10-11 November 2020

PaNOSC & ExPaNDS
Annual Meeting

Supporting the PaN EOSC services

11th November, 2020

Author: J-F. Perrin on behalf of the WP6 team.

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.

PaNOSC WP6 aims

- **Support the PaNOSC services (WP3/4/5/8) to integrate EOSC.**
- **Users' productivity: remove technical hurdles, simplify users access to data and services.**
- **Prepare RIs' IT infrastructure for EOSC.**

Today's agenda

- Data transfer: bringing data to services or vice versa, quest for the right model (depending on the service type, data volume, …)
- Authentication and Authorisation Infrastructure (AAI): common user authentication for all PaNOSC (EOSC?) services
- Other ongoing activities

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.

PaN RI typical environment

- Large user community (50 000 users/18 Months) with heterogenous profiles:
  - Scientific field: Biology, Material science, chemistry … Archaeology, nuclear physic, HEP
  - Academic and Industrial users
  - Quite often very limited IT support in the users' home organisation
- Datasets volume vary from 10s of GB to 100s of TB
- Yearly data production from 300TB to 10s of PB
- Data are openly accessible after a 3 years embargo period
- Relatively small IT teams, mainly focused on data production (i.e. experiments) support and integration of existing solutions

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.

Technical RI environment regarding data storage at RIs.

- Central archive on RI network: Large disk storage (typically GPFS) + Tape (cold storage)
- ACLs (Typically NFSv4) at the filesystem level
- Local user ID that needs to be mapped to federated AAI ID
- Users access to data store through gateways (NFS, SMB, Rsync, FTP, ...)

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.

PaNOSC data transfer activities

Support from UKRI-STFC and CESNET (EGI.eu Linked Third Parties)

Participation in EGI data transfer WG
https://wiki.egi.eu/wiki/TCB:Data_Transfer_WG

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.

PaNOSC data transfer use cases

Three clearly identified use cases:

1. An RI wants to archive its experimental data in a remote data centre. Cold backup

2. A user wants to access a data analysis service, data has to be available “transparently”

3. A facility user wants to transfer a large dataset from an RI’s archive to a remote compute center or her/his home PC

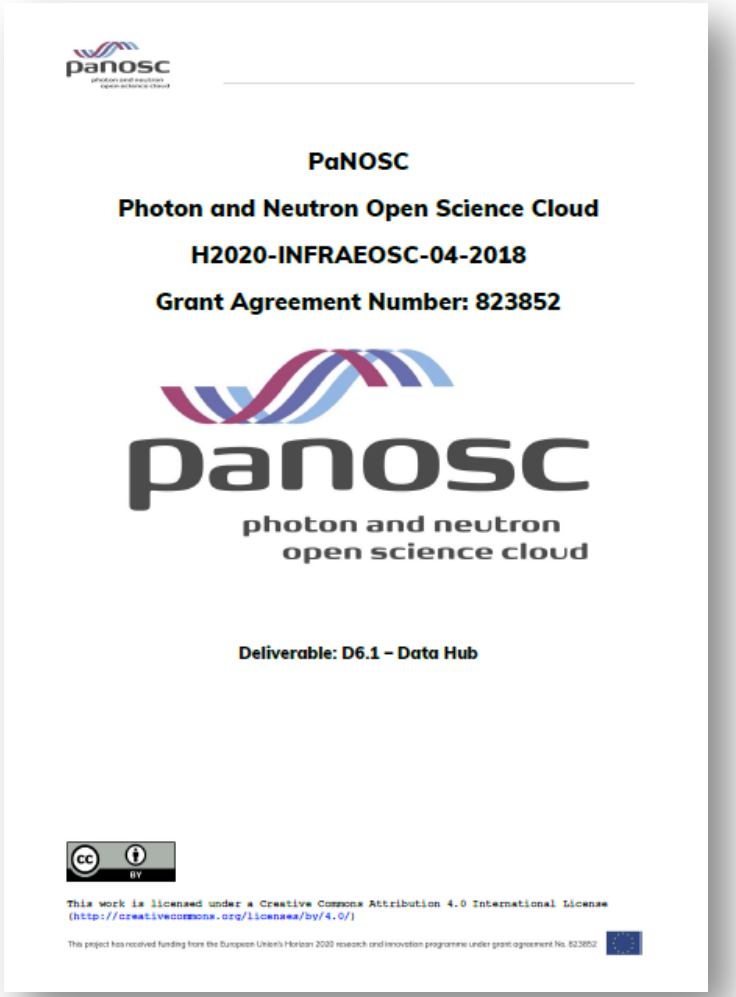
PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

1st Use case : Archiving

- Aims
  - Automate cold backup of raw data after experiment
  - Restore data
- Pilot
  - ILL as RI (data producer) (GPFS + NFS gateway)
  - UKRI-STFC as archive centre (Ceph S3 interface)
- Solutions evaluated
  - FTS / GridFTP
  - Rucio
  - **Rclone** (https://rclone.org/)

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.

2nd Use case: EGI Data-Hub



PaNOSC

Photon and Neutron Open Science Cloud

H2020-INFRAEOSC-04-2018

Grant Agreement Number: 823852

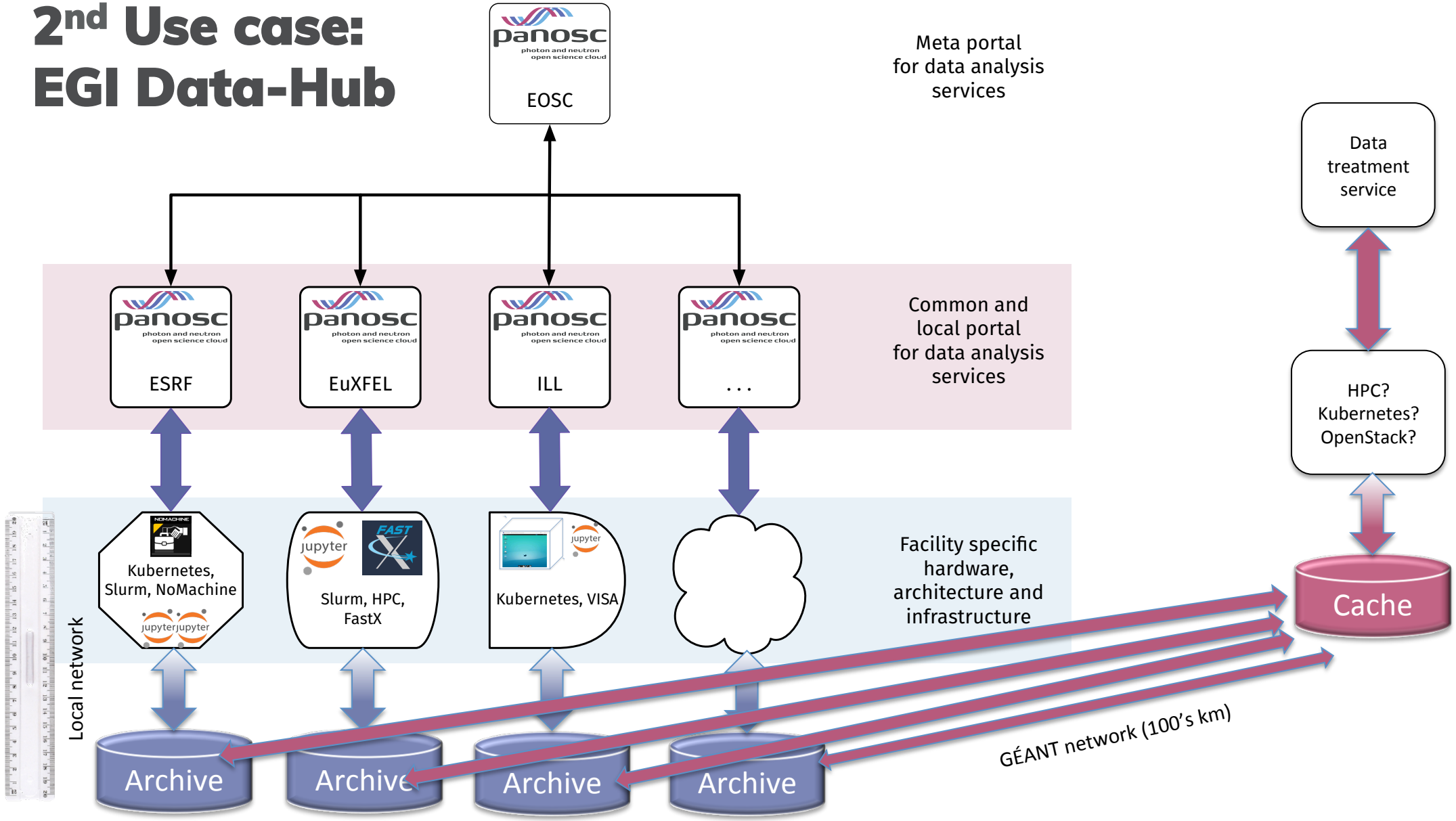
Deliverable: D6.1 – Data Hub

Objectives

- Transfer data to “EOSC” (i.e. not necessarily from the PaN community) data treatment services
- Transfer data transparently from the point of view of the users
- Archive the results back to the originating RI
- Authenticate users using EOSC ready AAI (UmbrellaID)
- Authorise data access (open or embargo data)

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.

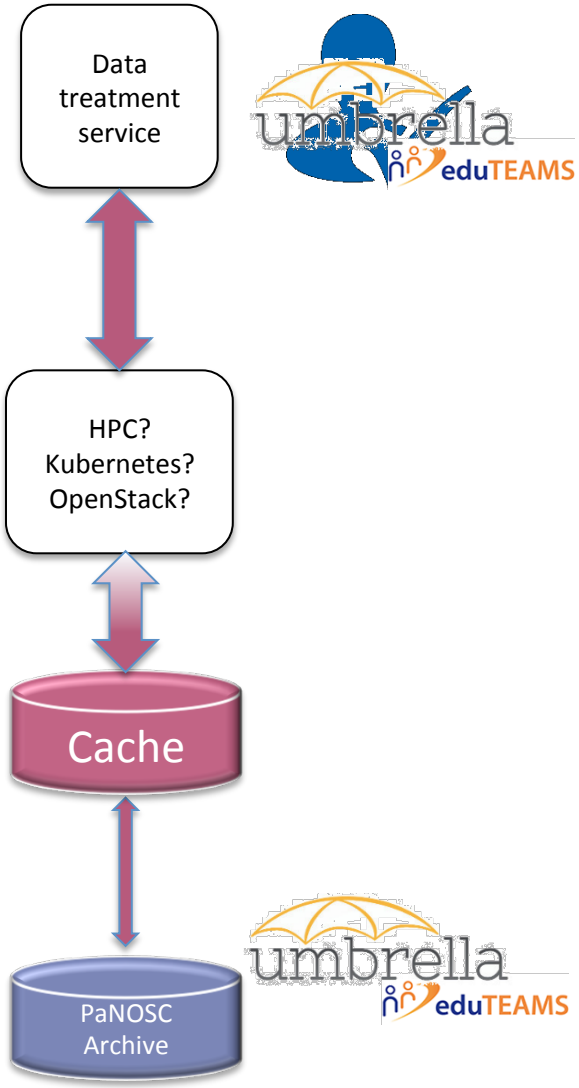
2nd Use case: EGI Data-Hub



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.

Initial results

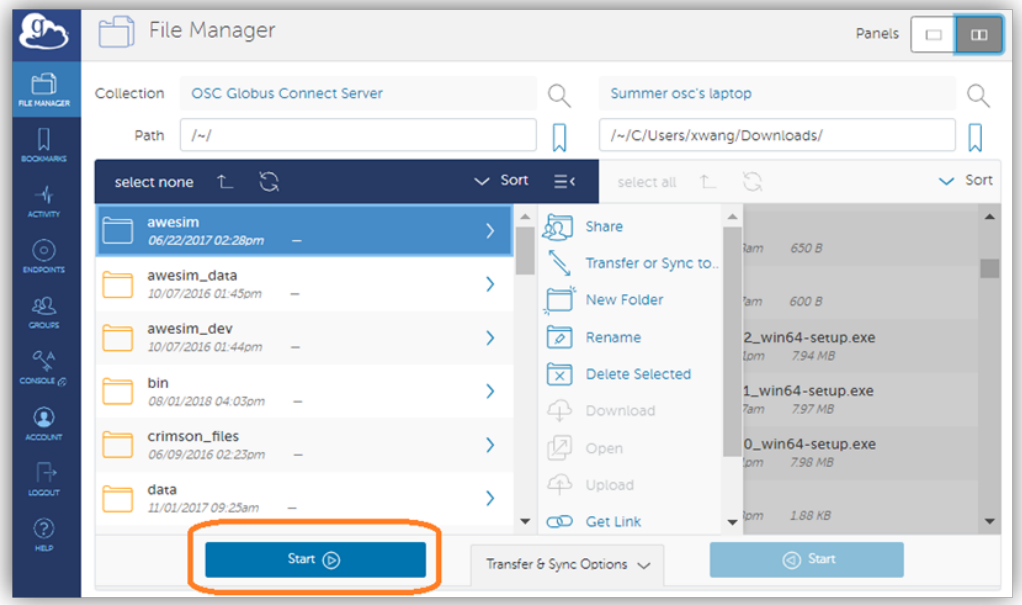
- Promising pilot:
  - Provide the same interfaces for data independently of the archive location
  - Transfer is transparent from the users' point of view
- Performance with small datasets seems acceptable
- Still need to understand the limits (dataset size, …) of the model
- Integration difficulties when trying to deploy the pilot on production environment: hidden files (.snapshot) are creating pb, ACLs, Mapping of users ID (local to UmbrellaID)…
- EOSC AAI model not yet ready for such a cross community use case
- Acceptance tests to be done



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.

3rd use case: transfer of data by RI users

- Aims
  - Provide a (super easy, reliable, able to resume transfer, without tricky configuration on the client side …) solution for RI **users** to transfer large datasets (10s of GB to 10s of TB) to their home lab/company
  - Potential user base: 10 000 users
- Pilot
  - ESRF as RI (data producer) (GPFS + NFS gateway)
  - ESRF users as receivers
- Solution envisaged
  - Globus
  - IBM Aspera



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.

PaNOSC & ExPaNDS
Annual Meeting

AAI - UmbrellaID

Strong collaboration with GÉANT and PSI

Participation in AEGIS (AARC engagement group)

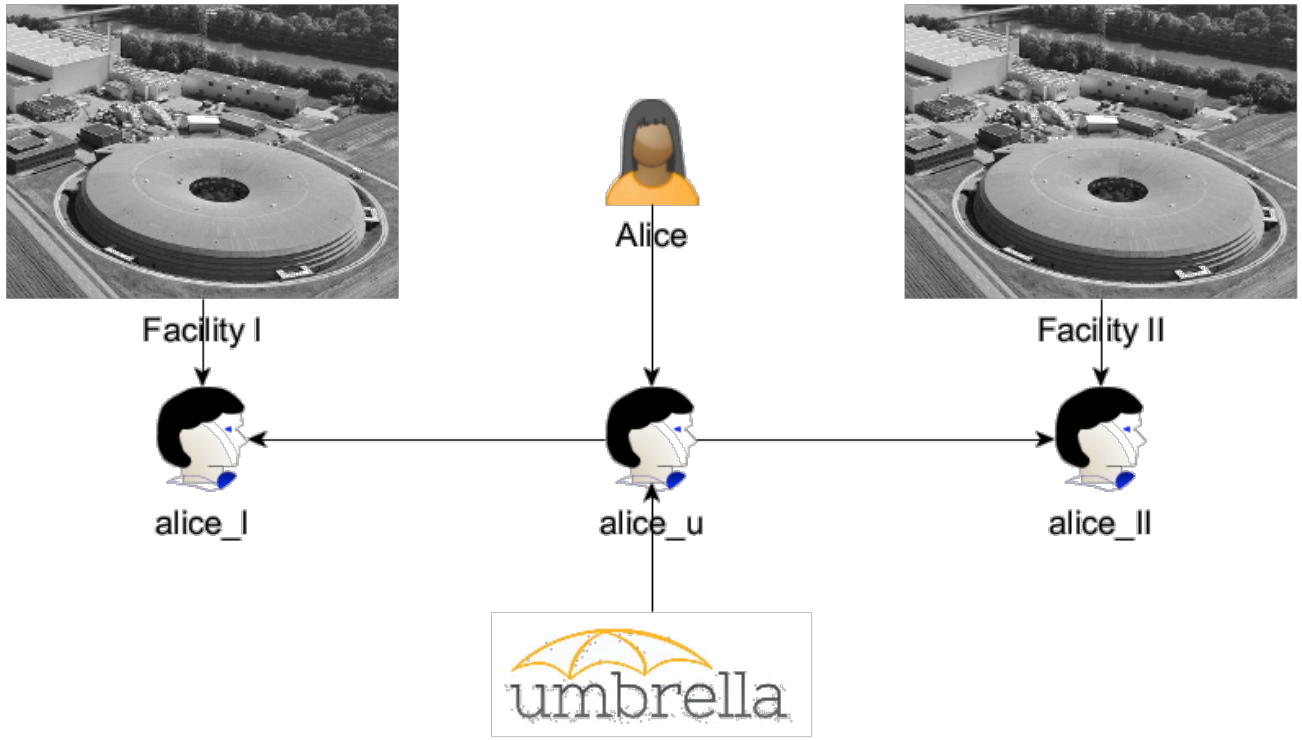
https://wiki.geant.org/display/AARC/AEGIS

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.

AAI

UmbrellaID is the AAI of the PaN community (PaNOSC + ExPaNDS + …) since 2012

- Provide a common Authentication and Authorisation Infrastructure that allows users of all PaN facilities to connect seamlessly to digital facility services with a **single** and **unique ID** (i.e. Common ID)
- Authorisations remain fully under control of each facility providing the services
- Users are in full control of their personal data



https://umbrellaid.org/what.html

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.

AAI

PaNOSC objectives:

- Integrate EOSC – i.e. move to the AARC BluePrint Architecture (BPA)
- Reach global and comprehensive interoperability with services outside of the Photon and Neutron domain
- Take into account growing security constraints and legal requirements
- Increase capacity to respond to user needs in a stable, secure and trusted environment (training and support)
- Collaborate closely with GÉANT to achieve the above objectives

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.

AAI

Looking for collaboration

Specifications

Initial contact with GÉANT

July 2019

Validation by the UmbrellaID mgmt. team.

Autumn 2019

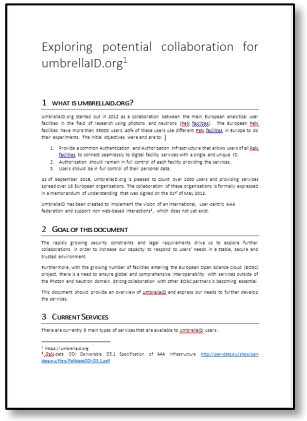
Set up of the dedicated eduTEAMS platform

Validation

Spring 2020

Production ready

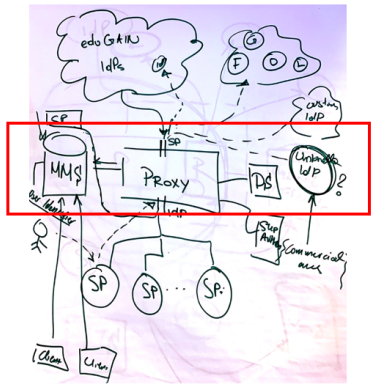
New services introduced


Exploring potential collaboration for umbrellaID.org[1]

1 WHAT IS UMBRELLAID.ORG?

2 GOAL OF THIS DOCUMENT

3 CURRENT SERVICES



Summer 2020

Migration of all SP to the new infrastructure

Additional metadata

Autumn 2020

New IdPs: Orcid, eduGain

Migration of all users

Spring 2021

Decommisioning UmbrellaID IdPs

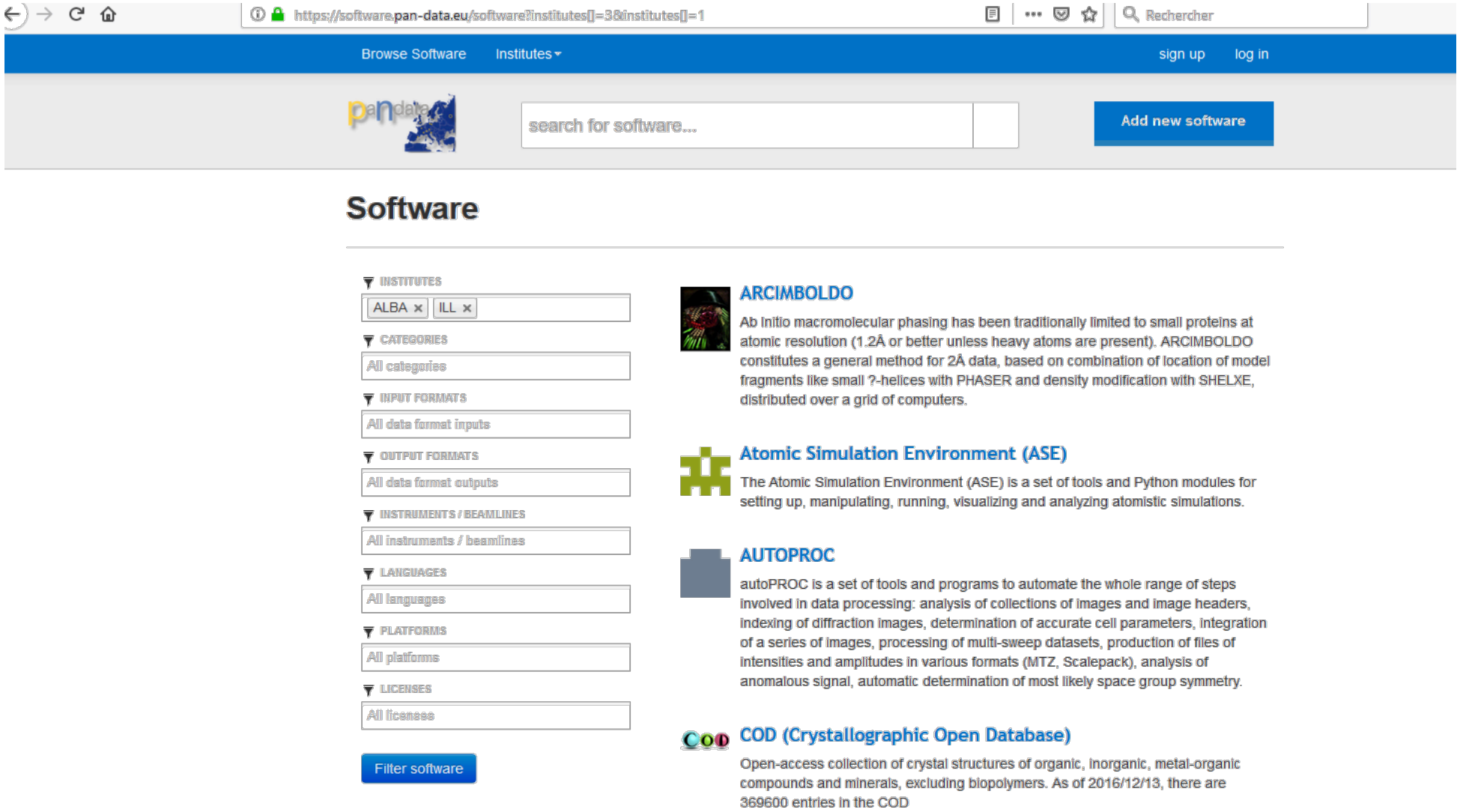
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaNOSC & ExPaNDS
Annual Meeting

PaN software catalogue

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.

PaN software catalogue



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.

2020 Survey community outcomes

https://software.pan-data.eu

- Is de facto the reference software catalogue for the PaN community, but not used yet by all facilities
- It references analysis and simulation software supported by the PaN facilities
- Provides complete examples with data sets
- Practical example from scientific instruments/beamlines

Further requirements:

- Provide reference to Docker/VM image registries
- Creation of a container registry
  - Currently investigate by ILL (is it really necessary? Benefits? Tech solutions?)
- Creation of APIs to access the content of the catalogue
  - Possibility to further link with other catalogues (EOSC, facilities data catalogues/portals)
- Integrate EOSC monitoring/accounting tools when ready

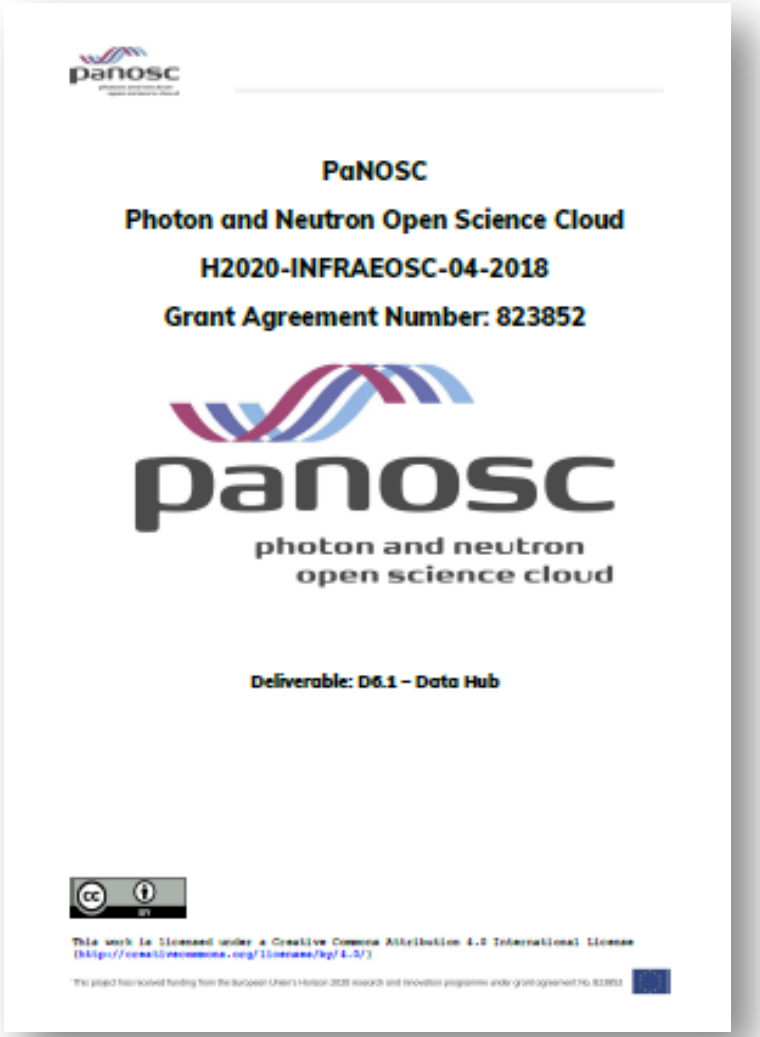
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Current status

D6.1 Report has been delivered in due time
MS19 slightly delayed (2/3 weeks)

Challenges:

- COVID crisis since Jan/Feb 2020
- System/network admin job market is very tight (extremely difficult to hire)



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.

Work ahead 1/2

- Integrate new colleagues (François, Gregory, Teodor, William … and more to come)
- AAI: Take benefits of the UmbrellaID platform
  - Move the existing services to the new infrastructure (SPs)
  - Organise trainings for RI staff (SAML/Oauth, Keycloack) and documentation to lower the technical hurdles
  - Integrate EduGain and take benefits of the new metadata possibilities
  - Connect the new PaN services to UmbrellaID
- Data Transfer
  - Go to production (when the solution fits) and provide feedback to the community
  - Remain open: explore new ideas (e.g. data lake), new technical solutions
  - Identify the limits of the different data transfer and processing scenarios

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.

Work ahead 2/2

- Software catalogue
  - Implement the requested features (mid 2021)
  - Encourage uptake
- EOSC service integration
  - 2 services to publish, more expected (which one? when?)
  - Expecting guidelines from EOSC portal centric projects (EOSC-Enhance, ...)
- Benefits from the project and its participants for working on current or future challenges
  - Software provisioning (CERNVMFS, …) tbd
  - EOSC-Future
- Cloud procurement
  - Gather needs from partners
  - Tender vs collaboration with existing initiative (e.g. OCRE)

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.

PaNOSC & ExPaNDS Annual Meeting

Thank you

Join with your solution or challenges of common interest

wp6@panosc.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.