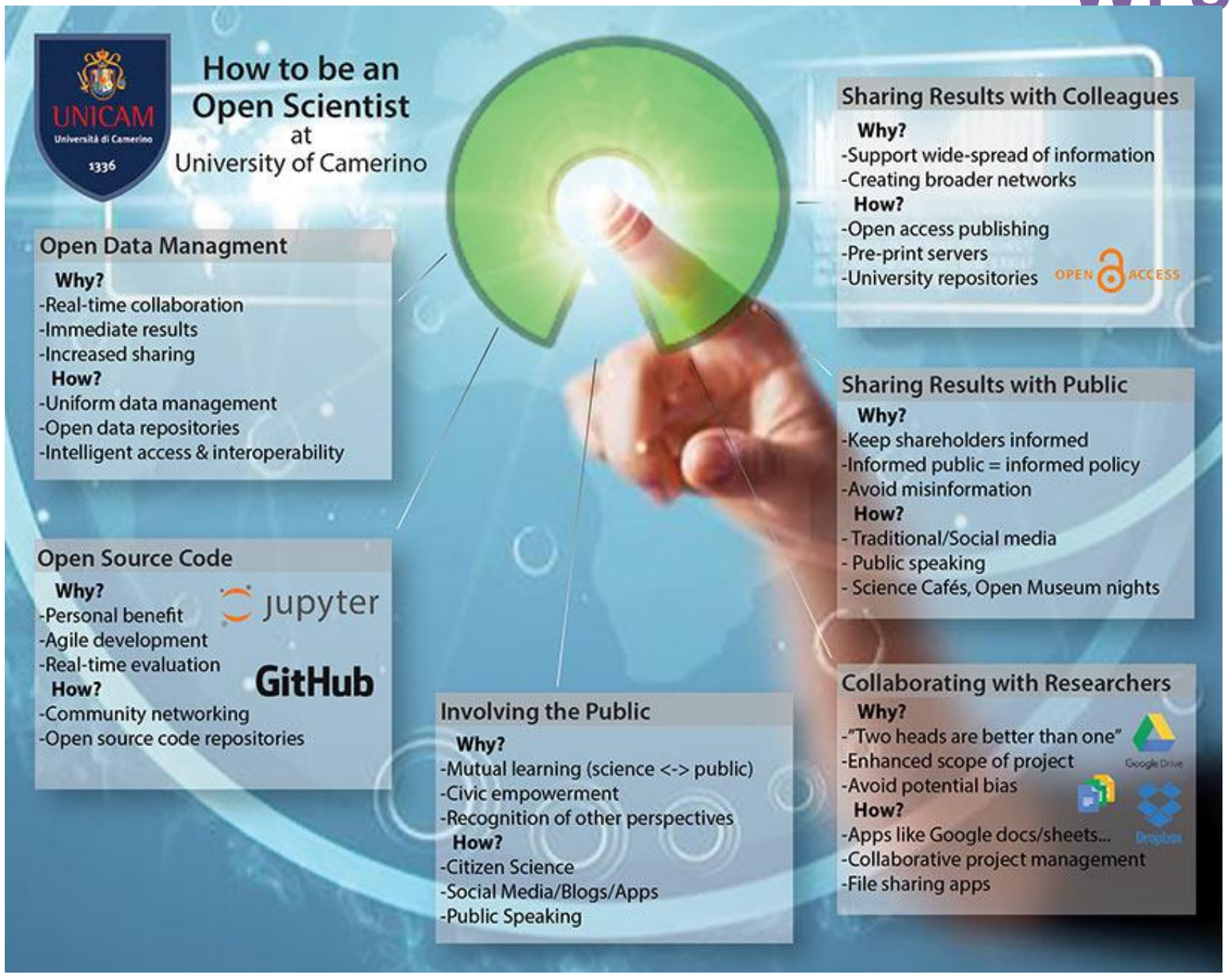
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

e-learning on Open Science

- Train young scientists + PhDs on how to do Open Science and make Data Open and FAIR

- Team up with Eurodoc[1] Hercules school, local universities[2], FAIRsFAIR, EOSC, etc.

[1] http://eurodoc.net/
[2] https://www.datacc.org/

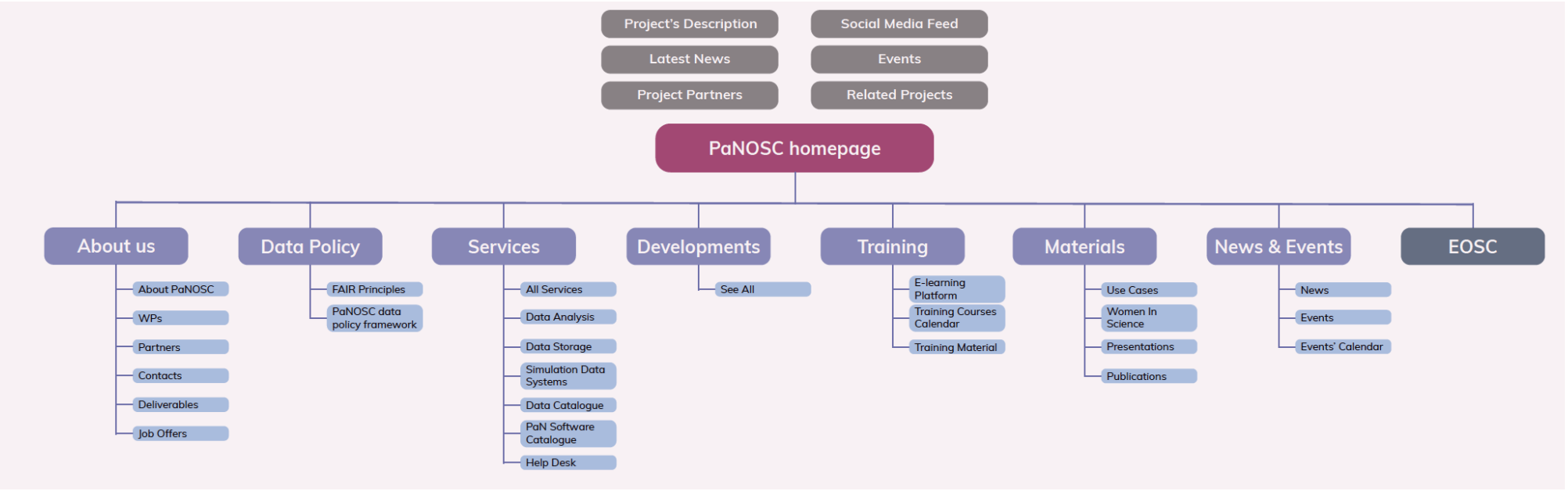


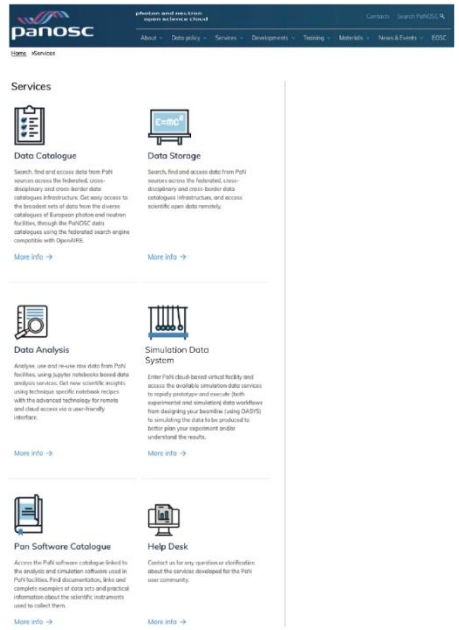
http://eurodoc.net/news/2018/handbook-on-how-to-be-an-open-scientist-for-early-career-researchers

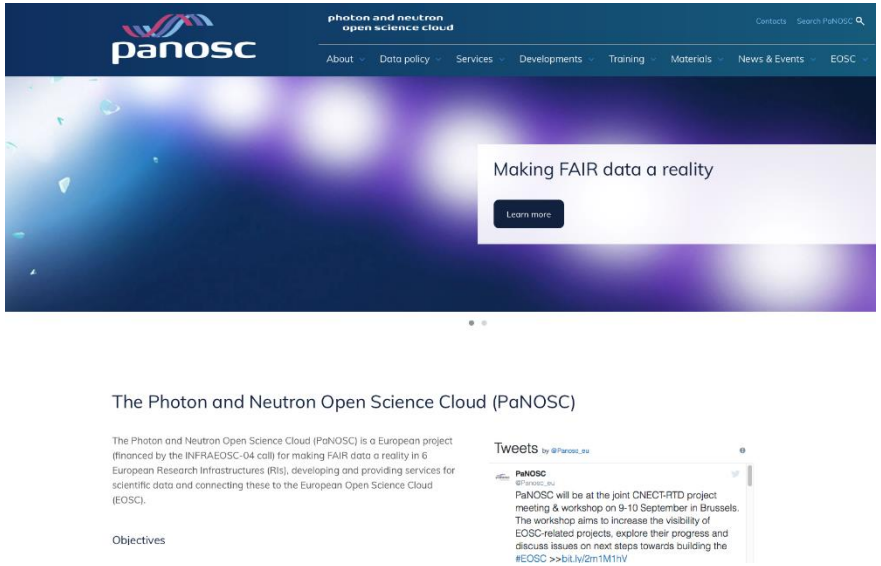
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

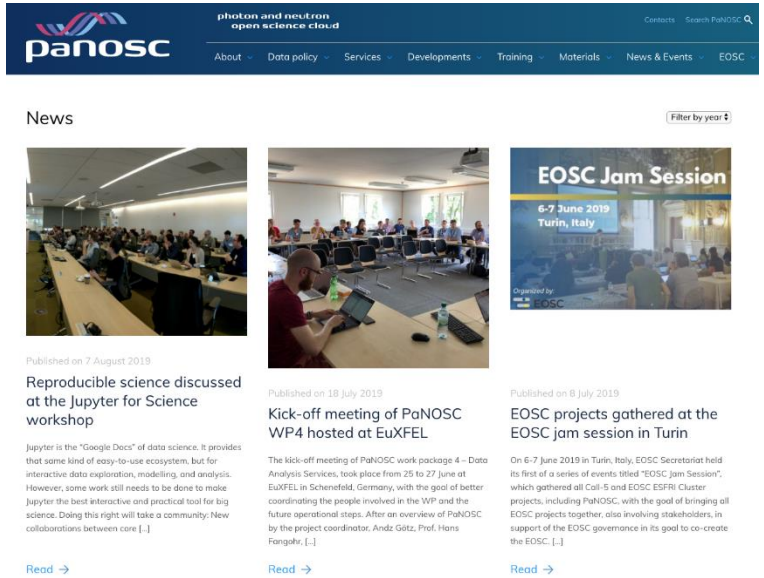
PaNOSC website – https://panosc.eu

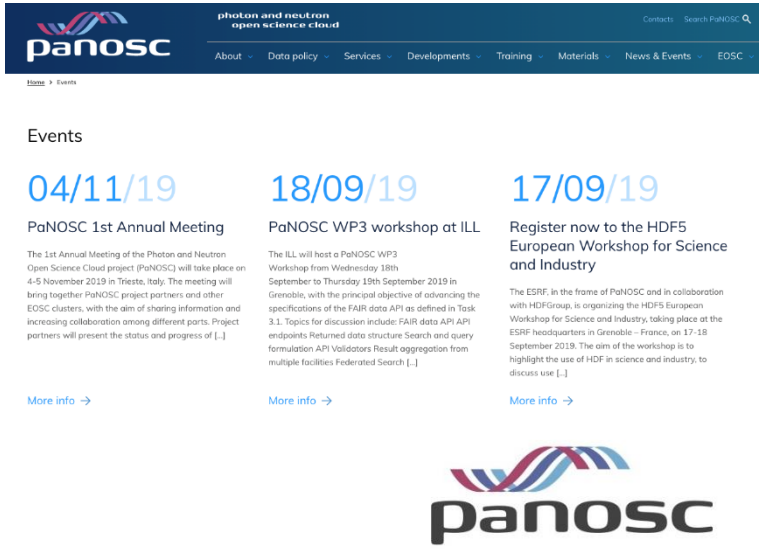
WP9





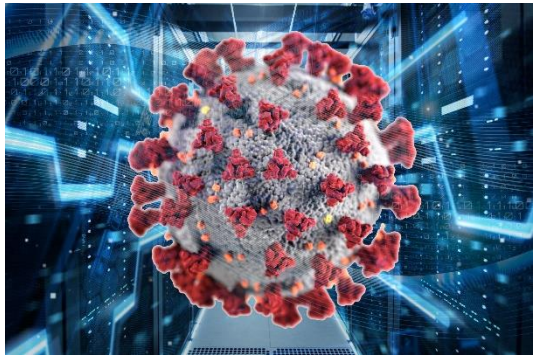




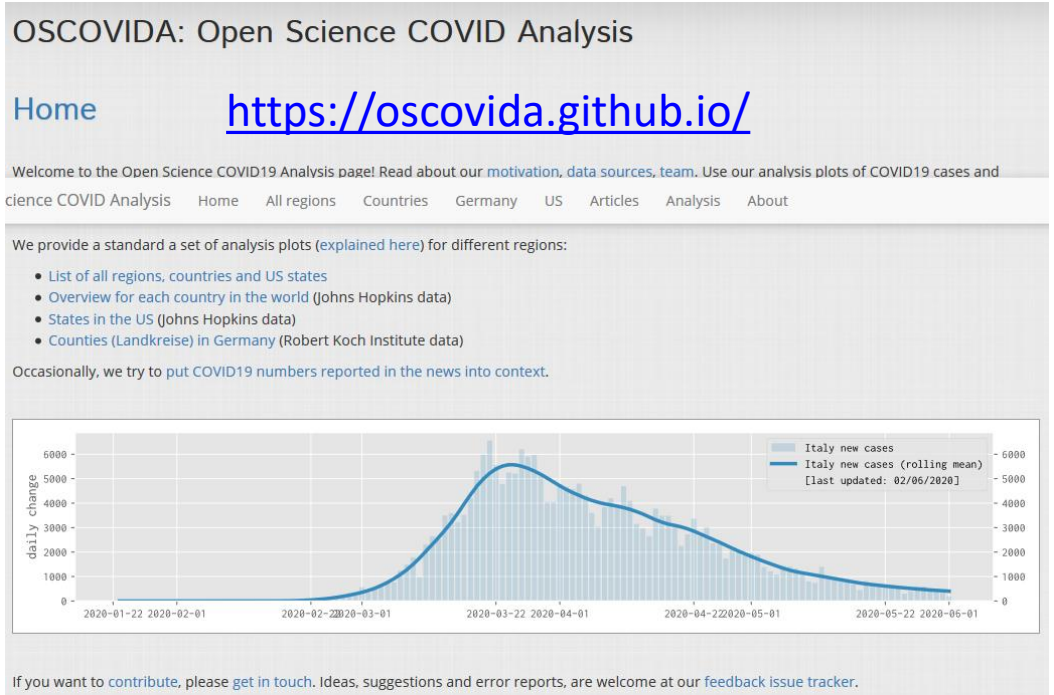


COVID-19 : a test for PaNOSC and EOSC

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 PANGEO project does (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)

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; with Users during remaining 30 months of project is key to the success of the project

- 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 to all PaNOSC project members for their contributions !

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

Watch these videos:

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

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

Presentation & Demo – Towards Reproducible Publications with FAIR Data

Demo on the e-learning platform www.pan-learning.org

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