



# Outcomes of the PaN EOSC projects for users and user offices

**European User Offices Meeting (EUOM) 14 June 2022**

Presenters: Jean-François Perrin (ESRF) + Patrick Fuhrmann (DESY)  
on behalf of PaNOSC + ExPaNDS partners



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.

# Quick reminder : The Projects Cheat Sheet



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 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, G. 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.



# Evolution of PaN with respect to EU projects.

<b>Policies</b>	Common data policy	<b>FAIR</b> data policy	Data Management Plans	
<b>Analysis</b>	Software Catalogue	Remote analysis	Jupyter	
<b>AAI</b>	UmbrellaID	AARC Blueprint	eduTeams	
<b>Training</b>	e-neutron	Training platform		
	2010	2015	2018	2021



Caterina's Talk



DiTARI  
And more  
Proposal



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.

# What did we achieve and what will we have achieved until the end of our projects?

**RESULTS**



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.



# 10 Primary Outcomes of PaNOSC and ExPaNDS

1. **FAIR data policy** and **DMPs**
2. **FAIR assessment** and common **PID** framework
3. Standardised metadata (**Nexus/HDF5**, PaN ontologies)
4. **Federated search API** for PaN data catalogues
5. **Open Data portal** for searching + downloading data
6. Community **AAI UmbrellaId**
7. **JupyterLab notebooks** and **Nexus/HDF5** files visualisation
8. **Remote data analysis** with **VISA** + data analysis pipelines
9. **Simulation** software for simulating experimental data (**SIMEX**)
10. **PaN-learning** platform ([pan-learning.org](http://pan-learning.org) + [pan-training.org](http://pan-training.org))



# Data Policy Frameworks

Feb 2011 : [https://www.panosc.eu/wp-content/uploads/2019/05/PaN-data-D2.1\\_PolicyFramework.pdf](https://www.panosc.eu/wp-content/uploads/2019/05/PaN-data-D2.1_PolicyFramework.pdf)

Since then DPs have been published in our RIs (ILL, ISIS, ESRF, ...). They are all based on the key principles of the PaN-Data Framework

But since the environment and comprehension has evolved.

**The new texts feature: FAIR, processed/analyzed data, electronic logbook, DMPs, ... and reflect on best practices for writing DP and implementing.**

PaNOSC: <https://doi.org/10.5281/zenodo.3738497>

ExPaNDS: <https://doi.org/10.5281/zenodo.5205825>

Guidance Note, Key Policy Elements: <https://zenodo.org/record/6090282>

The process of updating the DP has started or will start in most of our RIs. User Offices are key stakeholders in this process, it is important that you get access to these documents.



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.



# Data Management Plan (DMP)

A DMP is document that should help the scientists to properly curate and manage their data  
DMPs are more and more frequently requested by funders (H2020, ANR, DFG, ...) .

They could also be of interest for RI to raise the awareness of our users regarding data management questions and to collect information on the potential re-use of the data.

**With DMP users should know what to expect for the RI and what are there responsibilities.**

We (ESRF, ESS, ...) are opening new DMP services for users to:

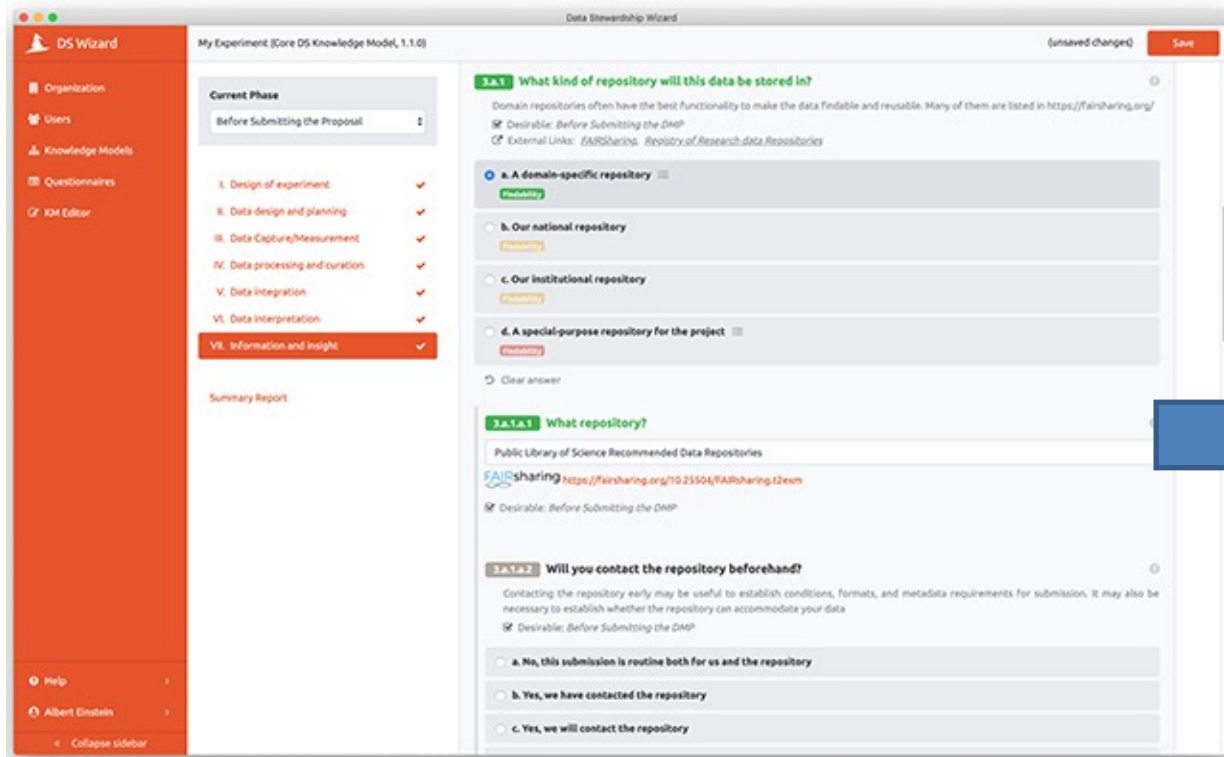
- Help them fill their duty in regards of the funders
- Improve their understanding of our DP and implementation

User offices may receive request from users asking help for filling their funders DMP, a DMP service will definitely help to clarify what users should expect from RI and what do we expect from them.



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.





- A DMP service is a web form
- Prefilled based on RI's Data Policy and practice
- This is a live document, enriched by information from the experiment
- Users have just to fill in a limited set of information (was the experiment successful?, Who could be interested by the data, ...)
- They can generate the document for their funder.

### 3a. How will data and metadata be stored and backed up during the research?

- ✗ *This question has not been answered yet!*
- ✗ *This question has not been answered yet!*

#### Backups

All data are archived and ingested into the metadata catalogue automatically right immediately they are acquired.

Beyond 2 months the data is migrated to tape storage and remains accessible via the ESRF data portal. Two copies are stored on separate tape systems. Writing data back from tape to disk is fully automatic when access to the data is requested via the data portal ([data.esrf.fr](http://data.esrf.fr)). Writing data back from tape to disk can also be done manually by the system administrators. This may be requested via a JIRA helpdesk ticket.

Backup Responsible: The IT infrastructure management is in the remit of the ESRF UNIX team. This team also makes sure that disk storage systems and tape storage are upgraded and maintained as required.

### 3b. How will data security and protection of sensitive data be taken care during the research

#### People who can access the data during the project

Access to the data is restricted to the experimental team for the duration of the embargo period (3 years). During this period the principal investigator can add other users as collaborators (via the data portal). Collaborators have access to the data and the electronic logbook as any other member of the experiment. After the embargo period data are made accessible to everyone i.e. open data. The PI can shorten the embargo period on request.

Identity is managed by the ESRF Single Sign On (SSO) platform. The implementation of Access Control Lists (ACLs) on the ESRF file systems allows fine-grained access control to directories and files depending on the user's identity. Access to



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.





# Data transfer services for users

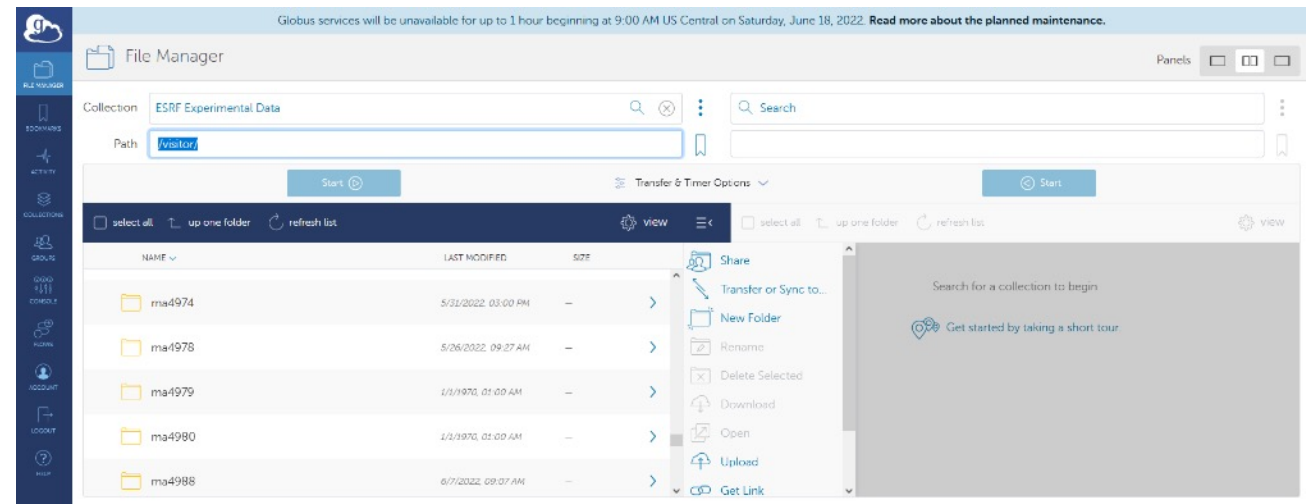
**Solution for users to transfer data (especially large datasets) to their home lab/univ**

We all usually propose data transfer solutions like:

- Data portal download
- SSH Based (RSYNC, SFTP, SCP)

What about very large data? data larger than 1TB, transfers spanning many days:

- Globus Connect is the solution that was identified to address these needs

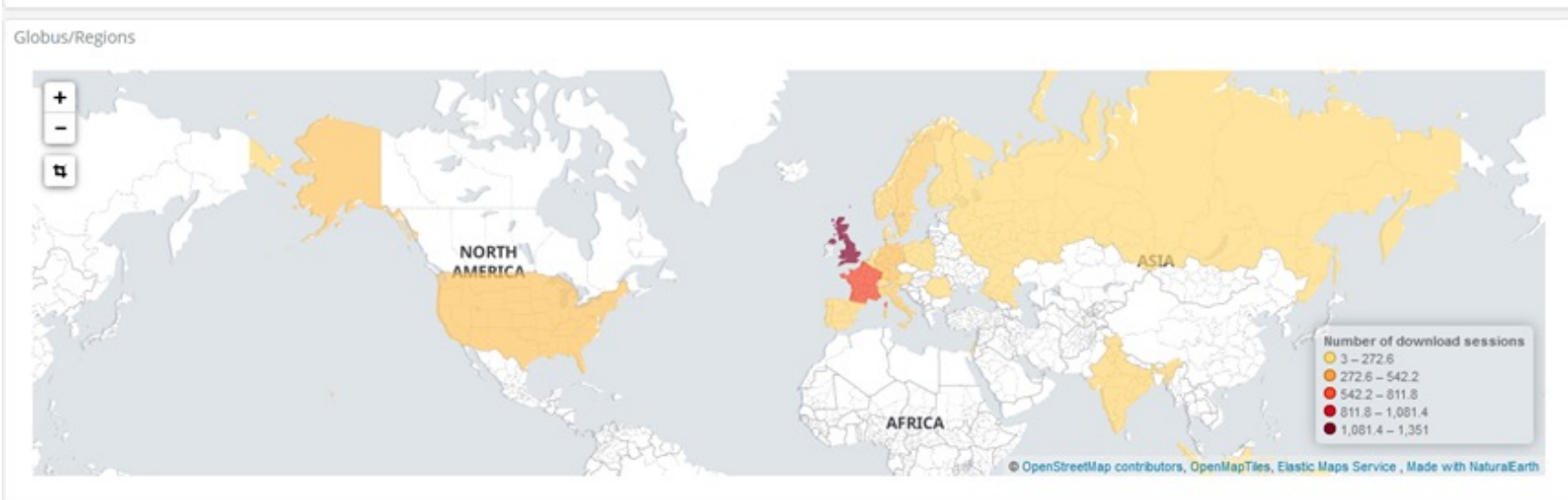
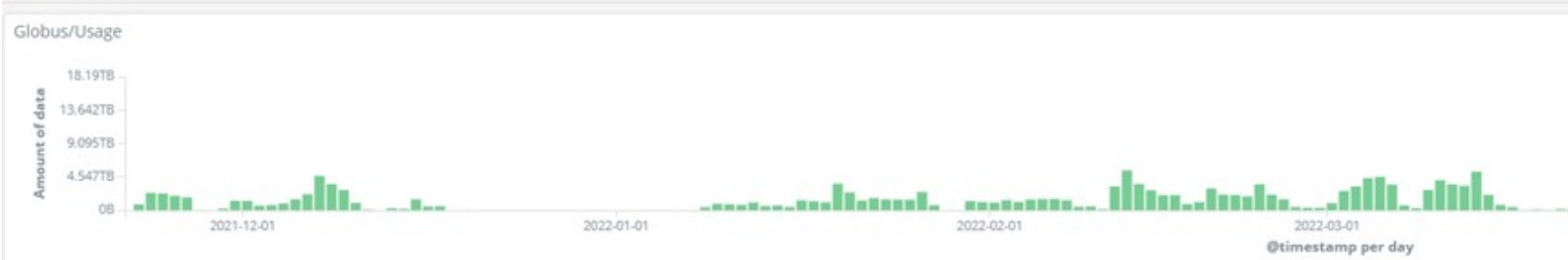


Users can transfer very large data sets reliably (+100TB) if they have the capacity to host these data. Graphical and user friendly solution. Fits a large number of use cases Already implemented by DESY, diamond, ESRF, EU-XFEL, PSI, Max IV, SOLEIL,



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.





### Globus/Top users

username.keyword: Descending ▾

username.keyword	Sum of nbytes
[REDACTED]	40.187TB
[REDACTED]	23.645TB
[REDACTED]	21.049TB
[REDACTED]	20.73TB
[REDACTED]	20.036TB
[REDACTED]	19.273TB
total	17.730TB

### Globus/Top file

file.keyword: Descending ▾

/data/visitor/mx2333/id30b/20220227/RAW_DATA/baduv/baduv-101_x11_cs/ref-baduv-101_x11_cs_1
/data/visitor/mx2333/id30b/20220227/RAW_DATA/baduv/baduv-101_x11_cs/ref-baduv-101_x11_cs_1
/data/visitor/im62/id16b/20201210/WO2C1_ht_120ms_25nm_bis_/WO2C1_ht_120ms_25nm_bis_v1_0
/data/visitor/ma4857/id19/RR2_3/stitched/RR2_3_z0016bits/z00_0010.tif
/data/visitor/ma4857/id19/RR2_3/stitched/RR2_3_z0016bits/z00_0003.tif
/data/visitor/ma4857/id19/RR2_3/stitched/RR2_3_z0016bits/z00_0001.tif
/data/visitor/ma4857/id19/RR2_3/stitched/RR2_3_z0016bits/z00_0004.tif

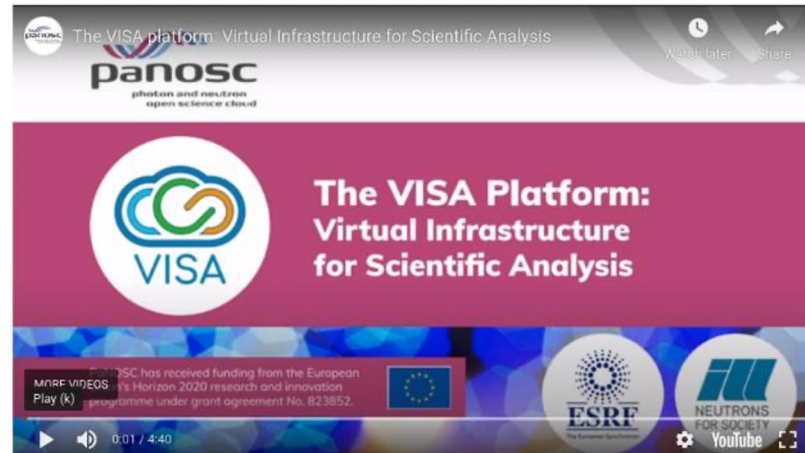


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.



# Users need remote data analysis

- Why it's important
  - next generation of data analysis in globalised research
  - **remote access** to facilities
  - **Mitigate the impact of large datasets** (time to scientific articles, difficulties processing data, ...)
- What we do
  - develop and deploy **VISA** platform
  - make **Jupyter** notebooks available at all sites
- Provides remote data analysis services in a web browser with access to
  - Experimental data
  - Scientific software
  - Compute resources
  - Support (IT and Scientific)

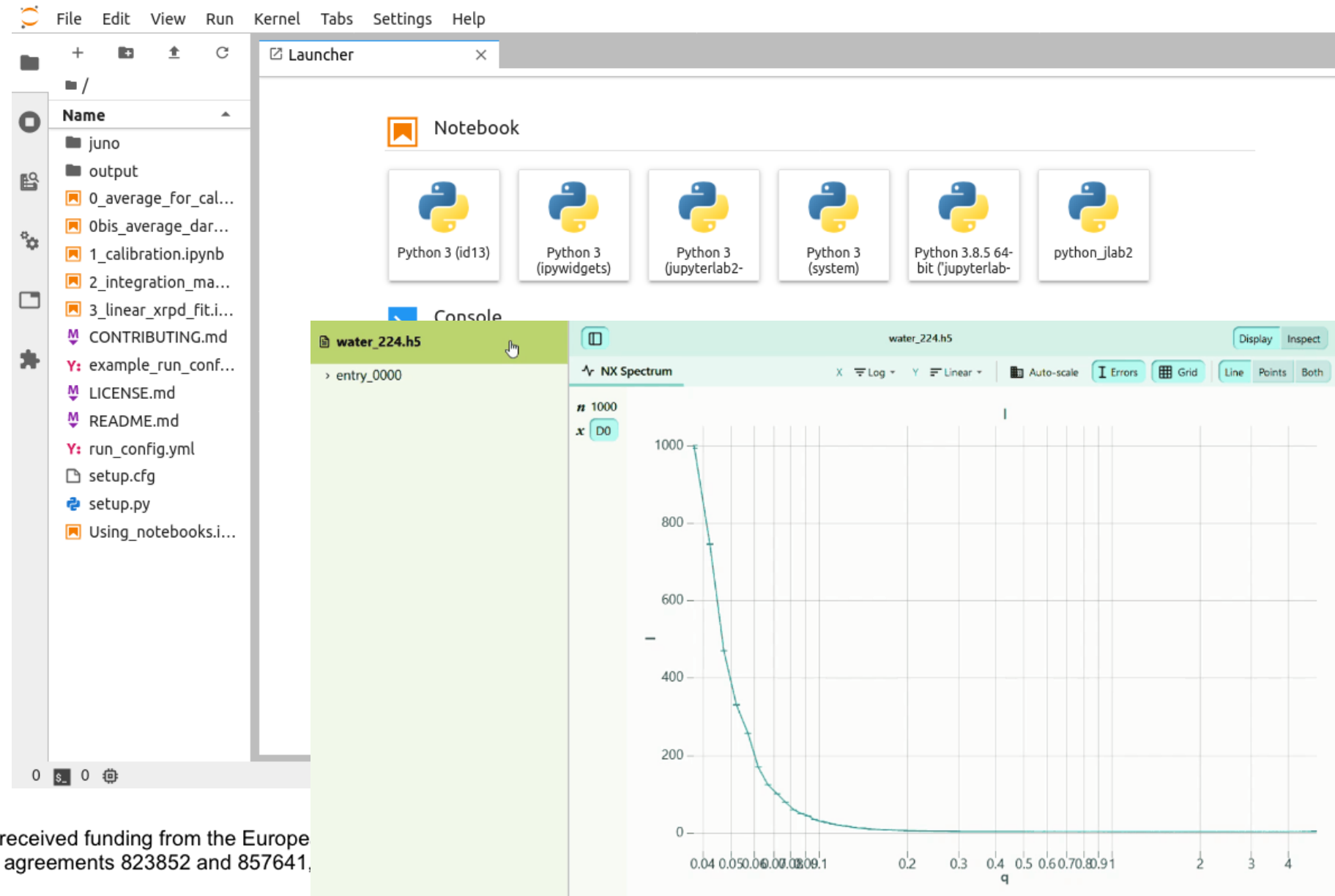


<https://bit.ly/VISA-video>



# Remote analysis + visualization with Jupyter notebooks

JupyterLab has been widely adopted as remote analysis tool



The screenshot displays the JupyterLab interface. On the left is a file browser showing a directory structure with files like 'juno', 'output', and various notebooks. The main area is the 'Launcher' view, which offers several Python kernels: 'Python 3 (id13)', 'Python 3 (ipywidgets)', 'Python 3 (jupyterlab-)', 'Python 3 (system)', 'Python 3.8.5 64-bit (jupyterlab-)', and 'python\_jlab2'. Below the launcher is a 'Console' window for the file 'water\_224.h5', showing a plot titled 'NX Spectrum'. The plot has a y-axis labeled 'I' ranging from 0 to 1000 and an x-axis labeled 'q' ranging from 0.04 to 0.91. The data points show a sharp peak at low q values that decays rapidly as q increases.



# VISA

Users get access to a desktop, like if they were on RI site.

They can exchange with other scientist and receive support through screen sharing

Fully automated workflow, VMs are created by the users and destroyed after N days, data are preserved.

The screenshot displays a remote desktop interface for VISA. At the top, a status bar shows 'Connected to: cycle 211 any (Full control)' and 'Connection time: 5 minutes'. The desktop includes a blue sidebar with navigation icons (Home, Support, Help, Sign out) and a top navigation bar with 'Take screenshot', 'Clipboard', and 'Keyboard' options. A 'Proposal 3-17-52' window shows 'Alarms' for 'PURGE FIPPS' and 'PURGE IFIN'. A terminal window displays a file listing for 'michelagnolic@visa-instance-3434'. A 'Nomad plots' window shows a histogram of 'Count' vs 'Channels' for 'ADCs: 167009'. A 'REC' window shows acquisition settings for 'Count' with a 'Summary' panel. A 'Title' window displays user information for 'michelagnolic'. The bottom right corner features the 'NEUTRONS FOR SOCIETY' logo and 'Members connected: 1'.

Remote data analyses services are being deployed in the different RI  
They should facilitate the processing of the data for the users and help to mitigate the impact of the growing data sets.

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.



# Authentication and Identity Management (AAI)

- Originally
  - Scientist needed a different identity at each facilities to access their services.
    - Lots of passwords to remember and to loose
    - Difficult with cross facility services.
- UmbrellaID
  - Scientist only needs one identity with UmbrellaID (on top of home identity)
  - 'Catch all' identity providers are no longer state of the art and not trusted.
- Very soon (Fall 2022)
  - Scientist only needs the **one identity** from his/her home facility!
  - **Single Sign On**: for cross facility services and beyond (EOSC AAI federation)
  - **Security**: We are part of the European wide CERT system.
  - Extended metadata (email, affiliation, ORCID, ...) will be available when the users login.



# training and learning platform

- The **lack of a central platform for PaN Teaching and Learning** has been identified by the facilities
- We evaluated available, **state of the art technologies for teaching and data collection platforms**
- We introduced a PaN training platform
  - **to create/store** courses and
  - **to collect** existing material
- We are **re-using** successful projects developed by
  - **Elixir (TeSS)**
  - **SINE2020 e-neutrons** and
  - **AAI UmbrellaID/eduTeams**

<https://pan-training.eu/>

The screenshot shows the PaN Training platform interface. At the top, there's a dark blue header with the logo 'Photon and Neutron Training' and navigation links: Catalogue, E-learning, Events, About. A search bar is on the right. Below the header, there are four icons: Materials (Find documents, videos and git repos), Events (Browse events provided by our community), Workflows (Guided processes for specific scientific management), and Providers (Browse by the events providing content). The main content area is titled 'E-Learning Courses' and includes a sidebar with 'Access Courses', 'PaN-wiki', and 'Log on' options. The background features a network diagram with binary code.



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



# Training and Learning Platform

<https://pan-training.eu/>

Apply filters

Add your own materials

Search by keyword

Subscribe to mail alerts

Register training material | Subscribe

What a... | Search materials... | Most recent

127 materials found

Report on status, gap analysis and roadmap towards harmonised and federated metadata catalogues for EU national Photon and Neutron RIs

Report on status, gap analysis and roadmap towards harmonised and federated metadata catalogues for EU national Photon and Neutron RIs

**Keywords:** expands

**Resource type:** Document

PaNOSC search scoring

It illustrates how the Scoring service is designed, how to deploy it and integrate it with the local data catalogue system and the PaN Search api

**Keywords:** expands, PaNOSC, scoring

**Resource type:** github

ExPaNDS mid-term review

Slides of the presentations done during the ExPaNDS midterm review.

**Keywords:** expands

**Resource type:** slides

Scientific topic	Count
powder diffraction	2
absorption spectroscopy	1
crystallography	1
extended x-ray absorption f...	1
imaging	1
inelastic scattering	1
macromolecular crystallography	1
neutron diffraction	1
neutron scattering	1

Target audience	Count
PaN Community	51
PhD students	38
PaN users	33
scientists	14
ExPaNDS and PaNOSC project ...	12
researchers	10
beamline users	8
data curator	6

Catalogue Log in

Search PaN training...

**Materials**  
Find documents, videos and git repos

**Events**  
Browse events provided by our community

**Workflows**  
Guided processes for specific scientific management

**Providers**  
Browse by the institute providing content

nce is.

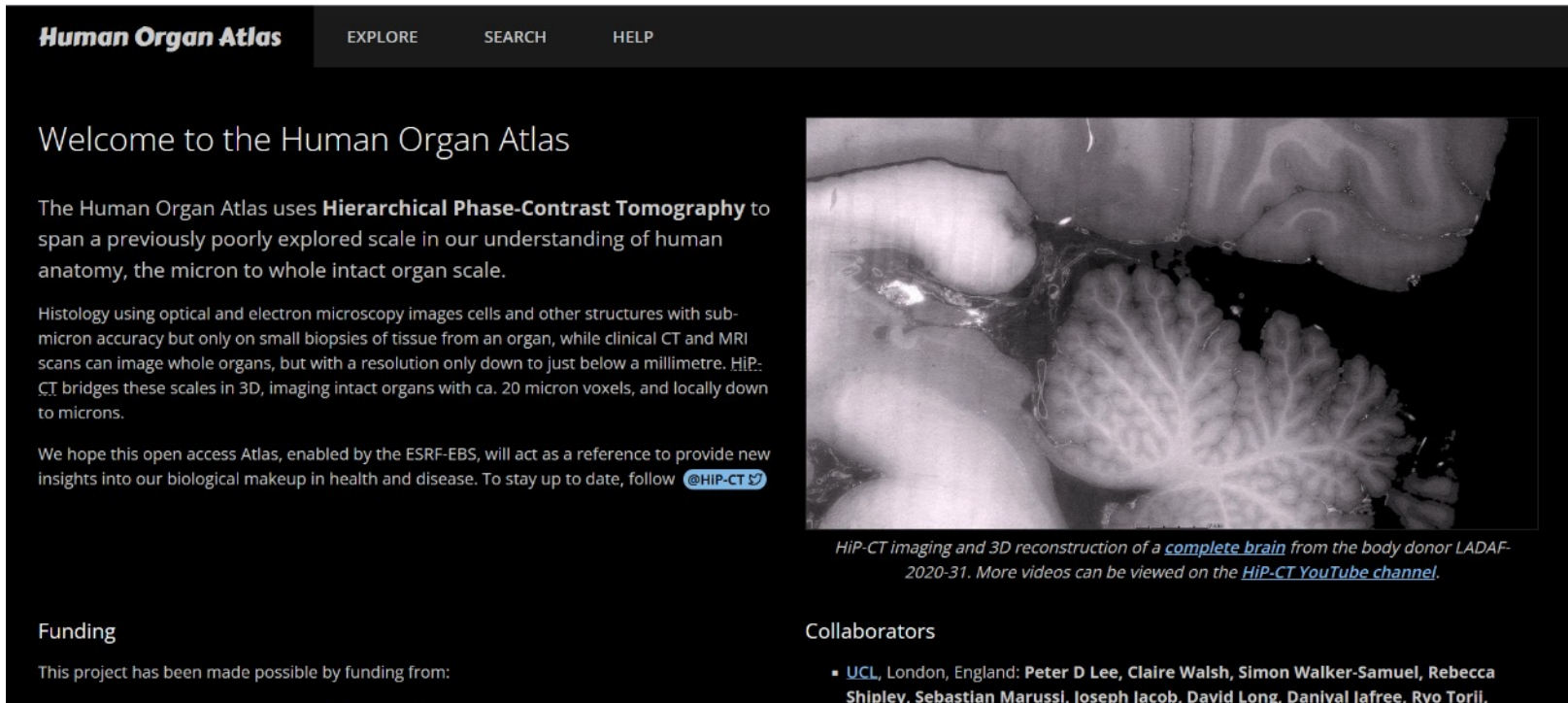


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.





# Open data brings a new type of users



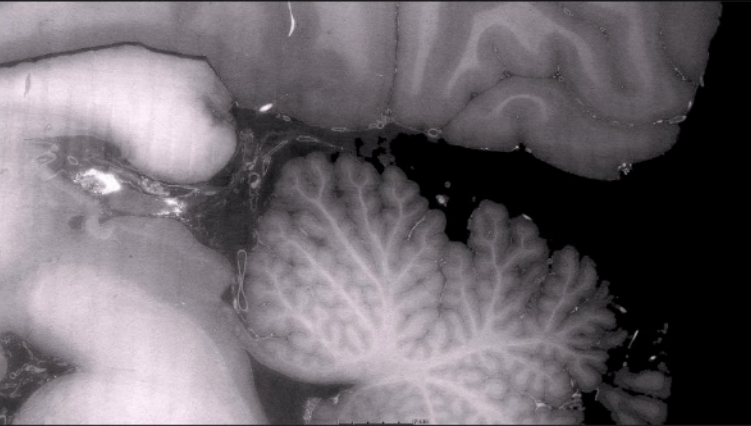
**Human Organ Atlas** EXPLORE SEARCH HELP

Welcome to the Human Organ Atlas

The Human Organ Atlas uses **Hierarchical Phase-Contrast Tomography** to span a previously poorly explored scale in our understanding of human anatomy, the micron to whole intact organ scale.

Histology using optical and electron microscopy images cells and other structures with sub-micron accuracy but only on small biopsies of tissue from an organ, while clinical CT and MRI scans can image whole organs, but with a resolution only down to just below a millimetre. **HiP-CT** bridges these scales in 3D, imaging intact organs with ca. 20 micron voxels, and locally down to microns.

We hope this open access Atlas, enabled by the ESRF-EBS, will act as a reference to provide new insights into our biological makeup in health and disease. To stay up to date, follow [@HiP-CT](#)



*HiP-CT imaging and 3D reconstruction of a [complete brain](#) from the body donor LADAF-2020-31. More videos can be viewed on the [HiP-CT YouTube channel](#).*

**Funding**  
This project has been made possible by funding from:

**Collaborators**

- UCL, London, England: Peter D Lee, Claire Walsh, Simon Walker-Samuel, Rebecca Shipley, Sebastian Marussi, Joseph Jacob, David Long, Daniyal Jafree, Ryo Torii,

<https://human-organ-atlas.esrf.eu/>

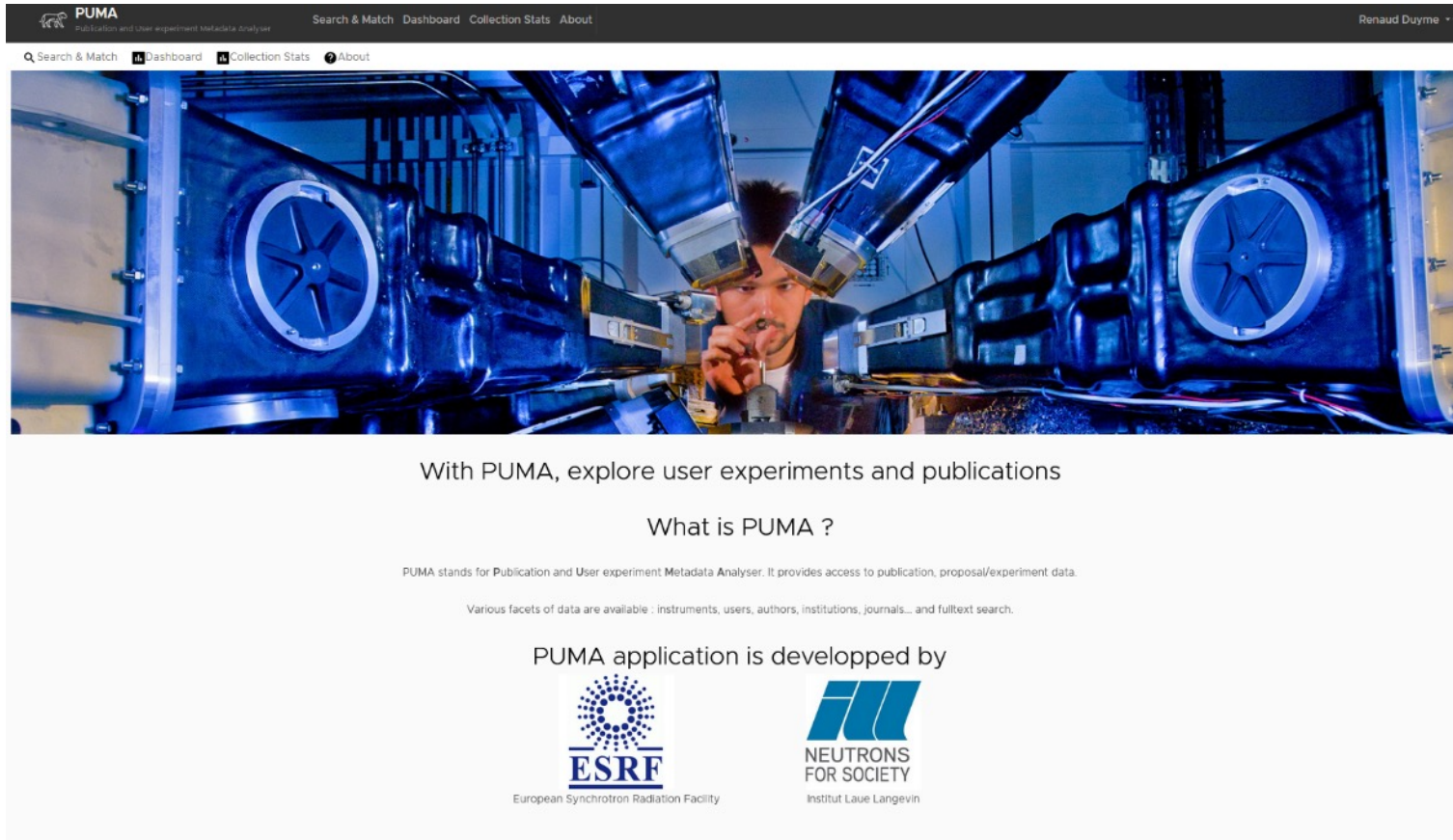
How do we support these new users that do not submit scientific proposals?  
What is the role of user offices?  
How do we count these publications?



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.



- PUMA V1 : Built for the ILL through H2020 funded project FILL2030 –
  - PUMA BI – Business Intelligence (**P**ublication **M**atcher)
- PUMA V2 : ESRF Funded project STREAMLINE : sept 2019 => sept 2022. Migration and enhancement.
  - **P**ublication and **U**ser experiment **M**etadata **A**nalysier



The screenshot shows the PUMA application interface. At the top, there is a navigation bar with the PUMA logo and the text "Publication and User experiment Metadata Analyser". The navigation bar includes links for "Search & Match", "Dashboard", "Collection Stats", and "About". The user's name "Renard Duyne" is displayed in the top right corner. Below the navigation bar, there is a search bar and a main content area. The main content area features a large image of a person working in a laboratory setting, with the text "With PUMA, explore user experiments and publications" overlaid. Below the image, there is a section titled "What is PUMA ?" followed by a paragraph explaining that PUMA stands for Publication and User experiment Metadata Analyser. Below this, there is a line of text stating "Various facets of data are available : instruments, users, authors, institutions, journals... and fulltext search." At the bottom, there is a section titled "PUMA application is developed by" with logos for ESRF (European Synchrotron Radiation Facility) and ILL (Neutrons for Society, Institut Laue Langevin).



## Need for reporting on **activity and scientific impact of research facilities.**

- Current stats are usually relying on institutional repositories tools : **Publication statistics** (citations, journals) using publications list coming from library repositories. **Proposals stats** using data from User Office databases.
- **Instrument scientists** also have to report on **their instrument:** publications, proposals, science topics covered, and techniques used.
- **Facility management board, business/industrial offices** regularly request specific reports based on countries, user institutions (educational/gov/corporate...)

 STREAMLINE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870313.

These analyses are **time-consuming** and not always possible to achieve.



Required solution : **CRIS (Current Research Information System).**

A CRIS is a database or information system to store, manage and exchange contextual metadata for the research activity funded by a research funder or conducted at a research-performing organisation.  
Ex: Clarivates InCites (TM), Elsevier SciVal (TM).

These allow **publication** analysis but are however **not tailored to hold specific entities** used by research facilities : **instruments, proposals, datasets.**





## PUMA is a CRIS for research facilities

Managers, director of research, instrument staff need an efficient way to get overview of science produced by their instruments.

PUMA aggregates meta data of proposals/experiments and publications documents : instruments, authors, institutions, abstract, topics, journals, citations.

PUMA provides :

- dashboards for facility instruments (proposal and publications metrics, citation/journal metrics, institution maps...)
- advanced search feature within all indexed documents.



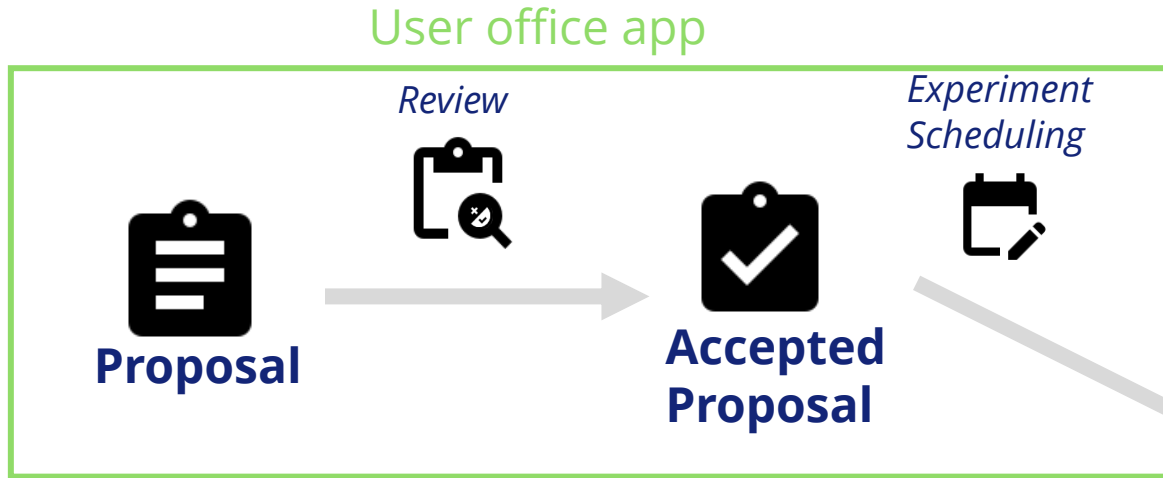


STREAMLINE

# PUMA in research facility architecture



**Institutions :**  
Universities,  
Companies,  
Research institutes,  
Museums... ?



**Sample management app**



**Library app**  
(FLORA)



*Instrument upgrade impact ?*

**Experiment**



*Publication editing*

*Time between  
experiment/publication ?*

*Data  
Analysis*



**Data management app**  
ICAT



Search: "van gogh" painting\*

Year: From [any year] to [any year]

Type: Proposals Accepted

Full text search:  -Uncheck search in: title, abstract, keywords, journal title.  -Check search in: title, abstract, keywords, journal title, fulltext read more

Search result : 56 documents (0.30s). Showing results 1 to 5

Active filters - Type: proposal accepted Expression: "van gogh" painting\*

FIRST PREVIOUS 1 2 3 4 5 6 7 8 9 10 NEXT LAST

**puma record : 90391.json**

**PROPOSAL (ACCEPTED) : co speciation in small pigment, in van gogh paintings**

Proposal IH-CH-659 Deposit : October 2012, Operation Year : 2013 id21 ( ESRF )

Authors : cotte marine; nuys gert; pouyet emeline; cagno simone; hellemans kevin

Keywords : ESRF\_PROPOSAL\_TYPE : inhouse ESRF\_SCIENTIFIC\_AREA : chemistry USER : ihr migration PUMA : new keyword

Abstract : no abstract available for this document.

Institutions : (show affiliations)

- European Synchrotron Radiation Facility, Grenoble, FR
- University of Antwerp, Antwerpen, BE

fullText : (maximum) CO speciation in small pigment, in Van Gogh "co speciation in small pigment, in van gogh" paintings ESRF/CRG Proposer (to whom correspondence) "paintings" embedded in resin and polished Single"

Files : PROPOSAL.pdf smis.esrf.fr

Ids : ESRF\_SMIS\_PROPOSAL\_ID : IH-CH-659 ESRF\_SMIS\_PROPOSAL\_SBM\_ID : 65367

**puma record : 131867.json**

**PROPOSAL (ACCEPTED) : ftir on van gogh samples: looking for painting techniques**

Proposal IH-CH-518 Deposit : April 2010, Operation Year : 2010 id21 ( ESRF )

Authors : cotte marine; salvant johanna; walbert charlotte; megens luc; radepont-kolin marie; geldof muriel

Keywords : ESRF\_PROPOSAL\_TYPE : inhouse ESRF\_SCIENTIFIC\_AREA : chemistry USER : ihr migration PUMA : new keyword

Abstract : no abstract available for this document.

Institutions : (show affiliations)

- European Synchrotron Radiation Facility, Grenoble, FR
- French National Centre for Scientific Research, Paris, FR
- Center for Research and Restoration of Museums of France, Paris, FR
- Cultural Heritage Agency of Netherlands, Amersfoort, NL

fullText : (maximum) FTIR on Van Gogh samples: looking for "painting" techniques ESRF/CRG Proposer (to whom "Description Substance and formula "painting" fragments"

Files : PROPOSAL.pdf smis.esrf.fr

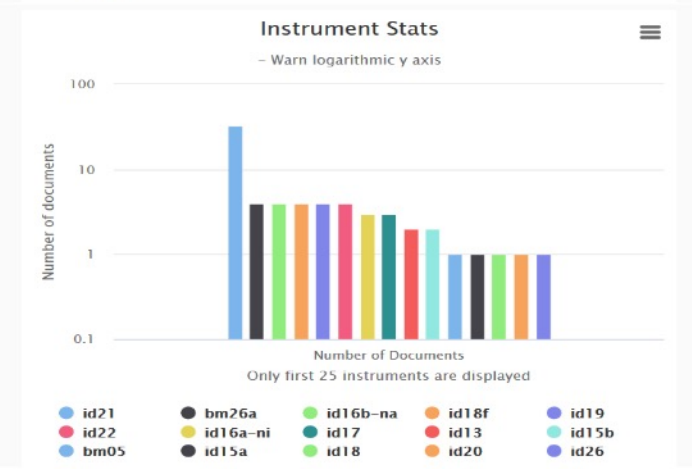
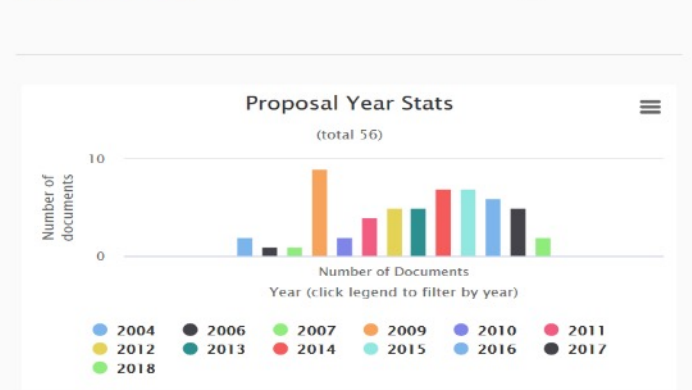
Ids : ESRF\_SMIS\_PROPOSAL\_ID : IH-CH-518 ESRF\_SMIS\_PROPOSAL\_SBM\_ID : 65221

Included in collections : esrf\_prop Target collection : Select new target collection

Search result : 56 documents found (with active search filters) .

Export : XLS DOI DOI (WOS)

ADD 56 DOCS to



- Search proposal or publication
- Full text search within pdf

- Left Column (result):
- Record detail, highlight search match, keywords, instrument, institution, pdf link...

- Right Column (aggregated graphs):
- Document by year
  - Instruments used
  - Main institutions
  - Main authors
  - Topics/Keywords...



## Instrument Group Dashboard : GRP ID01

Microdiffraction imaging Filter : 2015 - 2018

summary publication impact documents keywords journals users & institutions data rules

### Instrument Group History : "GRP ID01"

id01 1997/07/01-2013/12/20 : "Anomalous scattering Beamline"  
 id01 2014/12/01-2018/12/10 : "Nano/Micro-diffraction Imaging Beamline"  
 id01 2020/03/01- : "Nano/Micro-diffraction Imaging Beamline"

Instrument start date should be official opening date (should not include commissioning time)

ID01. Citations loaded

### Timeline



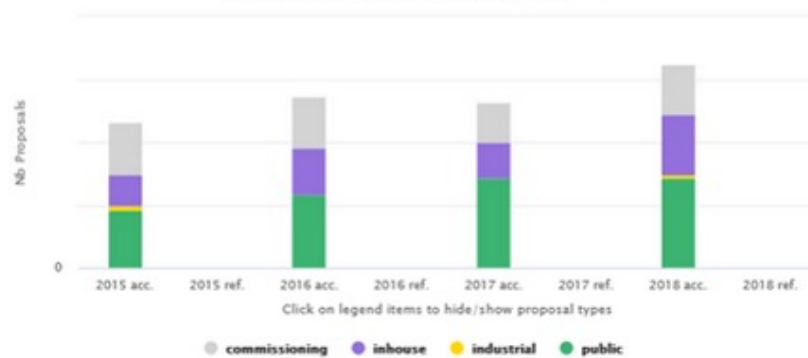
Source PUMA 2021/09/29 12:39

## Publication & Proposals stats

For more info on how documents are added to graph below please see "Data Rules" tab.

### Accepted/Refused proposals by types and proposal year

filter : 2015-2018 - "GRP ID01", (total "Nb Proposals" : )



Source PUMA 2021/09/29 12:39

### Yearly Publications Statistics

filter : 2015-2018 - "GRP ID01", (total "Stacked document type numbers" : )



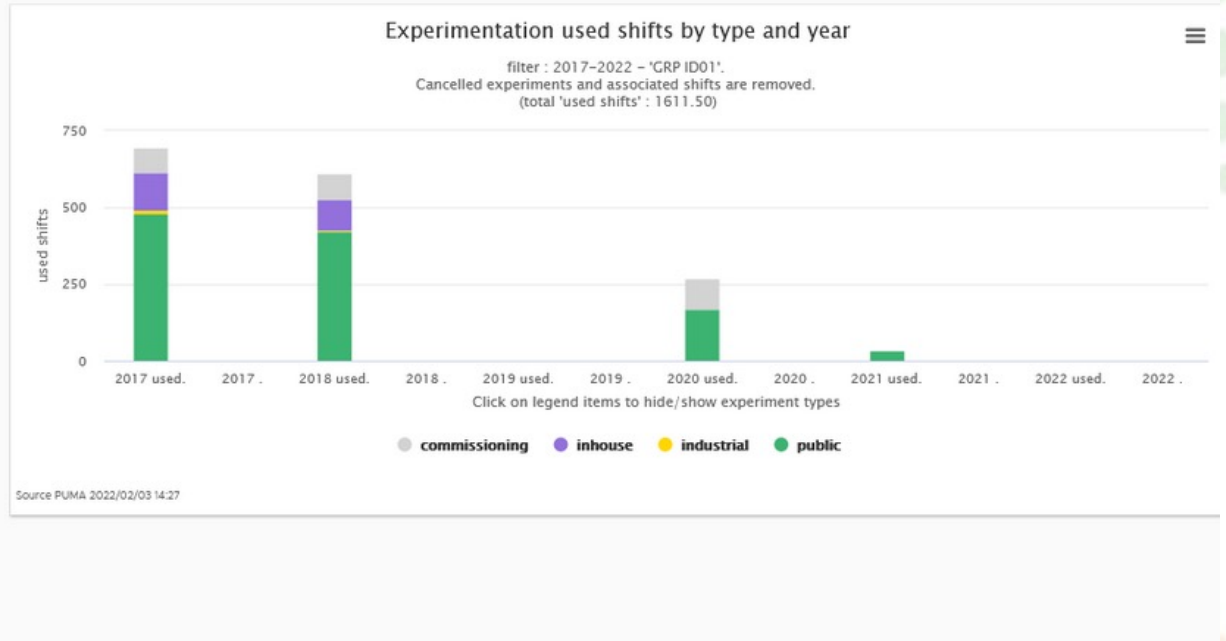
Source PUMA 2021/09/29 12:39





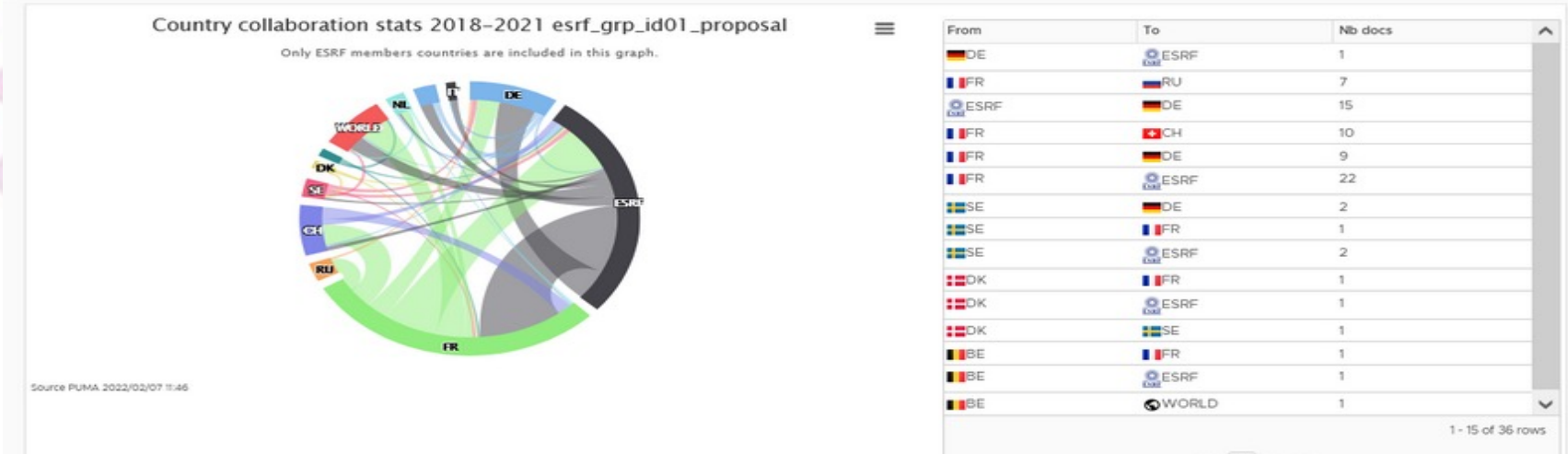
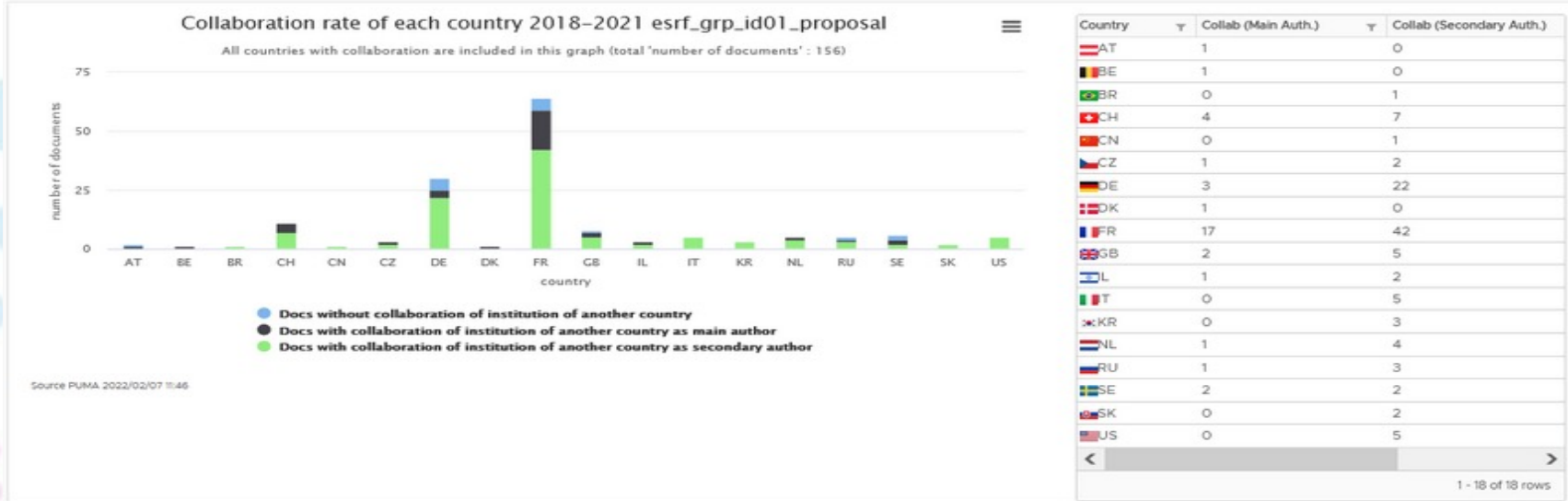


## User countries - Experiments



## Country collaboration in collection

A document in collection can have authors coming from multiple countries. Below are displayed country collaboration figures





ILL and ESRF are willing to set up a collaboration to extend the use of PUMA, are you interested?

If yes do not hesitate to contact us.

Thanks you.

