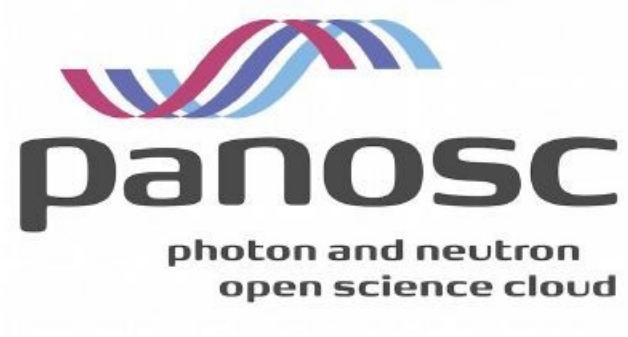


ExPaNDS / PaNOSC Technical Workshop

Survey Results

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

9th October 2020

Introduction

Context and timeline

- Nov. 2019: PaNOSC surveys its facilities

  → get a view of data analysis needs and services

  → develop requirements for the services in WP4

- Sept. 2020: ExPaNDS extends the survey to its facilities

  → Understand how prepared each of our facilities is for the challenges we will face within ExPaNDS (and beyond)

- Today: Presenting the results for all ExPaNDS & PaNOSC facilities *(15 because ELETTRA is included in ERIC-CERIC)*

Contents

- Main aggregated results per section + a few questions raised
- Raw ExPaNDS responses for reference (+ *see PaNOSC confluence* for complete picture)

Scientists & Data

Data generation and user-community of each facility

Data Analysis & Reduction

Tools and services concerning data analysis and reduction

Technology

IT infrastructure available for data analysis purposes

Security

Security requirements and solutions at each facility

Other

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Scientists & Data (1/2)

Scientists & Data

Data generation and user-community of each facility

User communities and portals

- Number of scientific **visits** *(N.B: ≠ unique individuals)* per year:

Facilities	Current situation	Forecast 2023
PaN	**2 347** ±96% visits per year, max: **6 548**	**3 206** ±84% visits per year, max: **10 000**

Currently, a bit more visitors for ExPaNDS facilities on average but very well balanced in 2023 (PaNOSC has 2 new facilities). *N.B: the standard deviation for ExPaNDS' # of visitors is **twice as low** as PaNOSC's.*

- **13/15** facilities have a User Portal, **download services offered to users** are diverse although converging:

PaN	**in-house dev (6/15)**, sftp (5), Globus (4), rsync (3), disk (3), (sc)icat (2), https (2), ssh (1), IRODs (1), TBD (1)	**(sc)icat (5)**, TBD (5), cloud services (3), sftp (2), disk (2), Globus (2), rsync (1), https (1), ~~ssh~~, ~~in-house dev~~

Data generation

- Yearly, PaNOSC produces on average **twice as much** data as ExPaNDS and foresees a **x10** for 2023

ExPaNDS	**1.7 PB** ±98% per year, max: **5.4 PB**	*Not enough data*
PaNOSC	**3.5 PB** ±136% per year, max: **10 PB**	**35 PB** ±111% per year, max: **100 PB**

- And the difference is even more important for **datasets sizes** where the largest dataset is **850 TB vs 50 TB** (x17). *N.B: ~3x more experiments at ExPaNDS facilities*

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Scientists & Data (2/2)

Open data generation

- **Much more open data in 2023** with a trend for 3 years embargo periods (sometimes extendable to 5 years on request)

	Current situation	Forecast 2023
ExPaNDS	**3/9** facilities have **open data** **4/9** have a **3-years embargo** period, **2** 5-years	**8/8** will have **open data** by 2023 No foreseen change in embargo periods
PaNOSC	**3/6** facilities have **open data** **4/6** have a **3-years embargo** period	**6/6** will have **open data** by 2023 **6/6** will have a 3-years default embargo period

FAIR data and curation

- **Nexus** is already widely used, and there is no real foreseen convergence of formats

PaN	**Nexus (10)**, **hdf5 (8)**, **instrument-specific (5)**, edf, cbf, dat (3), tif/json/txt/ascii/csv/ssv (1-2), raw, jpeg, mrc, pxt, jupyter-nb, pxt (1)	**Nexus (11)**, **hdf5 (8)**, instru-specific (4), cbf, dat (3), edf (2), tif/json/txt/ascii, raw, jpeg, mrc, pxt, jupyter-nb, pxt (1), openPMD (1), ~~csv/ssv~~

- **All but one planning to provide DOIs by 2023** and trend for a min **10-years curation** duration

PaN	**7/15** already providing DOIs **6/15** curating data for 10 years, **4** unlimited, **2** 5-years, **1** 3-months	**14/15** providing DOIs (and 1 also handles) **7/14** curating data for 10 years, **3** unlimited, **3** 5-years

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Data Analysis & Reduction (1/1)

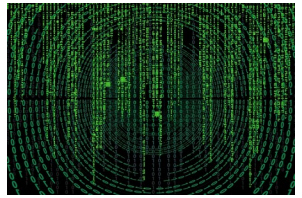
Remote access services for data analysis

- **8/15** facilities provide Jupyter Notebook / JupyterLab services to visitors + **4** are developing it
- **Many users** (>100) only for **3/12**, and still **very few users** (~10) for **7**, **1** in between (~60)
- Almost all (**13/15**) provide some sort of remote access, mostly **remote desktop (7)**, ssh access (5), remote shell (1), web app (1)

Training and support to users on data analysis

- Most offer training (**12/15**) but in most cases (**9/12**) it is only on request/not centrally organised/during experiments. Only 2 reported offering e-learning.
- All offer support to users (**13/13**) although often (6) not officially nor centralised

Access to processed data and e-logs



- All preserve processed data at the facility (**14/14**), incl. restrictions:
  1 on-demand only, 1 data generated on site only and 1 for in-house experiments only
- The curation period is often the same as the raw data (**6/15**, 5@ExPaNDS), it is shorter for **3** facilities and to be defined for **6**
- **7/15** facilities offer remote access to experimental logs after the experiment, 5@PaNOSC and only 2@ExPaNDS. In the 8 remaining, 4 don't at all and for 4 it is instrument-specific.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Technology (1/4)

Technology

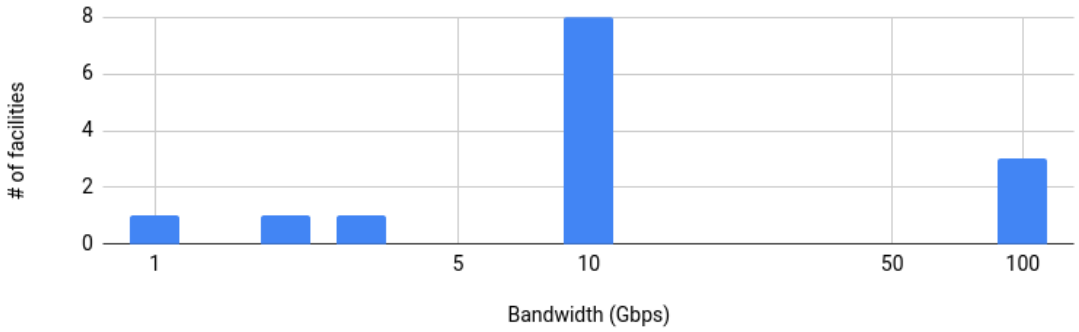
IT infrastructure available for data analysis purposes

Data analysis OS and software distribution

- Supported operating systems for data analysis: **CentOS (9), Ubuntu (8)**, Windows (6), Debian (3), RedHat (3), MacOS (2), Scientific Linux (1)
- SW distribution: **NFS/GPFS/AFS shared file system installations (6), custom Linux repositories (5), Git repositories (5)**, ubuntu/debian/conda/pypi packages (4), CERN-VMFS (3), Singularity containers, Mantid-like, Dockerhub and release downloads (1)
- Please refer to raw responses for commonalities

Compute infrastructure

- All facilities have a HPC cluster for scientific computation (general purpose vs. dedicated) with a very wide range of performances – from **24 cores / 4224 GB memory** to more than **18000 cores / 150 TB memory**
- Main Linux distribution for the server infrastructures: **CentOS (9)**, Debian (5), RedHat (4), Ubuntu (2), Suse (1)
- Bandwidth ranges from **1 to 100 GB/s** (in = out)



The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Technology (2/4)

Emerging tools

- Notable interest for cloud technologies in general and **containerisation** in particular

Data Analysis & Reduction

JupyterHub (3) BinderHub DAaaS

AI/ML pipelines (2)

DASK Py-bind-11

High-end compression tools

Technology

FPGAs (2) Keycloak SSO

RDMA over converged Ethernet

BlueSky (machine group)

Globus, Message queues

Common workflow language

Remote desktop technologies OpenStack

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Technology (3/4)

Remote desktops

- **10/15** offer remote desktop access to users during and after the experiment (N.B: more than for DA).
- Mentioned tools are NX (4), VNC (2), FastX, ISIS DAaaS, VISA

Cloud infrastructure toolbox

- **5/9** @ExPaNDS and **5/6** @PaNOSC have a **cloud infrastructure in place**. *Bigger challenge here for ExPaNDS.*
- Tools for machine virtualisation: **VMWare (10)**, KVM (4), xen, Citrix Hypervisor, OpenStack, Ovirt (1)
- Tools for containerisation: **Docker (14)**, **Singularity (8)**, Podman (3), Kubernetes (3), Rancher
- Protocols to access data: **NFS (13)**, SFTP, GPFS (7), SMBFS (5), SMB, CIFS (3), remote FTPS, GridFTP, SSH-tunnels (1)
- Only **1/15** provides FaaS (**PaaS**) but mentioning it is not much used by users yet, 4 have / are investigating it
- Most (**10/14**) use **internal** resources to host cloud infrastructure, see raw results to know who to call for outsourcing tips (spoiler: PSI, ALBA, Diamond, SOLEIL)

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Technology (4/4)

Governance of computational resources for experiments

- 3 types of governance: fixed resources allocated **by instrument / experiment** + batch scheduler (**6**), **ungoverned (6/15)**, or **‘first come first served’ (3)**
- **6/15** allow users to **choose the amount of resources** for their DA
- **7/15** have **quotas** in place (number of concurrent usage, number of nodes per job…)
- **Slurm** is the job submission system used by **13/15** facilities. There is also **1** using **Torque** but looking into Slurm and **1** using **Univa Grid Engine**. OAR mentioned as legacy.

The ExPaNDS project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 823852.

Security (1/1)

Security

Security requirements and solutions at each facility

Authentication

- Federated AAI used: **Umbrella** is supported by **10/15** facilities (incl. 2 for user office only), **Keycloak** by **3**+1, Helmholtz AAI by 1 – internal only for **4** (3@PaNOSC).
- Protocols: **CAS, SAML, LDAP (4)**, Kerberos (3), OpenID Connect (2+1), ssh, OpenLDAP, Moonshot (1)

Access to data or analysis services for non-users

- **10/15** allow non-users to access data or services, among which: **4** offer open data **without authentication**, **5** require to **create a user account**, **2** allow authentication through a **federated AAI**
- Only **2/15** provide **means to create an account** outside of the user office

Access to data or analysis services for users

- **10/15** offer a user home at the facility, usually Linux home on NFS
- **6/15** require **VPN connection** before accessing internal services, not required for the other **9**
- Users are associated to their experimental data via **LDAP groups** for **12/13**, **NFS ACL** attributes are used by **3**

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Some conclusion and questions raised

? How to account for remote users

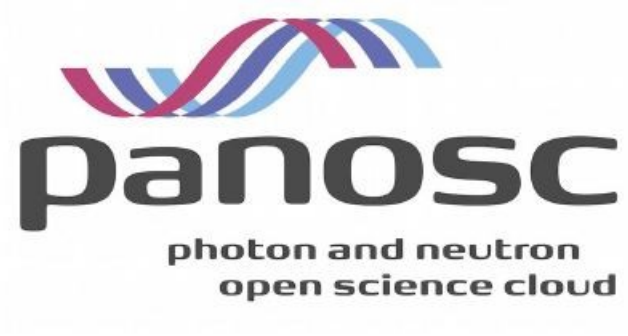
? DOI per proposal / per sub data-set / all data / only published data - *cf. next WP2 deliverable*

? Differentiate published and raw data in terms of curation in data policies

? Dealing with root access to remote VMs **-** *cf. discussions on uid/gid credentials for data access*

- Common tools and services – *cf. discussion from yesterday on an 'adaptation repository' to share e.g. portal plug-ins, but also Docker images*
- Tools that the facilities would like to explore in more depth: **container technologies** win
- Consensus on the necessity of a **Slurm module** – *cf. discussion from yesterday on a more general need for requirements gathering and 'upvoting' on Confluence?*
- **Bottlenecks** for future services: HPC *dedicated* resources and bandwidth
- ~ all needing to "upskill" on cloud technologies – PaN cloud masterclasses?

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.



Raw results
Scientists and Data

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

9th October 2020

How many scientific visitors do you receive per year?

Facility	Current situation (2018/2019)	Forecast 2023
DESY	3667 (2019)	> 3667 (we'll have more beamlines by then)
PSI	-	-
Diamond	~6000	Unknown
ISIS	In 2018 ISIS had 2512 scientific visitors.	>2500
SOLEIL	In 2019: 4986 users visits	Forecast for 2023: 4000 on site and 1500 in remote or mail in
ESRF	2018 (the last year of operation): 6548 user visits. Some users are coming several times. The estimation of unique individuals visiting ESRF in 2018 is about 3 000.	~ 10 000
ALBA	2200	3100
HZB	ca. 3000 user visits at BESSY II, these are about 1600 individuals (most people come twice a year since we work in semesters)	maybe a few more, since beamtimes become shorter due to better infrastructure, detectors, sample environment, IT... we also expect "remote users", but it is not clear, how to count them yet. LEAPS etc. are working on definitions and a common metrics
HZDR	100	120
MAXIV	700	1500

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How many experiments are performed per year?

Facility	Current situation (2018/2019)	Forecast 2023
DESY	1572 (2019) incl. # of granted beamtimes + # of internal experiments	> 1572
PSI	-	-
Diamond	1191	Unknown
ISIS	In 2018 ISIS had 1085 proposals from 32 countries and 264 Xpress proposals	More than pre-Covid.
SOLEIL	In 2019: 1315 experiments	Forecast: 1500 experiments
ESRF	2018: 924 experiments	1200 experiments
ALBA	710	1050
HZB	roughly 800 experimental campaigns in the last years	maybe 5% more due to new infrastructures, shorter beamtimes, more remote access; but Corona might also lead to a significant decrease even in the long run (Budget cuts at universities, less money in Horizon Europe, no Trans National Access Funding after 2021, ...)
HZDR	80	100
MAXIV	312	563

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How much data is produced per year (in TB, PB,...)?

Facility	Current situation (2018/2019)	Forecast 2023
DESY	around 2 PB / year (01.2019-12.2019)	
PSI	3PB	
Diamond	2019: 5.35 PB	
ISIS	About ~15 TB of raw data from April 2017 to April 2018.	
SOLEIL	700 TB in 2019	
ESRF	2018: 9 PB	2023: > 50 PB. The error bar on this number is large. It could be much more if our IT infrastructure is capable of dealing with it.
ALBA	We expect a big increase in the quantity of data produced in following years: 2020 < 200TB/year, 2021-950 TB/year, 2022 - 1.7PB/year, 2023-4PB/year	
HZB	The HZB Data Policy is not yet fully implemented and thus we do not have the full picture yet. The planning for the storage capacity were based on a rough estimate of 2 PB data production per year for all HZB facilities.	
HZDR	about 1 PB (2019)	
MAXIV	1 PB	

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What are the mean and maximum data file sizes for an experiment?

Facility	Current situation	Forecast 2023
DESY	Few bytes to few 100 GB. Currently we store 4.5 PB in 450.000.000 files (so about 10 MB per file in average)	Few bytes to few 100 GB
PSI	50TB max, mean 1GB mean dataset size	similar.
Diamond	The file sizes will depend on the beamline and the experimental techniques (especially detector rate).Max file size allowed: 4 TB (currently tapes have a limitation of 5 TB). Tomography hdf5 files can easily reach a few TB size.	Unknown
ISIS	Range from hundreds of KB to ~63 GB currently.	Maximum data file size is expected to be of the order of 100GB.
SOLEIL	Mean data file size = 1 to 300 MB depending on the type of experiment	Mean data file size = 1 to 1GB depending on the type of experiment
ESRF	Data file size is typically from 1MB to 16MB. Yet an experiment consists of many datasets, and a dataset consists of many files. For instance, mean and range for the (few) already publicly available experimental data: Mean: 500GB, Range: [2.4-3902] GB But that can be much more.	We will have more beamlines producing multi-image files in hdf5. Assuming an average of 100 images per file and an average of 4MB per image for an average camera we could say on average 400 MB with a maximum of 100 GB (this is a guess).
ALBA	That's in fact a complex question. It depends on the type of experiment and the beamline. For example in SAXS-WAXS experiment in BL11 the maximum size is 750MB but infact most of the experiments are simpler with just 5MB size. In MX (BL13) the actual data file size is up to 6MB but up to 3600 files could be generated. It is difficult to calculate the mean. In other beamlines the quantity of data generated is lower.	In the new beamlines of Microfocus MX and Tomography that will be operative in 2023 the size of the dataset will be much higher: 1.6GB in MX and 16GB in tomography scans.
HZB	We anticipate a large variation in size, ranging from some kB up to hundreds of GB per dataset.	Same as last question.
HZDR	mean: 100 MB, max. 2 TB	mean 1 GB, max. 10 TB
MAXIV	mean: 6MB, maximum: 500GB	mean: 12 MB, maximum: 2TB

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you have any open data (freely accessible by anyone)?

Facility	Current situation	Forecast 2023
DESY	Centrally managed: no. Some datasets are published at https://www.cxidb.org/	yes
PSI	yes	Yes
Diamond	Not through the DLS data catalogue. Staff members use Zenodo to mint DOIs.	Unknown
ISIS	Yes, see https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx .	Yes, see https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx .
SOLEIL	Not yet	Yes
ESRF	Yes, see https://data.esrf.fr/ 2019: 9 experiments comprised of 100datasets	Yes. 2023: 3000 experiments with 100s of thousands of datasets
ALBA	We do have some experiments were the embargo period has finished and therefore the data is available on request by anyone. That process currently needs manual interactions that could limit in practice the application of the word "freely" to the current situation.	Yes. We have to discuss internally a detailed plan of this implementation and its different stages.
HZB	We don’t have any open raw data yet. The first dataset with expiring embargo will become open in November 2020. We do have two open data publications, though, e.g. datasets having been created for the purpose of publication.	Yes, see last question.
HZDR	yes (rodare.hzdr.de)	yes
MAXIV	No	Yes

The ExPaNDS project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 823852.

How long do you keep data for after an experiment?

Facility	Current situation	Forecast 2023
DESY	Currently not limited	Unlimited for published data. At least 10 yrs otherwise.
PSI	30 days - 10yrs	10yrs
Diamond	Keep data on high parallel file system for 40 days and then accessible by downloading data from tapes. One copy on tape for the lifetime of the tape.	Unknown
ISIS	Forever	Perpetuity
SOLEIL	5 years extendable to 10 years in the future	idem
ESRF	2018: 50 days on disk after the end of the experiment, one year on backup tapes, part of the data for 10 years in a tape archive	2023: 30 days on disk after the end of the experiment, all data for 10 years in a tape archive
ALBA	Currently no experimental data have been deleted in practice from our data archiving. The continuity if this policy is in compromise due to the budget implications with the new beamlines with increased quantity of data generated.	The current commitment according our approved data policy is 5 years. We expect to increase this time but, depending on the technique, a careful analysis or definition of quotas much probably will have to be established.
HZB	According to the HZB Data Policy we keep raw data for at least ten years after the experiment. It is not decided yet what will happen after these ten years, e.g. there are currently no plans to delete any data. It is at least conceivable to keep selected data much longer then this.	Same as last question.
HZDR	10 years and longer on request	10 years and longer on request
MAXIV	Policy: 3 months, in practice indefinitely, funded for long term storage... under construction.	10+ years

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How long is the embargo period for data produced at your facility?

Facility	Current situation	Forecast 2023
DESY	Not defined yet. Aim is a PaNdata compliant policy.	Not defined yet. Aim is a PaNdata compliant policy.
PSI	3yrs	3yrs
Diamond	Unknown	Unknown
ISIS	3 years	3 years (very likely)
SOLEIL	3 (three) years by default with the possibility of up to 5 (five) years	idem
ESRF	3 years, extendable upon request	3 years, extendable upon request
ALBA	Three years	Three years. There is no plan to change this period.
HZB	Five years.	Five years.
HZDR	5 years	5 years
MAXIV	No MAXIV embargo, data is owned by PI.	Likely same as present.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What data formats are used at your facility (nexus, institute-specific, ...)?

Facility	Current situation	Forecast 2023
DESY	Nexus, HDF5, ASCII, EDF, tif, jupyter-nb, json, CBF, custom HDF5	unknown.
PSI	all the above....	all the above.
Diamond	Nexus and HDF5, domain specific formats (.dat, .cbf ...) For Electron microscopes: .mrc	Unknown
ISIS	Nexus, and for those instruments who are not yet using this format they use an ISIS binary .raw format	Nexus, and for those instruments who are not yet using this format they use an ISIS binary .raw format
SOLEIL	NeXus/HDF5	idem
ESRF	2018: EDF, Specfile, HDF5, NeXus+HDF5, other data formats depending on detectors	majority of HDF5, NeXus+HDF5, and other data formats depending on detectors
ALBA	Multiple formats are used: HDF5 but also EDF,CBF or generic formats: JPEG, TXT, DAT generic data ...	No changes are expected respect current formats used.
HZB	institute/manufacturer specific	NeXus where appropriate, or technique specific standards
HZDR	Institute-specific	Institute-specific, OpenPMD
MAXIV	HDF5, Nexus/HDF5 Nxmx Gold, Scienta Omicron data formats (pxt) + others	Same as present

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Does your facility generate DOIs for your experimental data? If yes, when are they generated?

Facility	Current situation	Forecast 2023
DESY	Not yet.	Yes, most likely at date of publication
PSI	Yes when manually publishing data	Yes when manually publishing data
Diamond	No	Unknown
ISIS	Yes currently for all except Xpress experiments. They are generated before cycles start	Yes currently for all except Xpress experiments. They will be generated before cycles start
SOLEIL	Not yet	Yes, by end of 2020
ESRF	Yes. DOIs are generated at the beginning of the acquisition. 2018: Partially deployed on the beamlines	Yes. DOIs are generated at the beginning of the acquisition. 2023: Available for all beamlines
ALBA	Currently no DOIs are generated.	Yes. The exact date milestone is not defined yet.
HZB	For manually curated data meant for a data publication	DOIs for data publications, handles for the rest of measurement data
HZDR	yes, during upload onto the repository rodare.hzdr.de	yes, may be even earlier after data generation and quality check.
MAXIV	No	Yes, or a similar type of persisten data indentifier

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Does your facility have a User Portal (eg to submit proposals)?

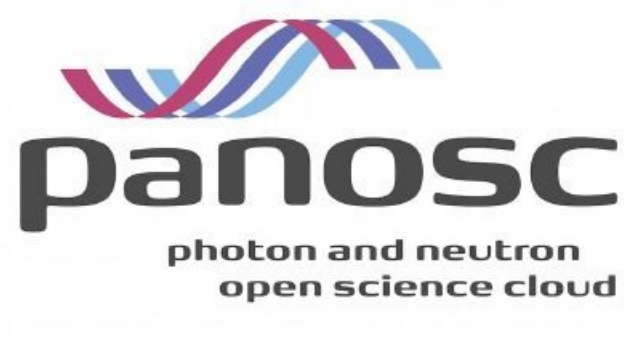
Facility	Current situation	Forecast 2023
DESY	yes	???
PSI	yes: duo.psi.ch	continued support
Diamond	Yes	Unknown
ISIS	Yes	It will continue to exist
SOLEIL	Yes SUNset	to be decided
ESRF	Yes, https://smis.esrf.fr/	100%
ALBA	The ALBA User Office Portal manages all the pre-experiment process (from proposal submission to visit scheduling) and post-experiment (e.g. experimental reports, feedback...)	In 2023, we expect to have 13 beamlines operating. That is 5 more than we currently have. Given the trend of proposals submitted (431 in 2016, 500 in 2017, 603 in 2018, 715 in 2019 and 713 in 2020), I would say the number or proposals to be received would be around 1150-1200 (considering 90 proposals per Beamline as the maximum capacity).
HZB	Yes (GATE)	GATE
HZDR	yes (gate.hzdr.de)	Will cover more HZDR facilities (at the moment three) in our proposal submission system and combine the portal with further data management workflows
MAXIV	Yes	Move to a new DUO

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How would a user download all of their data at your facility?

Facility	Current situation	Forecast 2023
DESY	Via the data catalogue: https://gamma-portal.desy.de	Via a (sc)icat data catalogue (or something interoperable)
PSI	currently via rsync or GlobusOnline from PSI disk	via http from archive service held at 3rd party site.
Diamond	Within 40 days after the visit: Globus, ftp, sftp, rsync 40 days after the visit: use the DLS data catalogue with the following available options: https, Globus, STFC cluster (for reprocessing), DLS cluster (for reprocessing)	Unknown
ISIS	Through the ISIS Data catalogue.	Through the ISIS Data catalogue or ISIS Data Analysis as a Service.
SOLEIL	Specific SOLEIL tool (SDR), soon SFTP	To be defined
ESRF	2018: - Via the Internet at https://data.esrf.fr/ or through ICAT API for ~10% of the data - Portable media (disks) for ~90% of the data	globes-online
ALBA	Locally using a external disk through USB3 connection. Remotely via a SFTP connection.	Locally using external disk through USB3.2/Thunderbolt4. Remotely via SFTP (also exploring Globus or Aspera) or via Cloud Services.
HZB	For each institute/instrument different, no central solution	From the ICAT repository
HZDR	Using rodare.hzdr.de or direct via ssh access	as now and using better toolkits (such as Globus)
MAXIV	Globus Connect, sftp, IRODs	sftp, Amazon S3

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.



Raw results
Data Analysis and Reduction

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you provide Jupyter Notebook/JupyterLab services to visitors?

Facility	Response
DESY	yes
PSI	yes
Diamond	Yes
ISIS	No, not directly. But we provide machine with Python on where users could setup Jupyter Notebooks.
SOLEIL	Experimental service. We provide Python, R, Julia, Matlab, Octave, Java, JavaScript, C and Fortran kernels. Users are identified with the site LDAP. Access to experimental data is being set-up.
ESRF	Yes: some beamlines have some jupyterhub servers running and jupyterhub+slurm is available at https://jupyter-slurm.esrf.fr.
ALBA	No. Jupyter shared services are currently under installation. Up to now these services have been published using Citrix application.
HZB	No
HZDR	yes
MAXIV	Yes

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How many staff and visiting scientists/end-users currently use Jupyter Notebook/JupyterLab at your facility?

Facility	Response
DESY	Up to 100 different/concurrent users per day, roughly 300-400 per month
PSI	unknown
Diamond	Between 8 - 12
ISIS	This is not recorded
SOLEIL	Some beamlines (typically 15% of 29 SOLEIL beamlines) have set-up such local Jupyter notebooks on a per-request basis.
ESRF	https://jupyter-slurm.esrf.fr: As of today, 3 days after ESRF User Mode restart: 43 unique users in the last 3 months. This does not account for beamline jupyterhub server users.
ALBA	We do not have a register about the use of Jupyter Notebook by users.
HZB	< 10?
HZDR	many users using the HPC analysis tools
MAXIV	25% staff scientists, minimum visiting scientists/end-users (service enabled for them recently), some visiting scientists/end-users may be using it during the beamtime via beamline generic users

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do external users currently have remote access to a data analysis/reduction services (remote desktops, grid computing, shell...)?

Facility	Response
DESY	yes
PSI	yes, NX
Diamond	Yes, via NX or SSH. Not clear for how long after the visit this is retained.
ISIS	Only for those instruments that are using the new ISIS Data Analysis as a Service platform.
SOLEIL	A NX/NoMachine service allows to access internal data treatment computers on the beamlines, for few users. But there is no rationale on this topic yet.
ESRF	2020: Jupyter server allowing to access the SLURM cluster: https://jupyter-slurm.esrf.fr; Remote desktop provided through NoMachine: https://nx.esrf.fr ; else connection using ssh. 2023: Specific portal?
ALBA	In MIRAS and MISTRAL we provide remote data analysis service of OPUS&Unescrambler and Thermo-Fisher Amira respectively. Both based in Virtual Machines. Occasionally some HPC resources have been provided on request also.
HZB	Not centralised
HZDR	no
MAXIV	Yes

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you offer data analysis/reduction training at your facility?

Facility	Response
DESY	To some extend in dedicated, instrument specific workshop
PSI	no
Diamond	The training provided will depend on the user communities.
ISIS	Yes, through annual Neutron training school and Muon training school. Plus e-learning material, documentation of software used on beamlines and through training by instrument scientists
SOLEIL	Not yet regularly, only punctually during some meetings (as annual Users' Meeting) or workshops.
ESRF	2018: We provide some Python trainings oriented towards data analysis (Training material: https://github.com/silx-kit/silx-training) 2023: Training material should be made available and possibly trainings should be complemented by tutoring services
ALBA	The Experiments Division has offered specific training to some users in the past.
HZB	Not centralised
HZDR	on individual request, not regularily
MAXIV	Yes

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do your scientists support users with their data analysis/reduction after their experiments?

Facility	Response
DESY	yes
PSI	yes
Diamond	Unknown
ISIS	In some cases if required
SOLEIL	Yes, on many of beamlines. Many data sets are too large or too complex to be handled at user home laboratories.
ESRF	2018: Yes, but not officially. This is done as an additional informal service by many of the local contacts on the beamlines. 2023: A clear mandate for DaaS should be in place and define the services provided by the ESRF
ALBA	Yes. We do support our users on a best effort basis but no systematic support or training is currently in place.
HZB	Not centralised
HZDR	yes
MAXIV	limited support

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Are data analysis/reduction results preserved at your facility?

Facility	Response
DESY	Yes if generated on site. Back-transfer of results from elsewhere is not very common.
PSI	up to researchers discretion.
Diamond	Yes (processed during the experiment and up to 40 days after the experiment).
ISIS	Not in perpetuity currently.
SOLEIL	Raw data and associated metadata are preserved at SOLEIL, currently on a central storage system. Unless otherwise specified, reduced data and the associated metadata are not curated for long-term by SOLEIL. Results issued from analyses performed on raw data and metadata using SOLEIL means are stored long-term by SOLEIL in case of in-house research, and only on a best effort basis in other cases.
ESRF	Yes for in-house experiments, because these storage areas are still covered by a backup. And no for external users, this is currently not part of the data policy 2023: It may be necessary to archive also processed data, if economically feasible
ALBA	Yes. Currently there is no differentiation in our data policy differentiating raw data, processed data or curated data.
HZB	Not centralised
HZDR	yes
MAXIV	

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How long do you keep the analysis results?

Facility	Response
DESY	At least as long as the data
PSI	connected to lifetime of raw data
Diamond	For processed during the experiment and up to 40 days after the experiment, we keep one single copy on tape for the lifetime of the tape.
ISIS	A policy for how long to keep analysed results created through the ISIS Data Analysis as a Service platform is currently being developed
SOLEIL	No general rule, see 6.
ESRF	One year for in-house experiments
ALBA	There is no differentiation in our data policy differentiating raw data, processed data or curated data. All data is preserved a minimum of 5 years and in practice no data has been deleted up to now.
HZB	Up to scientists
HZDR	about 6 month
MAXIV	Similar as for the stored data

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How much data is produced after data reduction / analysis?

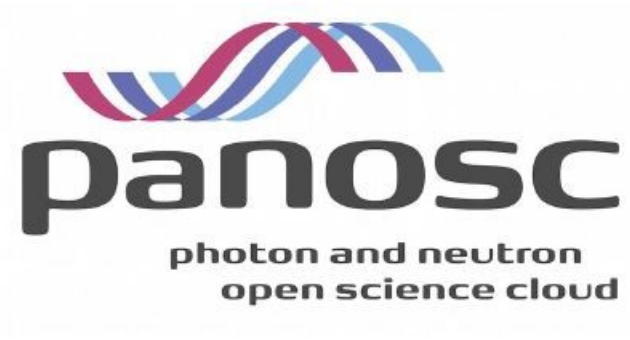
Facility	Response
DESY	No reliable estimate available
PSI	variable and experiment dependant
Diamond	Depend on the experimental technique.
ISIS	It varies a lot from instrument to instrument, but to a first approximation, about the same as the raw data volume: O(100KB) to O(100GB)
SOLEIL	Depending on the experiment. For MX and tomography, the processed data sets are larger than the raw data.
ESRF	N/A, this highly depends on the kind of experiment/analysis.
ALBA	That depends strongly on the technique. For current MX beamlines the data is doubles from the original RAW data generated. In the future projected MX beamlines this factor could be increased to 4. And in the future tomography a factor up to 6 is considered realistic.
HZB	?
HZDR	this strongly depends on the use case, TB up to PB, often similiar amount as before
MAXIV	This varies highly depending on the experiment type, in principle it can be comparable with size of raw data.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Can external users remotely access experimental logs after the experiment?

Facility	Response
DESY	There is currently no central e-log, but varies with instruments.
PSI	Yes in some cases and if still available
Diamond	Some beamlines have logs registered through ISPyB.
ISIS	The experimental logs are included in Nexus files (so yes).
SOLEIL	Not currently. Few users can access their data remotely, as well as attached logs via NX/NoMachine.
ESRF	An electronic logbook is being implemented for all beamlines as part of the data policy.
ALBA	The use of experimental logs depends strongly on the beamline. No general policy about the external access from users after the experiment has been stablished.
HZB	Not centralised
HZDR	only if stored in rodare.hzdr.de (which is frequently done)
MAXIV	Not currently although may be enabled in future.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.



Raw results Technology

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

9th October 2020

What operating systems for data analysis does your facility support (please specify any Linux distributions)?

Facility	Response
DESY	Debian, Centos, Ubuntu (typically the latest two releases)
PSI	RedHat 7, Windows 10
Diamond	RHEL 6 and RHEL7
ISIS	ISIS's support includes: Windows, Redhat, CentOS 7 and Scientific Linux
SOLEIL	CentOS Linux release 7.3.
ESRF	Debian 8, 9, Ubuntu 20.04
ALBA	Linux (CentOS , Debian) and Windows.
HZB	Not centralised
HZDR	Linux Ubuntu
MAXIV	Windows10, CentOS, Ubuntu, MacOs

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What is the main Linux distribution used for the server infrastructure at your facility?

Facility	Response
DESY	Centos 7
PSI	RedHat 7
Diamond	RHEL 6 and RHEL7
ISIS	Redhat and varieties
SOLEIL	CentOS Linux release 7.3.
ESRF	Debian and Ubuntu
ALBA	CentOS 7, Suse 11/12 and Debian 9
HZB	Debian, CentOS. The storage servers and the servers for the ICAT data catalogue use RHEL.
HZDR	Linux Ubuntu, Debian Linux
MAXIV	CentOS

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What compute infrastructure is dedicated to scientific computation at your facility?

Facility	Response
DESY	HPC platform
PSI	classic CPU and GPU cluster with batching system
Diamond	Kubernetes cluster for DASK. Univa Grid Engine clusters for HPC processing.
ISIS	SCARF and more recently the SCD Cloud as well as dedicated individual machines
SOLEIL	NEC cluster
ESRF	Compute clusters & dedicated machines for online data analysis
ALBA	Beamline Workstations, also VM's based workstations and HPC Cluster (Slurm). Also Adlink's Industrial PCs are used for certain computation during acquisition time.
HZB	HPC cluster
HZDR	Linux Cluster with 30% CPU capacity and 70% GPU capacity
MAXIV	Currently around 100 multicore servers in different systems

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you have an HPC cluster? If yes, what is the size of the cluster? I.e. number of cores, memory etc.

Facility	Response
DESY	yes. 450 nodes, typically 256-512 GB/node, but not dedicated
PSI	5000 cores
Diamond	Yes. 224 GPUs, ~2500 CPU Cores. Large GPFS of around 10PB.
ISIS	SCARF, see https://www.scarf.rl.ac.uk/hardware.
SOLEIL	1 Interactive partition with 64 cores INTERL XEON E7 and 4 TBytes RAM
ESRF	Yes: 2018: 3200 Intel cores and 28 GPUs 2020: 16 Power 9 computers with 2 GPUs each 2023: a substantial upgrade is required, ideally to 20k Intel (or AMD) cores and 150 GPUs
ALBA	Yes. 24 cores (18 nodes) with 4224 GB of memory using 25 GbE ethernet adapters
HZB	Yes
HZDR	Yes, 271 compute nodes, 60 of them attached with GPUs, with a theoretical computing power of approx. 2,85 PFlop/s
MAXIV	2000 cores, 128-376 GB RAM per node

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How is your scientific software distributed? (CERN-VMFS, custom Linux repositories, release downloads, artefact repositories...)

Facility	Response
DESY	packages, shared file installs, cvmfs, (spack)
PSI	module system.
Diamond	Resides on NFS as compiled binaries/libraries or Python modules. Some limited engagement with Singularity containers.
ISIS	Some as cross platform binary distributions such as Mantid https://www.mantidproject.org/Main_Page). Some through custom Linux repositories. Some through dockerhub, some through release downloads. Some through CERN-VMFS, but that is being reduced actively.
SOLEIL	Instllation of packages on the cluster directly
ESRF	As Debian/Ubuntu packages and Python packages in pypi and in conda (when other persons are packaging them)
ALBA	No major scientific software distribution is done. Some generic software is distributed at GitHub.
HZB	Not centralised
HZDR	custom repositories, gitlab.hzdr.de, github
MAXIV	NFS & GPFS, sw deployed by EasyBuild

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What emerging tools are you currently looking into or are interested in using?

Facility	Response
DESY	For software distribution? Spack (possibly on cvmfs)
PSI	k8s
Diamond	Singularity. Docker. Kubernetes. Knative. Kubevirt. HTCondor.
ISIS	Use of the cloud services for users.
SOLEIL	Cloud bases solution such as Kubernetes, OpenStack
ESRF	singularity
ALBA	JupyterHub, Containers running in HPC, Kubernetes orchestration, RDMA over Converged Ethernet...
HZB	- BlueSky (machine group!) - Data transfer: Globus - Message queues - Container orchestration: Kubernetes - Replace Docker by podman - ID management: Keycloak (see below)
HZDR	high end compression tools, AI based analysis
MAXIV	Jupyter, FPGAs

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you currently provide any remote desktop services to users internally at the institute and to users after their experiments?

Facility	Response
DESY	Yes, full graphical remote desktop access to HPC
PSI	yes NX
Diamond	Yes, NX
ISIS	ISIS Data Analysis as a Service
SOLEIL	NO remote access is provided after the experiment
ESRF	NoMachine: https://nx.esrf.fr
ALBA	During experiments we provide NX NoMachine remote Desktop. After the experiment for two beamlines (MISTRAL & MIRAS) we provide Virtual desktop for Data Analysis based in VMs.
HZB	Only to few, specific users. Not centralised.
HZDR	on request, not very often
MAXIV	Yes, both

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Are you currently running any Data Analysis as a Service pilot projects (such as VISA/CalipsoPlus at the ILL)?

Facility	Response
DESY	Yes, but it’s currently not in operation
PSI	yes but not in production
Diamond	Jupyterhub.
ISIS	ISIS Data Analysis as a Service
SOLEIL	NOT in production
ESRF	Calipsoplus and Jupyter
ALBA	The before mentioned based in VMs and CalipsoPlus pilot.
HZB	No
HZDR	no
MAXIV	No

The ExPaNDS project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you have a cloud infrastructure in place (OpenStack, Nebula, Orchestration etc.)?

Facility	Response
DESY	Openstack, k8s, rancher & co.
PSI	no
Diamond	PaaS layer in the shape of Kubernetes. No IaaS layer.
ISIS	SCD cloud uses OpenStack, SCARF uses Slurm
SOLEIL	NO
ESRF	OpenStack and KVM
ALBA	Yes. For Virtual Desktops/application publishing (Citrix based) and a Kubernetes orchestration pilot project under execution.
HZB	No
HZDR	yes, Open Stack
MAXIV	No

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What are you using for machine virtualisation (VMWare, KVM…) ?

Facility	Response
DESY	Vmware, xen, ...
PSI	VMWare
Diamond	Small VMWare for testing and infrastructure support. Everything else is bare metal.
ISIS	not sure, likely a mix
SOLEIL	VMWare AND KVM depending on applications
ESRF	N/A
ALBA	VMware and Citrix Hypervisor.
HZB	Not centralised
HZDR	VMware, OpenStack
MAXIV	VMWare

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What are you using for containerisation (Docker, Singularity…)?

Facility	Response
DESY	Docker, singularity, (podman)
PSI	Docker
Diamond	Singularity for HPC. Docker for Kubernetes. Podman on all desktopsto allow rootless containers.
ISIS	Docker and Singularity
SOLEIL	Few usage of Docker
ESRF	Docker
ALBA	Container managed by Rancher Orquestration platform.
HZB	Docker, others are looking into Singularity
HZDR	Docker
MAXIV	Docker and Singularity, Kubernetes

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Which protocol(s) are you using to access experimental data (NFS, SMBFS, SFTP…)?

Facility	Response
DESY	Mostly gpfs. Remote ftps, gridftp. On local windows desktops smb.
PSI	NFS, GPFS, SFTP, SMBFS,
Diamond	GPFS for HPC access. The GPFS is exported as NFS and CIFS to certain infrastructure, e.g. Kubernetes or Windows.
ISIS	not sure
SOLEIL	NFS for internal access. Soon an sftp access will be open
ESRF	NFS, SMBFS, GPFS
ALBA	NFS v4, CIFS. Remotely SFTP.
HZB	External access only through OpenCloud (https) or SSH-tunnels. Internal access via NFS and CIFS.
HZDR	NFS, SMBFS, SFTP
MAXIV	Your answer (NFS, GPFS, SMB, SFTP, Globus Connect) - Artur may reply better

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you provide functions-as-a-service?

Facility	Response
DESY	Yes, but not much used by users (yet)
PSI	no
Diamond	Not yet, but Knative on Kubernetes is planned.
ISIS	don't think so
SOLEIL	NO
ESRF	No
ALBA	No.
HZB	No
HZDR	no
MAXIV	No

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What is the incoming and outgoing bandwidth at your facility?

Facility	Response
DESY	2*100Gb
PSI	100GB and general internet 10GB
Diamond	Currently restricted to 10Gb/s to STFC which connects to JANET and the Internet. This will increase in the next few months to 40-80Gb/s
ISIS	not sure anyone knows
SOLEIL	10 Gb/s link to French Research network backbone
ESRF	2018: A shared 10Gbps Internet connection (to the metropolitan network TIGRE, then REANTER, then GEANT) 2023: 100Gbps
ALBA	Currently 2Gbps. It will be increased to 10 Gbps in February 2021. It could be increased without major cost up to 50Gbps from February 2021 in a straight way. All bandwith applied bidirectionally.
HZB	1Gbps
HZDR	2 x 3 GBit
MAXIV	2x10 Gbs, 4x40 Gbs to local university compute centre

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you use cloud providers for hosting external services or is everything hosted internally?

Facility	Response
DESY	internally
PSI	Yes academic cloud providers and some commercial for business applications.
Diamond	GCP used for Netapp replication. Some small GCP projects in use for ML.
ISIS	Currently internally but plans to expand to use external services also
SOLEIL	External service are used for standard IT services (Microsoft and soon emails)
ESRF	2018: everything is hosted on-site 2023: scale-out with commercial cloud providers
ALBA	AWS through Géant Cloud Framework (ends 2020): Providing Backup & archive repositories (S3+Glacier/DeepArchive)
HZB	No
HZDR	At the moment internally hosted, currently build up: www.hifis.net, a common Helmholtz platform
MAXIV	

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

If you were to give users access to remote virtual machines, would you give them root access?

Facility	Response
DESY	no.
PSI	yes
Diamond	No, we have no intention of running a IaaS cloud with virtualisationsupport. Containers orchestration through Kubernetes only.
ISIS	In general no, but exemptions can be made for very specific purposes to trusted users/staff, such as for example software testing/development or specific critical facility software.
SOLEIL	NO
ESRF	No
ALBA	No.
HZB	No
HZDR	no
MAXIV	Probably not

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How would you govern the computational resources required (i.e. CPUS, memory) for a given experiment?

Facility	Response
DESY	Batch scheduler, if requested dedicated reservations
PSI	TBD
Diamond	HPC resource allocation by beamline.
ISIS	Depends on the complexity of the instrument and the experiments. But vaguely: the type of experiment and spectrometer that is used is a rough guide.
SOLEIL	Through a bacth scheduling system such as SLURM
ESRF	2018: free access first come first served 2023: data intensive experiments ✉ Data Management Plan and accordingly resource allocation
ALBA	Usually resources are fixed and preadjusted, and available for users since experiment starts. From this definition we use SLURM to manage computational resources requests.
HZB	No centralised regulation
HZDR	According to queue priorities, realtime requirements first
MAXIV	We do not have accounting, although our partners for offline analysis do.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Can users choose the amount of resources required for data analysis at the facility?

Facility	Response
DESY	no. they can make wishes, but resources are limited
PSI	TBD
Diamond	HPC has allocation by beamline.
ISIS	Yes, but only a limited number of choices are given
SOLEIL	NO
ESRF	See previous question
ALBA	During the experiment the users have specific users with access to needed resources at workstations and HPC queues.After the experiment, when HPC service is requested, the user can control the number of nodes and RAM available in the assigned queue. Only the number of nodes are usually adjusted.
HZB	No centralised regulation
HZDR	yes
MAXIV	Yes

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you have any particular quotas in place related to computation processing (i.e. number of hours, cores, …)?

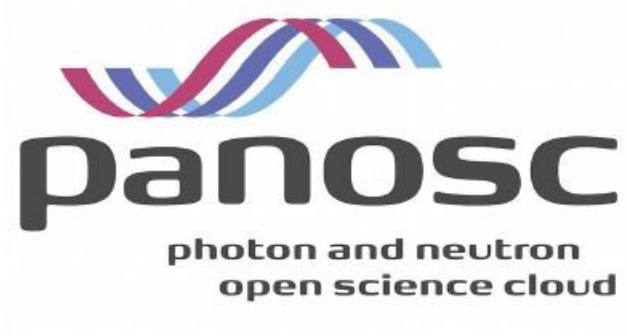
Facility	Response
DESY	slurm scheduler limits more the number of concurrent usage.
PSI	not as yet
Diamond	Yes, wallclock limit of 168 hours (8 days), CPU*seconds limit of 1512 hours.
ISIS	No
SOLEIL	NO
ESRF	2018: no, free access, but monitoring and complaints management. 2023: resource allocation methods have to be found which allow us to regulate the usage and to make sure everybody can work satisfactorily
ALBA	Not currently. The discussion about the need of quotas is starting at the facility.
HZB	Yes, number of hours, number of cores are restricted @HPC
HZDR	yes
MAXIV	No

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you have a job submission system for scientific computing resources (Sungrid, Torque, Slurm, Oar...)?

Facility	Response
DESY	Slurm (and htcondor)
PSI	slurm
Diamond	Yes, Univa Grid Engine (looking at HTCondor)
ISIS	SCARF uses Slurm
SOLEIL	SLURM is in palce
ESRF	SLURM (new one) and OAR (legacy one)
ALBA	Kubernetes based Data Analysis: worker jobs submitted on worker servers. Projected to be launched also via Slurm on HPC cluster
HZB	Yes (details follow)
HZDR	yes, Slurm
MAXIV	SLURM

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.



Raw results Security

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Which authentication provider(s) is used (i.e. umbrella, keycloak etc.) for user access (internal/external) ?

Facility	Response
DESY	Ssh, umbrella (only user office), keycloak in preparation
PSI	umbrella, active directory
Diamond	CAS for most web infra. This connects to AD. We run a keycloak instance that also uses AD for anything OIDC related. SSSD and VAS are used to provide client side auth to AD for Linux.
ISIS	Umbrella is supported
SOLEIL	OpenLDAP, Umbrella
ESRF	Umbrella and Keycloak
ALBA	Umbrella.
HZB	None for the moment.
HZDR	Umbrella, Helmholtz-AAI
MAXIV	Umbrella (currently only for user office), otherwise internal.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Which authentication protocols do you use to authenticate your users? I.e. OpenConnect, CAS, SAML etc.

Facility	Response
DESY	ssh, krb5
PSI	kerberos
Diamond	CAS. OIDC
ISIS	don't know
SOLEIL	SAML
ESRF	SAML
ALBA	CAS, OpenLDAP, and Active Directory.
HZB	Currently we use just LDAP. For the future, changes are under consideration, including support for OpenID Connect.
HZDR	SAML, Moonshot
MAXIV	CAS

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Can people who have never visited your facility access data or analysis services (ie to access open data)? If so how do they authenticate?

Facility	Response
DESY	They have to be registered users with an account. An actual visit is not required.
PSI	open access data, winch is open access.
Diamond	They need to request a login ID (Active Directory). This is managedby the User Office
ISIS	Open data is accessible for everybody. For non-open data access is only given to users who rESRFster with the user office and only for experiment they have participated in.
SOLEIL	Not implemented yet
ESRF	Open data is publicly accessible, data analysis services (Jupyter and remote desktop) are accessible after authentication which requires a user account.
ALBA	Yes, they can. But some manual actions are needed including the creation of specific local accounts.
HZB	Access to open data: yes, no authentication. Access to analysis services: no.
HZDR	Yes, if those are connected to umbrella or Helmholtz-AAI
MAXIV	They cannot currently

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you provide a means for people to apply for access to open data and analysis services?

Facility	Response
DESY	Not yet.
PSI	no
Diamond	Not yet.
ISIS	Open data is accessible for everybody. Analysis services are available to anyone rESRFstered with our User Office.
SOLEIL	Not implemented yet
ESRF	https://data.esrf.fr allows access to open data, but there is no plan to give access to analysis services publicly.
ALBA	We do not provide open data nor analysis services as of today.
HZB	Open data is accessible right away, no application needed. Analysis services: no.
HZDR	yes, as part of rodare.hzdr.de
MAXIV	Not currently

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

What IT services are provided to external users (ssh, sftp, specific web applications)?

Facility	Response
DESY	Ssh, ftps, https-access ...
PSI	ssh, sftp, web applications,
Diamond	Ssh, Nx, VPN
ISIS	For example the ISIS data catalogue and ISIS Data Analysis as a Service and SCARF.
SOLEIL	Specific Web app (SUNset), SFTP, NX
ESRF	2020: JupyterHub+SLURM, NoMachine, ssh 2023: JupyterHub + Remote Desktop
ALBA	SFTP, and specific data analysis services in two beamlines (MIRAS & MISTRAL) based in Remote Virtual Desktops with access to proposal data.
HZB	All mentioned
HZDR	Ssh, sftp, specific web applications
MAXIV	sftp, Globus Connect, VPN, ssh, ThinLinc HPC remote desktop, JupyterHub, MAX IV Wiki, HDF5view, Digital User Office (DUO), FragMAX (MX fragment screening) portal

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do you have any security restrictions at your facility (i.e. is a VPN connection required before accessing services)?

Facility	Response
DESY	no.
PSI	VPN needed for some services.
Diamond	Yes, services such as Jupyterhub currently require VPN access although this may change. Intranet is via VPN only.
ISIS	Authentication with the user office account.
SOLEIL	No
ESRF	A moving target. IT security has to be enforced. VPN access will most likely disappear.
ALBA	Some of the services are after linux firewall/proxies or Citrix Application Delivery Controller, so that user services can be accessed just using Web/cloud interfaces or remote clients.
HZB	Not in general
HZDR	Depending on service, usually not
MAXIV	A VPN is required to access all internal resources

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

How are users associated to their experimental data? (i.e. LDAP groups, POSIX attributes…)

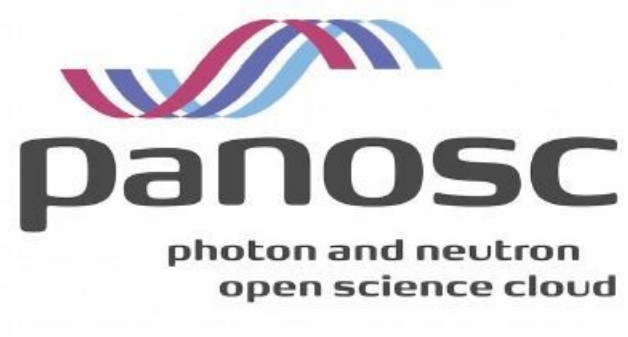
Facility	Response
DESY	ldap groups, nfs4 acls, ...
PSI	LDAP groups
Diamond	Database and LDAP groups
ISIS	LDAP groups
SOLEIL	LDAP groups
ESRF	ACLs
ALBA	LDAP accounts and groups.
HZB	LDAP (AD) groups according to proposal
HZDR	via LDAP groups
MAXIV	Microsoft AD

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Do scientific visitors have a user home at your facility (linux home, windows home…)?

Facility	Response
DESY	Yes, linux homes on HPC and in afs (outside HPC)
PSI	yes on linux
Diamond	Yes Linux home on NFS
ISIS	Those who use the ISIS Data Analysis as a Service have
SOLEIL	No
ESRF	2019: No 2023: Yes
ALBA	The scientific data proposal folders are configured as proposal user homes and automatically mounted for Linux (NFS4) and Windows (CIFS).
HZB	Currently: no even not necessarily an account. Next future: accounts for all, home if necessary
HZDR	Yes, temporarily
MAXIV	No, only as part of a visit.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.



Raw results
Other information

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

Please add any other comments you feel relevant for the survey.

Facility	Response
DESY	-
PSI	-
Diamond	-
ISIS	-
SOLEIL	-
ESRF	Data analysis and data analysis software developed/maintained at ESRF owes a lot to open source software, including: • The Python “scientific stack”: numpy, scipy, matplotlib, scikit-*, ... • HDF5 library, h5py, h5glance • Qt/PyQt • Jupyter/IPython ecosystem • Many other python packages and other softwares (e.g., ImageJ)
ALBA	-
HZB	-
HZDR	-
MAXIV	-

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.