

Photon and Neutron Open Science Cloud

**Andy Götz (ESRF)**  
**on behalf of the PaNOSC consortium**

**7 December 2018**



## Recall: New science = a lot of data!

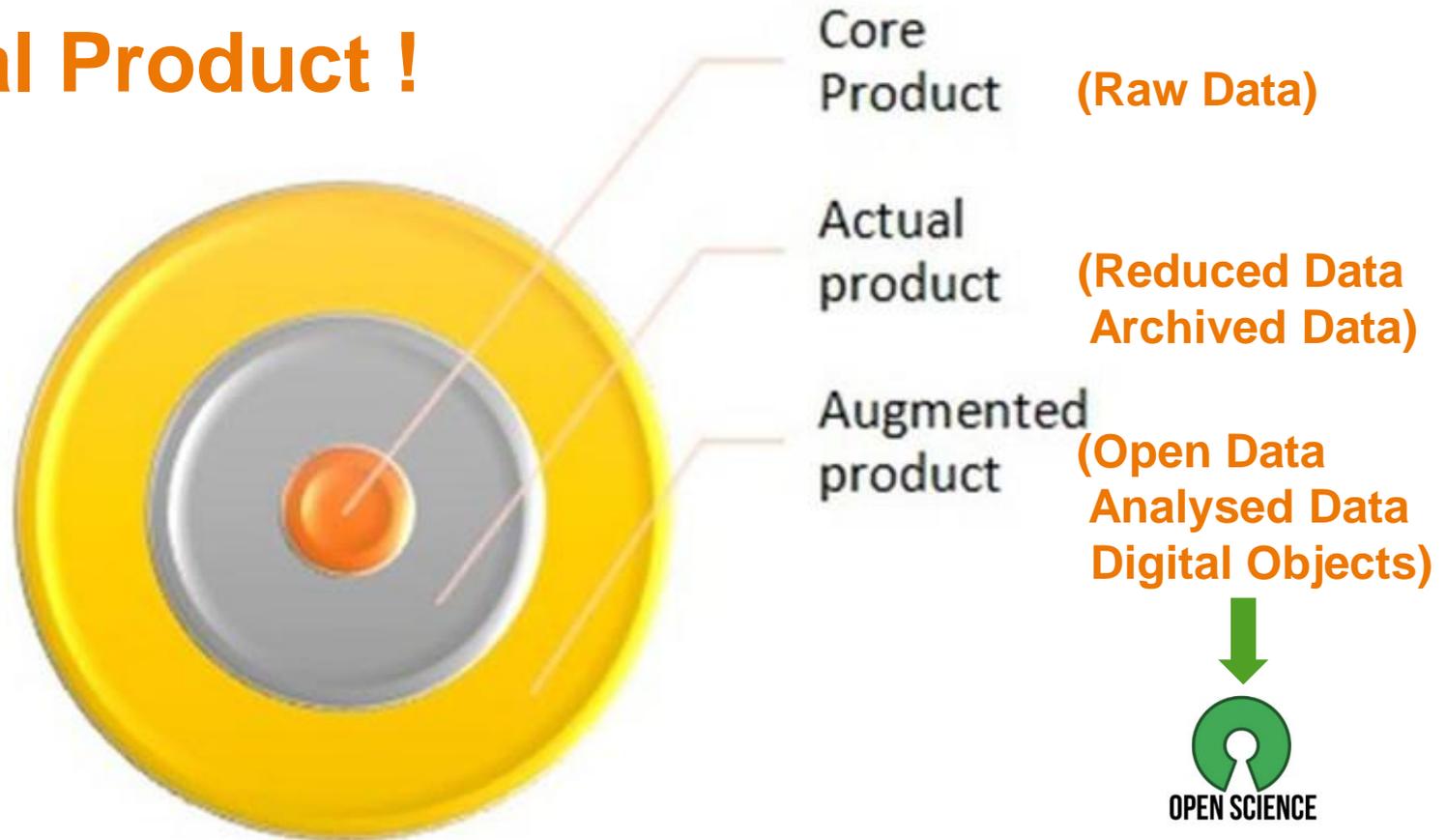
Hurdles for users of our RIs:

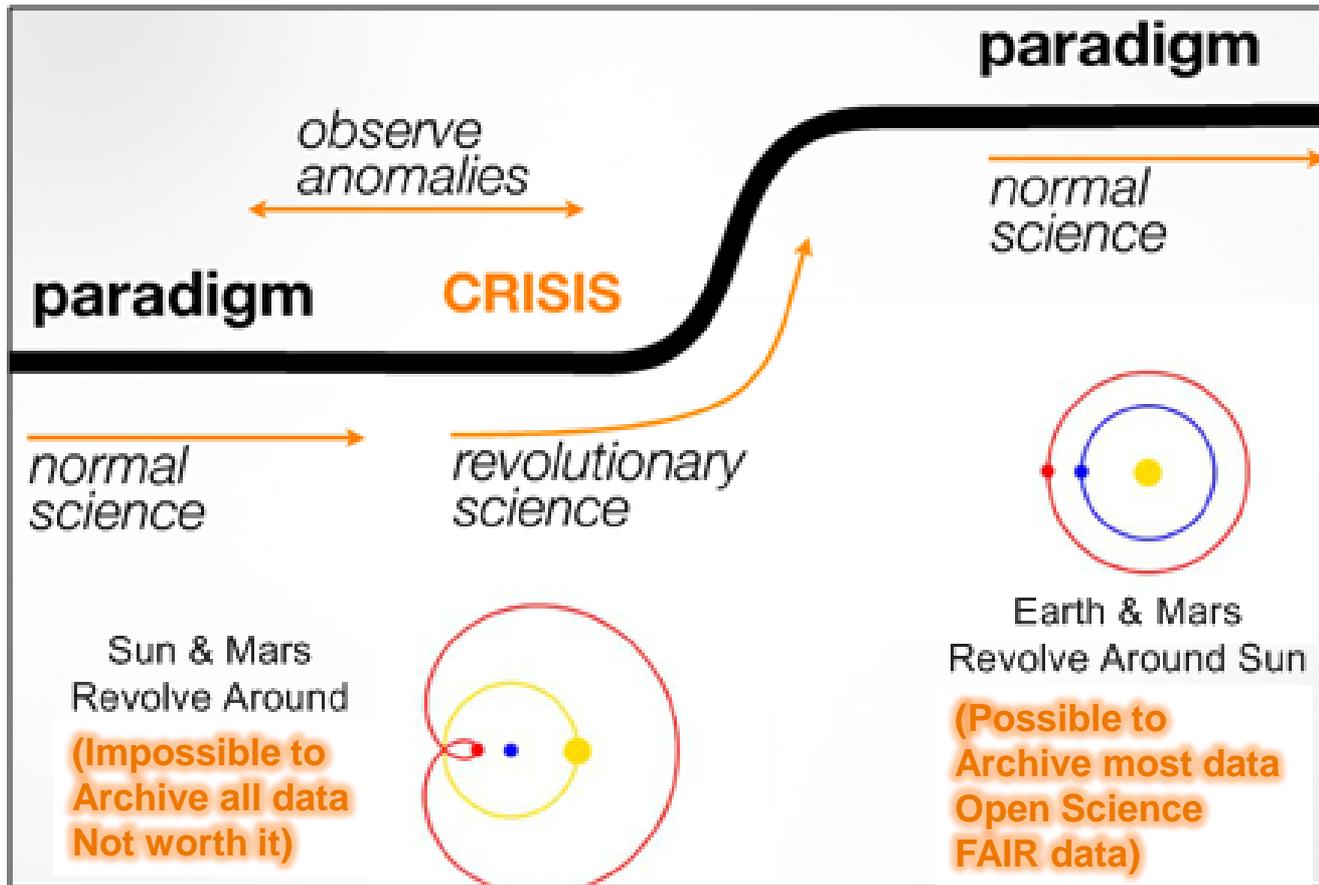
- Quantity of data generated too big to transfer
- Availability of adequate software can be problematic
- Big data need big computing resources
- New users: non-experts, need help in their data analysis
- Long term storage of the data (regulations and requirements)

**The PaNOSC partners operate large-scale RIs that are part of Europe's big-data science factories. They face common challenges, shared by many other RIs in Europe.**

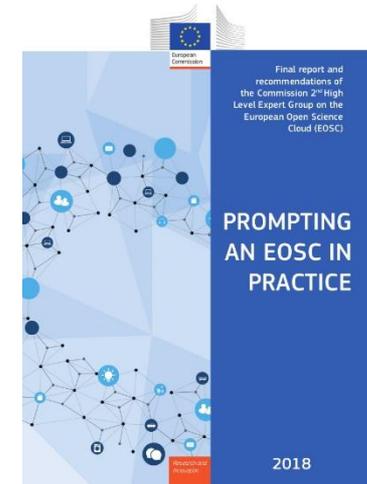


## Data is our Actual Product !





	Europe	
<p><b>Domain cluster projects (H2020-INFRAEOSC-04-2018 call)</b></p>	<ul style="list-style-type: none"> <li>» Interoperability and service architecture</li> <li>» Service provisioning</li> <li>» Data provisioning</li> <li>» Community building</li> <li>» Policy development</li> <li>» Skills &amp; capabilities development</li> <li>» Discipline specific approach</li> </ul> <p><b>NB:</b> Ensure the connection of the research infrastructures identified in the ESFRI Roadmap to the EOSC. Support will be provided through cluster projects gathering ESFRI projects and landmarks in each of the following large thematic domains: Biomedical Science, Environment and Earth Sciences, Physics and Analytical Facilities, Social Science and Humanities, Astronomy, Energy. Proposals will address the stewardship of data handled by the involved research infrastructures according to the FAIR[1] principles and in line with the objectives of open science.</p>	<ul style="list-style-type: none"> <li>» Research performing organisations</li> <li>» Scientific communities and individual researchers</li> <li>» Funding bodies</li> </ul>



**EOSC-LIFE, ENVRI-FAIR, ESCAPE, SSHOC, PANOSC**



## Data Stewardship

- Generalise the adoption of **FAIR** open data principles

## EOSC

- Federate data catalogs and integrate in EOSC and OpenAire meta-catalog, AAI, download services

## Data Services

- Develop new data analysis, modelling and simulation services

## User Training

- Train users how to write Data management plans, improve metadata, using DOIs, manage data better



## H2020 proposal to INFRAEOSC-4

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

**Decision:** Accepted on 13/8/2018

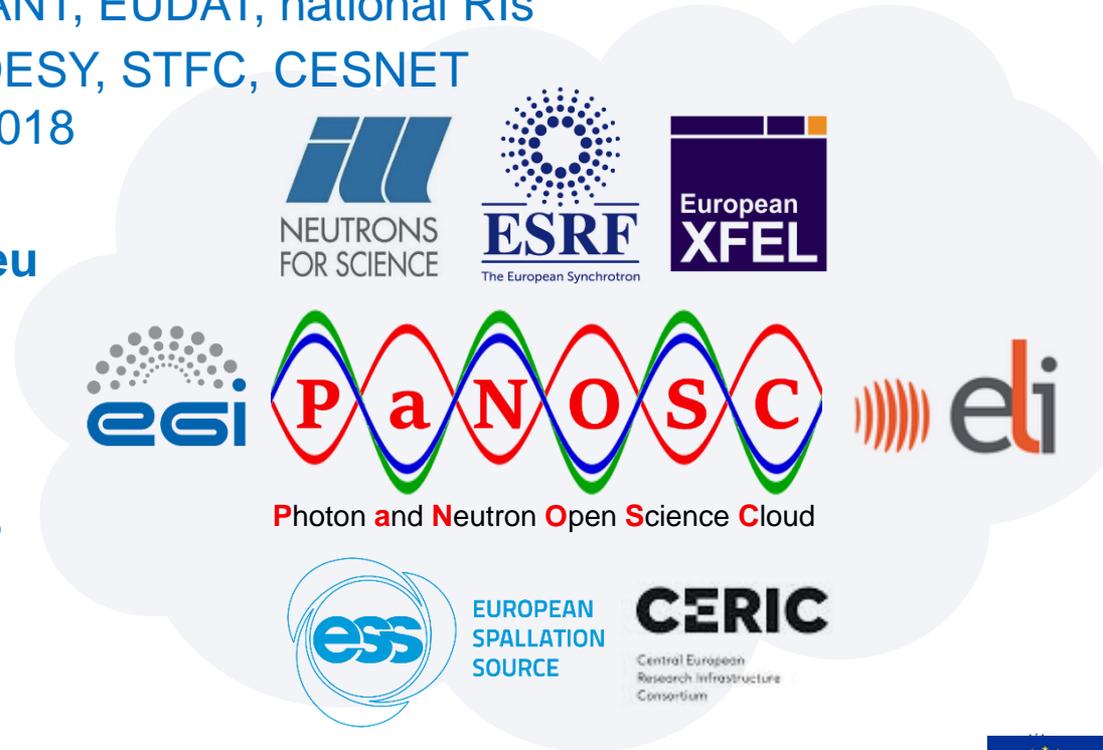
**Status:** Started 1/12/2018

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

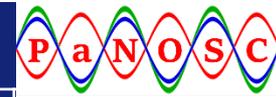
- **Budget:** 12 M€
- **Coordinator:** ESRF
- **Planned started:** 1/12/2018
- **Project duration:** 4 years

**Kick-off meeting**

**15-16 January at ESRF**



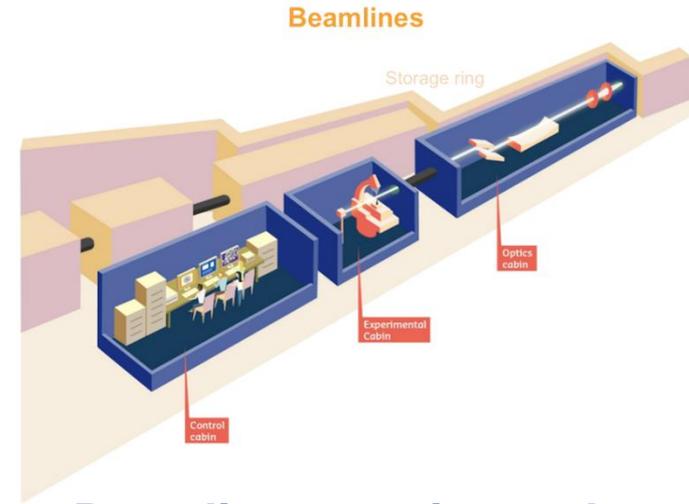
# Who are the PaNOSC ESFRI Ris ?



	ILL	ESRF	CERIC	XFEL	ELI	ESS
Operating since	1972	1994	2014	2017	2018	2022
Users / yr	1200	6000	500	850	100	100
Beamlines	40+	40+	40+	5	20+	20+



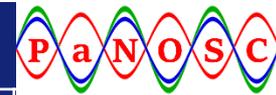
**ESRF + ILL (Grenoble, France)**



**Beamline experimental station**



# Who are the PaNOSC ESFRI Ris ?



	ILL	ESRF	CERIC	XFEL	ELI	ESS
Operating since	1972	1994	2014	2017	2018	2022
Users / yr	1200	6000	500	850	?	100
Beamlines	40+	40+	40+	5	20+	20+

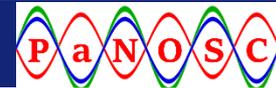
## ELI (Czech Republic, Hungary, Romania)



## XFEL (Hamburg, Germany)



# Before PaNOSC - 2018



	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data / yr	200 TB	8 PB	1 PB	3PB	?	0
Data Policy	2011	2016	2014 (3/8)	2017	?	2017
Metadata catalogue	Local	Icat	Local	myMdC	?	SciCat
Metadata definitions	Nexus	Nexus	custom	myMdC	?	Nexus
DOI	2012	2018	No	2018	?	2018
Open Data	yes	yes	No	yes	no	yes
Data Services	pilot	in progress	remote ops	in progress	?	in progress
Common data API	No	No	No	No	No	No
User training	No	No	No	No	No	No



## Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data Policy	2011	2016	2014 (3/8)	2017	?	2017

## After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Effort (PMs)	10	17	12	3	20	14
Common Framework Data Policy	2011	2016	2019	2017	2019	2017
Data Archiving	YES	YES	YES	YES	YES	YES
DOIs	YES	YES	YES	YES	YES	YES
Open Data	YES	YES	YES	YES	YES	YES
DMP templates	YES	YES	YES	YES	YES	YES

**ESRF is custodian of data and metadata**

**ESRF to collect high quality metadata to facilitate reuse of data**

**ESRF will keep metadata forever**

**ESRF will keep raw (or reduced) data for 10 years**

**Data will be registered in a data catalogue (ICAT)**

**Data will be published with a Digital Object Identifier (DOI)**

**The experimental team has exclusive access to data during the embargo period (3 years which can be extended on request)**

**Data will be made public after the embargo period under CC-BY**

**Data Policy will be implemented on all beamlines by 2020**

**<https://www.esrf.eu/datapolicy>**



- **Data Policy is not so difficult to write**
- **A number of examples exist e.g. PaNdata framework has spawned at least 7 Data Policies**
- **Data ownership is a question of definition and not law**
- **Data needs to be managed and curated not owned**
- **The minimum possible is to implement a Metadata Policy**
- **Adopting a Data Policy is necessary if data is your core product**
- **Implementing a Data Policy improves the data quality and management**
- **Scientists are the main people to profit from these improvements**
- **Publishing text is not enough any more, data needs to be published too**
- **The changing landscape (move to Open Science) makes it essential to develop and adopt a Data Policy**



## Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data Catalog	Local	Icat	-	myMdC	-	SciCat

## After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Effort (PMs)	21	25	88	36	78	43
Automated Collection of Metadata	2015	2020	2023	2019	2023	2023
Metadata anthologies	Nexus	Nexus	Nexus	Nexus	YES	Nexus
Common API	YES	YES	YES	YES	YES	YES
Federated search	YES	YES	YES	YES	YES	YES

AFAMIN-revi-B5-1\_1\_1719747

Crystallography

Summary

Crystallography

Metadata List

Files **1640**

DOI

Name AFAMIN-revi-B5-1\_1\_1719747

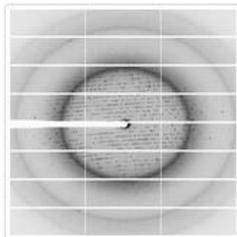
Definition

Start 7:34:00 PM

End 4:52:00 PM

Sample

Description



/data/id30a1/inhouse/opid30a1/20160408/RAW\_DATA/AFAMIN/AFAMIN-revi-B5-1

Download





## Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data Analysis Services	ssh	ssh	-	Jupyter	-	-

## After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Effort (PMs)	71	36	60	60	50	32
Jupyter notebook service	YES	YES	YES	YES	YES	YES
Remote desktop	YES	YES	YES	YES	YES	YES
Preconfigured VMs	YES	YES	YES	YES	YES	YES
Web applications	YES	YES	YES	YES	YES	YES



Select items to perform actions on them.

Upload New ↕ ↻

0 /		Name ↓	Last Modified	File size
<input type="checkbox"/>	ResultsScan0000		16 minutes ago	
<input type="checkbox"/>	ResultsScan0013		5 months ago	
<input type="checkbox"/>	ResultsScan0014		5 months ago	
<input type="checkbox"/>	ResultsScan0047		5 months ago	
<input type="checkbox"/>	vincent		an hour ago	
<input type="checkbox"/>	 Ptycho-CXI-id01-known-probe.ipynb		2 hours ago	1.27 MB
<input type="checkbox"/>	 Ptycho-CXI-id01-unknown-probe.ipynb		seconds ago	4.83 kB
<input type="checkbox"/>	 cuda_profile_0.log		4 months ago	7.92 kB
<input type="checkbox"/>	 cuda_profile_1.log		4 months ago	314 B
<input type="checkbox"/>	 cuda_profile_2.log		4 months ago	7.92 kB
<input type="checkbox"/>	 cuda_profile_3.log		4 months ago	3.21 kB
<input type="checkbox"/>	 cuda_profile_4.log		4 months ago	4.04 MB
<input type="checkbox"/>	 go_cu.sh		4 months ago	214 B
<input type="checkbox"/>	 params.nvvp		4 months ago	758 B
<input type="checkbox"/>	 params2.nvvp		4 months ago	4.78 MB
<input type="checkbox"/>	 S013.cxi		5 months ago	11.3 MB
<input type="checkbox"/>	 S014.cxi		5 months ago	11.1 MB
<input type="checkbox"/>	 S047.cxi		5 months ago	11.8 MB

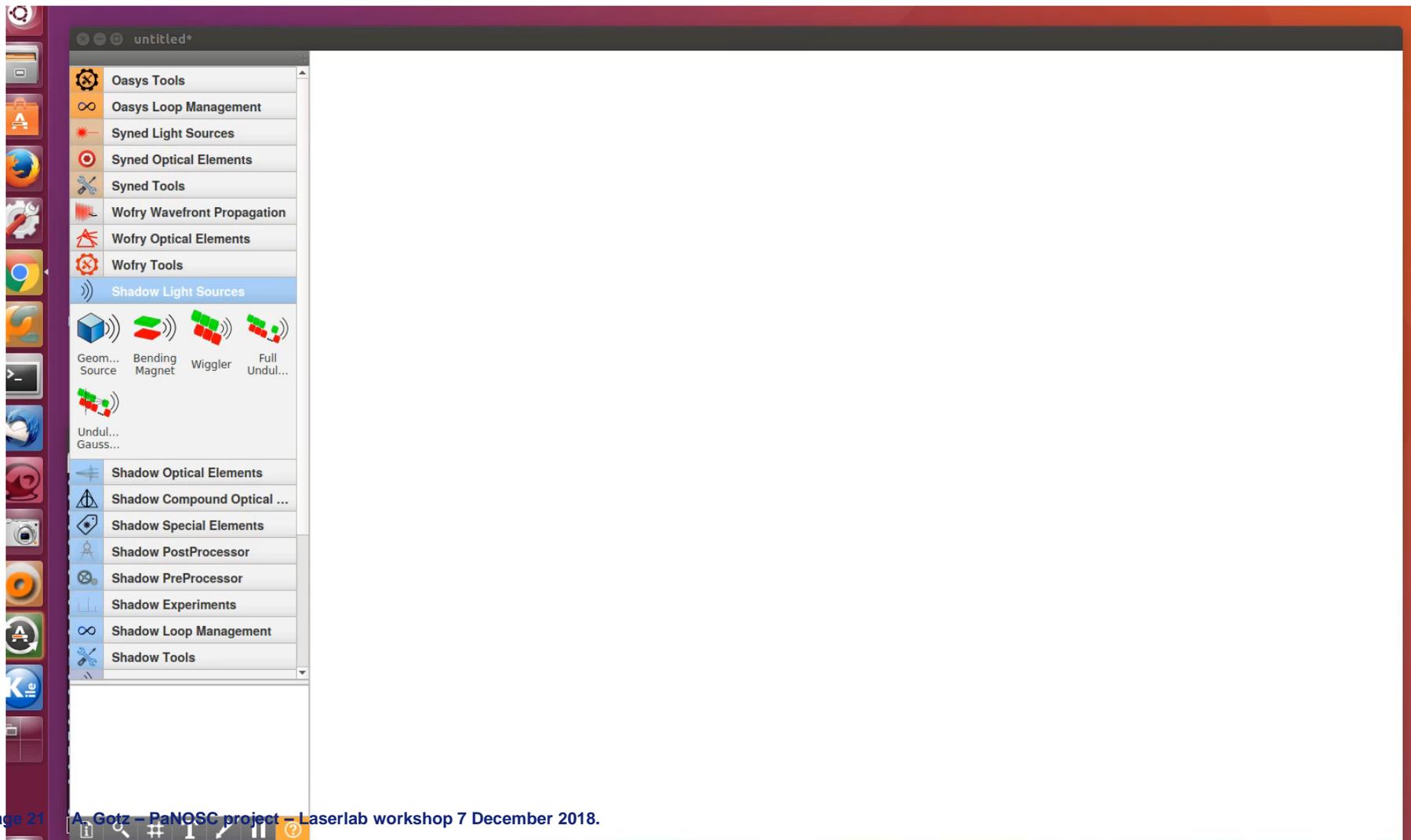


## Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data Simulation Services	local	OASYS	?	SIMEX	?	QENS, INS

## After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Effort (PMs)	36	40	40	48	24	36
Simulation service	YES	YES	YES	YES	YES	YES
OASYS packaging	-	YES	YES	YES	-	YES
Simulation notebooks	YES	YES	YES	YES	YES	YES
Documented API	YES	YES	YES	YES	-	YES



## Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
EOSC Integration	-	-	-	-	-	-

## After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS	EGL.eu	GEANT
Effort (PMs)	38	21	13	13	12	13	82	free
Define needs	YES	YES	YES	YES	YES	YES	-	-
Integrate PaN S/W in EOSC	YES	YES	YES	YES	YES	YES	YES	-
AAI	YES	YES	YES	YES	YES	YES	YES	YES
Open Data	YES	YES	YES	YES	YES	YES	-	-
Commercial cloud services	YES	YES	YES	YES	YES	YES	-	broker



## What is the European Open Science Cloud?

Iulia Popescu on what we know so far about the ambitious European initiative



**EUROPEAN OPEN  
SCIENCE CLOUD**

The idea of a [European Open Science Cloud](#) (EOSC) took shape in 2015, as a vision of the European Commission of a large infrastructure to support and develop open science and open innovation in Europe and beyond.

The EOSC is projected to become a reality by 2020 and will be Europe's virtual environment for all researchers to store, manage, analyse and re-use data for research, innovation and educational purposes.

”

**The European Open Science Cloud vision : “to give Europe a global lead in scientific data infrastructures and to ensure that European scientists reap the full benefits of data-driven science”. European Cloud Initiative publication**

### More information

[Iulia Popescu](#) is a Communications Officer at the EGI Foundation.

[The European Open Science Cloud was officially launched in November 2018.](#)

EOSC-related publications:

[European Cloud Initiative – Building a competitive data and knowledge economy in Europe](#) (2016)





## Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
User training	-	-	-	-	-	MOOC

## After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS	EGL.eu
Effort (PMs)	9	6	6	4	48	30	platform
E-learning platform	YES	YES	YES	YES	YES	YES	-
Teaching material	YES	YES	YES	YES	YES	YES	-
Staff training	YES	YES	YES	YES	YES	YES	YES
User training	YES	YES	YES	YES	YES	YES	YES
Hercules school	YES	YES	YES	YES	YES	YES	-





## Courses

[\(login required\)](#)

### My courses

Quick access to the courses you have already enrolled in.



### HERCULES

#### Introduction to Neutron Scattering

An introductory course to neutron scattering, intended for students at master level or above.

HERCULES school provides training for students, postdoctoral and senior scientists from European and non-European universities.

#### Muon Spin Spectroscopy

A course on a complementary technique to neutron scattering



### Introduction to neutron scattering

#### MY COURSES

Quick access to the courses you have already enrolled in.

#### Library

A collection of link to external resources.

course you will learn about the basic techniques used to perform experiments in neutron scattering. [READ MORE](#)

## Science cases

[\(login required\)](#)



### Finding crystal structure

Chemistry of materials



### Characterising liposomes in suspension

Life sciences



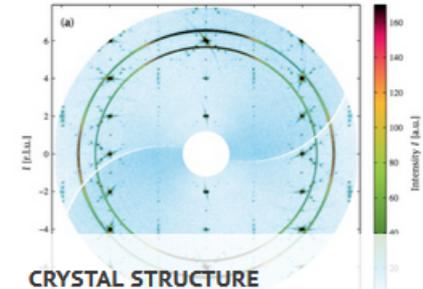
### Characterising magnetic order

Magnetic and electronic phenomena



### Characterising atomic lattice vibrations

Energy research



### CRYSTAL STRUCTURE

Try module "Diffraction from crystalline materials" in course "Introduction to Neutron Scattering"

[READ MORE](#)



- 1. FAIR data** – more difficult to implement than most believe
  - ✓ Implementing an electronic logbook as part of the RICH metadata capture
  - ✓ Promote use of Jupyter notebooks and workflows to capture data analysis
- 2. Integration** - services linked by a supported federated identity scheme covering more of the research life cycle where users access data, software, IT capacity and the expertise for performing analysis
  - ✓ GEANT will help PaNOSC by hosting AAI, ESFRIs to provide expertise
- 3. Hybrid model** - should not compete with but rather profit from ease of use and rates of innovation of commercial service providers
  - ✓ PaNOSC will procure and integrate commercial services
- 4. Provenance, citation and use of data & software**
  - ✓ How to get users to cite DOIs and provide Open Data
- 5. Business model** of how to provide services to all scientists and general public
  - ✓ ESFRI Photon and Neutron RIs have funding for Users who come to the source, they do not have funding for providing services for Open Data



**PaNOSC** is a joint effort to help the ESFRI Photon and Neutron sources to **adopt** and **implement data management, simulation and analysis services** for dealing with the increasing data rates and volumes and making **Open Data** available to the **EOSC**.



