photon and neutron
open science cloud

Review of Data Catalog Services (WP3)

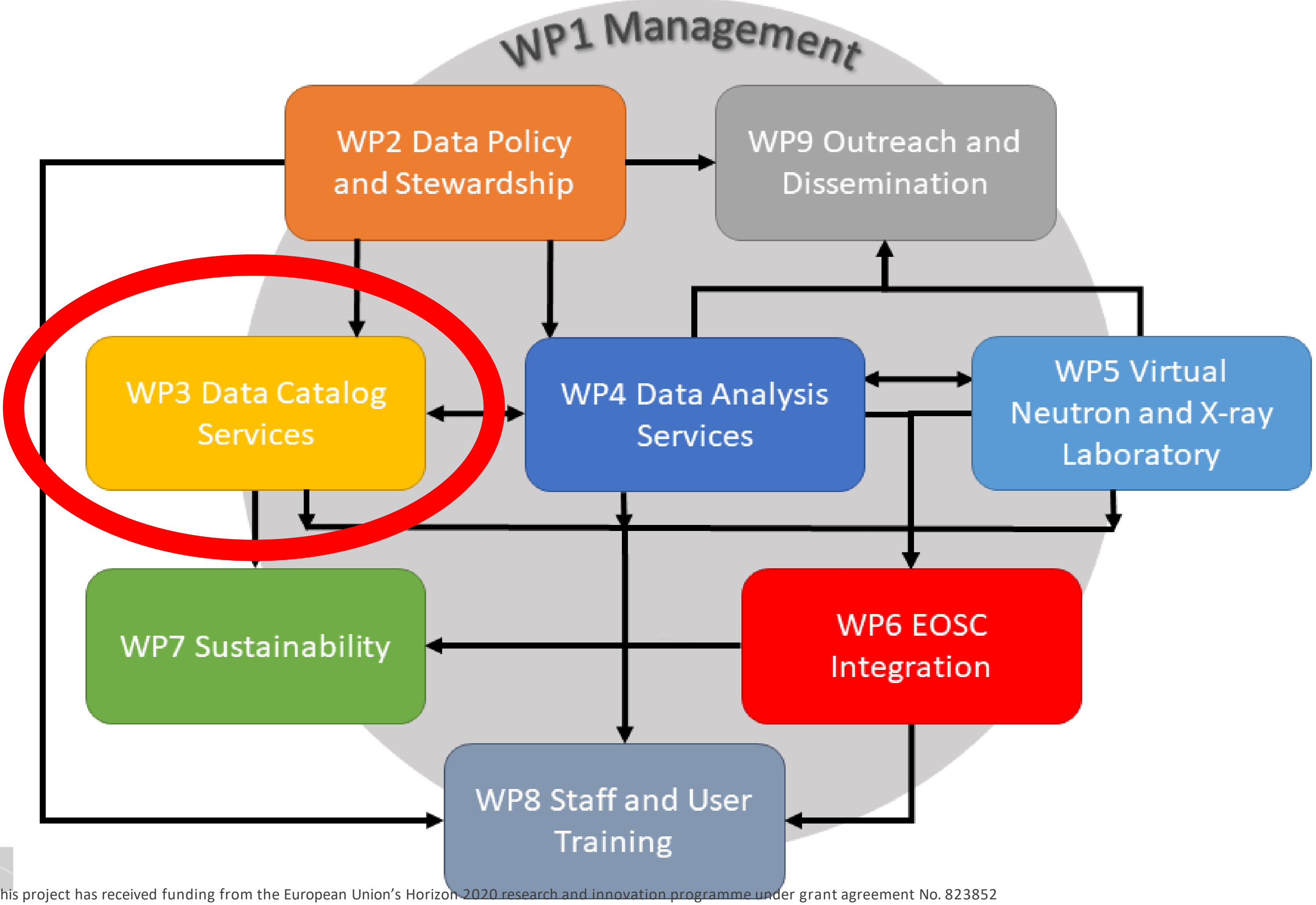
12 May 2021

Authors: Andy Götz (coordinator)+ Tobias Richter (WP3 leader)

Place: PaNOSC Project Management Committee zoom meeting

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

Data Catalog Services – their role in PaNOSC



WP1 Management

WP2 Data Policy and Stewardship

WP9 Outreach and Dissemination

WP3 Data Catalog Services

WP4 Data Analysis Services

WP5 Virtual Neutron and X-ray Laboratory

WP7 Sustainability

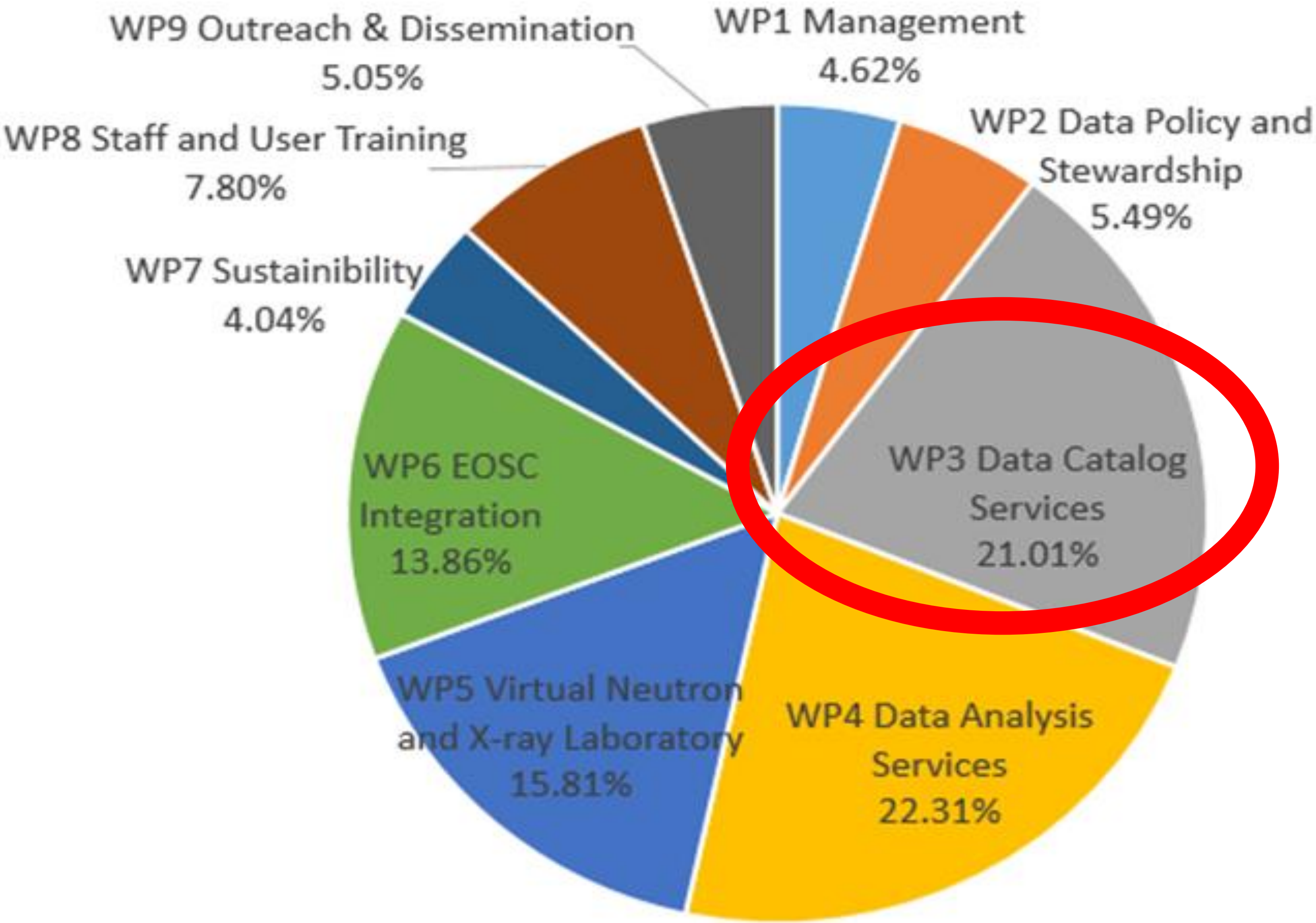
WP6 EOSC Integration

WP8 Staff and User Training

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

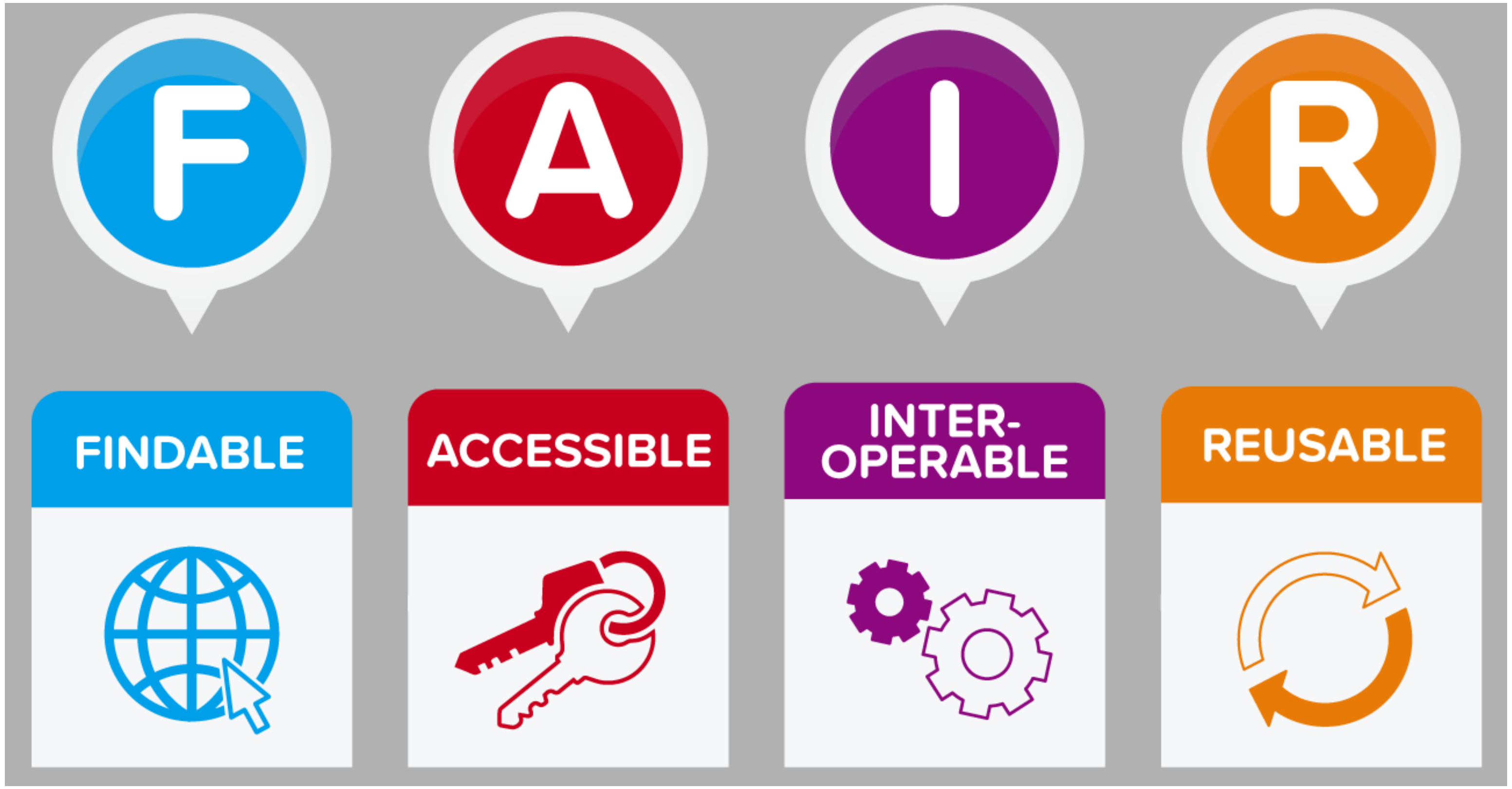
Data Catalog Services – 2nd largest WP of PaNOSC

WP3



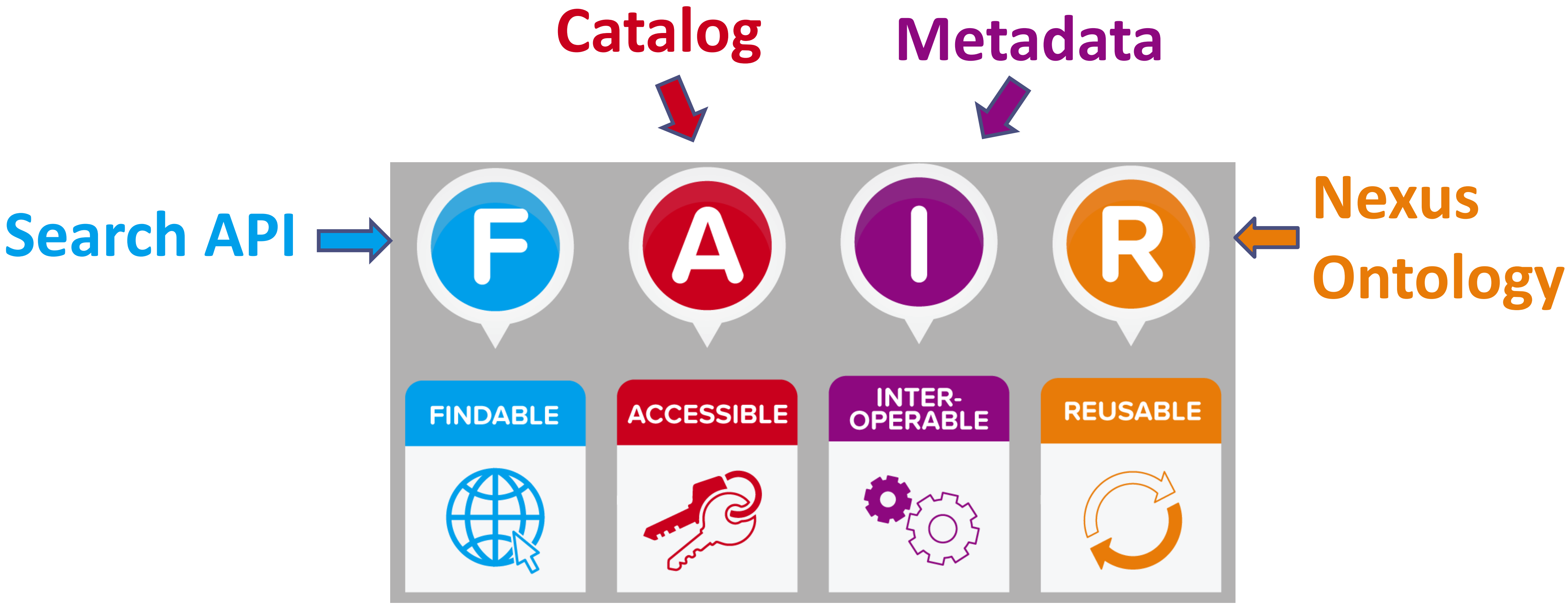
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823852

Data Catalog Services – essential for all FOUR principles of FAIR data



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

WP3 present in ALL 4 FAIR principles



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

Data Analysis Services – Effort per partner

1 - CERIC-ERIC	**88.00**
2 - ELI-DC	**78.00**
3 - ESS	**43.00** (WP leader)
4 - European XFEL	**36.00**
5 - ESRF	**25.00**
6 - ILL	**21.00**

Total 291.00 (~ 24 years)

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

Partner	CERIC	ESS	ELI	ESRF	ILL	XFEL
Catalogue	VUO (online storage NOT a catalogue)	SciCat	TBD	ICAT	ILL Own	myMdC
URL	https://vuo.elettra.trieste.it	https://scicat.esss.se	---	https://datahub.esrf.fr	https://data.ill.eu	https://in.xfel.eu/metadata
Login required	Yes	Yes	---	Yes	Yes	Yes
File formats	NeXus, HDF5, ASCII and many others	NeXus	---	EDF, SPEC, MCA, CBF, CCD, MCCD, HDF5, NeXus	NeXus and ILL Ascii	HDF5
Database	Oracle	MongoDB	---	Oracle and MongoDB	Oracle	MySQL and PostgreSQL
Language	Plsql, Python	Javascript	---	JAVA and Javascript	PHP	App: Ruby(onRails), Client: Python
Main technologies	WebDAV, Guacamole	Angular	---	React, NodeJS, EJB, JPA	Symfony, JQuery	Rails
Number of public datasets/files	0/0	181/250,000	---	~540K/157M	~250K/4M	0/0
Using OAI-PMH	No	Not yet installed	---	No	No	No
Minting DOIs	Yes	Yes	---	Yes	Yes	Yes
Data/embargo policy	Not defined	Embargoed for 3 years	---	Embargoed for 3 years, ESRF Data Policy	Embargoed for 3 to 5 years, ILL Data Policy	Embargoed for 3 with possible extension to 5 years, XFEL Data Policy
Number of instruments connected to data catalogue	None	1	---	17	54	16

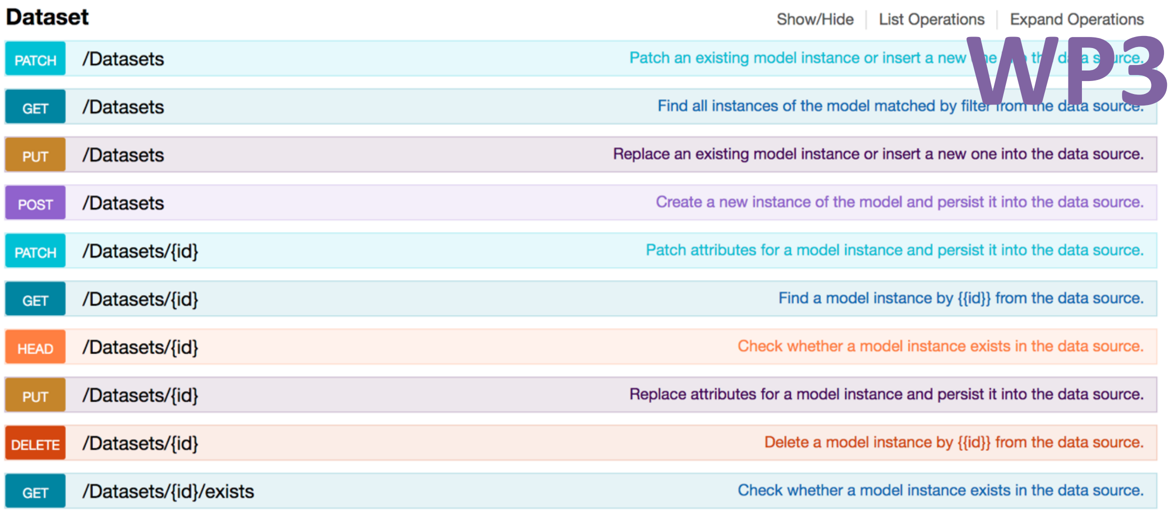
Data catalogue survey

WP3

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

Data catalogue progress

WP3



Dataset

Show/Hide | List Operations | Expand Operations

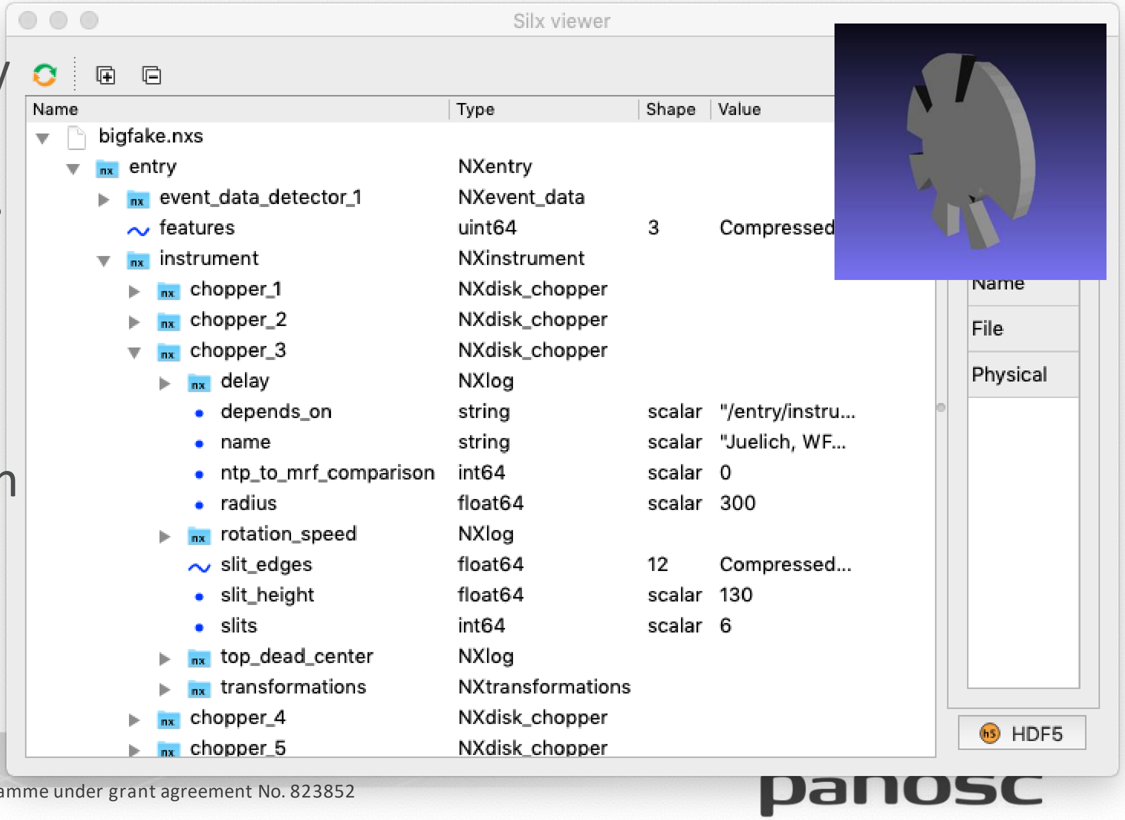
PATCH	/Datasets	Patch an existing model instance or insert a new one into the data source.
GET	/Datasets	Find all instances of the model matched by filter from the data source.
PUT	/Datasets	Replace an existing model instance or insert a new one into the data source.
POST	/Datasets	Create a new instance of the model and persist it into the data source.
PATCH	/Datasets/{id}	Patch attributes for a model instance and persist it into the data source.
GET	/Datasets/{id}	Find a model instance by {{id}} from the data source.
HEAD	/Datasets/{id}	Check whether a model instance exists in the data source.
PUT	/Datasets/{id}	Replace attributes for a model instance and persist it into the data source.
DELETE	/Datasets/{id}	Delete a model instance by {{id}} from the data source.
GET	/Datasets/{id}/exists	Check whether a model instance exists in the data source.

Fair Data API development

Identified two sides of EOSC Integration:

- Harvesting Data by EOSC Agents and Repos
- FAIR Data Search to be federated

Existing catalogues and their capabilities as well as third-party options for implementations have been surveyed. Prototypes are being discussed at a WP3 meeting in Grenoble next week.



Silx viewer

Name	Type	Shape	Value
bigfake.nxs			
entry	NXentry		
event_data_detector_1	NXevent_data		
features	uint64	3	Compressed
instrument	NXinstrument		
chopper_1	NXdisk_chopper		
chopper_2	NXdisk_chopper		
chopper_3	NXdisk_chopper		
delay	NXlog		
depends_on	string	scalar	"/entry/instru...
name	string	scalar	"Juelich, WF...
ntp_to_mrf_comparison	int64	scalar	0
radius	float64	scalar	300
rotation_speed	NXlog		
slit_edges	float64	12	Compressed...
slit_height	float64	scalar	130
slits	int64	scalar	6
top_dead_center	NXlog		
transformations	NXtransformations		
chopper_4	NXdisk_chopper		
chopper_5	NXdisk_chopper		

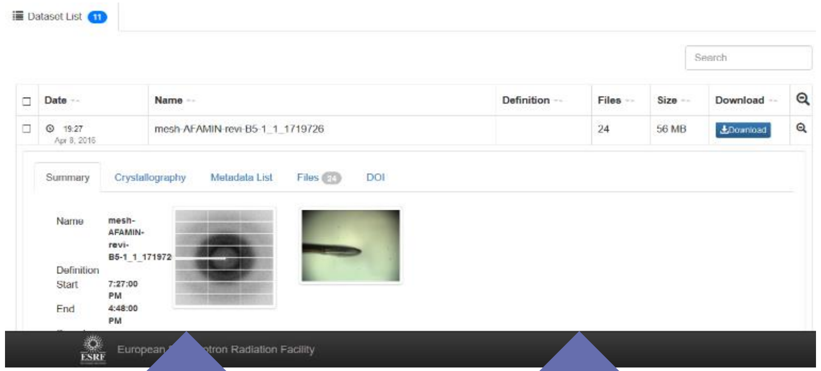
NeXus Survey for Ontologies

Facility practises and plans around file formats are being shared in sessions with partners. Identified commonalities can be applied everywhere and inform the catalogue search task.

Both Tasks l[illegible] well with ExPaNDS WP3 – Alun's talk.

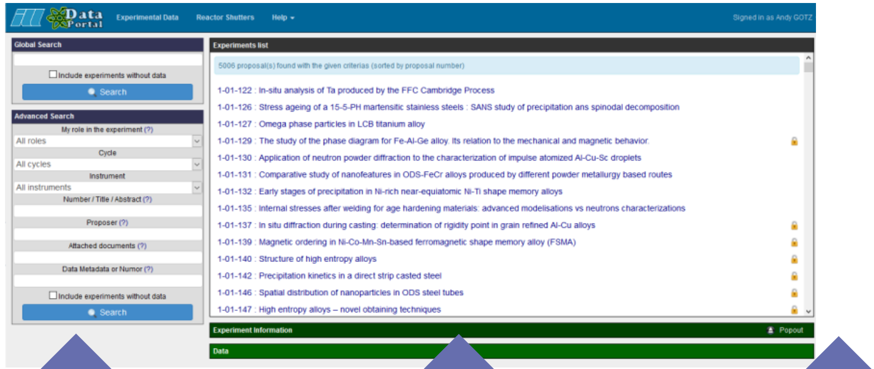
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

PaNOSC has 6 data catalogues with different APIs + UIs



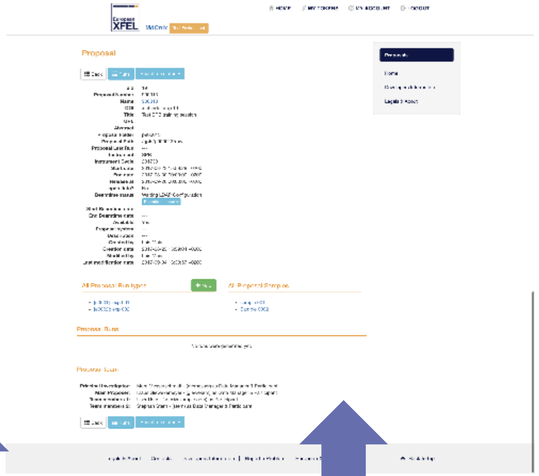
□	Date	Name	Definition	Files	Size	Download
□	19:27 Apr 8, 2016	mesh-AFAMIN-revi-B5-1_1_1719726		24	56 MB	Download

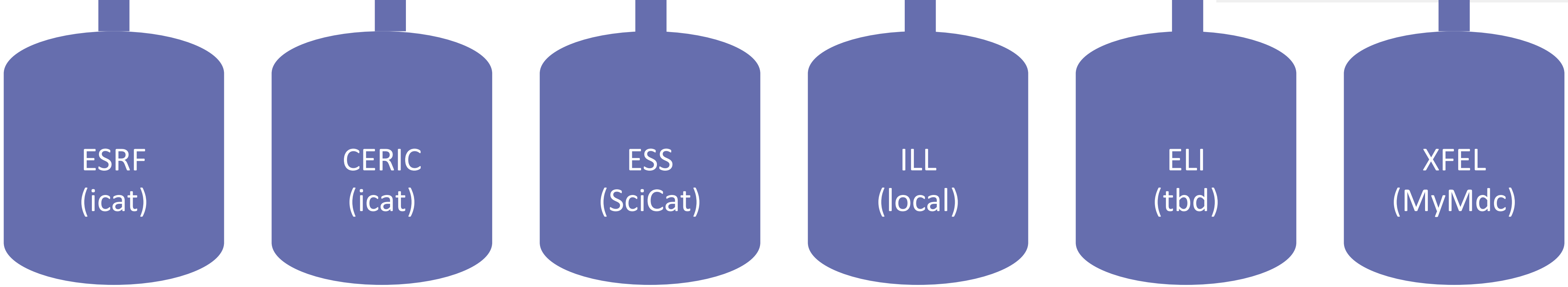
Name mesh-AFAMIN-revi-B5-1_1_171972
Definition
Start 7:27:00 PM
End 4:48:00 PM



5006 proposal(s) found with the given criterias (sorted by proposal number)

1-01-122 : In-situ analysis of Ta produced by the FFC Cambridge Process
1-01-126 : Stress ageing of a 15-5-PH martensitic stainless steels : SANS study of precipitation ans spinodal decomposition
1-01-127 : Omega phase particles in LCB titanium alloy
1-01-129 : The study of the phase diagram for Fe-Al-Ge alloy. Its relation to the mechanical and magnetic behavior.
1-01-130 : Application of neutron powder diffraction to the characterization of impulse atomized Al-Cu-Sc droplets
1-01-131 : Comparative study of nanofeatures in ODS-FeCr alloys produced by different powder metallurgy based routes
1-01-132 : Early stages of precipitation in Ni-rich near-equiatomic Ni-Ti shape memory alloys
1-01-135 : Internal stresses after welding for age hardening materials: advanced modelisations vs neutrons characterizations
1-01-137 : In situ diffraction during casting: determination of rigidity point in grain refined Al-Cu alloys
1-01-139 : Magnetic ordering in Ni-Co-Mn-Sn-based ferromagnetic shape memory alloy (FSMA)
1-01-140 : Structure of high entropy alloys
1-01-142 : Precipitation kinetics in a direct strip casted steel
1-01-146 : Spatial distribution of nanoparticles in ODS steel tubes
1-01-147 : High entropy alloys – novel obtaining techniques

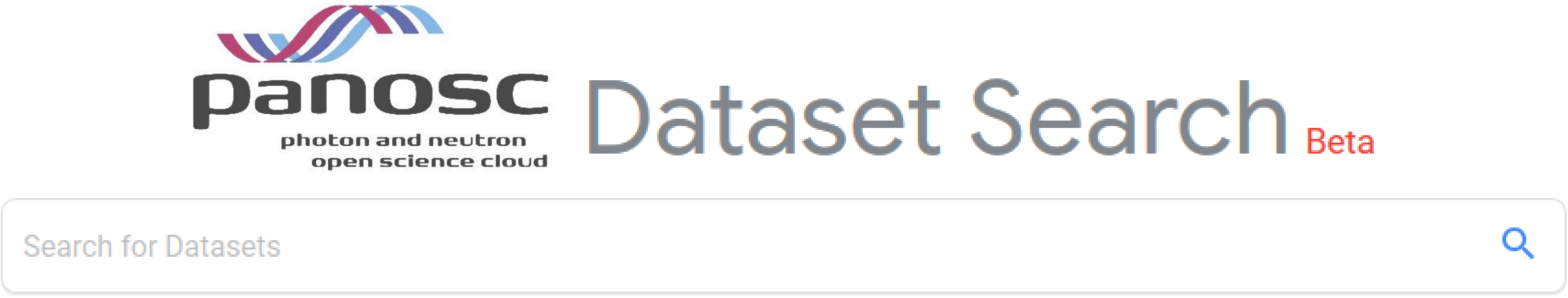




This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

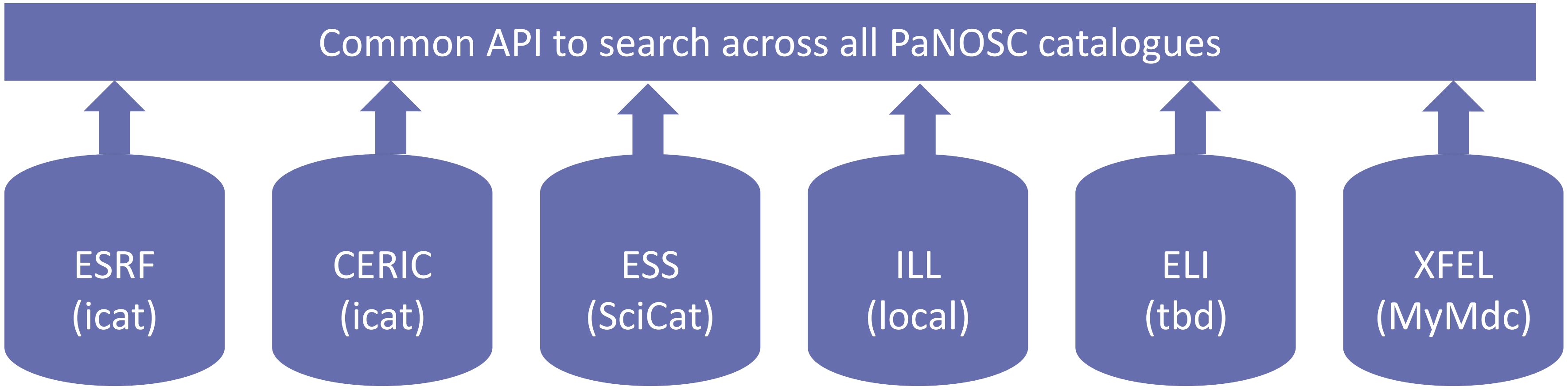
WP3

PaNOSC common API across all sites



Dataset Search Beta

Search for Datasets


Common API to search across all PaNOSC catalogues

Federated Search API

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

WP3

Data Catalog Services status in 5/2021

1. Recruitment of staff
2. **Define and develop needs for Search API (ALL)**
3. **Implement Search API locally (ESS, ELI, ILL, EuXFEL, ESRF)**
4. **Setup Federated Search demonstrator (ELI-ALPS)**
5. **Extend Nexus standard (EuXFEL, ELI)**
6. **Local deployment of metadata catalogs (ALL)**
7. **Collaborate with other Work Packages and ExPaNDS**
8. **Extend catalogs and e-logbooks (ESRF, ELI, EuXFEL, ESS)**
9. **Contribute to maintenance of HDF5 (h5py, EU HDF meeting)**
10. **Federated search demonstrator integrated in EOSC (ESS+ELI)**

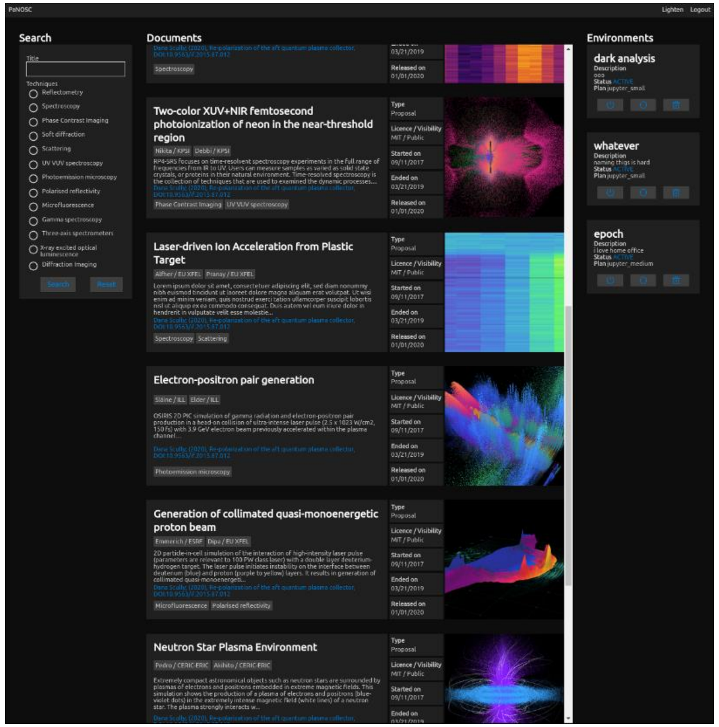
Data Catalog Services issues in 5/2021

1. Deployment of metadata catalogs (CERIC + ELI)

2. Deployment of federated search demonstrator (ELI/ESS)

3. Integrating search API in PaN portal (ELI, ESS, ILL)

4. Extending Search API to implement authentication + closed data (ILL)



Dane Scully; (2020), Re-polarization of the aft quantum plasma collector, DOI:10.9563/f.2015.87.012

Spectroscopy

Ended on 03/21/2019

Released on 01/01/2020

Two-color XUV+NIR femtosecond photoionization of neon in the near-threshold region

Nikita / KPSI Debbi / KPSI

RP4-SRS focuses on time-resolvent spectroscopy experiments in the full range of frequencies from IR to UV. Users can measure samples as varied as solid state crystals, or proteins in their natural environment. Time-resolved spectroscopy is the collection of techniques that are used to examined the dynamic processes...

Dane Scully; (2020), Re-polarization of the aft quantum plasma collector, DOI:10.9563/f.2015.87.012

Phase Contrast Imaging UV VUV spectroscopy

Type Proposal

Licence / Visibility MIT / Public

Started on 09/11/2017

Ended on 03/21/2019

Released on 01/01/2020

Laser-driven Ion Acceleration from Plastic Target

Alfher / EU XFEL Pranay / EU XFEL

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie...

Dane Scully; (2020), Re-polarization of the aft quantum plasma collector, DOI:10.9563/f.2015.87.012

Spectroscopy Scattering

Type Proposal

Licence / Visibility MIT / Public

Started on 09/11/2017

Ended on 03/21/2019

Released on 01/01/2020

Electron-positron pair generation

Sitine / ILL Elder / ILL

OSIRIS 2D PIC simulation of gamma radiation and electron-positron pair production in a head-on collision of ultra-intense laser pulse (2.5 x 1023 W/cm2, 150 fs) with 3.9 GeV electron beam previously accelerated within the plasme channel...

Dane Scully; (2020), Re-polarization of the aft quantum plasma collector, DOI:10.9563/f.2015.87.012

Photoemission microscopy

Type Proposal

Licence / Visibility MIT / Public

Started on 09/11/2017

Ended on 03/21/2019

Released on 01/01/2020

Generation of collimated quasi-monoenergetic proton beam

Emmerich / ESRF Dipa / EU XFEL

2D particle-in-cell simulation of the interaction of high-intensity laser pulse (parameters are relevant to 100 PW class laser) with a double layer deuterium-hydrogen target. The laser pulse initiates instability on the interface between deuterium (blue) and proton (purple to yellow) layers. It results in generation of collimated quasi-monoenergeti...

Dane Scully; (2020), Re-polarization of the aft quantum plasma collector, DOI:10.9563/f.2015.87.012

Microfluorescence Polarised reflectivity

Type Proposal

Licence / Visibility MIT / Public

Started on 09/11/2017

Ended on 03/21/2019

Released on 01/01/2020

Neutron Star Plasma Environment

Pedro / CERIC-ERIC Akihito / CERIC-ERIC

Extremely compact astronomical objects such as neutron stars are surrounded by plasmas of electrons and positrons embedded in extreme magnetic fields. This simulation shows the production of a plasma of electrons and positrons (blue-violet dots) in the extremely intense magnetic field (white lines) of a neutron star. The plasma strongly interacts w...

Dane Scully; (2020), Re-polarization of the aft quantum plasma collector,

Type Proposal

Licence / Visibility MIT / Public

Started on 09/11/2017

Ended on 03/21/2019

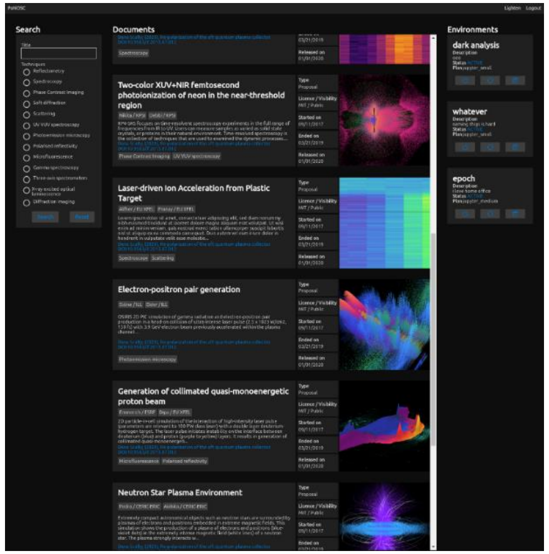
dark analysis
Description
ooo
Status ACTIVE
Plan jupyter_small

whatever
Description
naming thigs is hard
Status ACTIVE
Plan jupyter_small

epoch
Description
i love home office
Status ACTIVE
Plan jupyter_medium

https://github.com/panosc-eu/panosc/tree/master/Work%20Packages/WP4%20Data%20analysis%20services/use_cases

Data Catalog + Analysis services – big picture



PaNOSC

Lighten Logout

Search

Title

Techniques

- ○ Reflectometry
- ○ Spectroscopy
- ○ Phase Contrast Imaging
- ○ Soft diffraction
- ○ Scattering
- ○ UV VUV spectroscopy
- ○ Photoemission microscopy
- ○ Polarised reflectivity
- ○ Microfluorescence
- ○ Gamma spectroscopy
- ○ Three axis spectrometers
- ○ X-ray excited optical luminescence
- ○ Diffraction imaging

Search Reset

Documents

Spectroscopy

Released on 01/01/2020

Two-color XUV+NIR femtosecond photoionization of neon in the near-threshold region

Type: Proposal
Licence / Visibility: MIT / Public
Started on: 09/11/2017
Ended on: 03/21/2019
Released on: 01/01/2020

Phase Contrast Imaging · UV VUV spectroscopy

Laser-driven Ion Acceleration from Plastic Target

Type: Proposal
Licence / Visibility: MIT / Public
Started on: 09/11/2017
Ended on: 03/21/2019
Released on: 01/01/2020

Spectroscopy · Scattering

Electron-positron pair generation

OSIRIS 2D PIC simulation of gamma radiation and electron-positron pair production in a head-on collision of ultra-intense laser pulse ... with 3.9 GeV electron beam previously accelerated within the plasma channel...

Type: Proposal
Licence / Visibility: MIT / Public
Started on: 09/11/2017
Ended on: 03/21/2019
Released on: 01/01/2020

Photoemission microscopy

Generation of collimated quasi-monoenergetic proton beam

Type: Proposal
Licence / Visibility: MIT / Public
Started on: 09/11/2017
Ended on: 03/21/2019
Released on: 01/01/2020

Microfluorescence · Polarised reflectivity

Neutron Star Plasma Environment

Extremely compact astronomical objects such as neutron stars are surrounded by plasmas of electrons and positrons embedded in extreme magnetic fields. This simulation shows the production of a plasma of electrons and positrons (blue-violet dots) in the extremely intense magnetic field (white lines) of a neutron star. The plasma strongly interacts w...

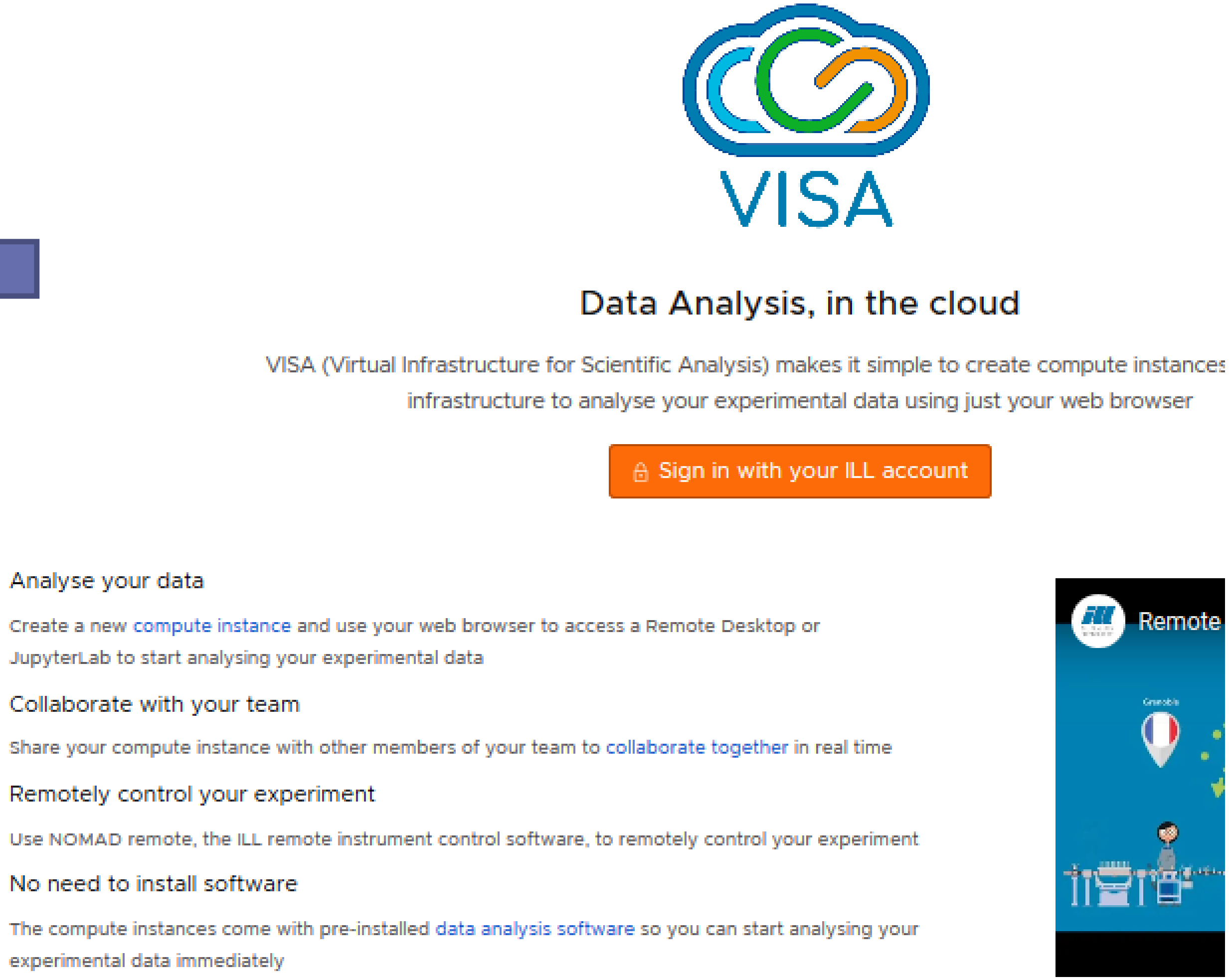
Type: Proposal
Licence / Visibility: MIT / Public
Started on: 09/11/2017

Environments

dark analysis
Description
Status ACTIVE
Plan jupyter_small

whatever
Description
naming things is hard
Status ACTIVE
Plan jupyter_small

epoch
Description
i love home office
Status ACTIVE
Plan jupyter_medium



VISA

Data Analysis, in the cloud

VISA (Virtual Infrastructure for Scientific Analysis) makes it simple to create compute instances
infrastructure to analyse your experimental data using just your web browser

Sign in with your ILL account

Analyse your data

Create a new compute instance and use your web browser to access a Remote Desktop or JupyterLab to start analysing your experimental data

Collaborate with your team

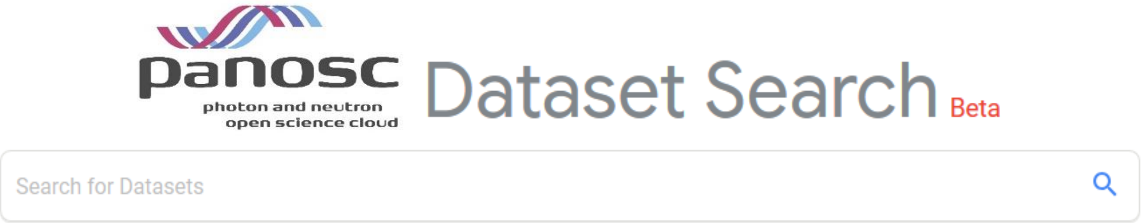
Share your compute instance with other members of your team to collaborate together in real time

Remotely control your experiment

Use NOMAD remote, the ILL remote instrument control software, to remotely control your experiment

No need to install software

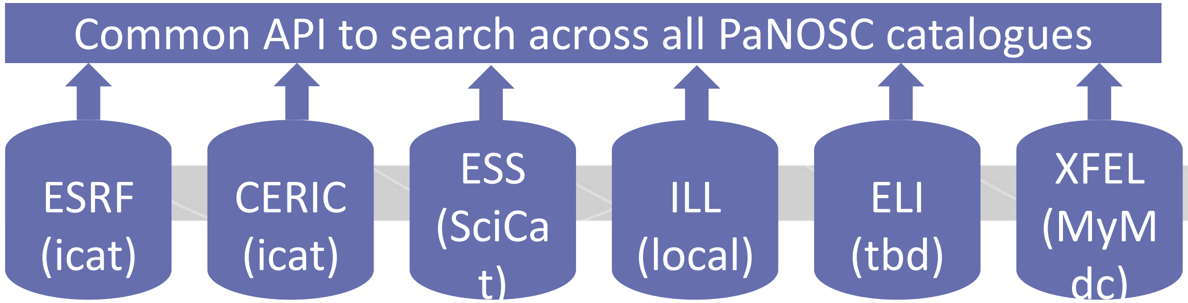
The compute instances come with pre-installed data analysis software so you can start analysing your experimental data immediately



panosc photon and neutron open science cloud

Dataset Search Beta

Search for Datasets



Common API to search across all PaNOSC catalogues

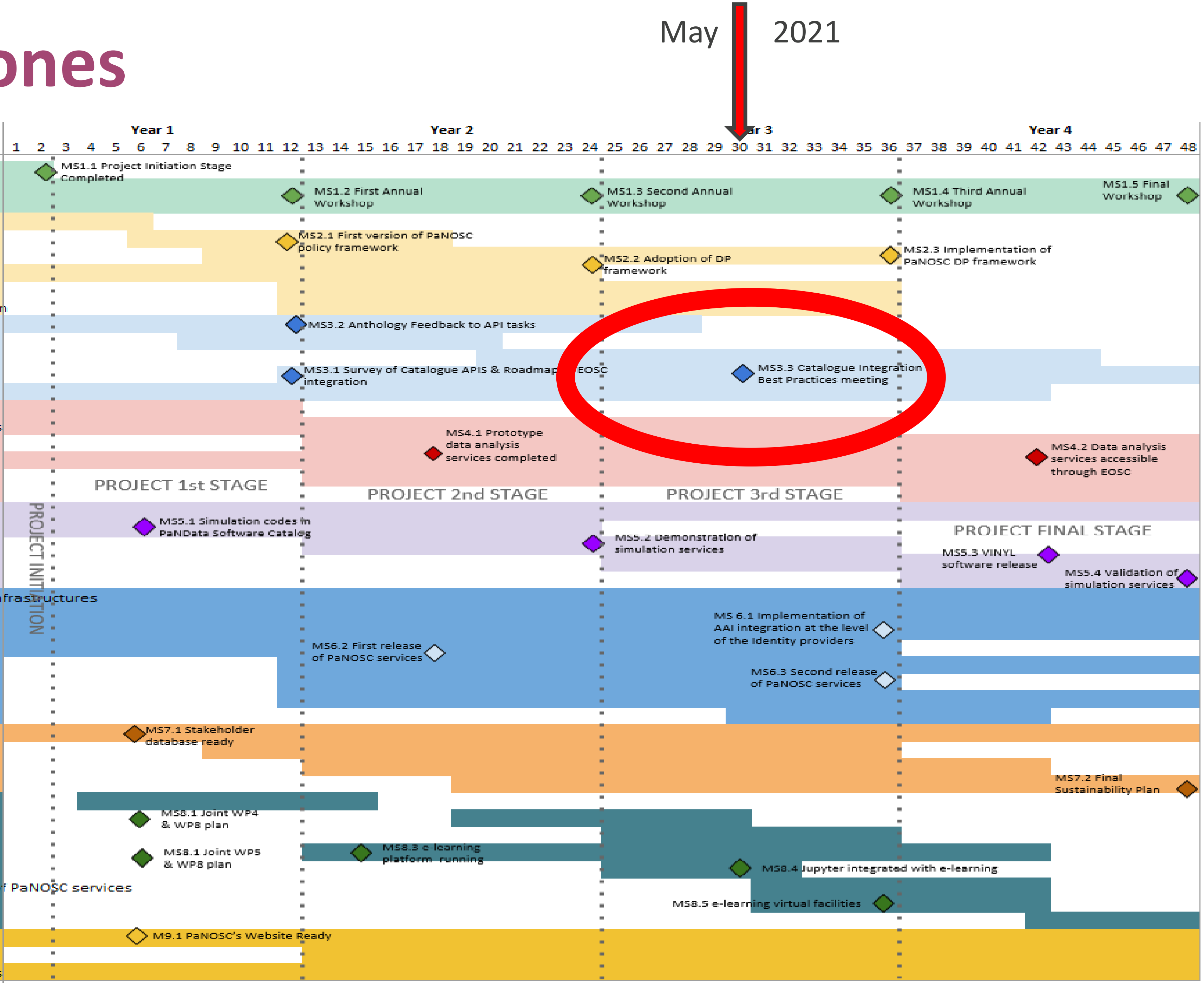
PaNOSC KPIs

	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

PaNOSC Milestones

Work Package	Id	Task Name
WP1 Management	T1.1	Creation of PID and Executive Committee
	T1.2	Project Management and coordination
	T1.3	Administration and local coordination
WP2 Data Policy and Stewardship	T2.1	Lesson learned and FAI Definitions
	T2.2	Update PaNOSC data policy framework
	T2.3	Aprove Data Policy framework
	T2.4	Create Guidelines
	T2.5	Implement DMP template
	T2.6	Validation of Data Policy implementation
WP3 Data Catalog Services	T3.1	Develop API
	T3.2	Provisioning Federated search
	T3.3	Local Deployment
	T3.4	Catalog integration
	T3.5	Meta Anthology for Catalogues
WP4 Data Analysis Services	T4.1	Survey
	T4.2	Remote Desktop based analysis services
	T4.3	EOSC integration and common portal
	T4.4	Jupyter based analysis services
	T4.5	Deployment
	T4.6	Demonstrator
WP5 Virtual Neutron and X-ray Laboratory	T5.1	Simulation Infrastructure
	T5.2	Source simulation services
	T5.3	Beamline simulation services
	T5.4	Signal generation simulation
	T5.5	Simulation data workflows
WP6 EOSC Integration	T6.1	Management of the integration with e-Infrastructures
	T6.2	EOSC Hub Service Catalogue
	T6.3	Data availability for services
	T6.4	AAI Integration
	T6.5	Services Support
	T6.6	PaN Software Catalogue
	T6.7	Data archiving pilot
	T6.8	Procure commercial cloud
WP7 Sustainibility	T7.1	Stakeholder definition
	T7.2	Metrics and costs
	T7.3	Business models
	T7.4	Sustainability plan
Staff training in data stewardship	T8.1	Sustainability and maintainability
	T8.2	Integrate Jupyter technology
	T8.3	Integrate ESFRIs
	T8.4	Staff training in e-learning platform
	T8.5	Staff training in data stewardship
	T8.6	Training materials to foster the uptake of PaNOSC services
	T8.7	Introductory course to materials science
	T8.8	PaNOSC schools for students
WP9 Outreach &Communication	T9.1	Internal and external communication
	T9.2	Dissemination of results
	T9.3	Collaboration actions with other clusters



PaNOSC WP3 Deliverables

D3.1 API definition (M18, R, PU, ESS)

D3.2 Demonstrator implementation (M28, Other, PU, ESS)

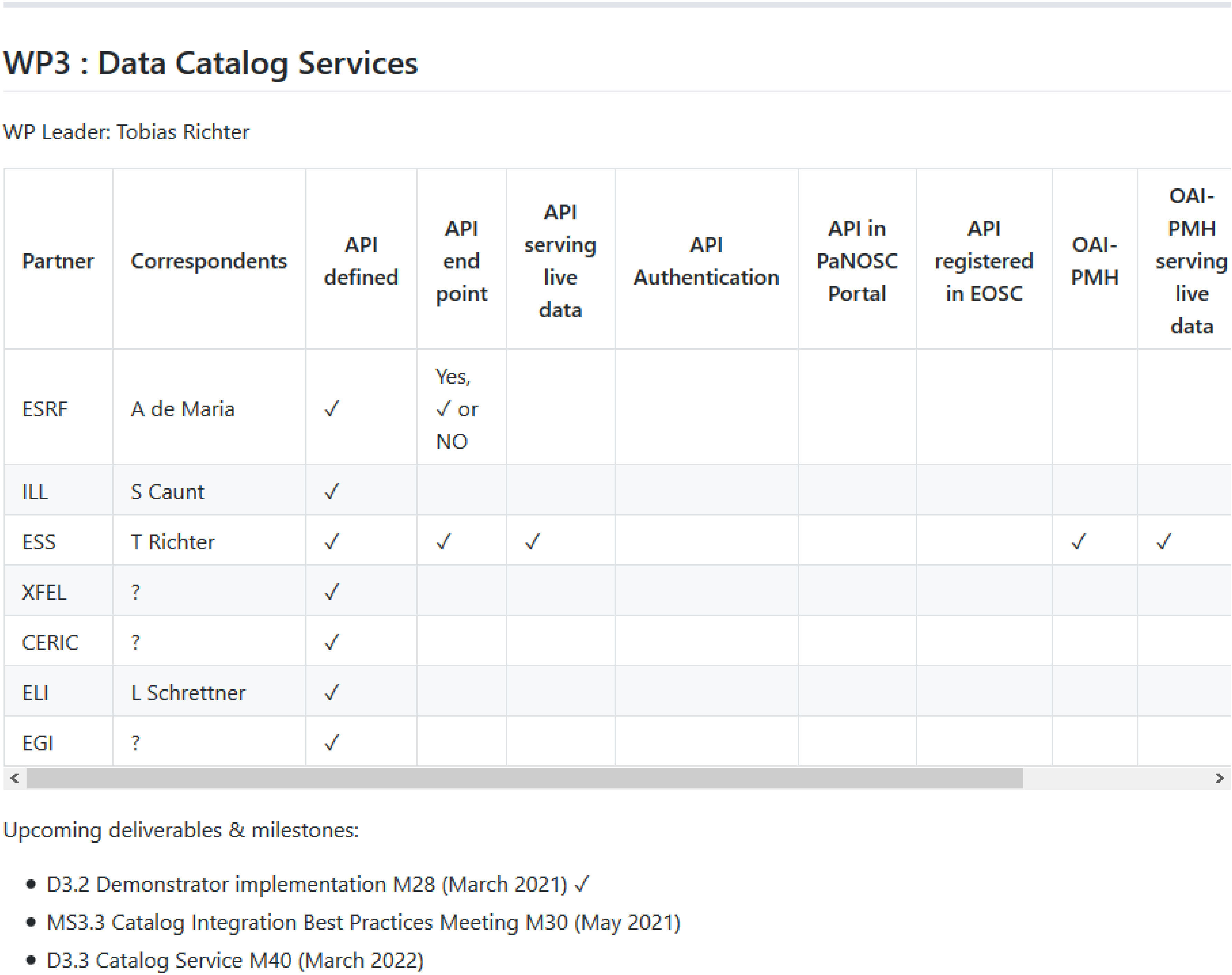
D3.3 Catalog service (M40, DEC, PU, ESS)

D3.4 Implementation Report from Facilities (M44, R, PU, ESS)

D3.5 NeXus Metadata Mapping Schema and Proposed New Definitions (M42, R, PU, ESS)

Data Catalog Services

Dashboard



WP3 : Data Catalog Services

WP Leader: Tobias Richter

Partner	Correspondents	API defined	API end point	API serving live data	API Authentication	API in PaNOSC Portal	API registered in EOSC	OAI-PMH	OAI-PMH serving live data
ESRF	A de Maria	✓	Yes, ✓ or NO						
ILL	S Caunt	✓							
ESS	T Richter	✓	✓	✓				✓	✓
XFEL	?	✓							
CERIC	?	✓							
ELI	L Schrettner	✓							
EGI	?	✓							

Upcoming deliverables & milestones:

- D3.2 Demonstrator implementation M28 (March 2021) ✓
- MS3.3 Catalog Integration Best Practices Meeting M30 (May 2021)
- D3.3 Catalog Service M40 (March 2022)