PaNOSC

Photon **a**nd **N**eutron **O**pen **S**cience **C**loud

Andy Götz (ESRF)
on behalf of the PaNOSC consortium

7 December 2018
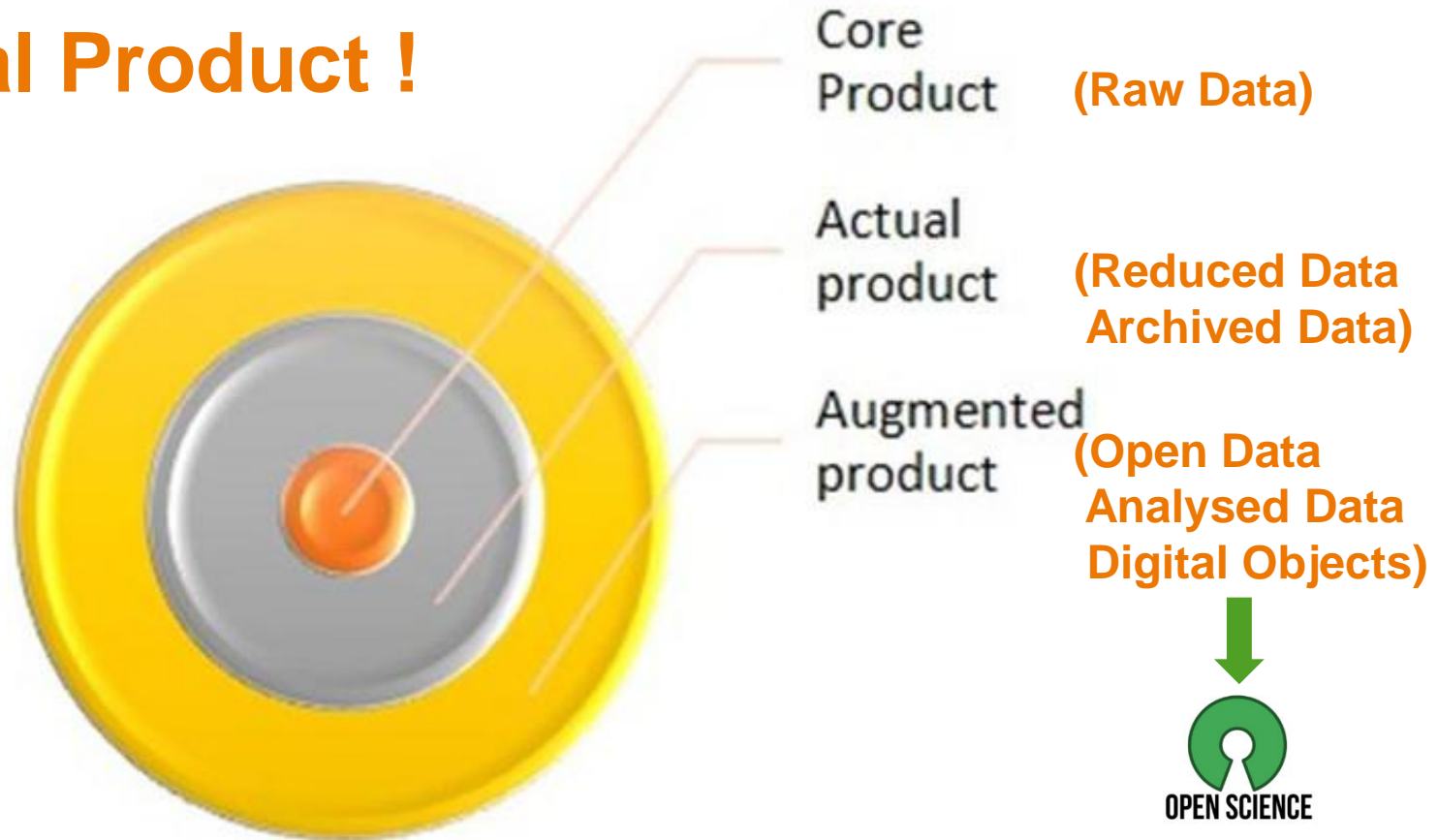
The European Synchrotron

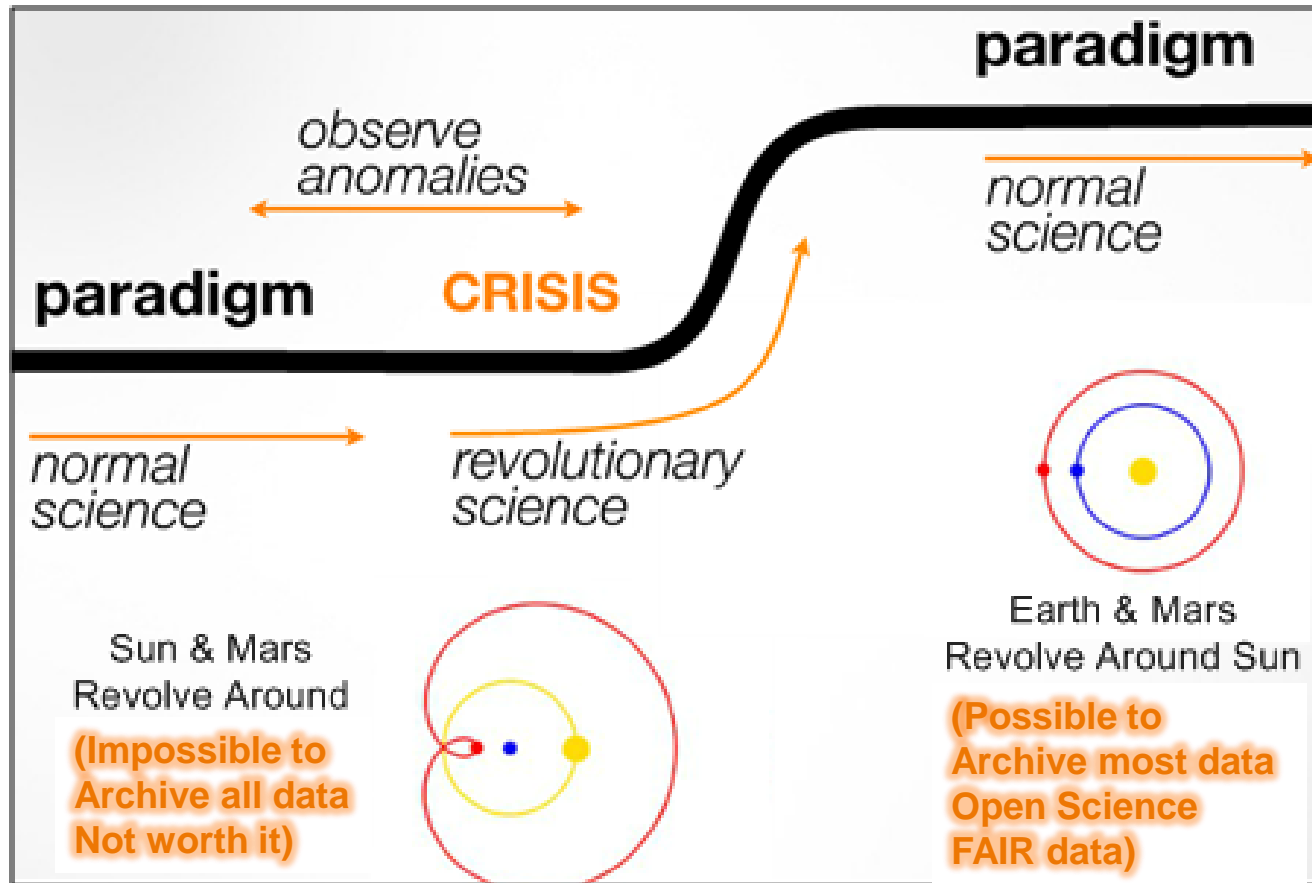## Recall: New science = a lot of data!

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
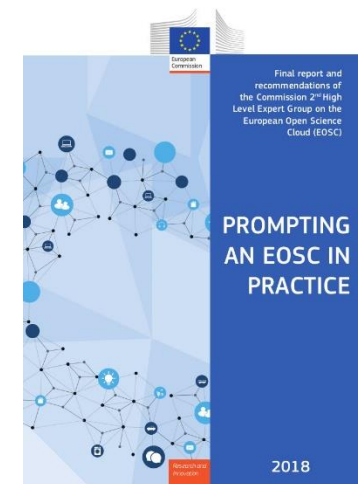- Long term storage of the data (regulations and requirements)

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.

The European Synchrotron

Data is our Actual Product !

Core Product **(Raw Data)**

Actual product **(Reduced Data Archived Data)**

Augmented product **(Open Data Analysed Data Digital Objects)**

OPEN SCIENCE

The European Synchrotron

| Domain cluster projects (H2020-INFRAEOSC-04-2018 call) | » Interoperability and service architecture
» Service provisioning
» Data provisioning
» Community building
» Policy development
» Skills & capabilities development
» Discipline specific approach
NB:
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. | » Research performing organisations
» Scientific communities and individual researchers
» Funding bodies |

PROMPTING AN EOSC IN PRACTICE

Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC)

2018

EOSC-LIFE, ENVRI-FAIR, ESCAPE, SSHOC, <u>PANOSC</u>

The European Synchrotron

Data Stewardship
- Generalise the adoption of **FAIR** open data principles
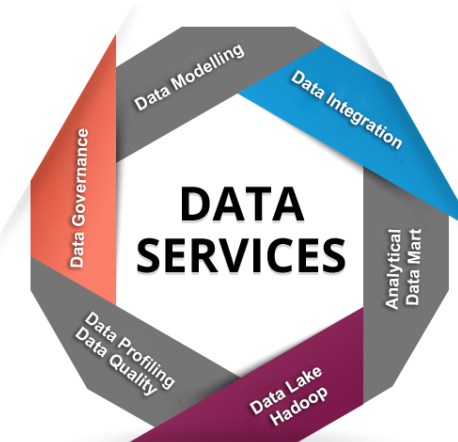
EOSC
- Federate data catalogs and integrate in EOSC and OpenAire meta-catalog, AAI, download services

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

The European Synchrotron

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 3rd parties via EGI: DESY, STFC, CESNET

Decision: Accepted on 13/8/2018

Status: Started 1/12/2018
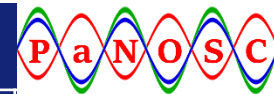
Home page: https://panosc.eu

- **Budget:** 12 M€
- **Coordinator:** ESRF
- **Planned started:** 1/12/2018
- **Project duration:** 4 years
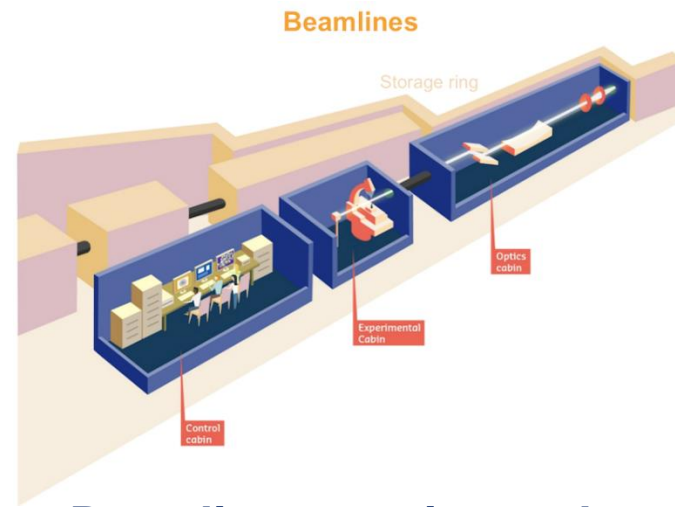
Kick-off meeting
15-16 January at ESRF

Photon and **N**eutron **O**pen **S**cience **C**loud

The European Synchrotron

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)

Beamlines

Storage ring

Optics cabin

Experimental Cabin

Control cabin

Beamline experimental station

The European Synchrotron

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)

THE EXTREME LIGHT INFRASTRUCTURE
A distributed pan-European laser user facility open to the world

XFEL (Hamburg, Germany)

The European Synchrotron

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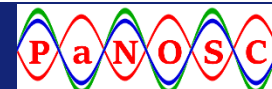
	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data / yr	200 TB	8 PB	1 PB	3PB	?	0
Data Policy	2011	2016	2014 (3/8)	2017	?	2017
Metadata catalogue	Local	Icat	Local	myMdC	?	SciCat
Metadata definitions	Nexus	Nexus	custom	myMdC	?	Nexus
DOI	2012	2018	No	2018	?	2018
Open Data	yes	yes	No	yes	no	yes
Data Services	pilot	in progress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

After PaNOSC - 2023
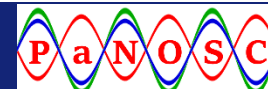
	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data / yr	600 TB	50 PB	15 PB	100PB	10 PB	<1PB
Data Policy	2011	2016	2019	2017	2019	2017
Metadata catalogue	Local	Icat	Icat	myMdC	[Icat]	SciCat
Metadata	Nexus	Nexus	Nexus	Nexus	[Nexus]	Nexus
DOI	Yes	Yes	Yes	Yes	Yes	Yes
Open Data	Yes	Yes	Yes	Yes	Yes	Yes
Common data API	Yes	Yes	Yes	Yes	Yes	Yes
User training	Yes	Yes	Yes	Yes	Yes	Yes
Data Services	Prod	Prod	Prod	Prod	Prod	Prod
EOSC	Integrated	Integrated	Integrated	Integrated	Integrated	Integrated

WP2 Data Policy – led by ESRF

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 Data Policy

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 Indentifier (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 guidelines

- **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**

The European Synchrotron

WP3 Data Catalog Services – led by ESS

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

https://datahub.esrf.fr

DOI > 10.15151/ESRF-DC-142893590

Data collection

Dataset Open access

STRUCTURAL EVIDENCE FOR A ROLE OF THE MULTI-FUNCTIONAL HUMAN GLYCOPROTEIN AFAMIN IN WNT TRANSPORT

Andreas Naschberger ; Matthew W. Bowler ; Bernhard Rupp.

DOI

DOI 10.15151/ESRF-DC-142893590

Licence (for files)

Creative Commons Attribution 4.0

Abstract

Afamin, a human plasma glycoprotein and putative transporter of hydrophobic molecules, has been shown to act as extracellular chaperone for poorly soluble, acylated Wnt proteins, forming a stable, soluble complex with functioning Wnt proteins. The 2.1-Å crystal structure of glycosylated human afamin reveals an almost exclusively hydrophobic binding cleft capable of harboring large hydrophobic moieties. Lipid analysis confirms the presence of lipids, and density in the primary binding pocket of afamin was modeled as palmitoleic acid, presenting the native O-acylation on serine 209 in human Wnt3a. The modeled complex between the experimental afamin structure and a Wnt3a homology model based on the XWnt8-Fz8-CRD fragment complex crystal structure is compelling, with favorable interactions comparable with the crystal structure complex. Afamin readily accommodates the conserved palmitoylated serine 209 of Wnt3a, providing a structural basis how afamin solubilizes hydrophobic and poorly soluble Wnt proteins.

Proposals	Beamlines	Publication year
OPID-1	ID30A1	2018

Experimental report

There is currently no experimental report.

Experimental data

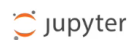The data can be accessed by clicking on the link below

Access data

WP4 Data Analysis Services – led by EuXFEL

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

A. Gotz – PaNOSC project – Laserlab workshop 7 December 2018.

The European Synchrotron

Jupyterhub – example running pynx

A. Gotz – PaNOSC project – Laserlab workshop 7 December 2018.

WP5 Virtual Neutron and x-raY Laboratory– led by EuXFEL

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

The European Synchrotron

WP6 EOSC integration – led by ILL
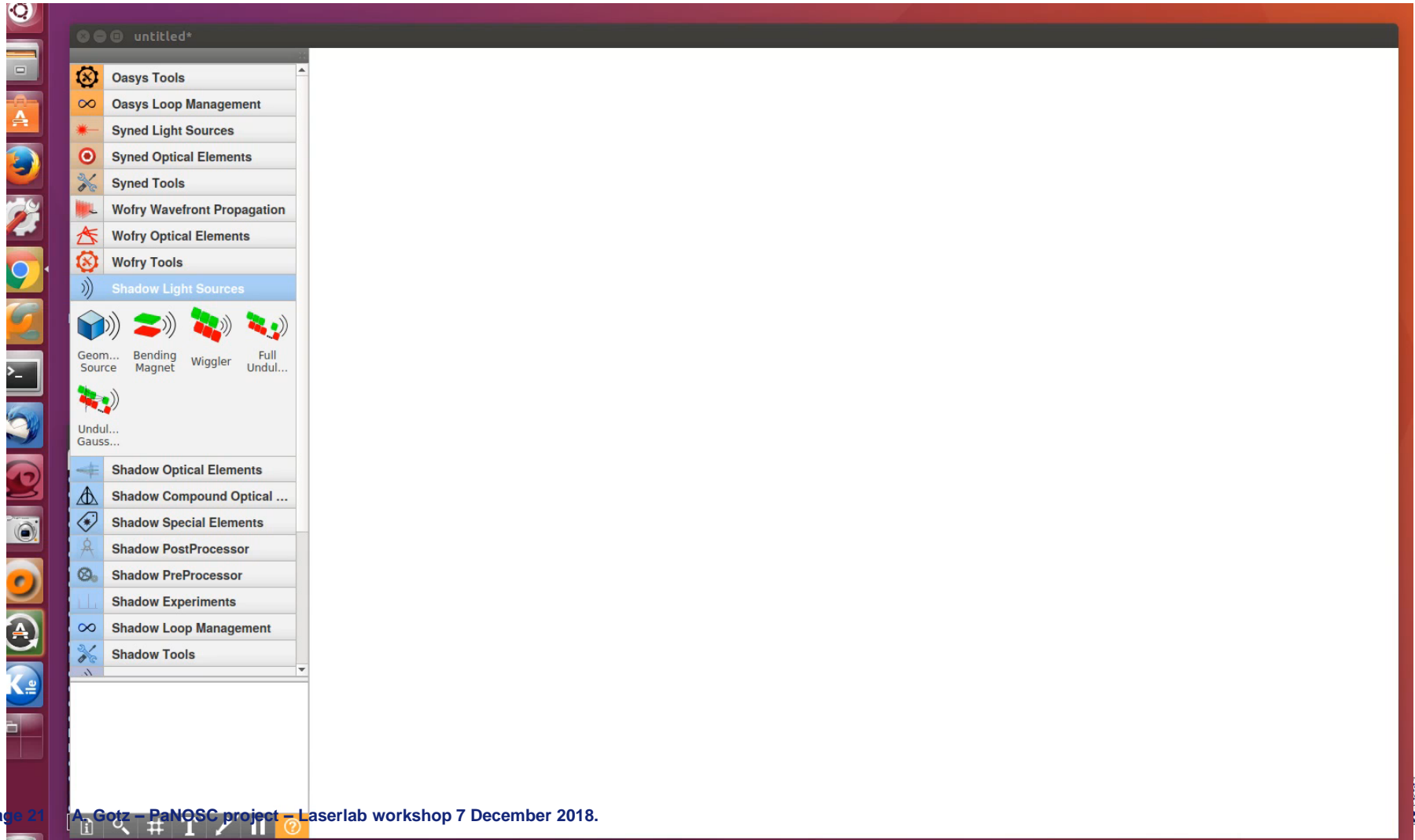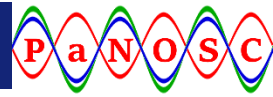
Before PaNOSC (2018)

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

After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS	EGI.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

PaNOSC

eGi

SERVICES | FEDERATION | USE CASES | BUSINESS | ABOUT

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)

The European Synchrotron

The European Synchrotron

WP8 Staff and User training – led by ESS

Before PaNOSC (2018)

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

After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS	EGI.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	-

e-neutrons.org online teaching platform (moodle)

A. Gotz – PaNOSC project – Laserlab workshop 7 December 2018.

The European Synchrotron

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

The European Synchrotron

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.**

Open Data & Open Science
builds on

Data Management, Simulation
& Analysis Services

Christian © www.hautesavoiephotos.com — Cordée sous l'Aiguille du Midi

The European Synchrotron