PaNOSC Overview and Status

13 May, 2019

Author: Andy Götz (ESRF)

Role: PaNOSC Coordinator

Place: LEAPS-IT meeting @ PSI

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

PaNOSC project - 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 Partners – ESFRI projects

NEUTRONS FOR SCIENCE

ESRF

European XFEL

egi

panosc

photon and neutron open science cloud

eli

ess EUROPEAN SPALLATION SOURCE

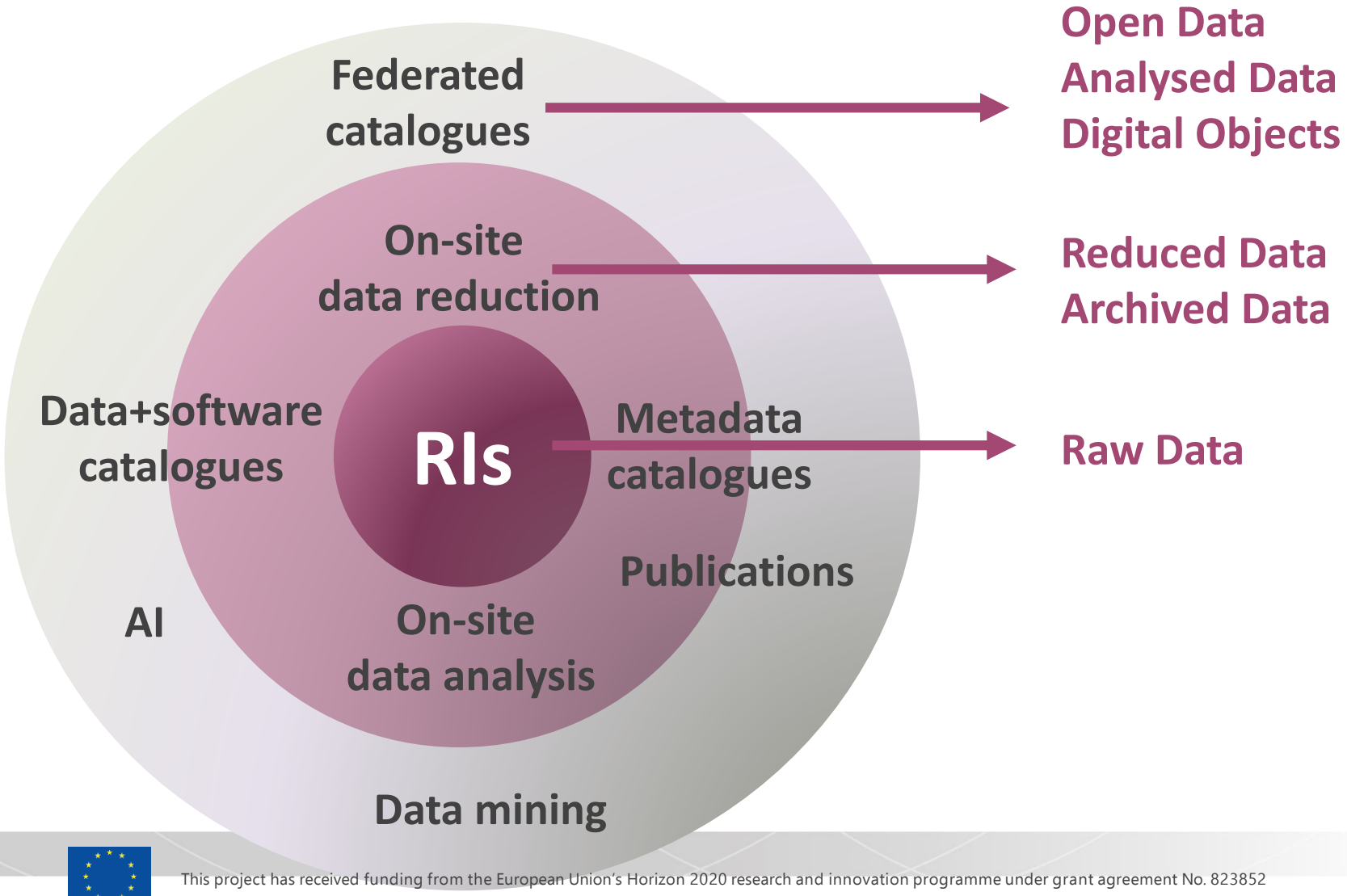
CERIC Central European Research Infrastructure Consortium

ESFRI European Strategy Forum on Research Infrastructures

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

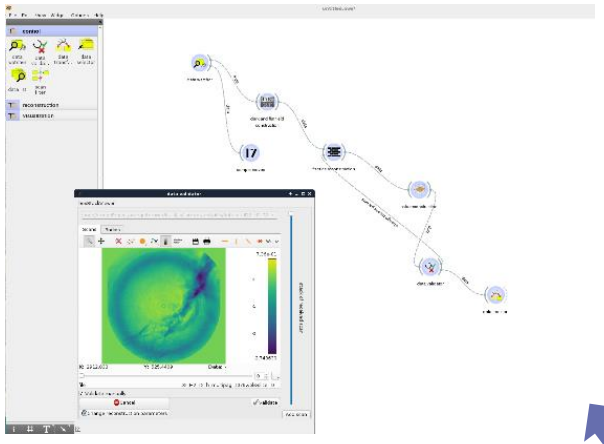
panosc

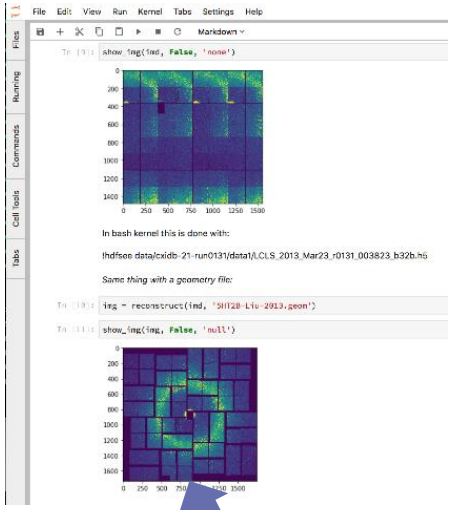
Data is our Product!

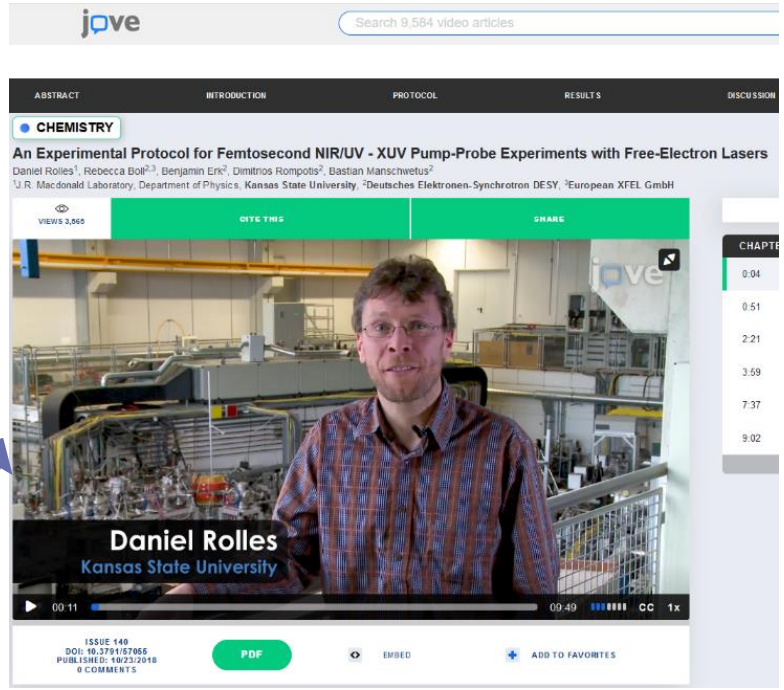


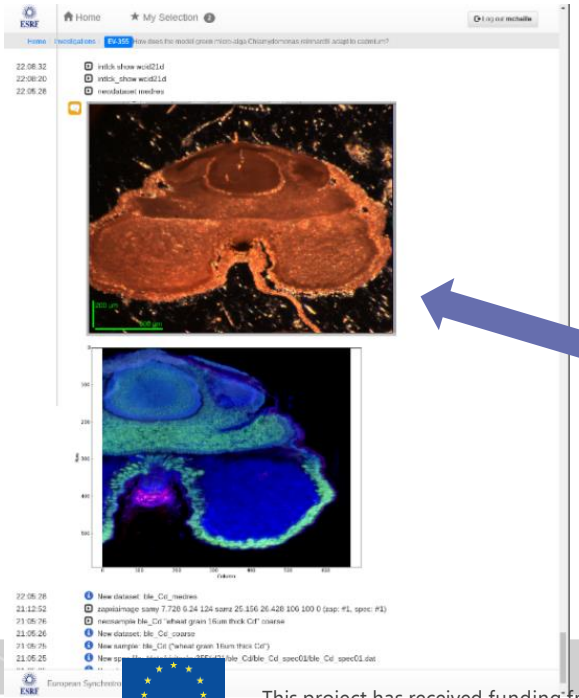
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Why – to link all scientific data and output together









notebooks
Workflow
video
Logbook
data
Software
preprints
papers

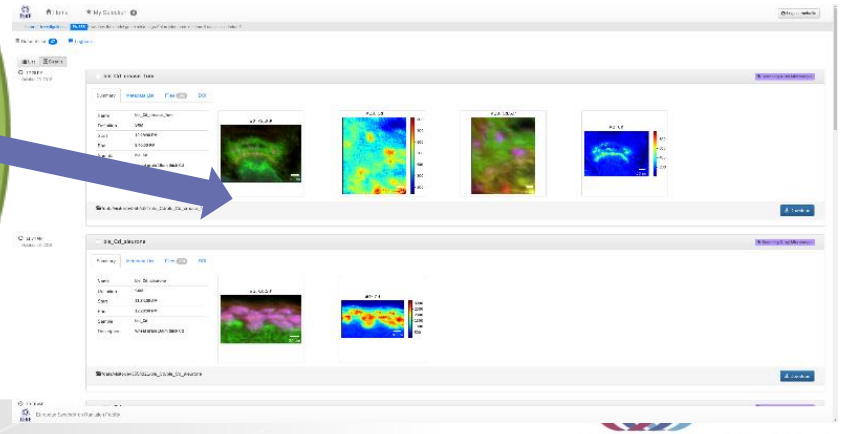


Image Source: http://michaelnielsen.org/blog/the-future-of-science-2/

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

Making FAIR a reality for PaNOSC

- How to make FAIR reality?
- How to make the EOSC reality?
- How to make Open Science reality?

- PaNOSC will build on and help make FAIR, EOSC and Open Science become reality for the Photon and Neutron community
- PaNOSC developments: new Data Policy framework, Nexus-compliant metadata, e-logbook, certified data catalogues, search API, data services, linking to EOSC, etc.

er grant agreement No. 823852

PaNOSC, ExPANDS and **EOSC** are about

making science

RELIABLE and REPRODUCIBLE

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Perspective

Cite this article: Hopf H, Krief A, Mehta G, Matlin SA. 2019 Fake science and the knowledge crisis: ignorance can be fatal. *R. Soc. open sci.* **6**: 190161.
http://dx.doi.org/10.1098/rsos.190161

Received: 25 January 2019
Accepted: 19 March 2019

Subject Category:
Chemistry

Subject Areas:
human-computer interaction

Keywords:
fake science, trust in science, knowledge crisis

Author for correspondence:
Stephen A. Matlin
e-mail: s.matlin@imperial.ac.uk

Fake science and the knowledge crisis: ignorance can be fatal

Henning Hopf[1,2], Alain Krief[1,3,4], Goverdhan Mehta[1,5] and Stephen A. Matlin[1,6]

[1]International Organization for Chemical Sciences in Development, 61 rue de Bruxelles, 5000 Namur, Belgium
[2]Institute of Organic Chemistry, Technische Universität Braunschweig, Braunschweig 38106, Germany
[3]Chemistry Department, Namur University, 5000 Namur, Belgium
[4]HEJ Research Institute, University of Karachi, Karachi City, Sindh 75270, Pakistan
[5]School of Chemistry, University of Hyderabad, Hyderabad 500046, India
[6]Institute of Global Health Innovation, Imperial College London, London SW7 2AZ, UK

HH, 0000-0001-7040-6506; AK, 0000-0002-9223-1644; GM, 0000-0001-6841-4267; SAM, 0000-0002-8001-1425

Computers, the Internet and social media enable every individual to be a publisher, communicating true or false information instantly and globally. In the 'post-truth' era, deception is commonplace at all levels of contemporary life. Fakery affects science and social information and the two have become highly interactive globally, undermining trust in science and the capacity of individuals and society to make evidence-informed choices, including on life-or-death issues. Ironically, drivers of fake science are embedded in the current science publishing system intended to disseminate evidenced knowledge, in which the intersection of science

editorial

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Findable Accessible Interoperable Re-usable (FAIR) diffraction data are coming to protein crystallography

John R. Helliwell,[a]‡ Wladek Minor,[b]§ Manfred S. Weiss,[c]¶ Elspeth F. Garman,[d]‡‡ Randy J. Read,[e]‡‡ Janet Newman,[f]§§ Mark J. van Raaij,[g]§§ Janos Hajdu[h,i]¶¶ and Edward N. Baker[j]‡‡‡

This project has received funding from the Europe

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	2011	2016	2019	2017	2019	2017
Metadata catalogue 2018	Local	Icat	Local	myMdC	No	SciCat
Metadata catalogue 2023	Local	Icat	Icat	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 KPIs

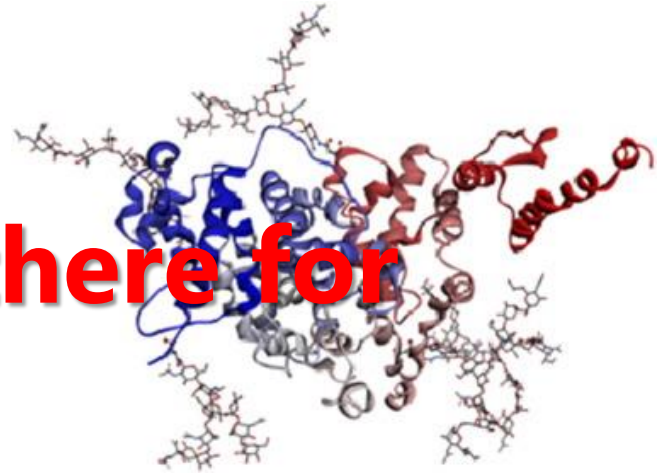
	ILL	ESRF	CERIC	XFEL	ELI	ESS
Open Data 2018	100s	2	0	10s	0	0
Open Data 2023	1000s	1000s	100s	1000s	100s	10s
Data Services 2018	Pilot	In progress	Remote	In progress	?	In progress
Data Services 2023	Desktop Jupyter	Jupyter Desktop	Jupyter Desktop	Jupyter Desktop	Desktop Jupyter	Jupyter Desktop
Common data 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

Sharing data across domains & sites

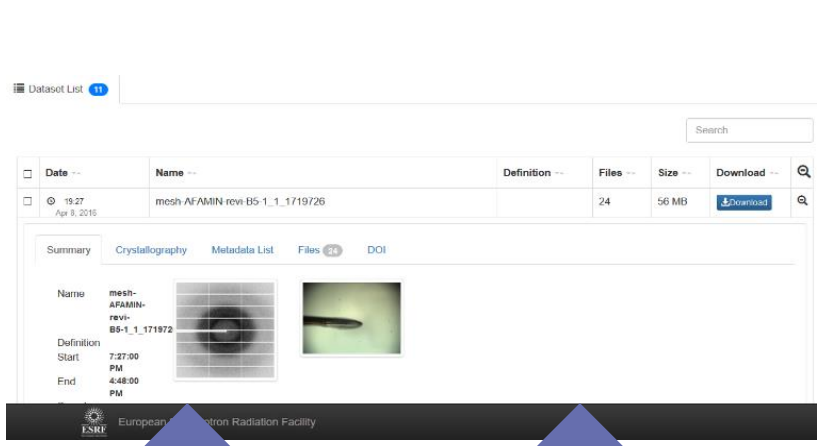
1. **Find all xray diffraction datasets of Afamin protein**

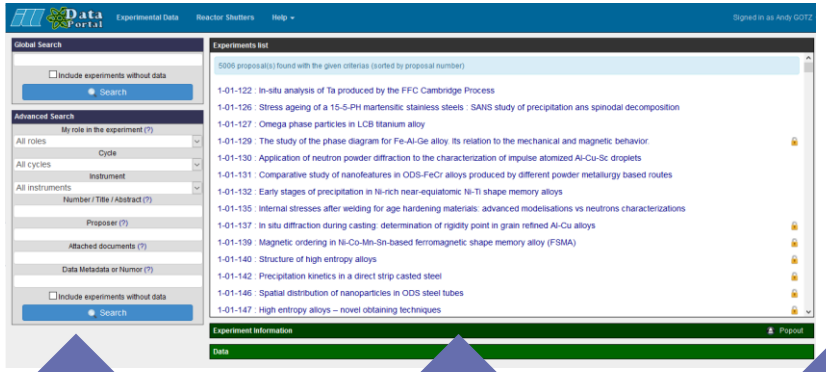
2. **How many NMR datasets are there for Afamin and Wnt proteins ?**

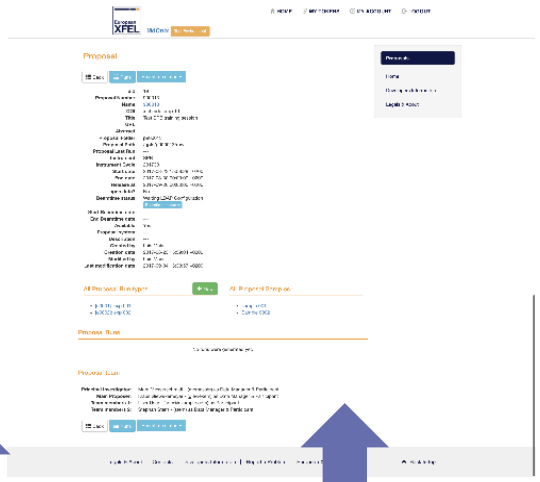


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

PaNOSC has 6 data catalogues with different APIs + UIs







ESRF (icat)

CERIC (icat)

ESS (SciCat)

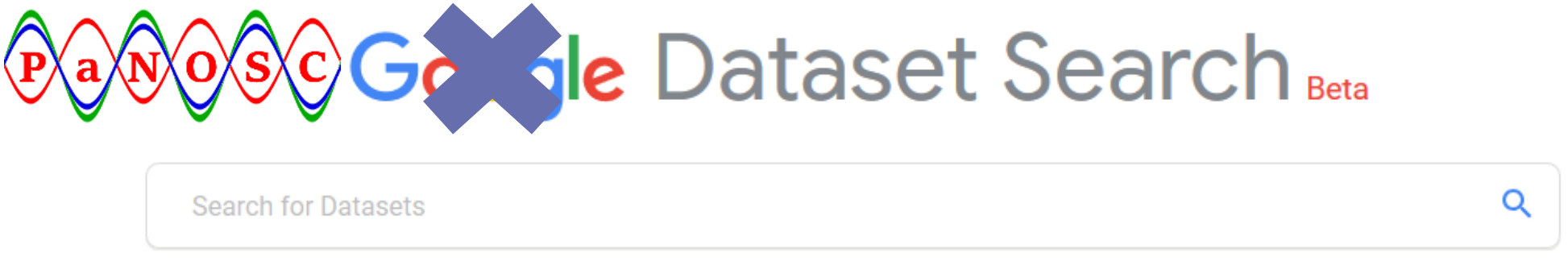
ILL (local)

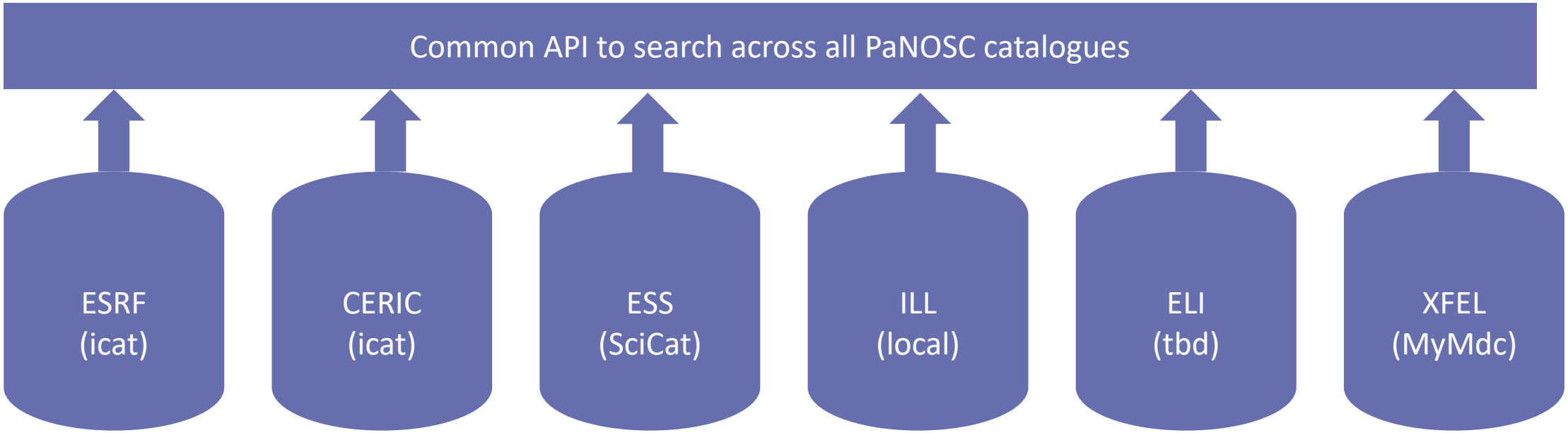
ELI (tbd)

XFEL (MyMdc)

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

PaNOSC is implementing a common API searchable across sites

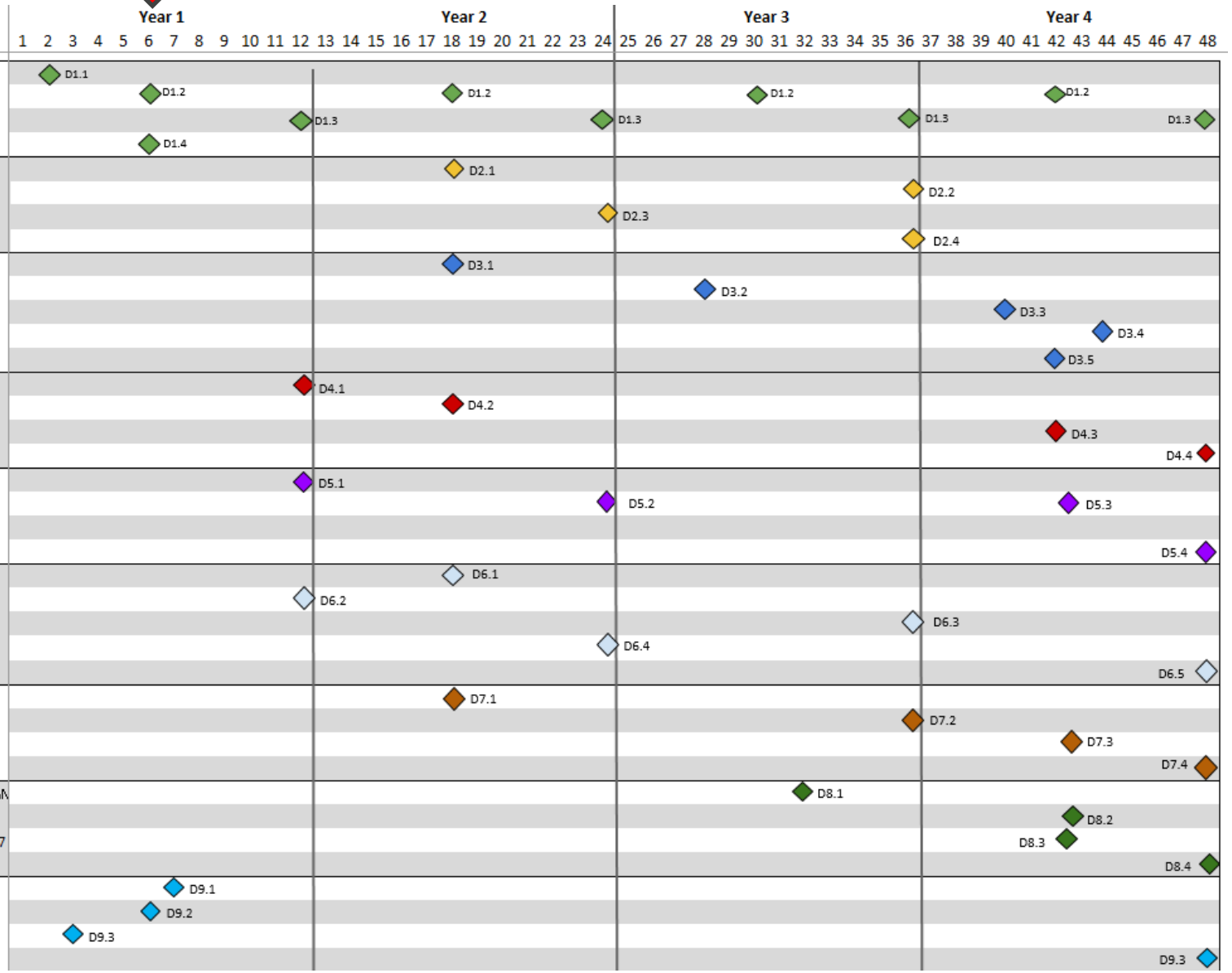




This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

PaNOSC Deliverables

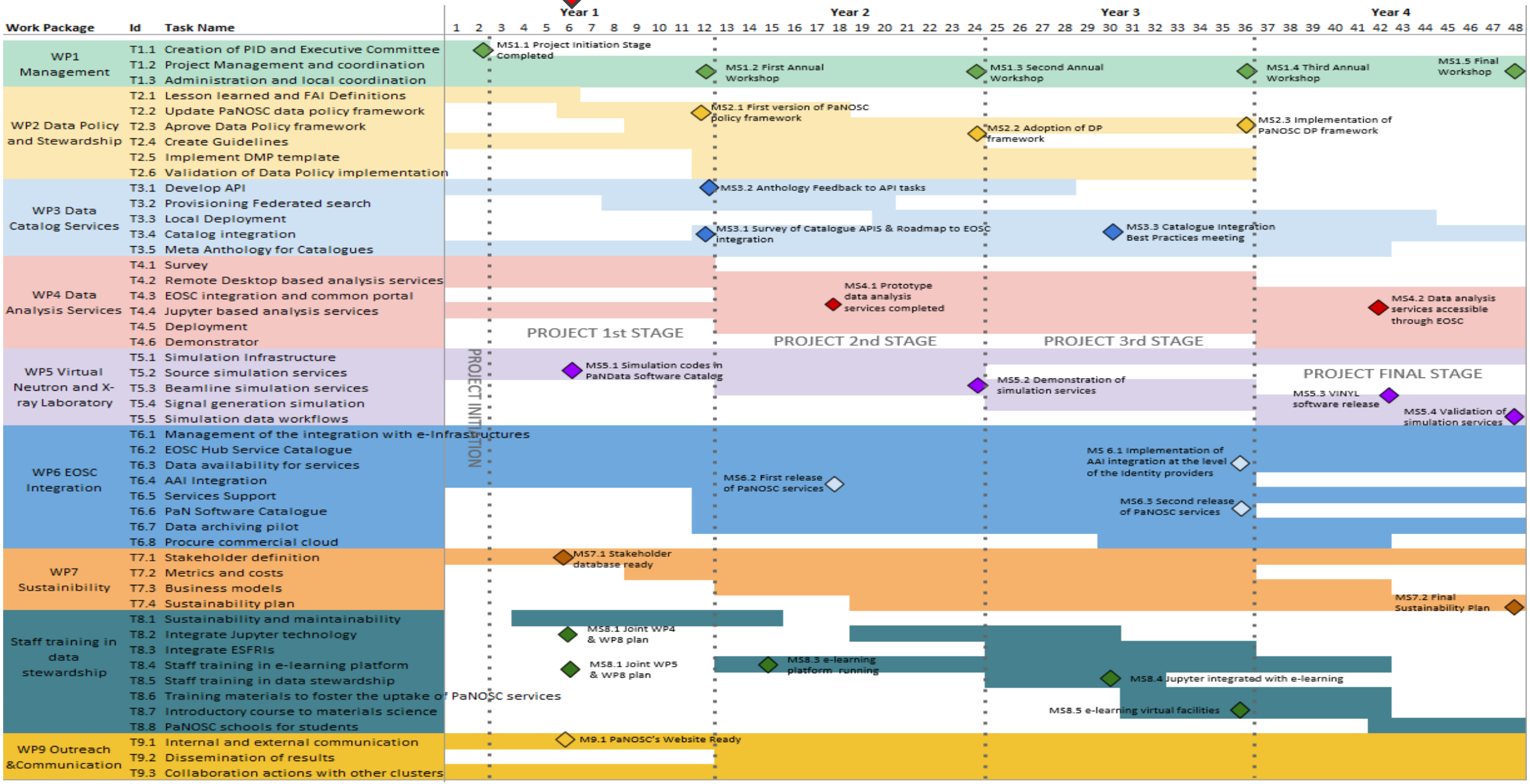
May 2019



Work Package	Id	Task Name
WP1 Management	D1.1	Project initiation Documentation
	D1.2	Mid-Year summary
	D1.3	Annual workshop report
	D1.4	Data management plan
WP2 Data Policy and Stewardship	D2.1	PaNOSC data policy framework updated
	D2.2	TMP Template published
	D2.3	Guidelines published
	D2.4	Integration of the policy in the User Access & Facility information systems
WP3 Data Catalog Services	D3.1	API Definition
	D3.2	Demonstrator implementation
	D3.3	Catalog service
	D3.4	Implementation report from facilities
	D3.5	NeXus Metadata Mapping Schema and proposed new Definitions
WP4 Data Analysis Services	D4.1	Report data analysis capture
	D4.2	Prototype remote desktop and Jupyter service
	D4.3	Remote desktop and Jupyter analysis service tested at EOSC
	D4.4	Jupyter based analysis services
WP5 Virtual Neutron and X-ray Laboratory	D5.1	Prototype simulation data formats
	D5.2	Documented simulation APIs
	D5.3	Documented simulation tasks executable
	T5.4	Software tested & released including interactive simulation & analysis workflow
WP6 EOSC Integration	D6.1	DataHub: EGI DataHub integration with the facilities data repositories
	D6.2	Compute cloud: integration of local compute resources into the EOSC cloud
	D6.3	AAI: Integration of the PaN AAI into the EOSC
	D6.4	Software catalogue: Demonstration of the PaN software catalogue integration into EOSC
	D6.5	Report on EOSC integration
WP7 Sustainibility	D7.1	Photon and Neutron EOSC Stakeholder Feedback
	D7.2	Photon and Neutron EOSC metrics and costs model
	D7.3	Photon and Neutron EOSC Business model reference document
	D7.4	Photon and Neutron EOSC Sustainability plan
Staff training in data stewardship	D8.1	Report on lessons learned and future prospects for adopting best practises data stewardship at the PaN
	D8.2	Report on lessons learned for adopting the e-learning platform at the PaNOSC facilities, task 8.4
	D8.3	Teaching material accessible in the e-learning platform at panlearning.org, task 8.5-7
	D8.4	Closing report including report from summer school, task 8.8
WP9 Outreach &Communication	D9.1	PaNOSC's Communication and Dissemination Plan
	D9.2	PaNOSC's Website
	D9.3	PaNOSC's repository for internal communication
	D9.4	Dissemination and Outreach activities

PaNOSC Milestones

May 2019



Deliverables status in May 2019

Completed

D1.1 – Project Initiation Documentation
D9.3 – Repository for internal communications

In progress

D1.2 – Mid-year summary
D1.4 – Data Management Plan
D9.2 – Website

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

Milestones status in May 2019

Completed

MS1.1 – Project Initiation Stage
MS5.1 – Simulation codes in PaNdata catalog

In progress

MS7.1 – Stakeholder database
MS8.1 – Joint WP4 and WP8 plan
MS8.1 – Joint WP5 and WP8 plan
MS9.1 – PaNOSC website ready

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

PaNOSC's Objectives

1. **Participate** in the construction of the EOSC by linking with the e-infrastructures and other ESFRI clusters.
2. **Make** scientific data produced at Europe's major Photon and Neutron sources fully compatible with the FAIR principles.
3. **Generalise** the adoption of open data policies, standard metadata and data stewardship from 15 photon and neutron RIs and physics institutes across Europe.
4. **Provide** innovative data services to the users of these facilities locally and the scientific community at large via the EOSC.
5. **Increase** the impact of RIs by ensuring data from user experiments can be used beyond the initial scope.
6. **Share** the outcomes with the national RIs who are observers in the proposal and the community at large to promote the adoption of FAIR data principles, data stewardship and the EOSC.

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

Objective 1 – Integrate RIs with EOSC

Participate in the construction of the EOSC by linking with the e-infrastructures and other ESFRI clusters.

Progress –

ILL (WP6 leader) has organised regular meetings (with minutes)
Good interaction with EGI, GEANT
GEANT working on prototype AAI based on eduTEAMS
CESNET ready to deploy Jupyter service for PaNOSC (waiting for AAI)
ILL, ESRF, CERIC participated in EOSC-hub week
XFEL, ESRF, ILL to participate in EGI-week

Weaknesses –

No involvement from ELI, CERIC, ESRF in WP6 (so far)
Problem of data transfer is not solved (Data One not the solution)

Next Steps –

Test FTS3 file transfer, Transfer large data volumes to STFC,
Align configuration of Jupyter+Kubernetes in PaNOSC with EGI, Provision GPUs
EOSC-hub to decide which solution to promote for AAI

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

Objective 2 – FAIR data

Make scientific data produced at Europe's major Photon and Neutron sources fully compatible with the FAIR principles.

Progress –

ESRF (WP2 leader) has organised regular meetings (with minutes)
Good interaction with EOSC-hub (recent deliverable D9.3 on data policy v. useful)
Analysis of current data policies, what to keep and what to enhance
CERIC started a Lessons learned document, ESRF wrote paper for SRN
ESS attended the FAIRsFAIR kickoff, presented PaNOSC questions and established contact

Weaknesses –

No involvement from CERIC partners (so far)
Adopting FAIR data policy for ELI (workshop in May)
Modifying existing data policies (ILL, ESRF, XFEL, ESS)

Next Steps –

Complete Lessons learned document, Prepare draft FAIR-compliant Data Policy, Analyse best practices for metrics

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

Objective 3 – Open data policies

Generalise the adoption of open data policies, standard metadata and data stewardship from 15 photon and neutron RIs and physics institutes across Europe.

Progress –

Activities happening at each site e.g. ILL developed PUMA metrics tool (April)
ESRF organising a plenary meeting with scientists (June)
ELI organising workshop on data management (May)

Weaknesses –

No activities at CERIC partner sites e.g. Solaris, BNC, LASDAM, …
Adopting FAIR data policy for ELI (workshop in May)
Modifying existing data policies (ILL, ESRF, XFEL, ESS)

Next Steps –

Define KPIs for FAIR data policies and monitor,
Analyse best practices for metrics e.g. Crossref, Altmetrics, PUMA, …
Solaris to visit ESRF to study solutions for data management

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

Objective 4 – Data services

Provide innovative data services to the users of these facilities locally and the scientific community at large via the European Open Science Cloud (EOSC).

Progress –

Jupyter service is first common data service to be provided
Work on setting up at ESRF, XFEL (DESY), ILL, ESS advancing
EGI setup a demo service with limited resources
ESRF organising OASYS school for designing+simulating beamlines

Weaknesses –

No common Jupyter installation (so far), each site doing their own
No clear commitment from all partners which services they will provide

Next Steps –

Align Jupyter installations, Develop data visualisation for Jupyter notebooks
Provide OASYS-as-a-service

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

Objective 5 – Sharing data

Increase the impact of RIs by ensuring data from user experiments can be used beyond the initial scope.

Progress –

ILL has developed PUMA service for tracking data production and (re)use
WP3 developing a common API for search Open Data

Weaknesses –

No metrics for how much open data is and will be available
Not clear how FAIRsFAIR, EUDAT, etc. will help with this

Next Steps –

Define and monitor metrics for Open Data

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

Objective 5 – Working with ExPANDS

Share the outcomes with the national RIs who are observers in the proposal and the community at large to promote the adoption of FAIR data principles, data stewardship and the EOSC.

Progress –

EXPaNDS, a consortium of national RIs, has been funded to provide FAIR data in EOSC
First informal contact took place at EOSC-hub

Weaknesses –

EXPaNDS and PaNOSC end up not working closely together

Next Steps –

Define how PaNOSC and EXPaNDS will work together and share outcomes

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

PaNOSC - Software developments

1. AAI → ILL with GÉANT
2. Data catalogues → ICAT (ESRF), SciCat (ESS), MDC (EuXFEL), ILL
3. Common search API → ESS
4. E-logbook → ICAT+ (ESRF)
5. Jupyter on Kubernetes → ESRF, ILL, ESS, …
6. DAAS portal → Calipsoplus (ESRF+ALBA), Visa (ILL), …
7. Software catalogue → PaNdata catalogue (ILL)
8. Simulation software → Simex (EuXFEL), OASYS (ESRF)
9. e-Training platform → ESS, ELI,

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

Challenges

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 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 user friendliness and innovation of commercial service providers
   - ✓ PaNOSC will procure and integrate commercial services
4. **Provenance, citation and use** of data & software
   - ✓ Train 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, **but no funding for providing services for using Open Data. Will EOSC provide resources?**

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