FAIR Data Analysis Services for Photon and Neutron Science:

16th September 2022

Jean-François Perrin (ESRF – PaNOSC)

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

Projects Cheat Sheet

	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, Soleil, UKRI
Observers	GEANT EU-DAT National RI's	
Linked 3rd 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, G. Bodera	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.

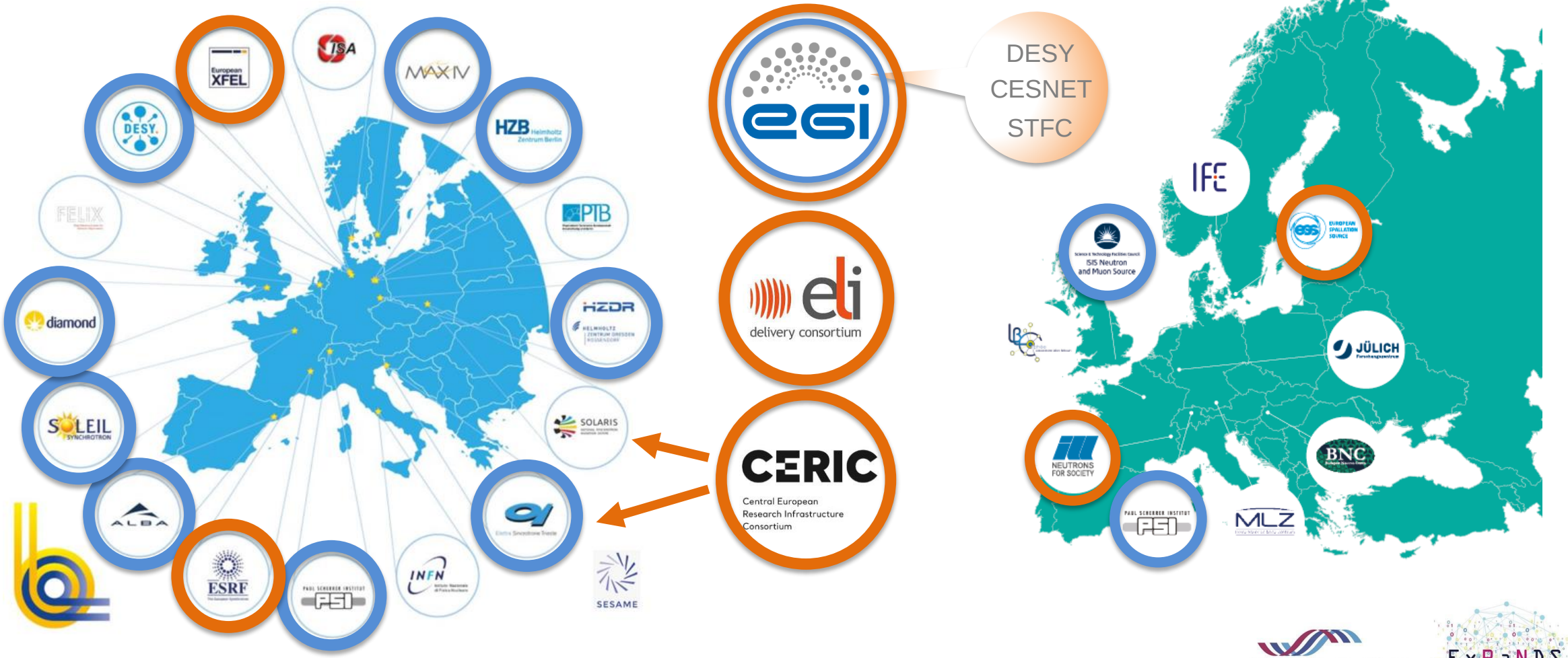
PaN facilities involved in PaNOSC or ExPaNDS

PaNOSC

ExPaNDS

Photon (LEAPS)

Neutron (LENS)



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
photon and neutron
open science cloud

ExPaNDS
European Open Science Cloud Photon
and Neutron Data Services

Some key achievements of these projects

- FAIR data policy framework
  - https://doi.org/10.5281/zenodo.3738497
- E-learning platform
  - https://pan-training.eu/
- FAIR data management practices
  - Data catalogue for all RI
  - Common APIs to access data
  - PIDs for data
  - DMPs
  - …
- Data transfer solutions
- Community AAI ready for EOSC
- **Development and deployment of Data Analysis Services**

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

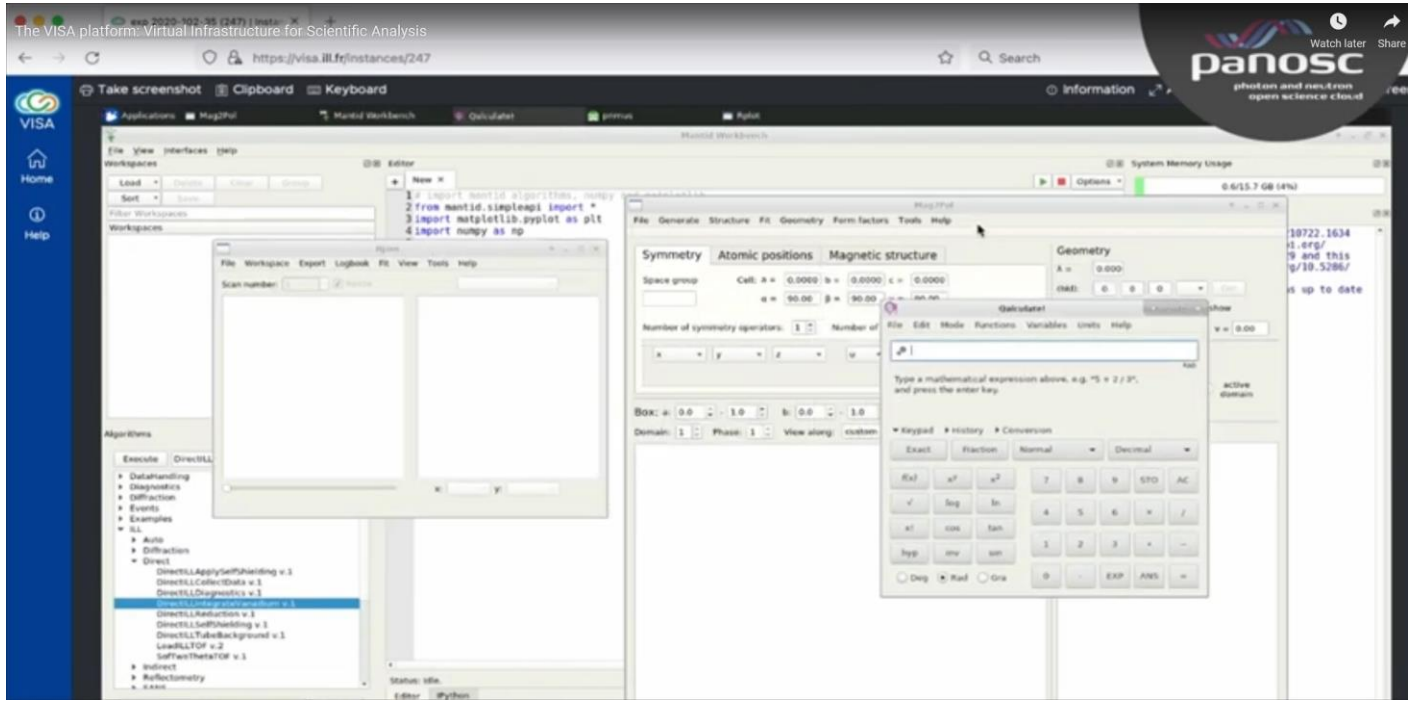
High level objectives for Data Analysis Services

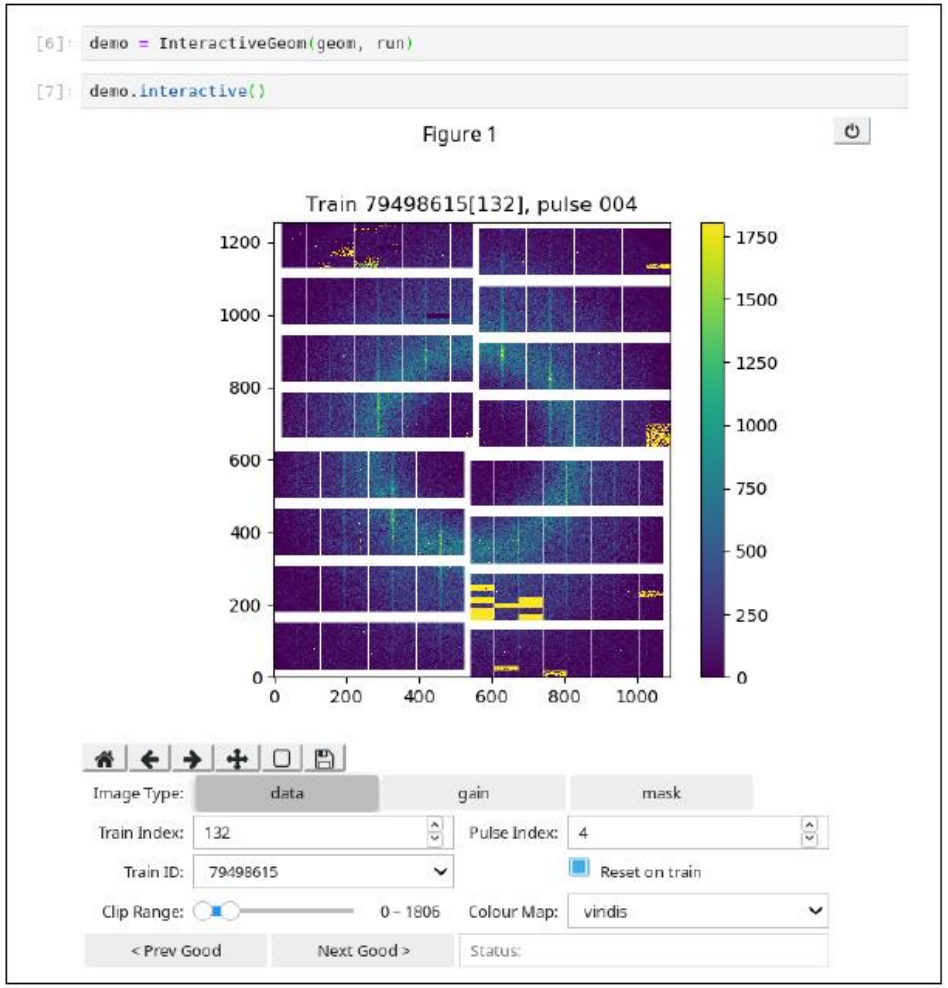
- Keep the high level of RI users' scientific articles production despite the growing complexity of experiments (volume of datasets, …)
- Contain the necessary time for RI users to publish their work after experiments
- Try to get data processing as FAIR as possible (especially Findable and Reproducible)
- Keep the RI computing infrastructure a safe place for research activities
- Running cost for RIs as low as possible

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

Data Analysis Services

- Jupyter Notebooks
- Desktop interface in a web browser





These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

Specifications …

- **Simplify the processing activities for RI's users**
  - Has to be a remote service (Web)
  - Should avoid transfer of data, data should be available for processing
  - All necessary processing software should be ready to be used
  - Computing resources should be available
  - Support from RI experts (Scientific and/or technical) should be simplify

- **Traceability and reproducibility of data processing (FAIR)**
  - Notebooks when possible (preservation of the processing workflow)
  - At least preservation of data and software

- **Cybersecurity**
  - Isolate these services as much as possible from the production RI networks.
  - Keep this service up to date in regards to security patches.
  - …

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

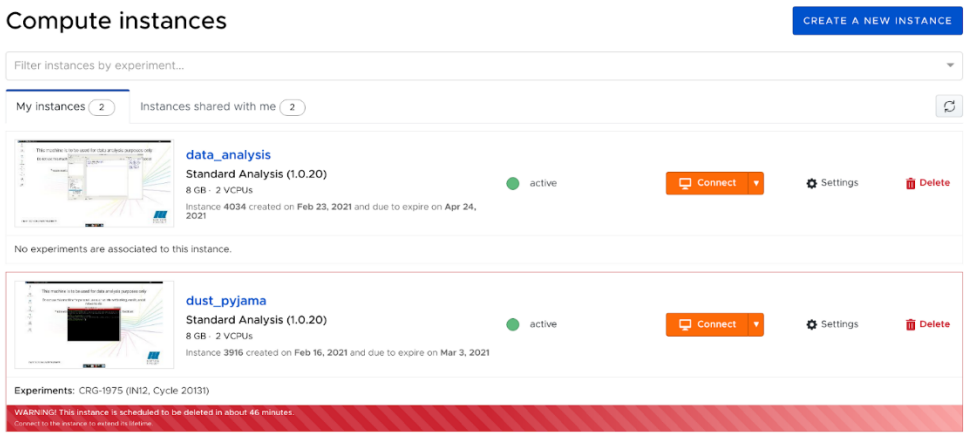
VISA: Virtual Infrastructure for Scientific Analysis

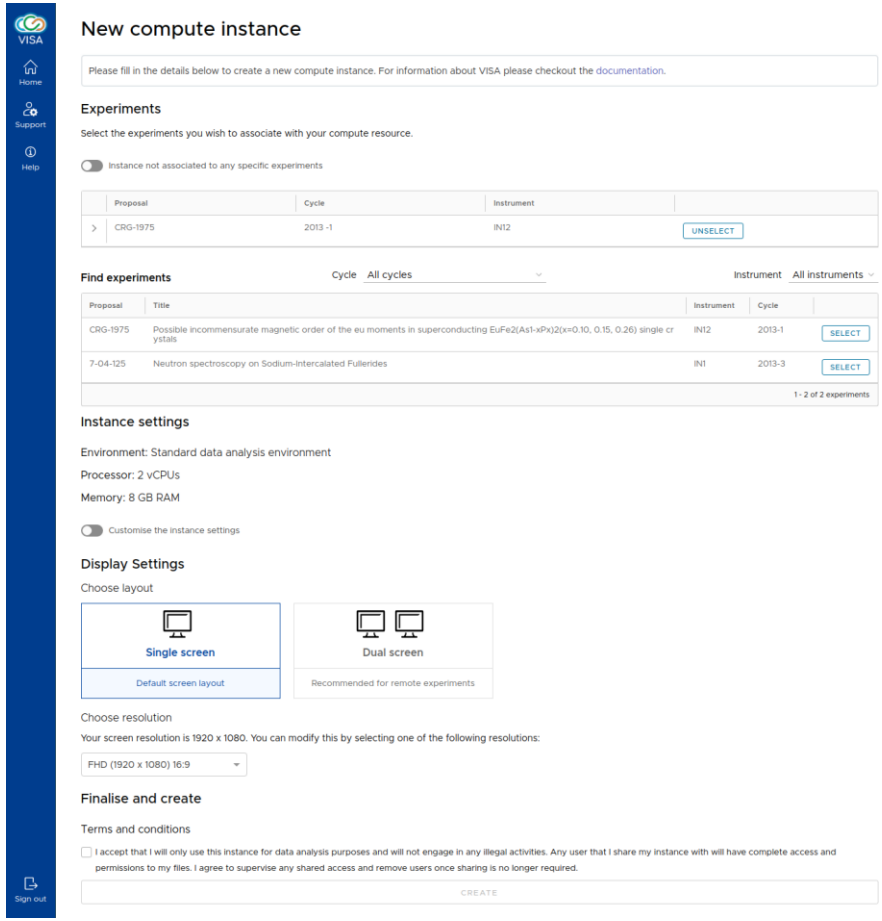
- Provides remote data analysis services in a web browser with access to
  - Experimental data
  - Analysis software
  - Compute infrastructure
  - Support (IT and Scientific)
- Makes access as simple as possible using a web browser
  - Remote desktop as if the user was sitting in front of an RI data treatment workstation
  - Jupyter Notebook environment
  - Easy and flexible machine management
- Allow scientific collaborations and support
  - Sharing remote desktops in real time
- Remote experiments (ILL Specific for now)
  - Access the Instrument Control Software to perform remote experiments

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

Users view – typical workflow

1. Login
2. Selects an experiment to associate to the machine
3. Customises the machine (CPUs, Memory, GPU, Screen resolution, …)
4. The machine is being created (10s to 20s)
5. Starts processing either using the desktop interface either using the Jupyter notebook one
6. Can share the interfaces with colleagues, or RI staff for support
7. Processing data (results) and notebooks are saved alongside the RAW data
8. The machine is deleted after N days or few days of inactivity, users are informed and can request extension.

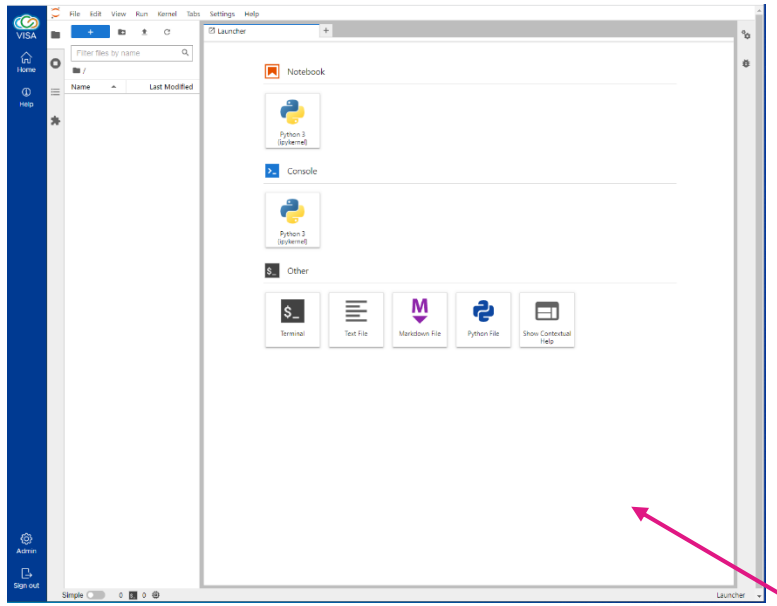


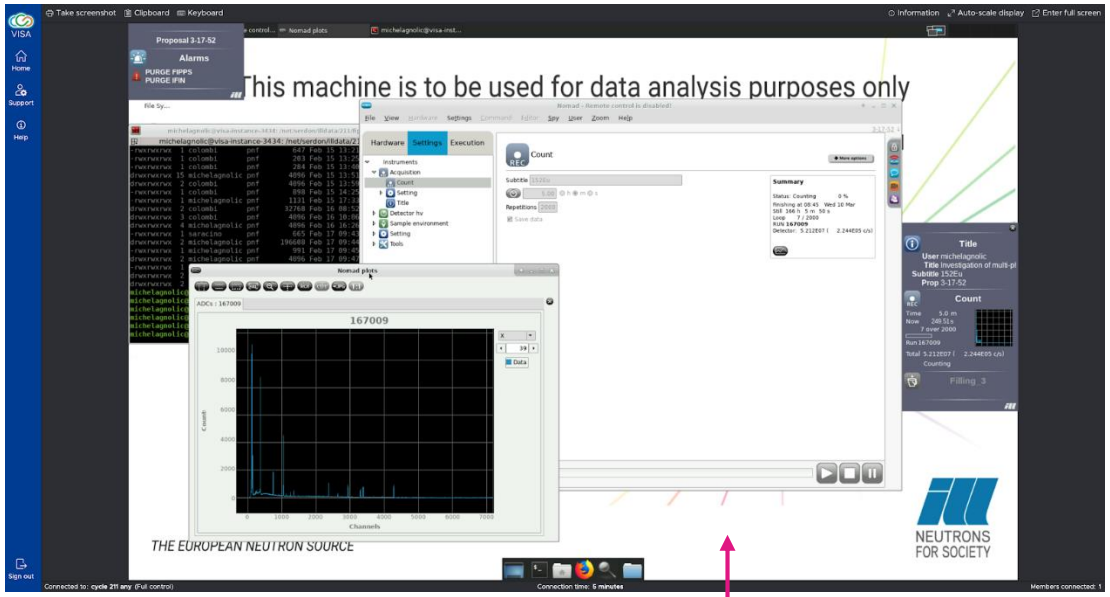


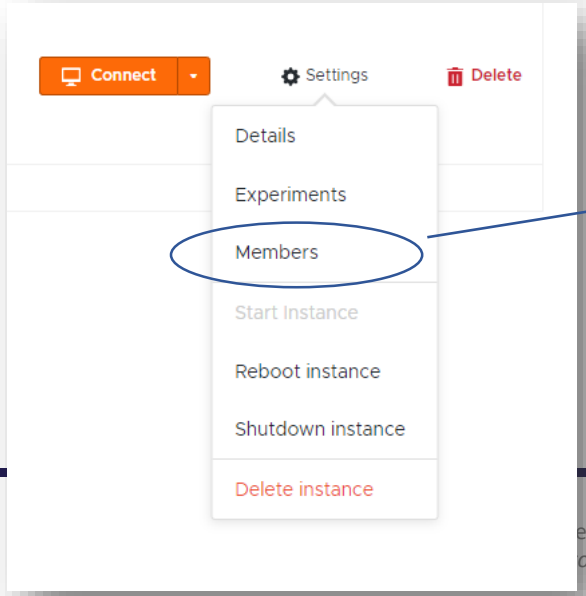
These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

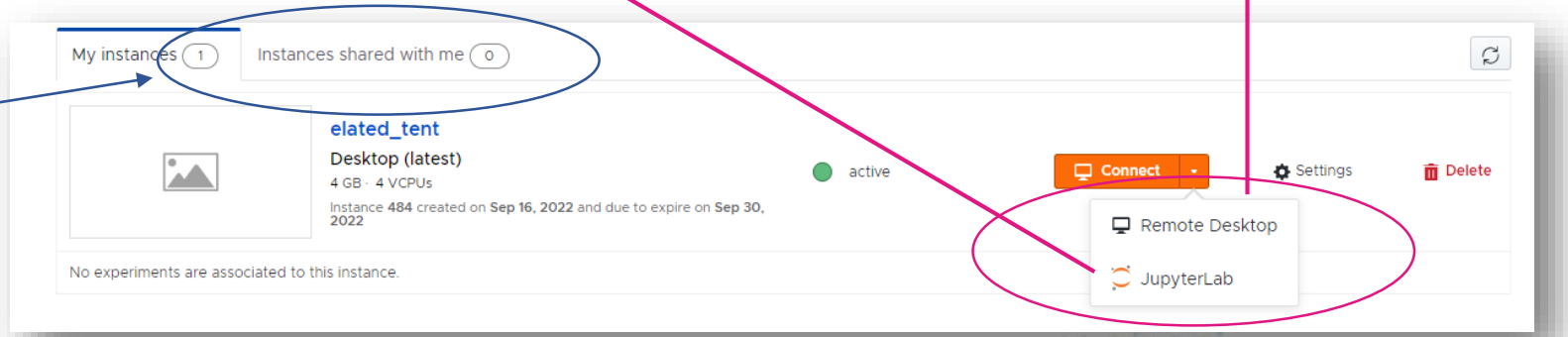
Users view

- Users get access to a desktop (like if they were on sitting on the RI's site) or jupyter interface.
- They can exchange with other scientists and receive support through screen sharing









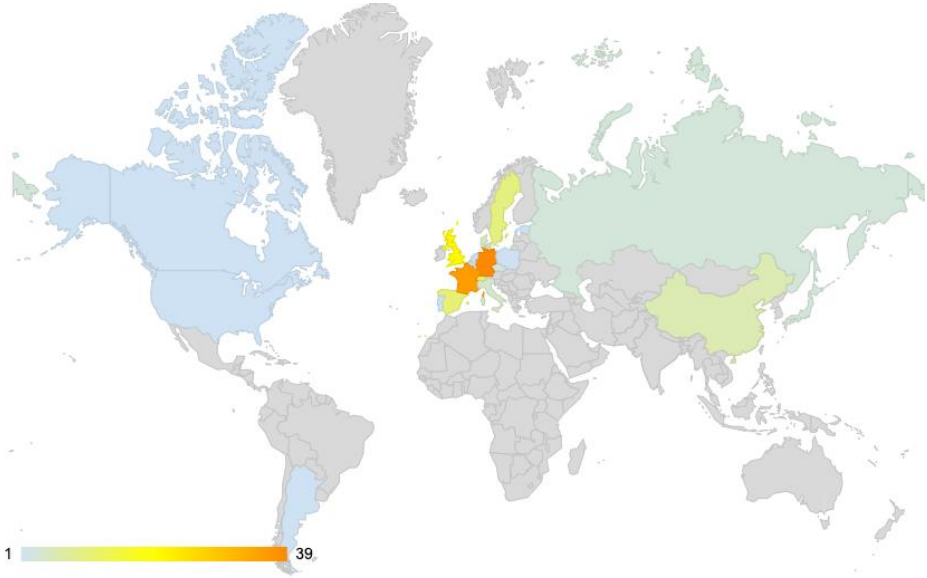
eceived funding from the *European Union's Horizon 2020 research ogramme* under grant agreements No 857641 and No 823852.

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

panosc
photon and neutron open science cloud

Deployment

- Initially developed by the ILL (Institut Laue – Langevin)
  - ILL rolled out VISA early 2020
  - **1,164** distinct users from **38** countries
  - **8,290** VM/instances created
  - **70,000** connections
  - https://visa.ill.fr/



- Further developed inside PaNOSC and ExPaNDS to fit the needs of the different partners
  - Access to HPC resources, GPU availability, …
  - Software provisioning and software preservation solutions (containers)
  - How to handle the need for windows software
  - What about using VISA for mass training?

- Currently, in deployment by all partners
  - Pilots with some BLs
  - Some are starting to open: https://visa.xfel.eu/ https://visa.esrf.fr/ …

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.

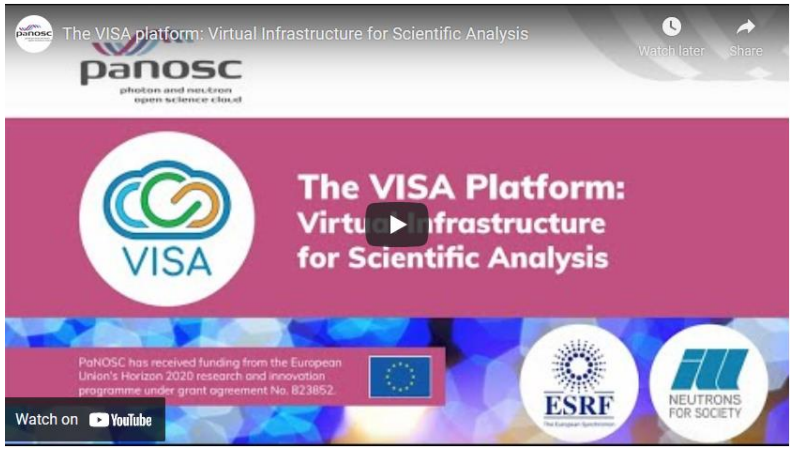
VISA is open source (GPL-3.0 License)
https://github.com/ILLGrenoble

VISA is fully documented
https://visa.readthedocs.io/en/latest/index.html

Thank you for your attention

Time for demonstrations

Special thanks to Jamie Hall & Stuart Caunt (ILL) who are the main developers and have largely contributed to this presentation.



https://www.panosc.eu/news/new-video-released-on-visa-virtual-infrastructure-for-scientific-analysis/

These projects have received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreements No 857641 and No 823852.