

# Open Science and Data at SOLEIL

**Brigitte GAGEY**

**Synchrotron SOLEIL – IT Strategy and Quality**

**October 17, 2022**

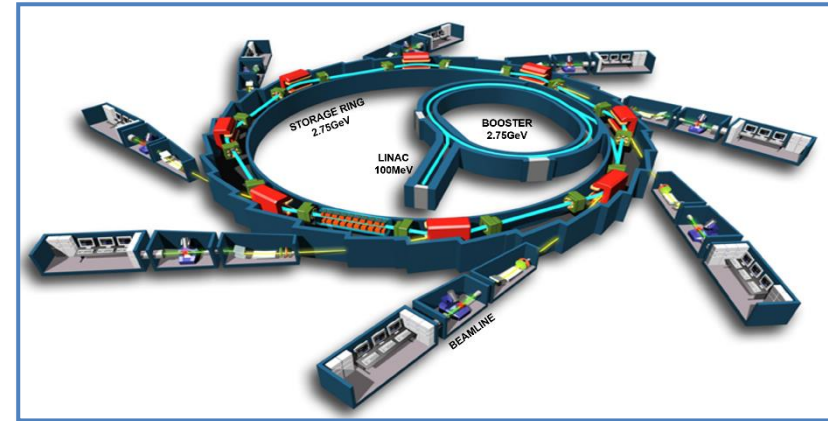


A high-technology facility at the service of research, major national challenges and industry



72%

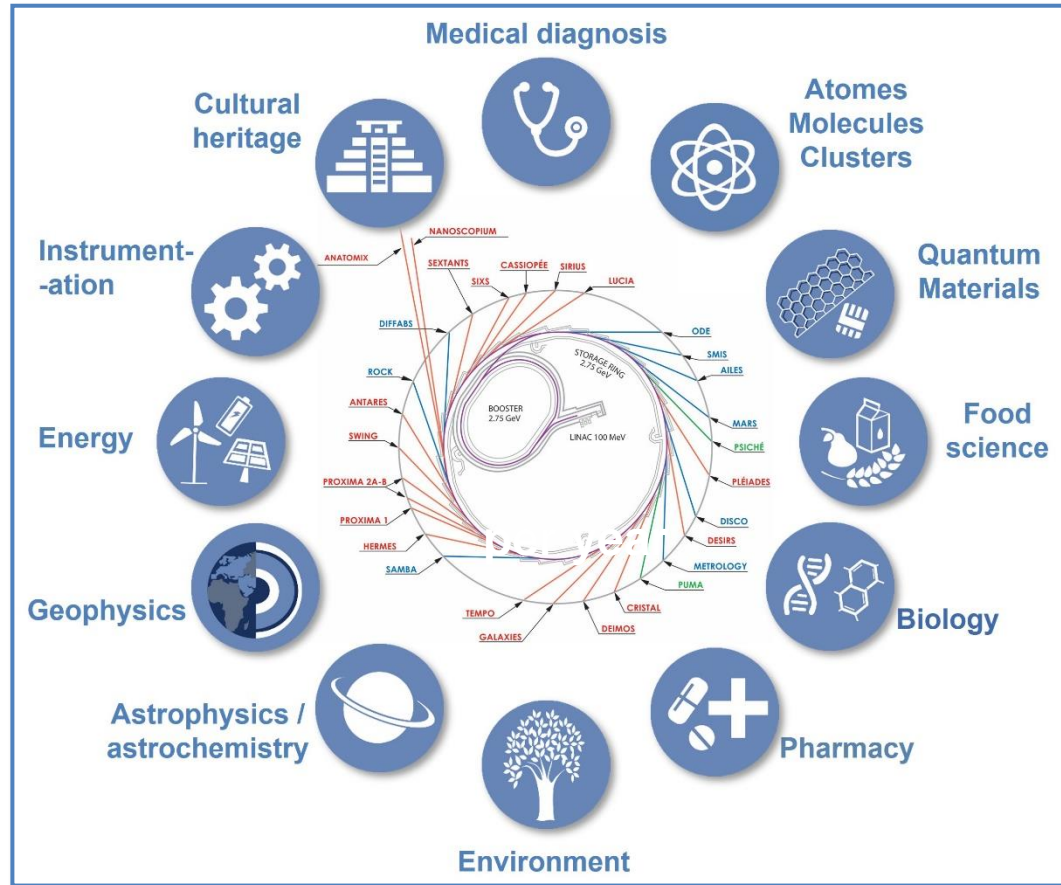
28%



- Storage ring 354m, 2.75GeV
- 29 Beamlines, covering 9 orders of magnitude in energy

- Decision to build the facility made in 2000
- Open to external users in 2008
- Overall construction budget ~600 M€
- Annual budget ~63 M€
- ~450 staff members





Users (2017-2021)	
Biology and Health	2000
Energy	650
Physical sciences and engineering	4300
Earth science and environment	350
Human and social sciences	300
Astrophysics and astrochemistry	300

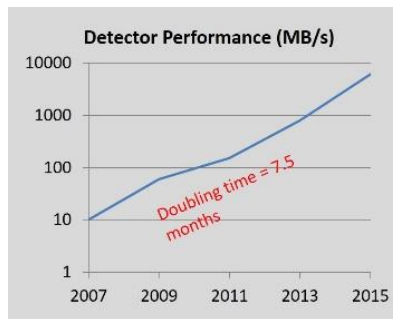
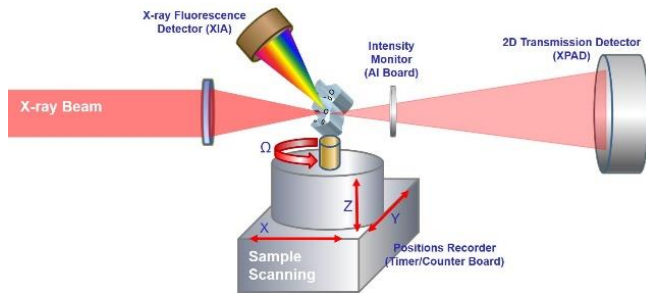
per year : ~5.000 hours of available beamtime for users  
 ~ 1.200 experiment proposals, <50 % accepted

2021 : 3.500 user visits + 650 remote  
 700 publications

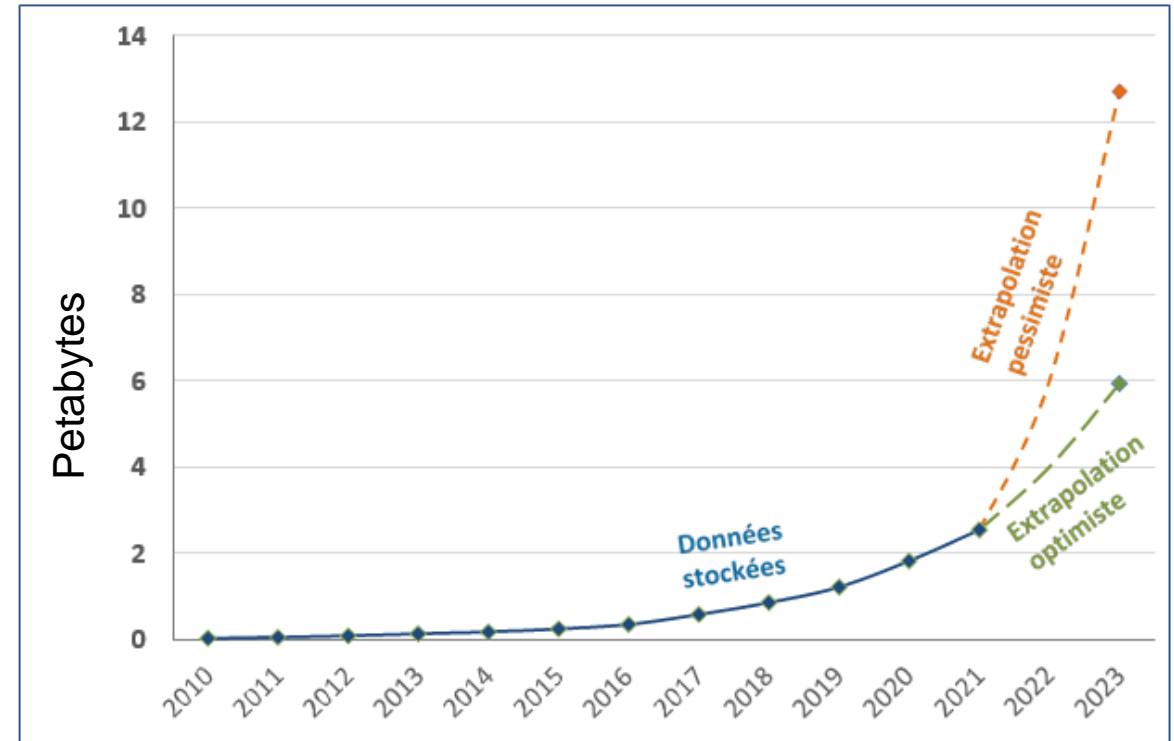


## Towards an explosion in the data production

- Flyscan acquisition : multi-techniques / multi-detectors, fast, synchronous
- High-rate and high-volume detectors
- More automated Beamlines



## Raw Data: Past and projected data production



2010-2020 : Amount of data stored

2021-2023 : Projection based on BL scientists' responses to a survey

### 4.5 Petabytes stored until 10 October 2022

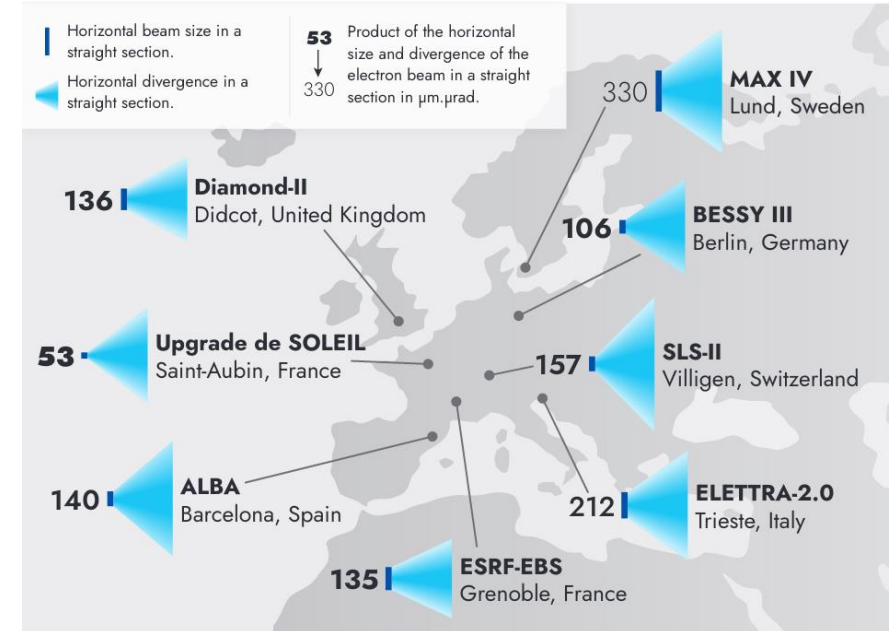
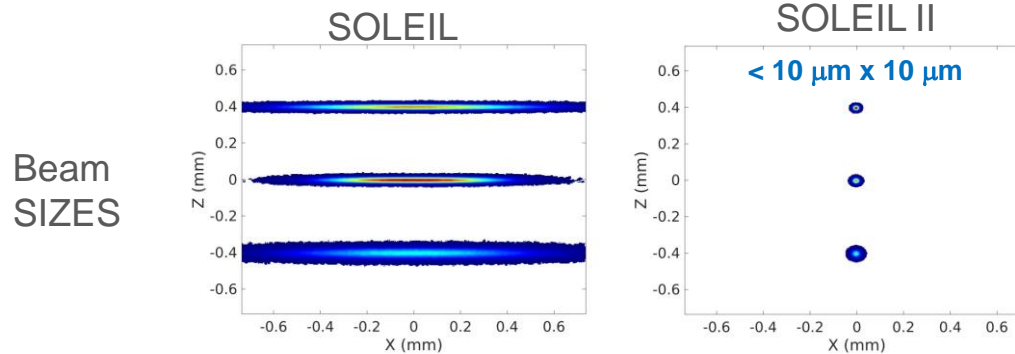
## → Addressing new scientific and societal challenges

**Advanced Materials**

**Sustainable Energy**

**Biology & Health**

**Earth & Environment**



**EXPERIMENTS UP TO 10,000 TIMES FASTER**



**NANOSCALE RESOLUTION**



**EXPERIMENTS UP TO 1000 TIMES MORE SENSITIVE**

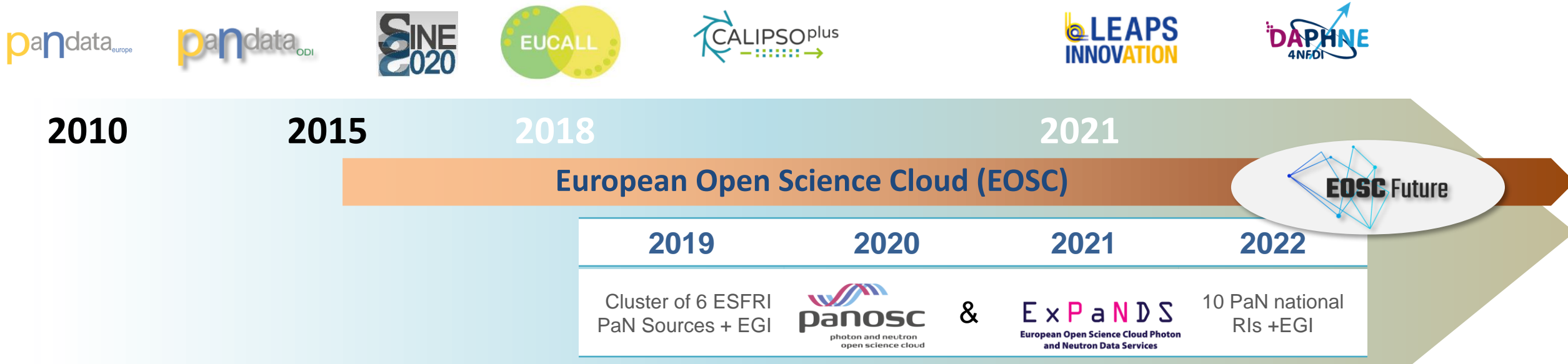


**STUDY OF DEVICES IN REAL OPERATING CONDITIONS**

# Photon and Neutron Facilities towards Open Science and Data

pandata<sub>europe</sub>  
(INFRA-2010-3.3)

→ A template for Data Policy (<https://doi.org/10.5281/zenodo.3738497>) establishing common principles, adopted progressively according to the context of each facility



Policies

Software Catalog

AAI

Data Access

Remote Analysis

Data Management Plans

Training



PaNOSC  
ExPaNDS



EU Call	HORIZON 2020 INFRA-EOSC-04	HORIZON 2020 INFRA-EOSC-5B
Description	Cluster of ESFRI PaN Sources	EOSC PaN Data Services
Partners	ESRF, ILL, ESS, EU-XFEL, CERIC-ERIC, ELI-DC, EGI	DESY, ALBA, DLS, ELETTRA, EGI, HZB, HZDDR, MAX IV, PSI, <b>SOLEIL</b> , UKRI
Observers	GEANT EU-DAT National RI's	
Linked 3 <sup>rd</sup> Party	DESY STFC CESNET	
Start – End (Duration)	2018-12-01 – 2022-11-30 [4 Years]	2019-09-01 – 2023-02-28 [ 3 ½ Years]
Coordinators	A. Götz, J. Boderà	P. Fuhrmann, S. Servan
Budget	12 M Euros	6 M Euros
Home Page	PaNOSC.EU	ExPaNDS.EU
Twitter	@PaNOSC_eu #PaNOSC	@ExPaNDS_eu #ExPaNDS
GitHUB	github.com/panosc-eu	Github.com/expands-eu

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

# SOLEIL towards Open Science and Data

## SOLEIL Data Management Policy established in consultation with SOLEIL Users and Beamlines' Scientists, with the support of CNRS experts in research data management.

Encouraging and helping SOLEIL Users and Scientists to produce **FAIR data**  
*Findable, Accessible, Interoperable, Reusable*

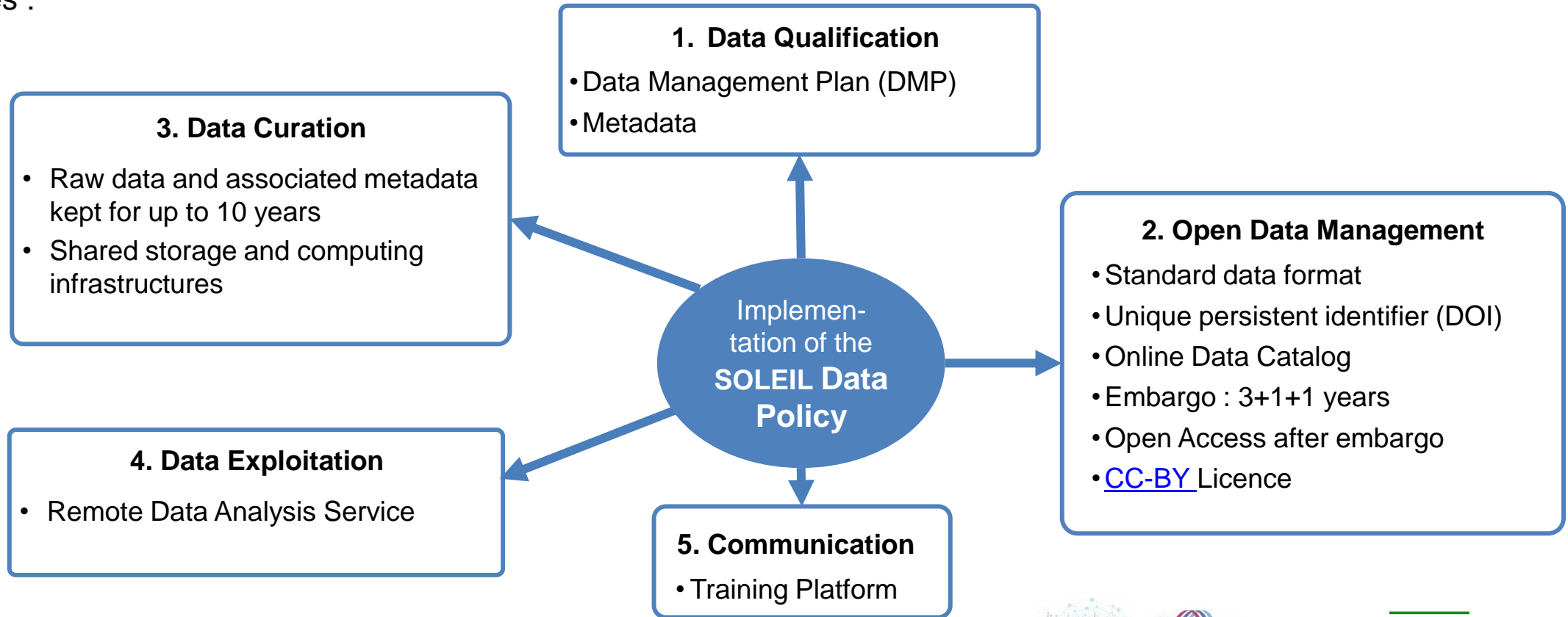
« Open data as possible, closed data as necessary »

- Applicable to data obtained via public research conducted at SOLEIL
- Long-term storage of raw data and metadata up to 5 years (striving for up to 10 years)
- Curated in well-defined formats (HDF5/NeXus)
- With a unique persistent Identifier, DOI
- Access via an online data catalogue
- 3+1+1 years embargo; open access after the embargo
- CC-BY-4.0 license



The experimental team has to establish a data management plan (DMP) and to ensure the completeness of the metadata. SOLEIL committed to provide means for the capture of metadata, for the reduction and processing of raw data.

→ 5 axes :



→ based on international and national collaborations, projects and services,

→ and in coherence with European and national initiatives



## Data Management Plan

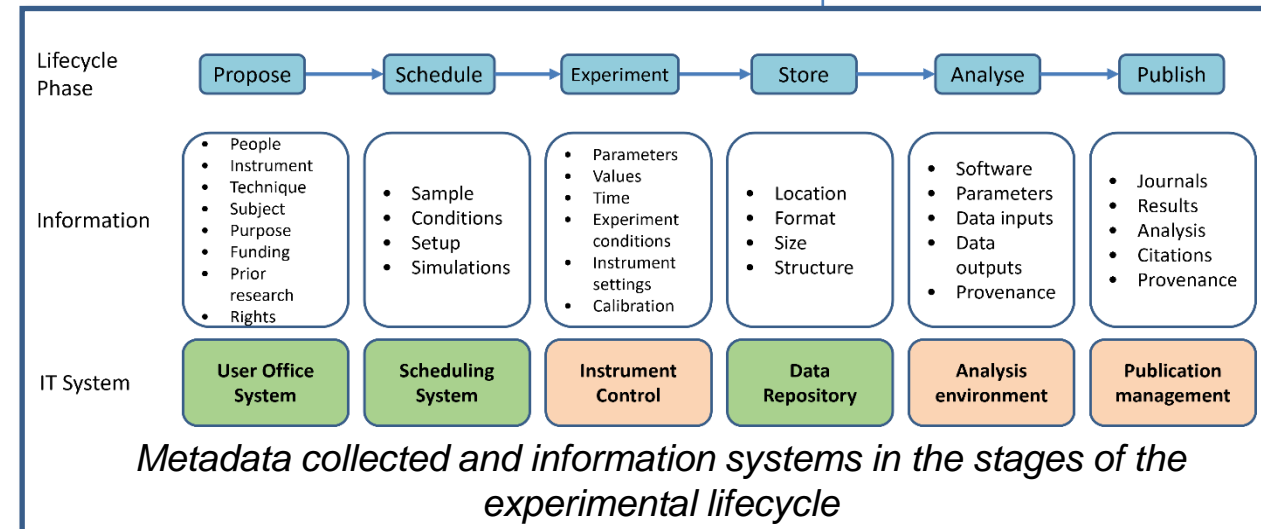
- ✓ Up to now, Recommendation to scientists to use the [DMP-OPIDoR](#) tool with existing templates
- ✓ Soon, a PaN DMP template
- ➔ a SOLEIL DMP template to be developed on PaN DMP basis



The 4 DMP phases	0 - Before proposal submission	Typically knowledge of instrument scientist or RDM team (static parameter)
	1 - Proposal submission	Typically knowledge of the user, with support by the facility administration and RDM team.
	2 - Accepted experiment planning	Typically knowledge of the user, with support from the facility administration and instrument scientist.
	3 - Data Collection / Data processing / analysis	Typically knowledge of the user, with support from the instrument scientist.

## MetaData

- ✓ Currently : Collected by different tools depending on the Exp. Lifecycle Stages; SOLEIL naming use
- ✓ Standardization of PaN MetaData at each Exp. Lifecycle Stage
- ➔ PaN Standardized Metadata to be implemented and gathered in the SOLEIL Data Catalog



## Data Format

### NeXus/HDF5 recommended since SOLEIL opening

- Used systematically at 15 SOLEIL Beamlines; depending on the context (instrument, etc.) at the other ones
- Now widely adopted at PaN facilities, and supported by



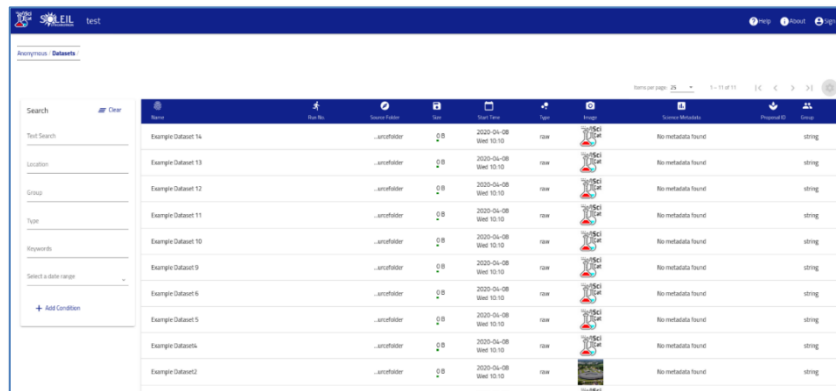
## On line Data Catalog

### SciCat under implementation at SOLEIL

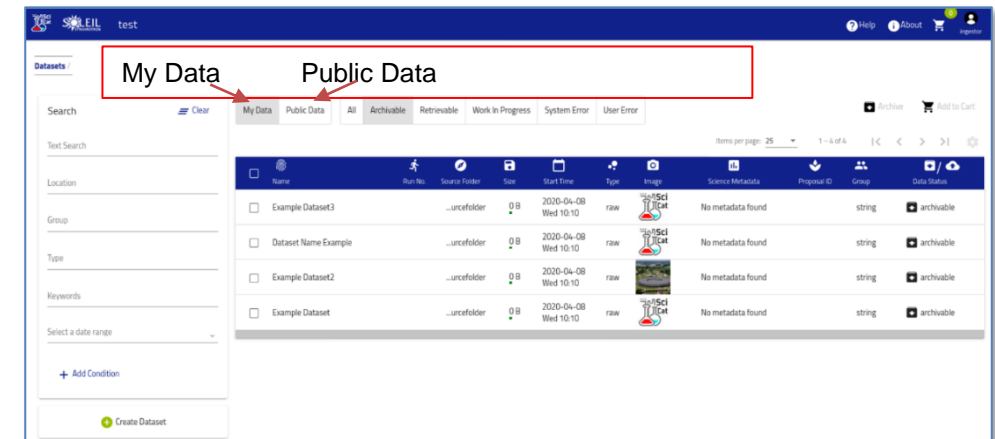
#### SciCat Open Source Project



- ✓ View Public Data (without authentication)



- ✓ Find, Filter, Generate DOI (after authentication)



➔ Unified access to PaN data catalogues ➔ EOSC Service

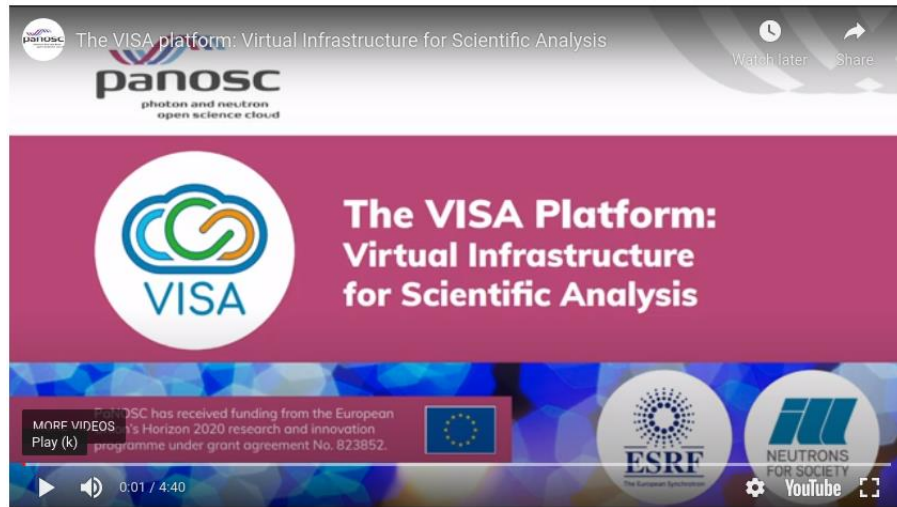
✓ VISA Platform originally developed by ILL



✓ Adopted by

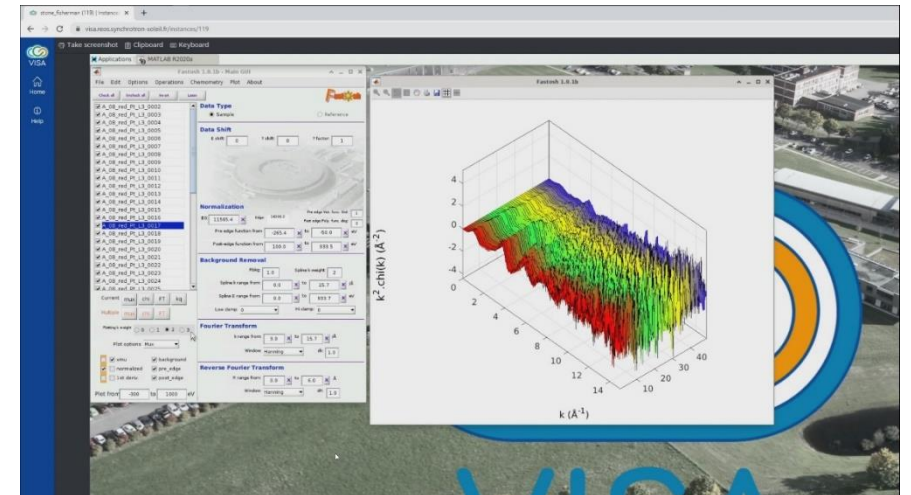


- ✓ Providing remote data analysis services in a web browser with access to :
  - Experimental data, Analysis software, Compute infrastructure, Support (IT & scientific)
- ✓ Users can :
  - create their own virtual working environment simply by using a web browser;
  - access datasets and benefit from the same computing resources as if working on site;
  - have the latest version of the analysis program without having to install it, download it, ...
  - Jupyter notebook environment
- ✓ Team members can also :
  - join a secure work environment, provide support and enjoy real-time collaboration.



<https://bit.ly/VISA-video>

➔ Being installed for SOLEIL Data Analysis Services



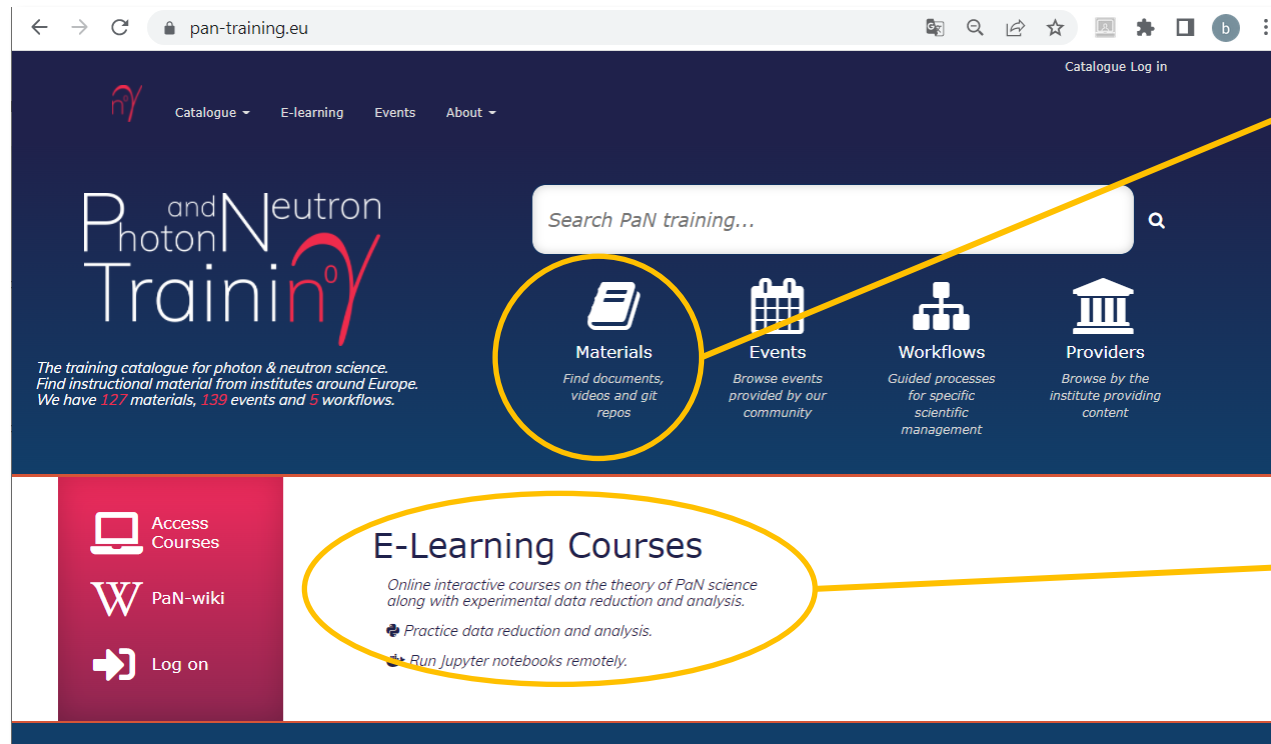
Testing SOLEIL (*SAMBA Beamline*) XAFS data analysis software on a VISA instance

**ExPaNDS**  
European Open Science Cloud Photon and Neutron Data Services

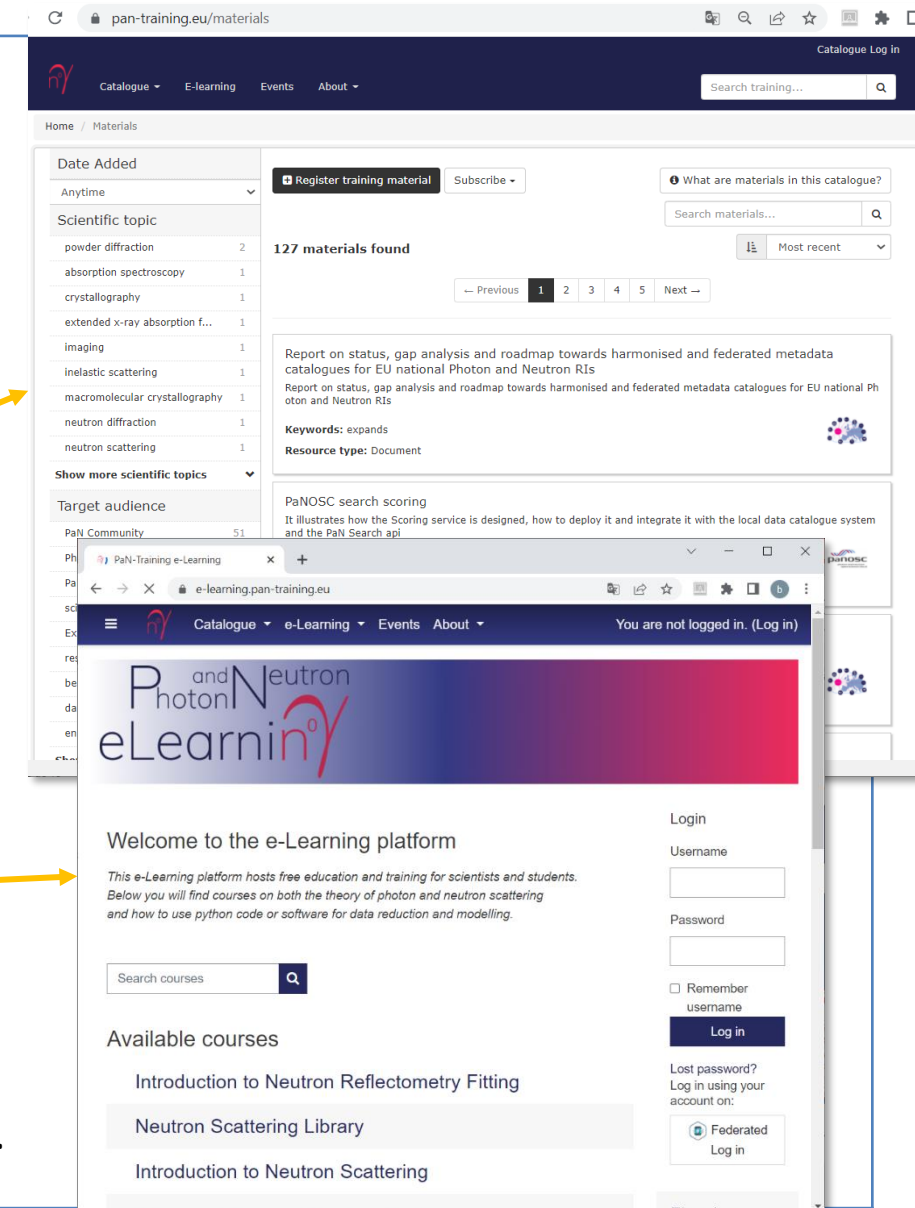
**panosc**  
photon and neutron open science cloud

**PaN-training** : A centralized platform to collect our distributed knowledge, documents, training, ... and share them

*From the ELIXIR and SINE2020 platforms*



and others : Workshops for Users (as [EOSC, an asset for Science](#)), etc.





Infrastructure for data storage, processing and making data available, dissemination and enhancement

## FITS Project

- ✓ CNRS & GENCI
- ✓ Equipex+ / PIA3
- ✓ Started in June 2021
- ✓ 2022, in "pre-FITS" phase

SOLEIL Data > 3 years :

- Copy → IDRIS
- Secure archive → CC-IN2P3



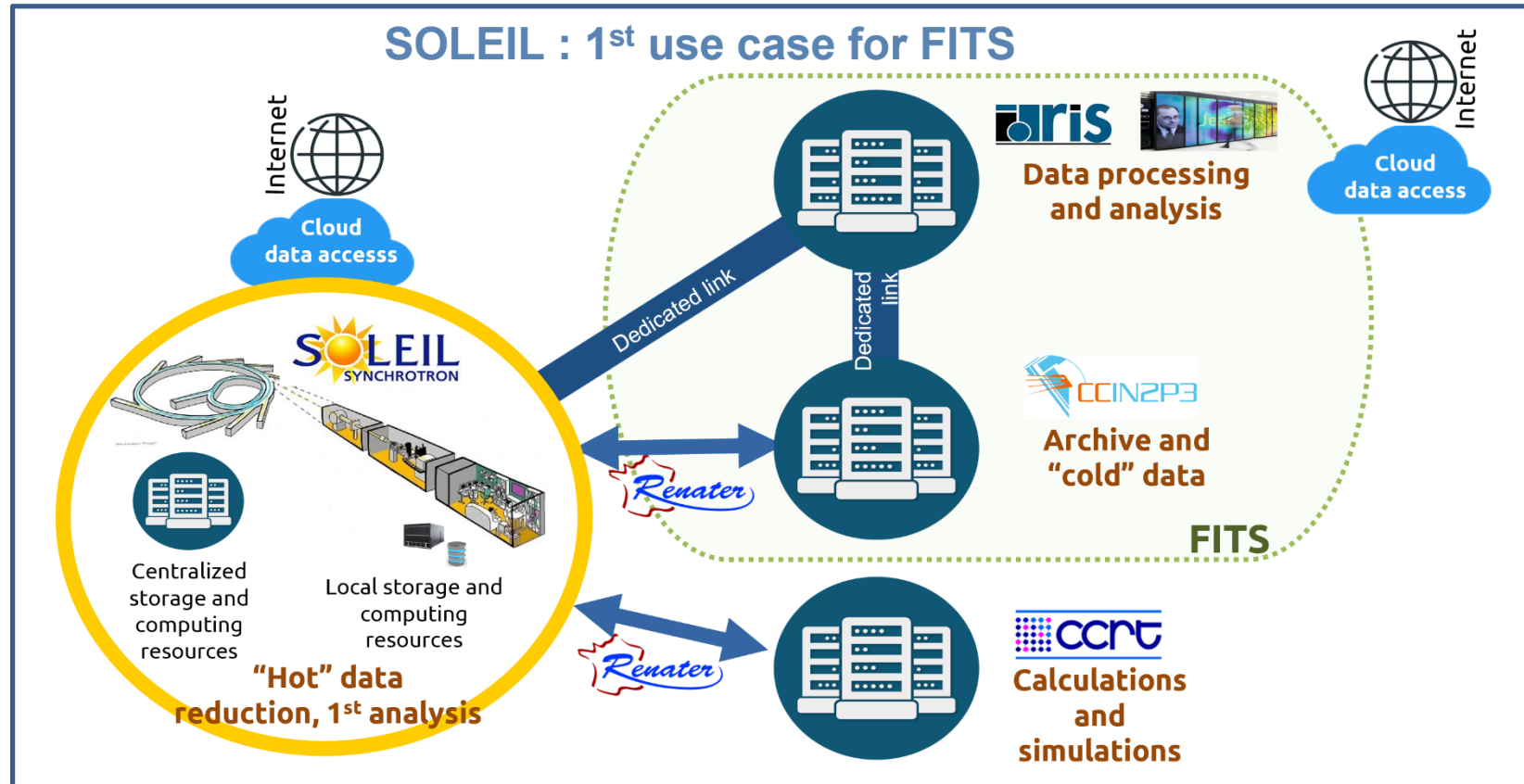
- ➔ Implementation of a distributed infrastructure hosted in low carbon environmental conditions
- ➔ Federate the services and know-how of the 2 CNRS computing centres (IDRIS and CC-IN2P3)

## LEAPS-INNOV project

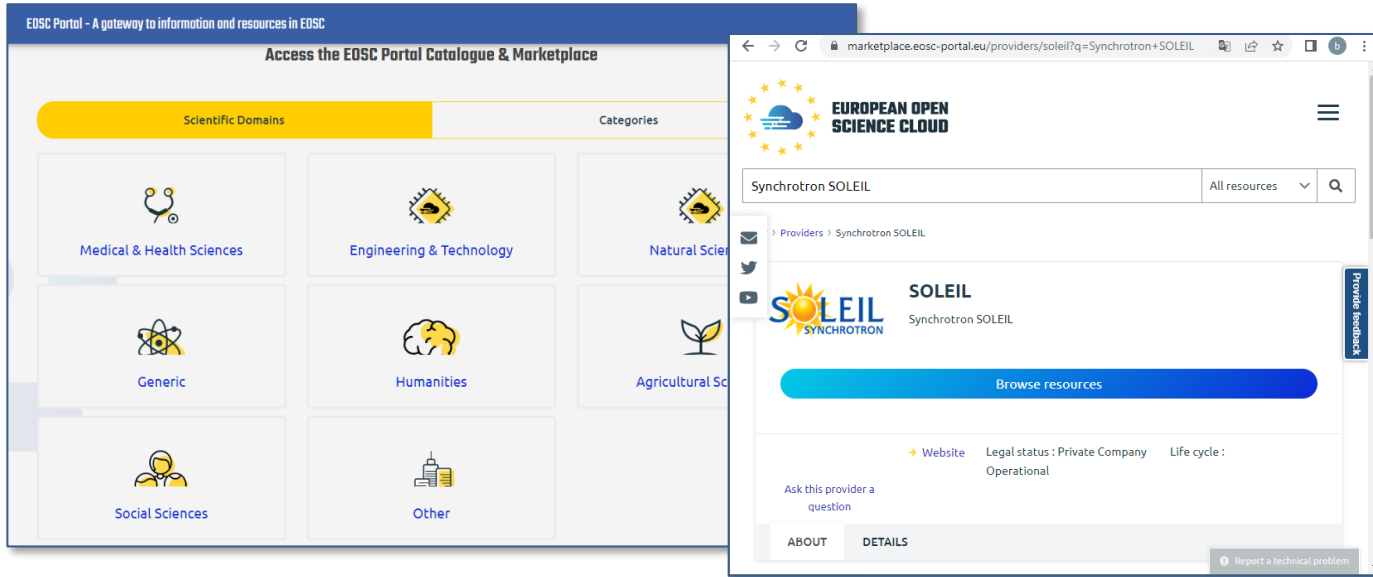
➔ new strategies for data reduction and compression



LEAPS-INNOV has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101004728

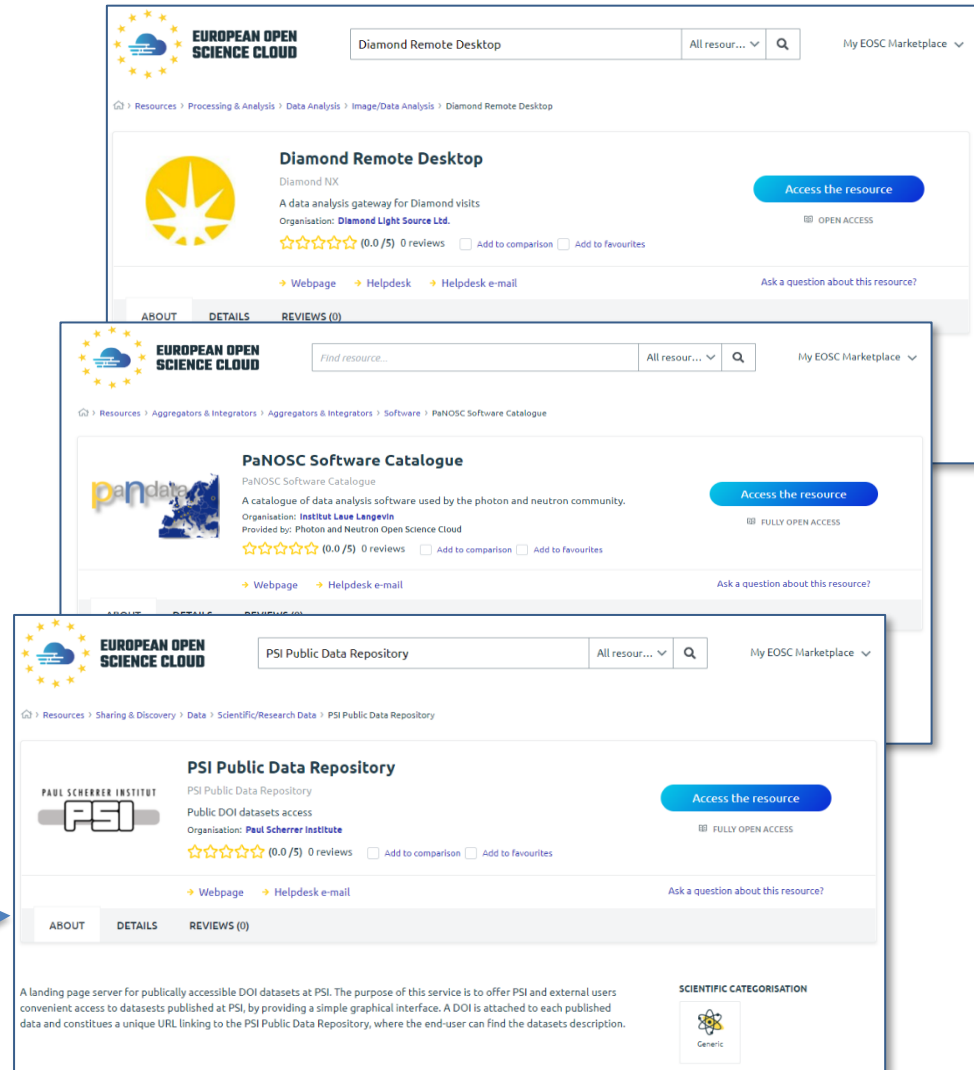


- ✓ EOSC portal (<https://eosc-portal.eu/>) :  
SOLEIL registered as a **potential service provider**



➔ SOLEIL data services to be available at the EOSC Portal, in the example of other PaN facilities

- ✓ EOSC Association (<https://eosc.eu/>) : SOLEIL **observer**
- ✓ EOSC-France College : SOLEIL **member**



*Examples of Data Services of PaN facilities available at the EOSC portal*

- ✓ A long road since the first discussions to establish a SOLEIL experimental data policy
- ✓ Involving Users, Scientists and IT staff at the early stages.
- ✓ An action plan based on international and national collaborations, projects and shared services, and in coherence with European and national initiatives :

*not reinventing what is done elsewhere and building on joint efforts*

## Next Steps :

- Make SOLEIL Experimental Data Policy a reality at all Beamlines
- Deploy SOLEIL Data Services at all Beamlines
- Make SOLEIL Data Services available via the EOSC portal
- Upcoming MoUs to continue the collaborative works
- Sustainability of PaN Data Commons could be managed within LEAPS (the League of European Accelerator-based Photon Sources ) and LENS (the League of European Neutron Sources) **frameworks**
- With EOSC support in the Open Science and Data process
- Data Repository certification under consideration

Thank you for your attention !

