RIC & friends seminar

🐸 Sophie 🐣 - 06.07.2020

Topic of the day: ExPaNDS

1. Some context
2. What it stands for
3. What we did so far and do right now
4. How DESY > RIC > I are involved

Context

An experiment by Herr Röntgen in 1901: ~10 kB in 5 days vs. an experiment today by a fellow DESY user: ~750 TB in 5 days

⇒ we need to share compute and storage resources

Peer-reviewed articles are among the most reliable sources of information

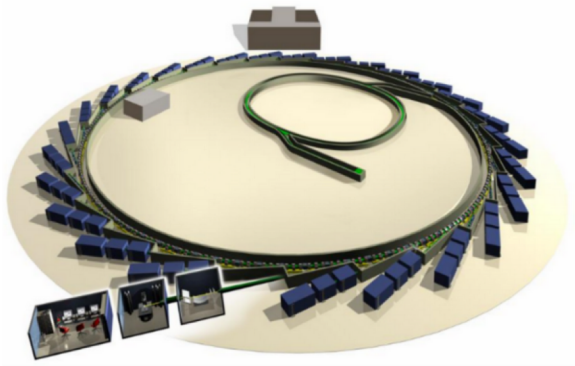
⇒ we need tools easily available to reproduce scientific results

More and more researchers use several facilities for their research

⇒ we need trans-facility search of data

AI is coming

⇒ we need machine-readable and -understandable data



Photon and Neutron

ExPaNDS

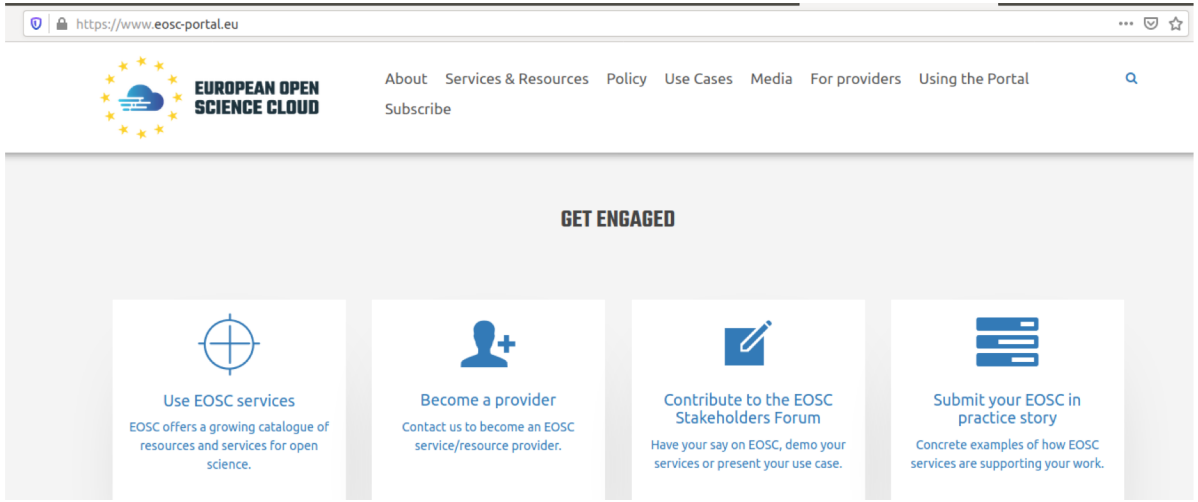
European Open Science Cloud

Data Services

European Open Science Cloud

EU project with funding from the H2020 R&I program, to add to the EOSC services that are valuable to researchers

What is the EOSC right now?



A marketplace with linked services

Trying to get people on board.

Trying to showcase its added value.

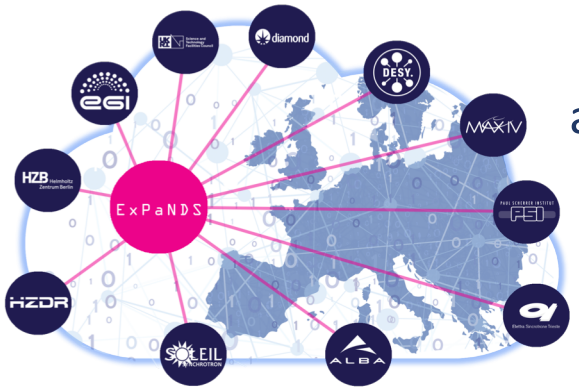
Photon and Neutron

ExPaNDS

European Open Science Cloud

Data Services

Photon and Neutron



Entire European Photon and Neutron Community

Community standards and tools

- PaN data policy framework
- NeXus data format for neutron, X-ray and muon science
- UmbrellaID, a digital identity for PaN users
- PaN software catalogue
- PaN e-learning platform

Photon and Neutron

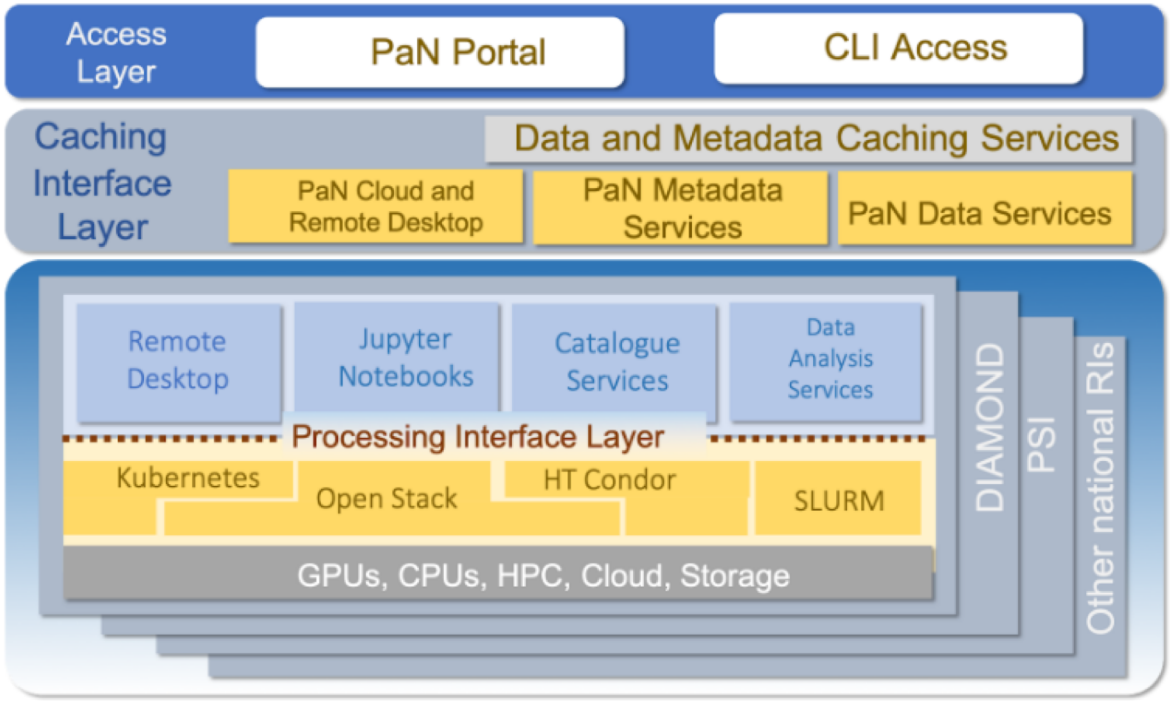
ExPaNDS

European Open Science Cloud

Data Services

Data Services

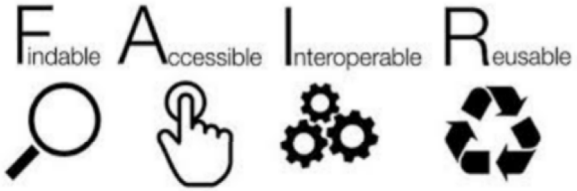
- Catalogue of available datasets and their location
- Catalogue of available data analysis services
- Correlation of which dataset works with which analysis service



Data Services

A few prerequisite

- Have open data generated in the first place
  - 3 out of our 10 facilities had 0 public datasets in Dec. 2019 [1]
- Make it FAIR
  - 2 out of our 10 facilities minted DOIs for their datasets [1]
  - 3 out of our 10 facilities used DMPs defining metadata standards and licences [1]



- Build the different stacks to abstract local catalogues and analysis services

[1] ExPaNDS data landscaping dating from Dec. 2019 DOI 10.5281/zenodo.3673811

Progress and ongoing works

FAIR

- Updating with PaNOSC the 2011 data policy framework to include FAIR principles
- Fostering its adoption by our facilities
-
- Working on PIDs for instruments (HZB)
- Working on maDMPs with RDA

Data catalogues

- Get a state-of-the-art picture of existing agreed vocabulary of techniques in PaN communities (e.g. Pankos, wayforlight)
- Investigate existing ontologies repositories (e.g. BioPortal)
- Survey researchers for their typical search use cases (results soon to be published in GitHub)

Data and analysis services catalogues

- Prepare the stacks to deploy the PaN portal at our facilities
- Select demonstration services to be implemented into EOSC, with matching datasets
  - Serial crystallography: publication
  - 4D full field tomography: publication
  - Terahertz Spectroscopy: publication

Sophie

My contribution

ExPaNDS @DESY

- IT: Patrick, Sophie, Frank Schlünzen, new recruit with Michael
- FS-CFEL: Anton Barty, new recruit
- EUP: Beatrix Bugla
- V-D4: Carsten Porthun (DPO)

- Make sure the quality of what we deliver meets the EU, and our own, expectations
- Follow technical and financial progress and try and make it consistent
- Provide internal communication channels
- Keep up with what's happening in the EOSC ecosystem
  - Other projects doing work we could re-use
  - Important documents we should give feedback to
  - Conferences where we can (be) present
- Write things like project's DMP and POPD
  - Feed Zenodo and GitHub with our public outputs and push for CC licences use
  - Being transparent on what we do with personal data