EUROPEAN
SPALLATION
SOURCE

Data Management and Software Centre

Common STAP – October 2024

PRESENTED BY TORBEN R NIELSEN (ON BEHALF OF THOMAS H. ROD)

2024-10-22

Agenda

1 DMSC

2 Updates form teams at DMCS

3 Integration testing

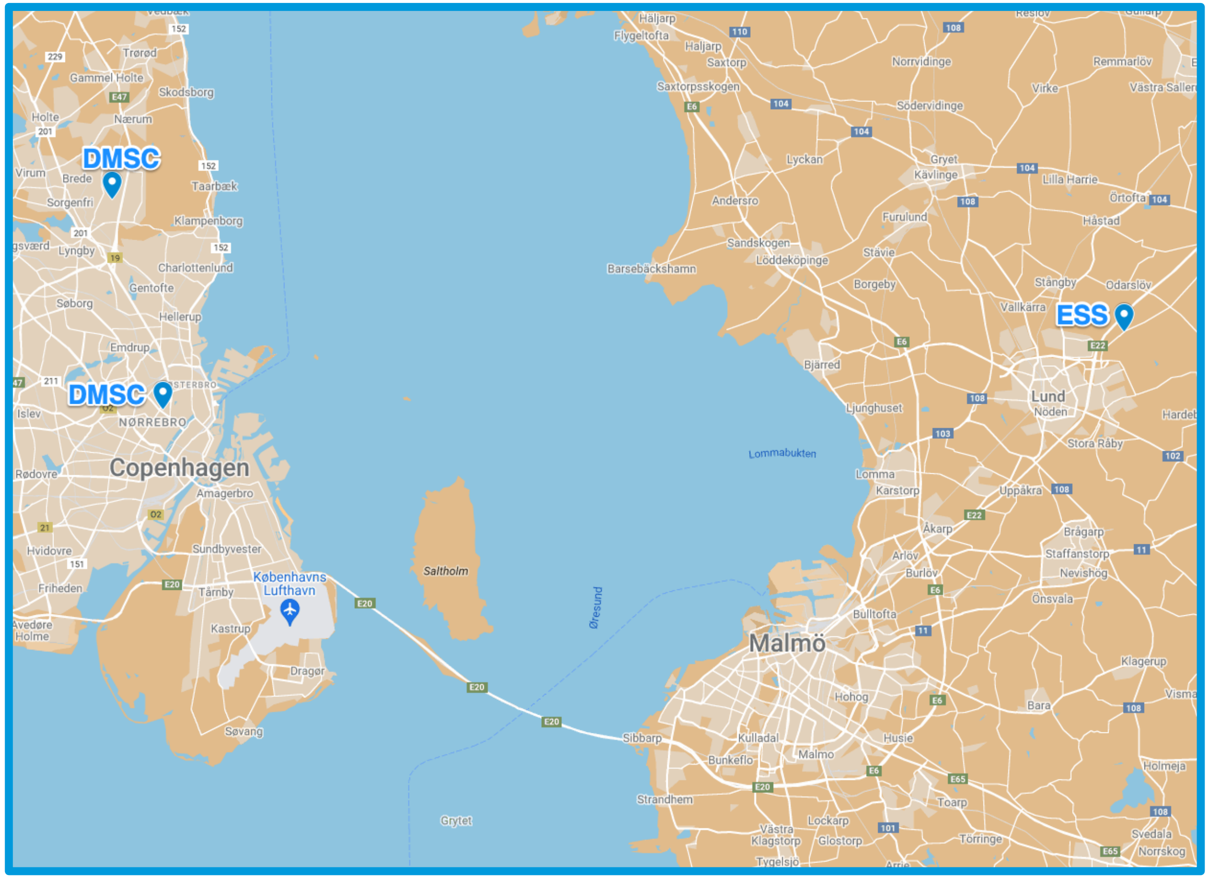
4 DMSC Summer School

5 Summary

DMSC

Data Management and Software Centre (DMSC)

New location: moved from Copenhagen (COBIS) to Lyngby (DTU)

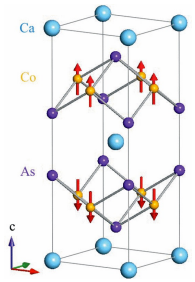


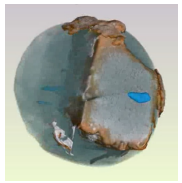
DMSC - Copenhagen area

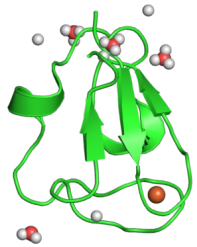
ESS - Lund

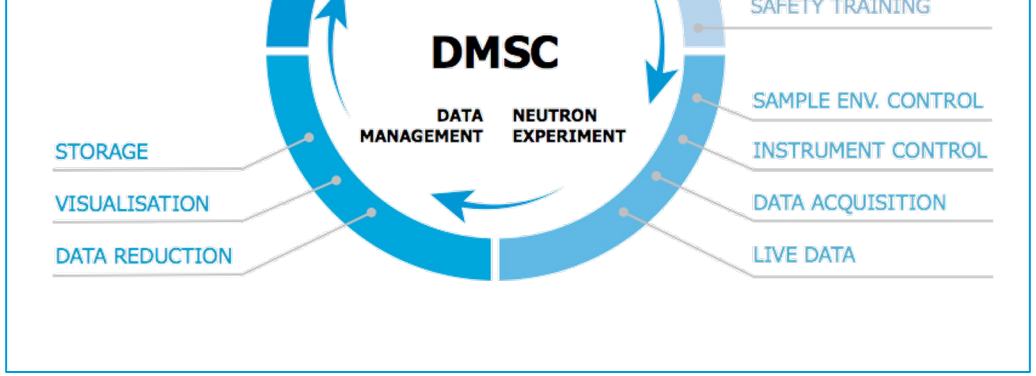
DMSC & Scientific computing

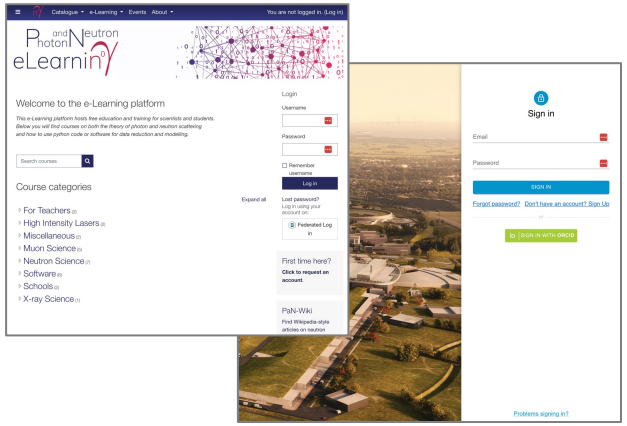
Support **user journey** from proposal to publication with scientific computing tools & services

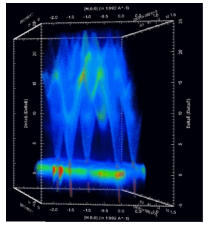


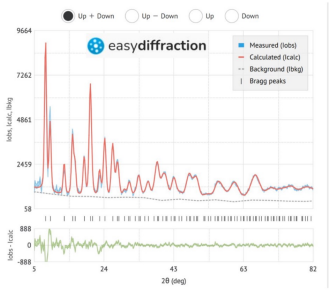


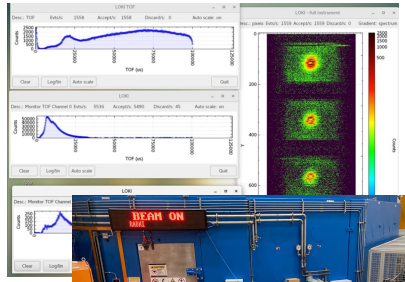


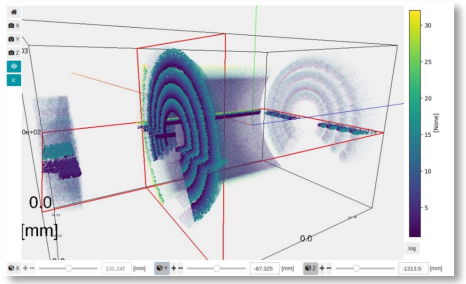


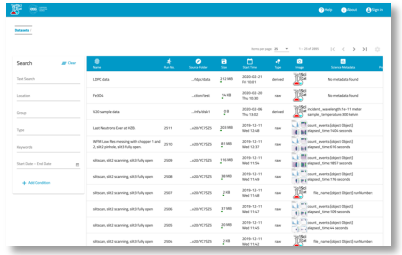


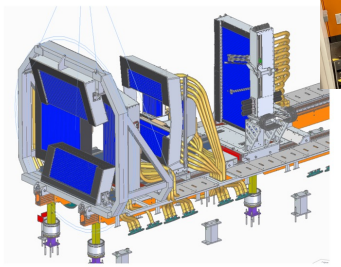






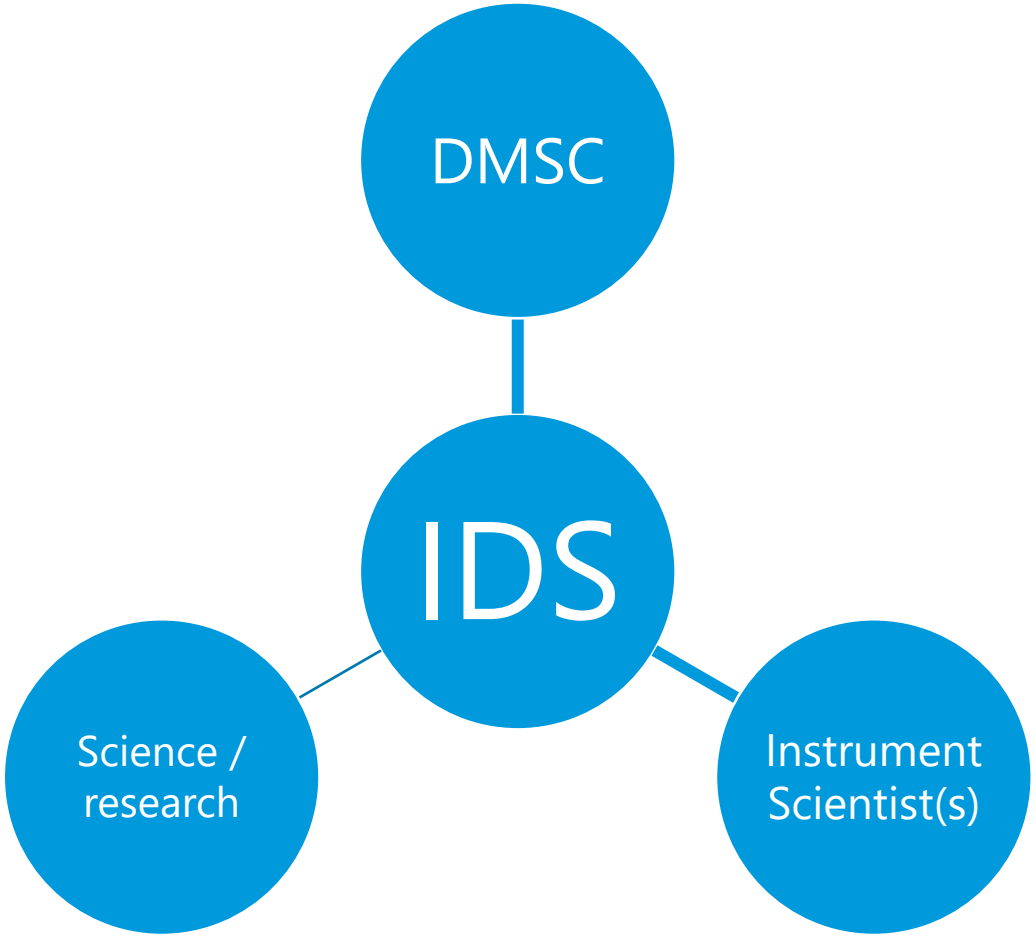


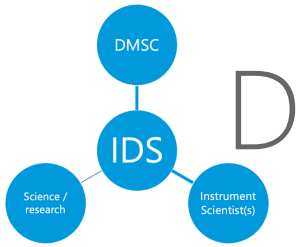




Instrument Data Scientists (IDS)

DMSC interface to instrument teams

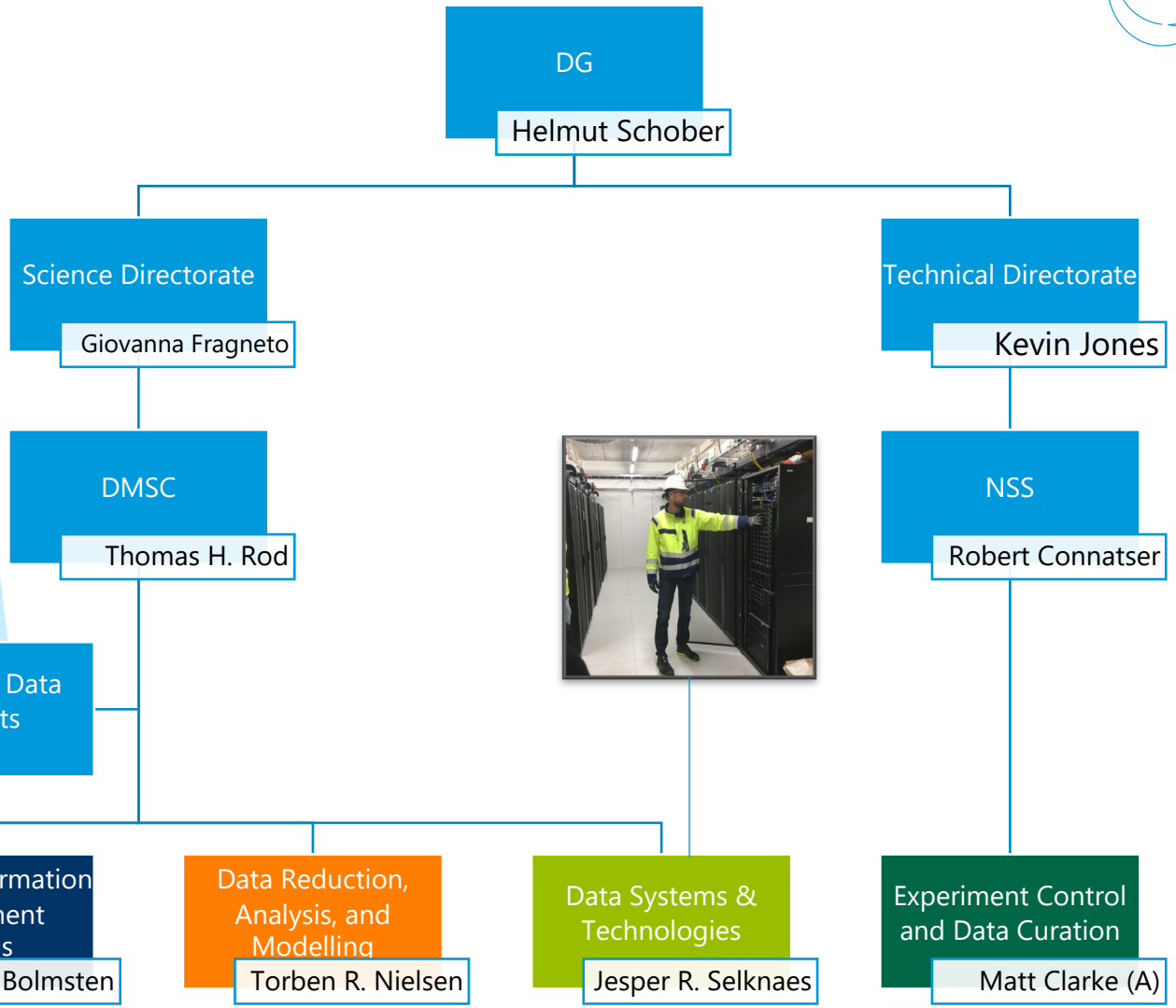




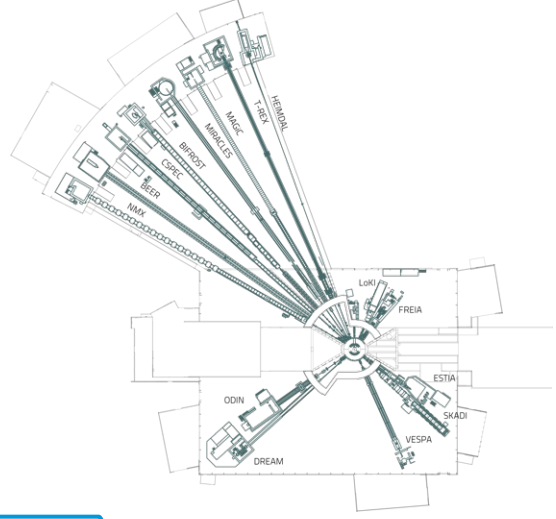
DMSC organisation

Instrument Data Scientists

LOKI (& SKADI)	Oliver Hammond**
DREAM (& MAGIC)	Céline Durniak
ODIN (& BEER)	Søren Schmidt
BIFROST (& CSPEC)	Gregory Tucker
ESTIA (& FREIA)	Nicolò Paracini**
NMX	Aaron Finke



Integrated data pipeline



IDS: Interface to instrument teams
Customization to instruments
Hands on support

UOS | EPICS | NICOS | kafka | scipp | Technique specific | SciCat | NeXus

User office software → Experiment control → Stream events & meta data → Data reduction & visualization → Data analysis → FAIR Data management

DST:

VISA | jupyter

Remote Access

Updates from DST

Infrastructure: HCP, GPU, storage, VISA, etc.

Data Systems and Technology

DST - Status report

What has been accomplished:

- New staff
  - Filled almost all vacant positions (7+1 FTEs) + 3 student worker
- **Considerable effort has gone into the future hosting situation**
- **Installed workstation for experimental control in the BIFROST hutch**
- Installation of additional virtualization equipment in both datacenters
- **Continues to work with VISA and software packages for reduction and analysis**
- Integration with SciCat and UserOffice well underway with good results so far
- Engaged in the post-PaNOSC VISA ecosystem

Terminal equipment in hutches

Status and plans

Status:

- ❑ Control systems workstations has been procured for tranche 1 instruments.
- ❑ Deployment and provisioning setup for tranche 1 instruments
- ❑ Physical workstation deployed in the BIFROST hutch and commissioning on going.
- ❑ Remaining T1 instruments are awaiting electricity and network to be available. LOKI, DREAM and TBL is upcoming.

Plans:

- ❑ It has been decided to move the scope to the IT division to provide local support since DST I physical located in Copenhagen.

DMSC server room

Status and plans

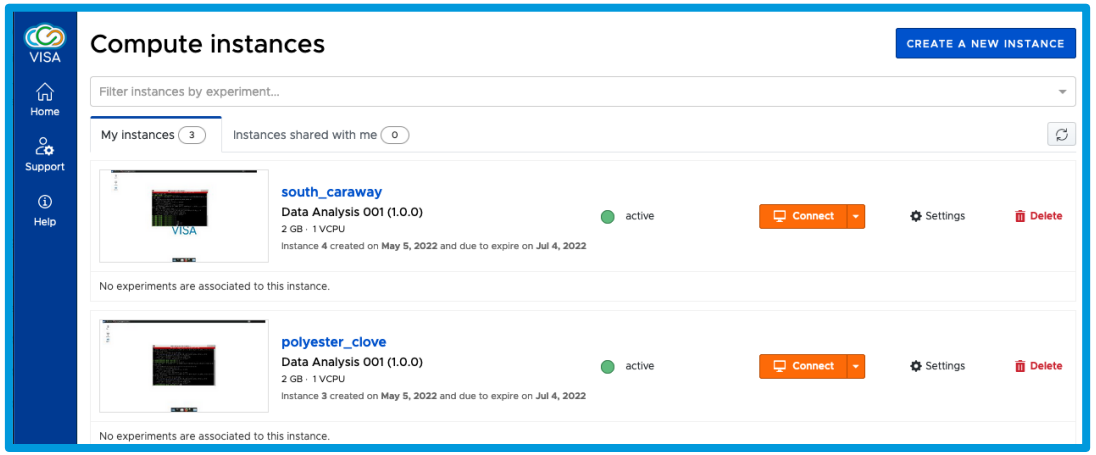
After moving out of the COBIS building DMSC will need to eventually find an alternative hosting solution apart the server room in he COBIS building. Current lease for the COBIS server room runs out in September 2025. A parallel effort is on going:

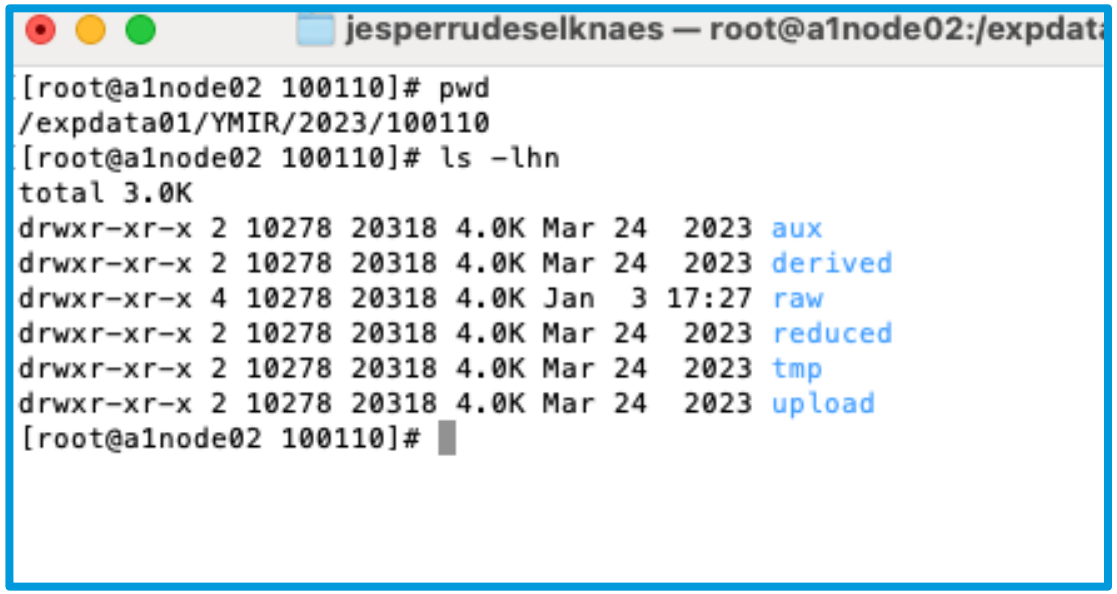
Effort	Description	Status
Extending lease in the COBIS building	Extending the lease for the current server room allows for a timely permanent solution.	Change request approved Building owner is drawing up contract for extension. We aim for an extension to December 2026.
ESS Lund options	Exploring options on the ESS site in Lund	Facility Management is exploring and pricing options in collaboration with DMSC and IT.
DTU Lyngby Campus solution	DTU and DMSC is exploring options for a solution on the DTU Lyngby Campus next to the DMSC address.	Considerable resources from consulting engineers from major Danish companies are being put to use by DTU.

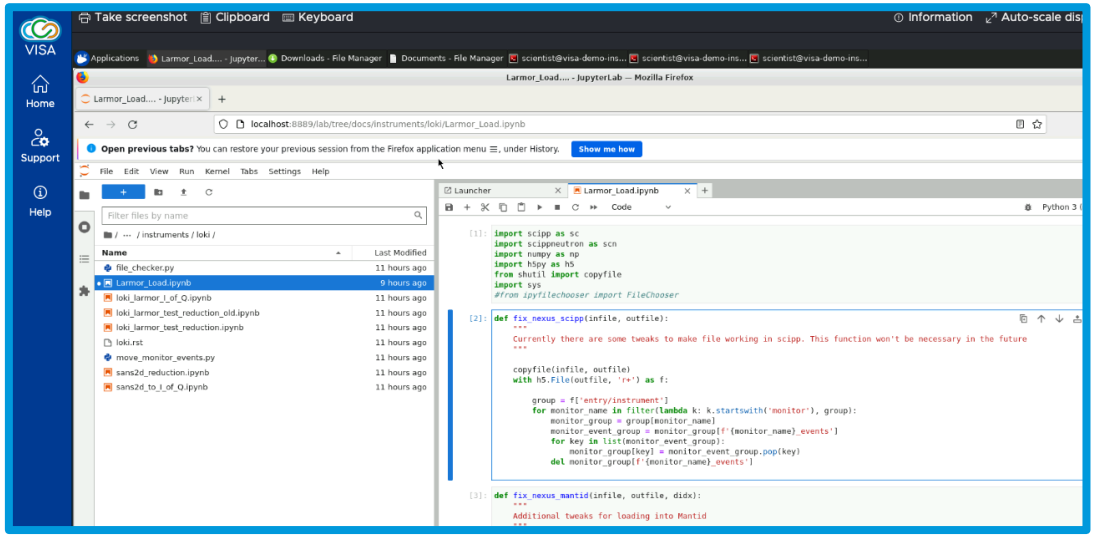
DST achievements

VISA and storage is now integrated with UserOffice

- Joint effort with the SIMS group
- An accepted proposal now results in:
  - A VISA instance
  - A file structure for the experiments data
  - Policies for data access and data transfer to Copenhagen.



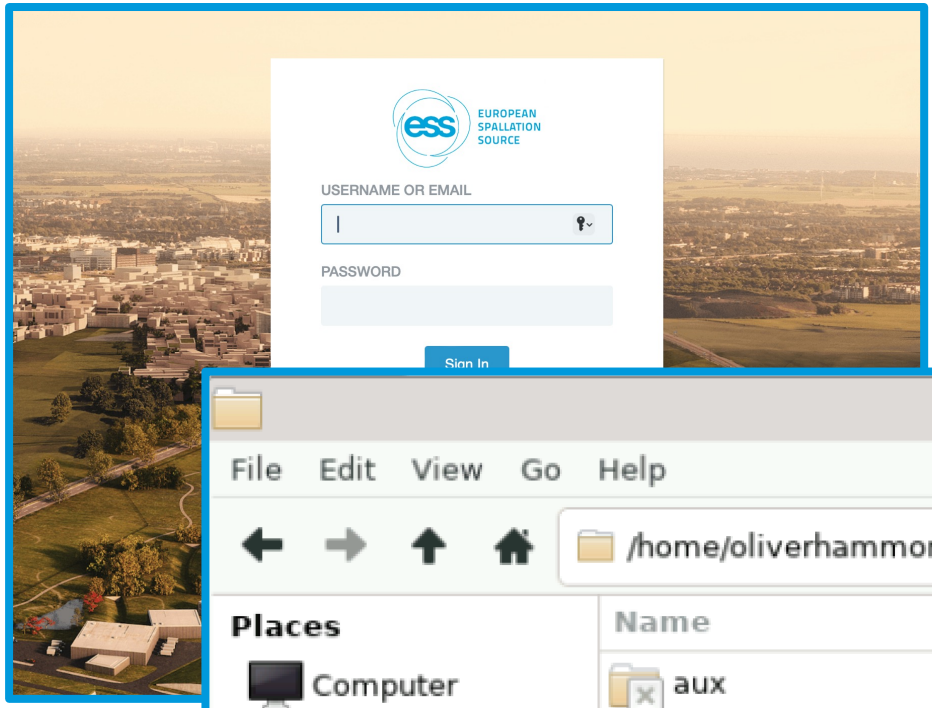


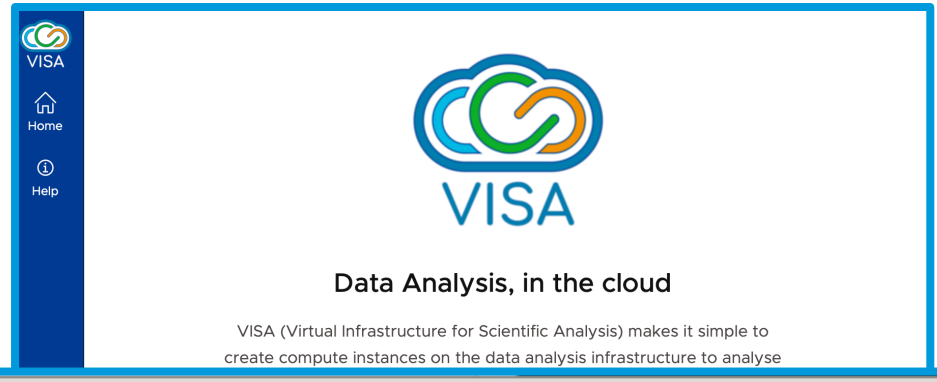


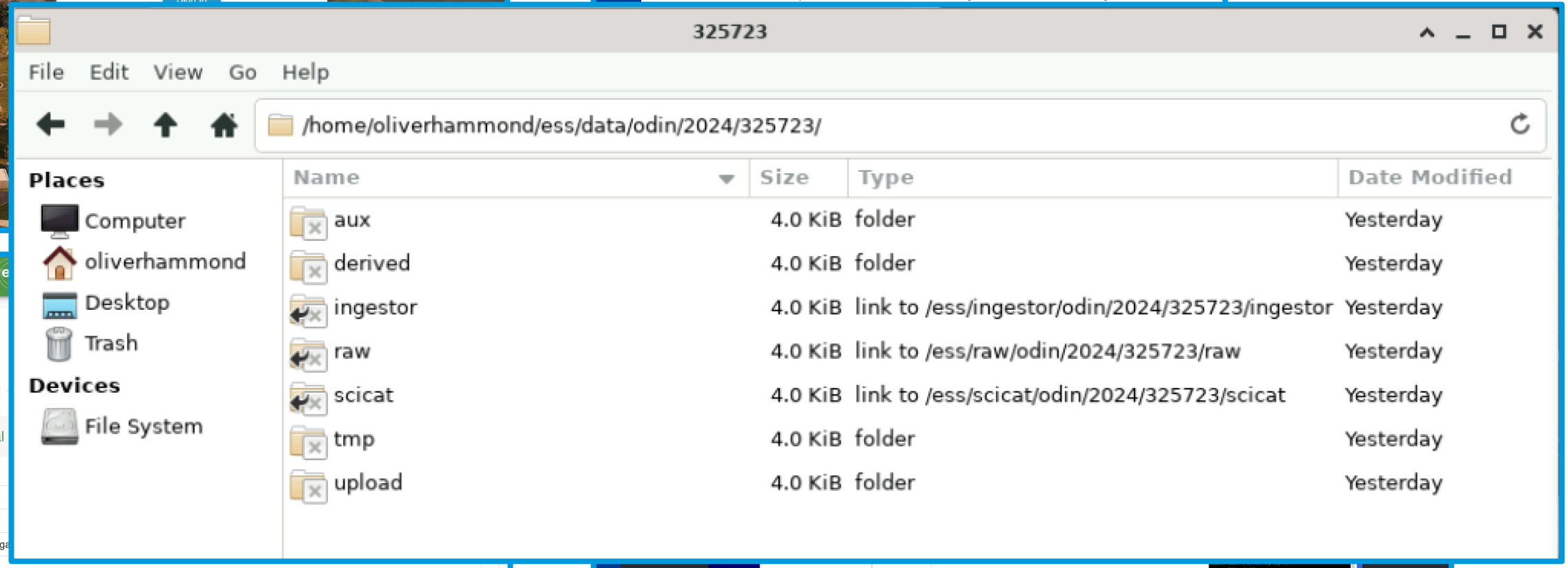
Updates from SIMS

User Office & Data Catalogue

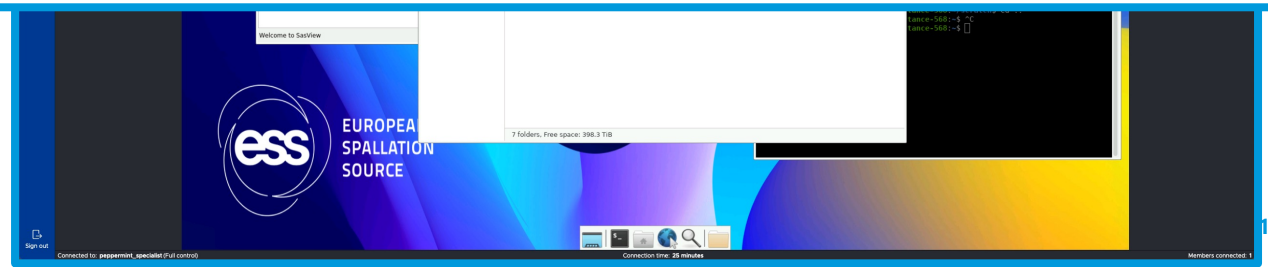
Scientific Information Management Systems


EUROPEAN SPALLATION SOURCE
USERNAME OR EMAIL
PASSWORD
Sign In


VISA
Home
Help
VISA
Data Analysis, in the cloud
VISA (Virtual Infrastructure for Scientific Analysis) makes it simple to
create compute instances on the data analysis infrastructure to analyse


325723
File Edit View Go Help
/home/oliverhammond/ess/data/odin/2024/325723/
Places
Computer
oliverhammond
Desktop
Trash
Devices
File System
Name
Size
Type
Date Modified
aux 4.0 KiB folder Yesterday
derived 4.0 KiB folder Yesterday
ingestor 4.0 KiB link to /ess/ingestor/odin/2024/325723/ingestor Yesterday
raw 4.0 KiB link to /ess/raw/odin/2024/325723/raw Yesterday
scicat 4.0 KiB link to /ess/scicat/odin/2024/325723/scicat Yesterday
tmp 4.0 KiB folder Yesterday
upload 4.0 KiB folder Yesterday

New proposal
New proposal
Proposal ID
Title
Abstract
Principal Investiga
Co-Proposers
Pick your Instrument.
LOKI
BACK
SUBMITTED
DOWNLOAD PDF
PRIVACY STATEMENT
FAQ


EUROPEAN SPALLATION SOURCE

Proposal Folder and File Structure (see Link)

Top level path: /ess/data/<instrument>/<year>/<proposal_id>

- **raw**
  This folder will contain the raw data **files created by the file** writer with the data collected during the experiment time.
  Files/Folder are only writable by the file writer (ECDC staff), **read-only for the proposal team** + instrument team.

- **reduced**
  This folder will contain the reduced data produced by the reduction pipeline specific for the instrument. The reduction pipeline should be running automatically, although the data instrument and instrument scientists might decide to run it additional times after the experiment.
  Files should only be writable by the automated reduction processes

- **derived**
  This folder will contain all the files produced by subsequent user run analysis pipelines.
  The pipelines producing this data might be triggered automatically or manually by internal and external users which decide to store data related to publication in the ESS catalogue.
  SciCat Ingested files need to be accessible and downloadable through SciCat, so they need to be protected from deletion and changes

- **auxiliary**
  This folder is intended to contain **all additional files** that are instrumental for the proposal and experiment but are not directly part of the datasets available in SciCat.
  These files can be produced before, during or after experiments.
  Users will have **write permissions** on this folder.
  SciCat Ingested files need to be accessible and downloadable through SciCat, so they need to be protected from deletion and changes

- **temporary**
  This folder is for users to store **any file** that can be recreated or are not important and can be deleted at any point
  Users will have full **write permissions** on this folder

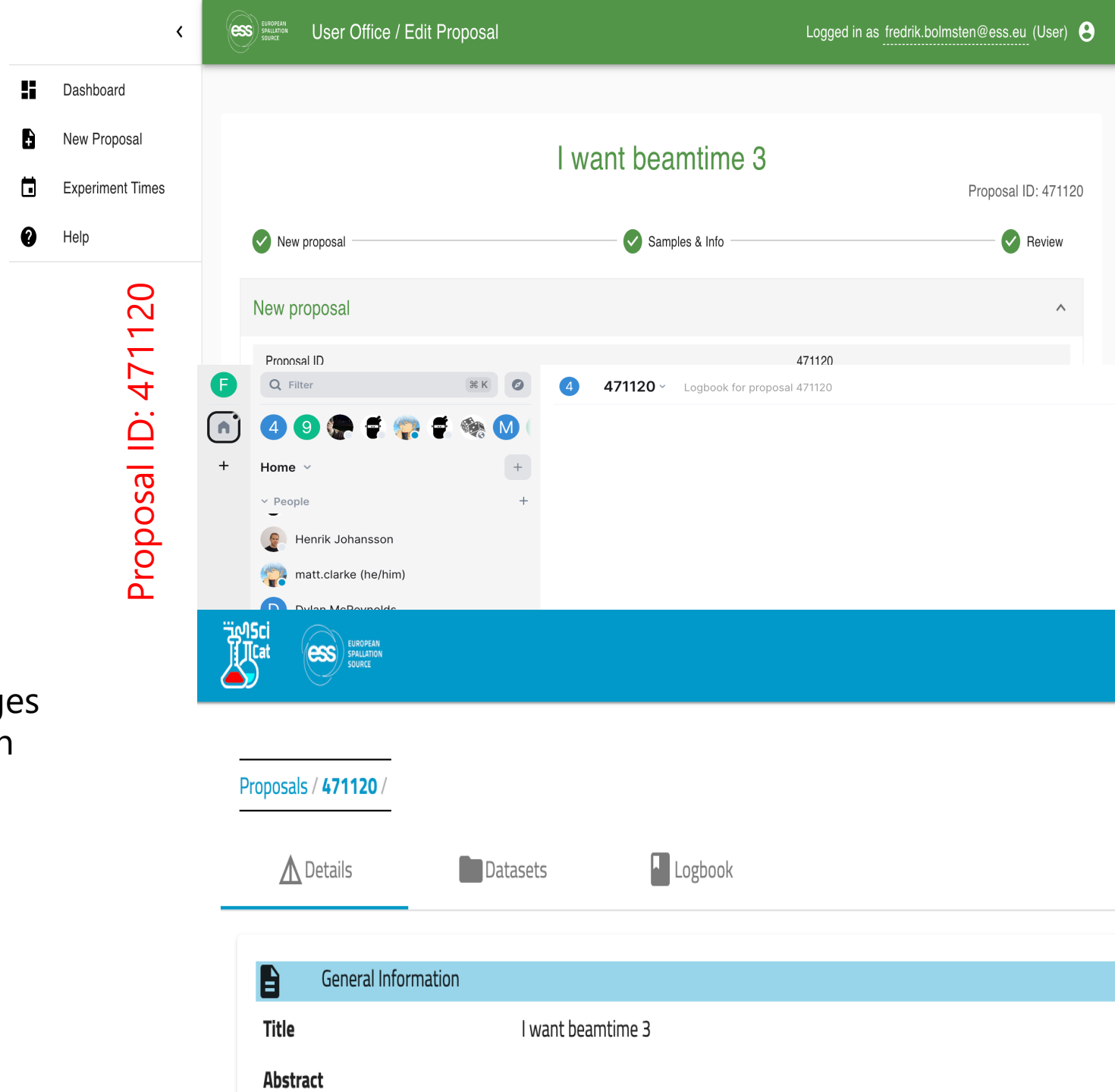
  The proposal folder structure is created as soon as the proposal is accepted and scheduled for beam time.

```
[root@sftpserver2 ~]# sudo -u nexus ls -l /sftproot/ess/data/ymir/2024/451867/
total 4
drwxr-x---     2 nexus ess_proposal_451867 4096 May  6 15:26 aux
dr-xr-x---     2 nexus ess_proposal_451867 4096 May  6 15:26 derived
drwx------ 65535 root  root               4096 May  6 15:26 ingestor
drwx------ 65535 root  root               4096 May  6 15:26 raw
dr-xr-x---     2 nexus ess_proposal_451867 4096 May  6 15:26 reduced
drwxr-x---     2 nexus ess_proposal_451867 4096 May  6 15:26 tmp
drwxr-x---     2 nexus ess_proposal_451867 4096 May  6 15:26 upload
```

YMIR

Testing it all

- Account creation
- Proposal creation
- Technical Review
- Scientific Review
- Scheduling
- Folder creation
- VISA Instance
- **SciChat room creation**
- **SciCat dataset creation**
- Simulating NiCOS to SciChat messages
- Jupyterhub to SciCat dataset creation
- Download



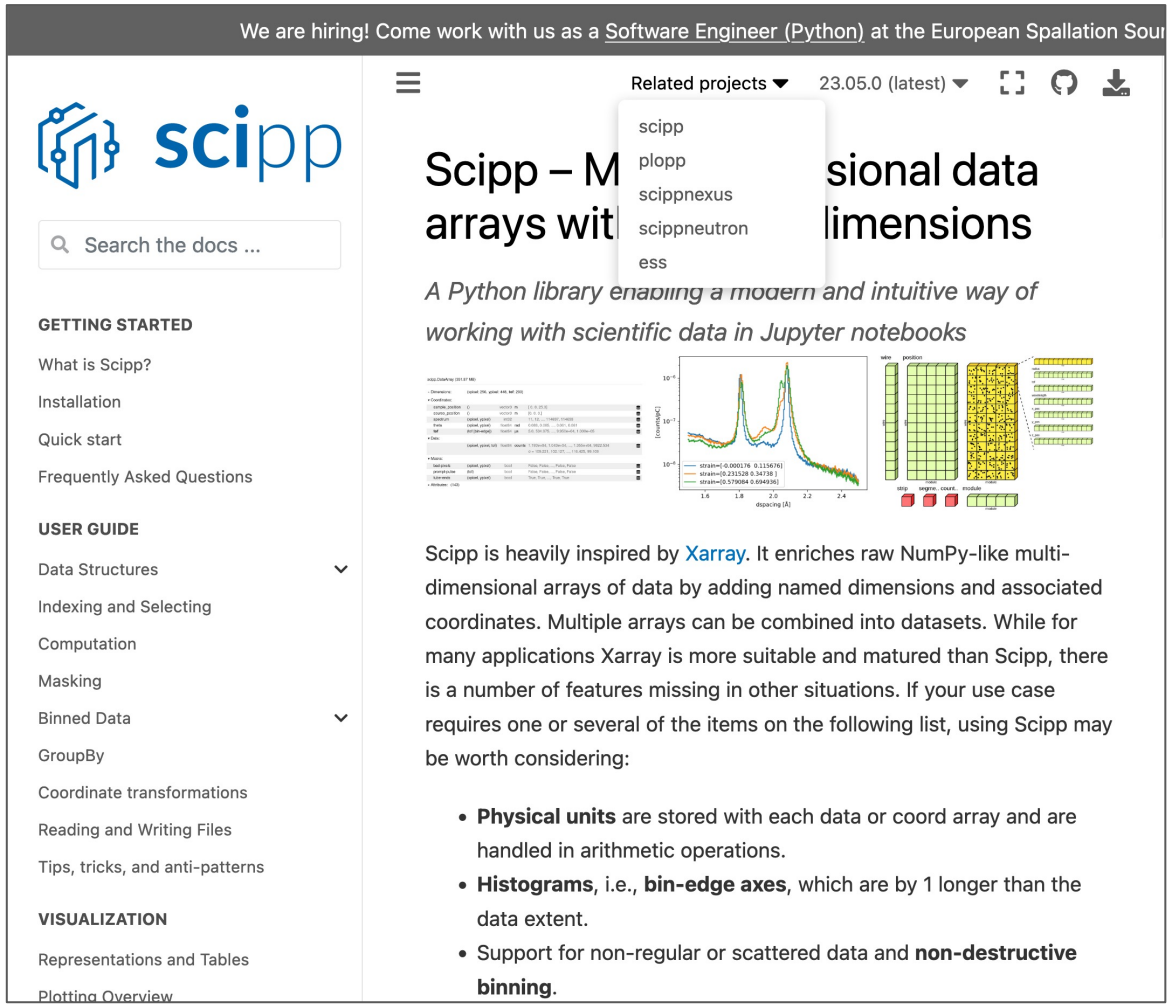
Updates from DRAM

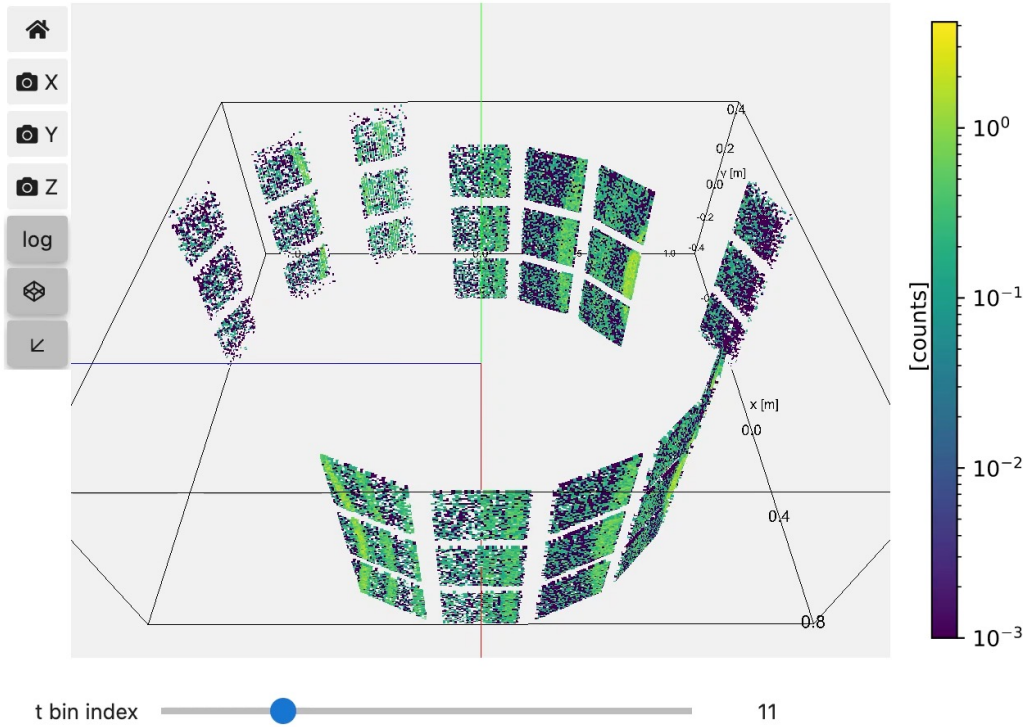
Scipp, EasyScience, McStas

Data Reduction, Analysis & Modelling

Data reduction & visualization software

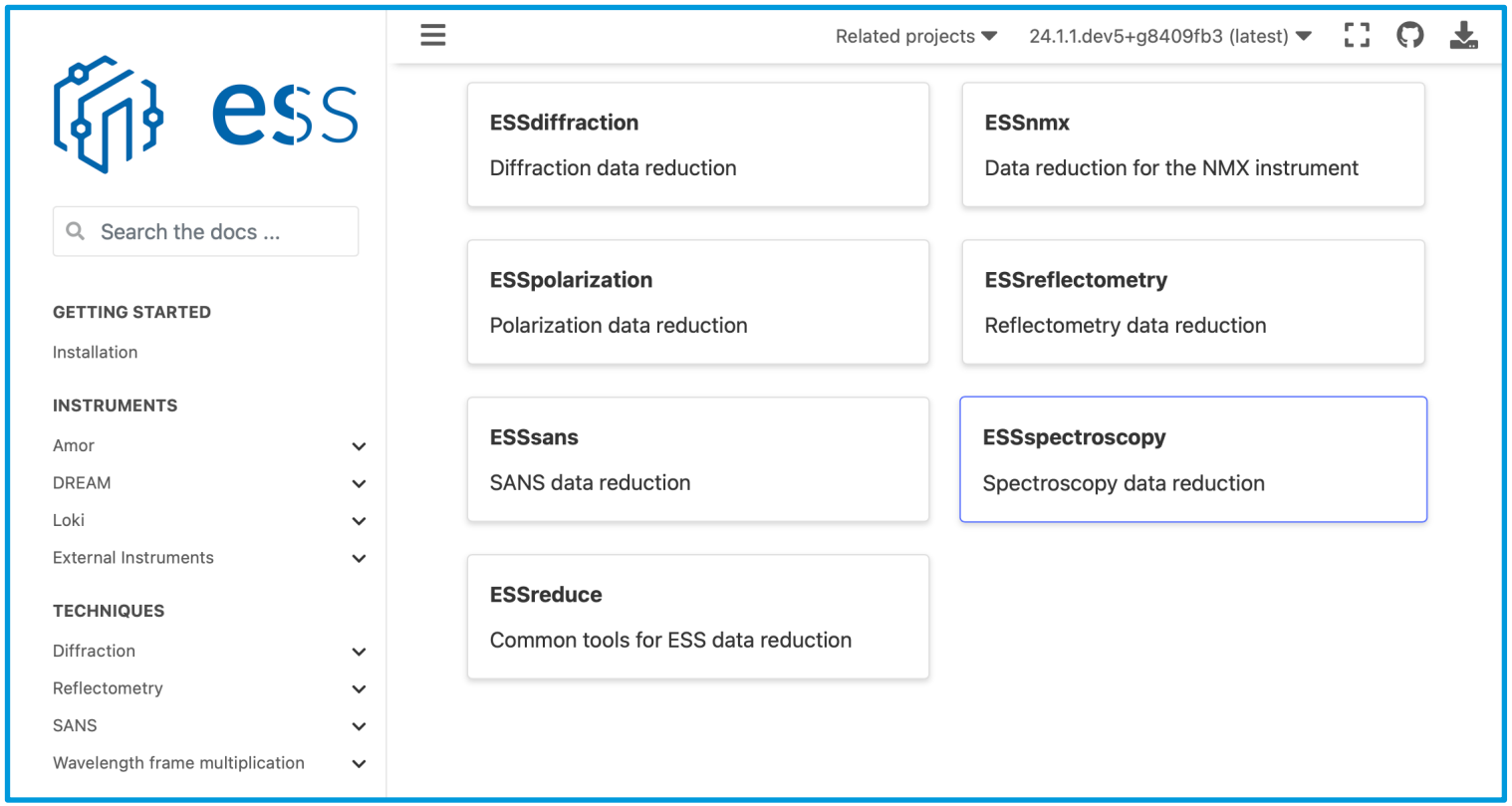
https://scipp.github.io





Data reduction workflows for ESS

On-line documentation // Getting started



- scipp.github.io/ess

python

jupyter

GitHub

- **IDS team <---> Scipp team**
- **Part of a demo for DMSC STAP next week (& P6 MS)**

Data reduction workflows for ESS

(Very) Early preview version of auto-generated GUI for selecting and running reduction workflows in ESSsans

[1]: Workflow: LokiAtLarmorTutorialWorkflow

▼ Typical Outputs

BackgroundSubtractedIofQ
BackgroundSubtractedIofQxy
IofQ[SampleRun]
IofQxy[SampleRun]
IofQ[BackgroundRun]
IofQxy[BackgroundRun]
MaskedData[BackgroundRun]
MaskedData[SampleRun]
WavelengthMonitor[SampleRun, Incident]
WavelengthMonitor[SampleRun, Transmission]
WavelengthMonitor[BackgroundRun, Incident]
WavelengthMonitor[BackgroundRun, Transmission]

▸ Extended Outputs

Run | Clear Output

scipp.DataArray (3.75 KB)

▸ Dimensions:	(Q: 100)			
▾ Coordinates:				
L1	()	float64	m	2
Q	(Q [bin-edg...	float64	1/Å	0
gravity	()	vector3	m/s^2	[
incident_beam	()	vector3	m	[
wavelength	(wavelengt...	float64	Å	2
▾ Data:				
	(Q)	float32	1	4

Refresh Parameters

CorrectForGravity

NeXusDetectorName: larmor_detector
NeXusMonitorName[Incident]: monitor_1
NeXusMonitorName[Transmission]: monitor_2
TransformationPath: transform
PixelMaskFilename: /home/runner/.cache/ess/loki/2/mask_new_
PixelShapePath: pixel_shape

ReturnEvents

<enum 'UncertaintyBroadcastMode'>: UncertaintyBroadcastM

Filename[SampleRun]: /home/runner/.cache/ess/loki/2/60339-2
Filename[TransmissionRun[SampleRun]]: /home/runner/.cache/e
Filename[EmptyBeamRun]: /home/runner/.cache/ess/loki/2/603

angstrom start: 2 stop: 12 nbins: 300

1/angstrom start: 0.01 stop: 0.3 nbins: 100

None DirectBeam

x = -0.02914868 y = -0.01816138 unit: m

Accessing the results

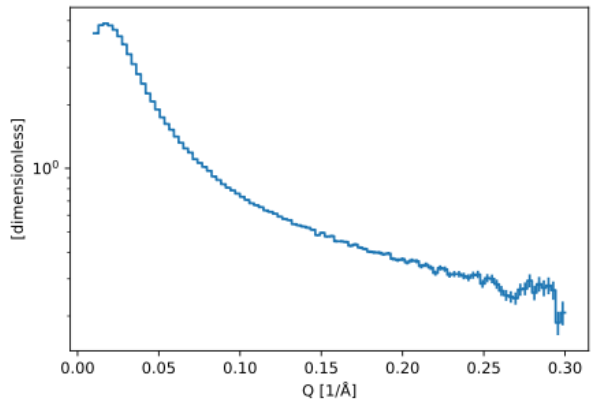
We can now access the computed result in the `results` dictionary:

[3]:
```
results
```

[3]:
```
{ess.sans.types.IofQ[ess.reduce.nexus.types.SampleRun]: <scipp.DataArray>
 Dimensions: Sizes[Q:100, ]
 Coordinates:
   L1                          float64              [m]  ()  25.61
 * Q                           float64           [1/Å]  (Q [bin-edge])  [0.01, 0.0129, ..., 0.2971, 0.3
 * gravity                     vector3          [m/s^2]  ()  (0, -9.80665, 0)
   incident_beam               vector3              [m]  ()  (0, 0, 25.61)
 * wavelength                  float64              [Å]  (wavelength [bin-edge])  [2, 12]
 Data:
                               float32  [dimensionless]  (Q)  [4.35242, 4.75759, ..., 0.186054, 0.20749
 }
```

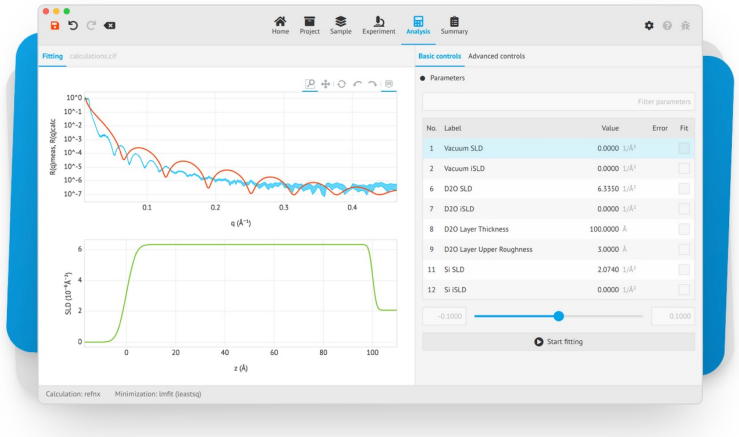
[4]:
```
(da,) = results.values()
da.plot(norm="log")
```

[4]:



Data analysis software

https://easyscience.software



User interfaces:

- Python
- Jupyter
- GUI

Multiple libraries (Calculators)

Other requests are:

- Data analysis for QENS
- Data analysis for TOF imaging

IDS team <---> Analysis team

Updates from integration test

Test the DMSC data pipe line. Can we read the files – and more .....

ECDC Data pipeline

--> (DMSC) Integration testing

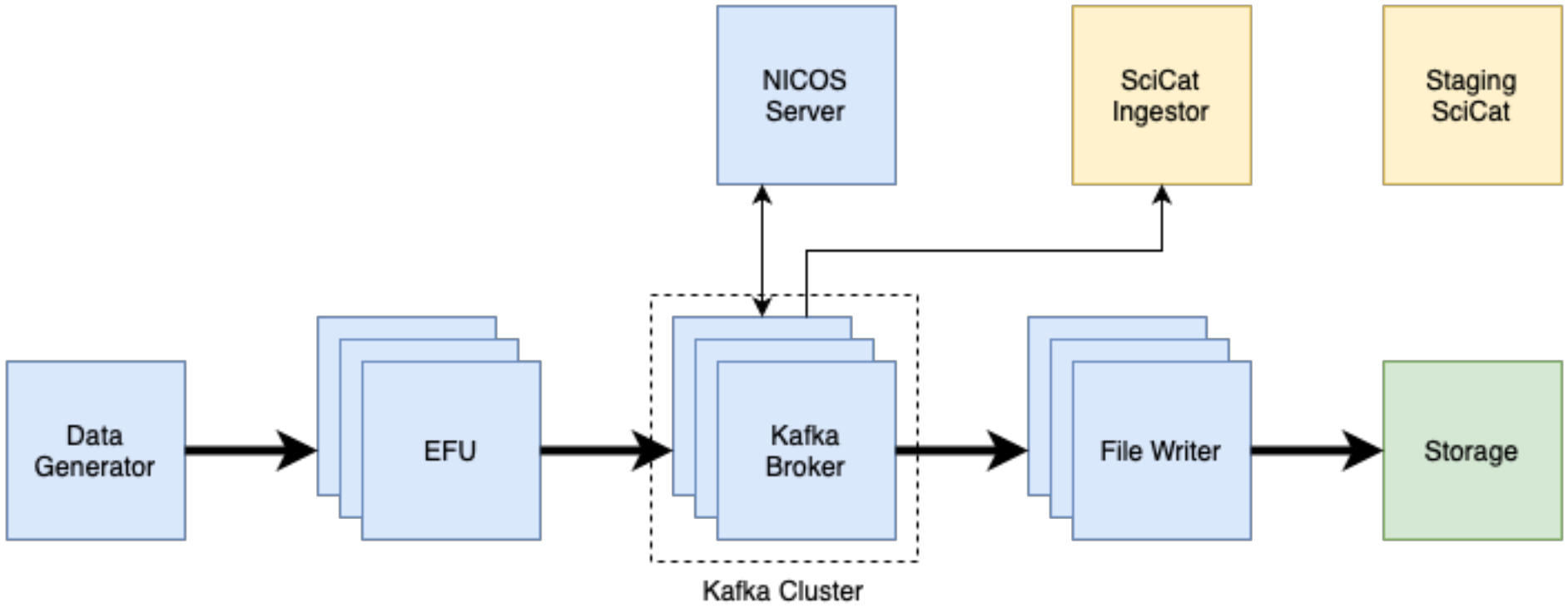
- **CODA** (COntinuous Data Acquisition) add for various instruments
  - Writes multiple datafiles per hour by **ECDC**
  - Datafiles are stored on **DST** hardware
  - Data catalogued in SciCat by **SIMS**
  - Data reduced and analysed by **DRAM**
  - Has been the driver for solving many issues

- Established monthly data surgeries
  - DRAM, ECDC, IDSs, SIMS
  - Discuss/resolve issues with data and the data pipeline

CODA

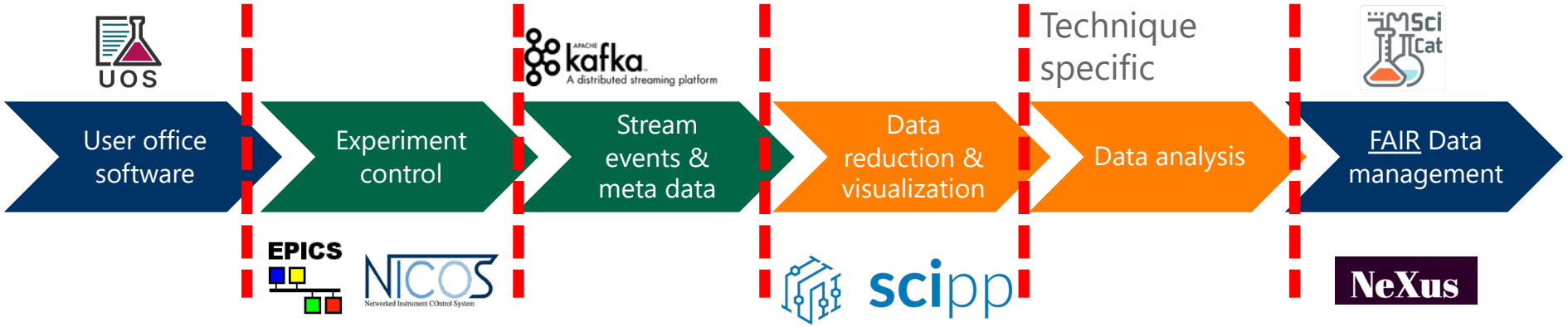
https://confluence.esss.lu.se/display/ECDC/CODA

- CODA (Continuous Data) continuously demonstrates the ECDC software infrastructure at work, from EFUs to file writing, through Kafka.



DMSC integration testing

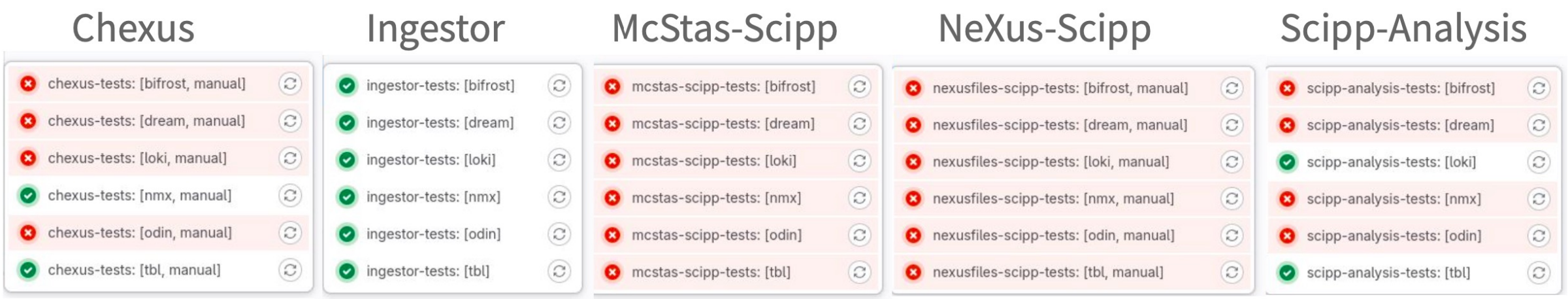
We are testing interfaces!



Overview

Different test set-up for different interfaces

- Gitlab project: https://git.esss.dk/dmsc-nightly/dmsc-nightly
- Pipeline runs every night (00:05)
- Currently have: **57** tests, with **29** failing, and **28** passing. (End August 2024)



What are we testing?

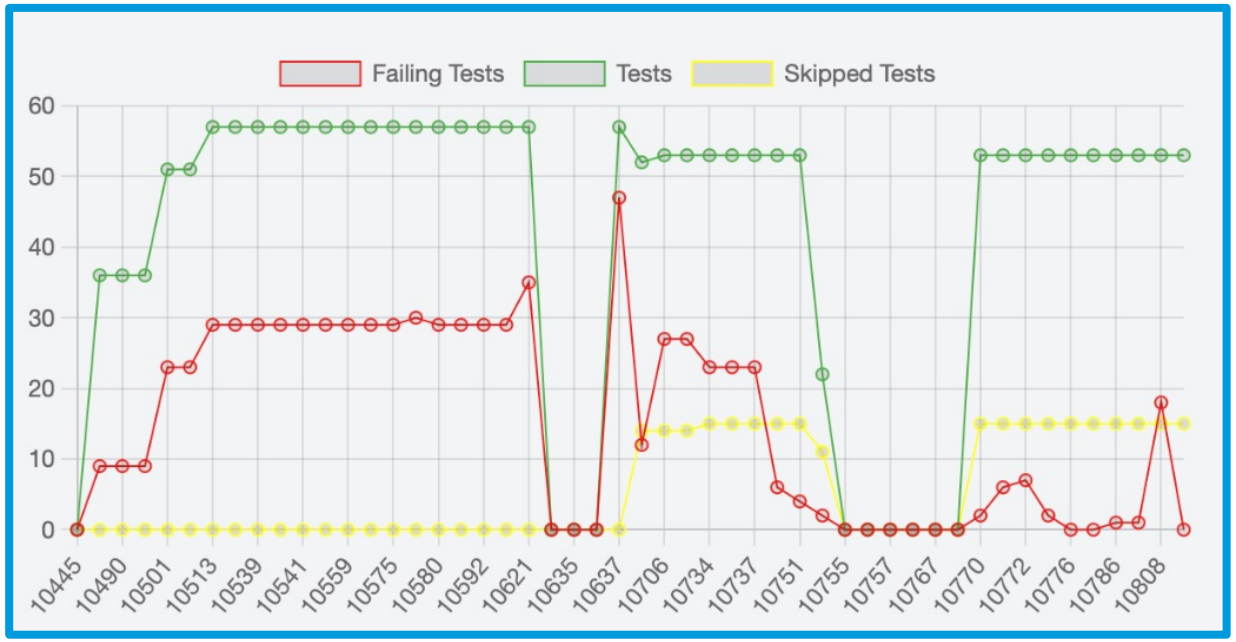
- **Chexus:**
  - Verify that files written by ECDC adhere to the **NeXus standard**

- **File ingestor:**
  - Check that the latest file found in the CODA folder is less than 24h old
  - Check that a Scicat query for the latest file in the CODA project is less than 24h old
  - Check that the file returned by the Scicat query is the same as the latest file found on disk

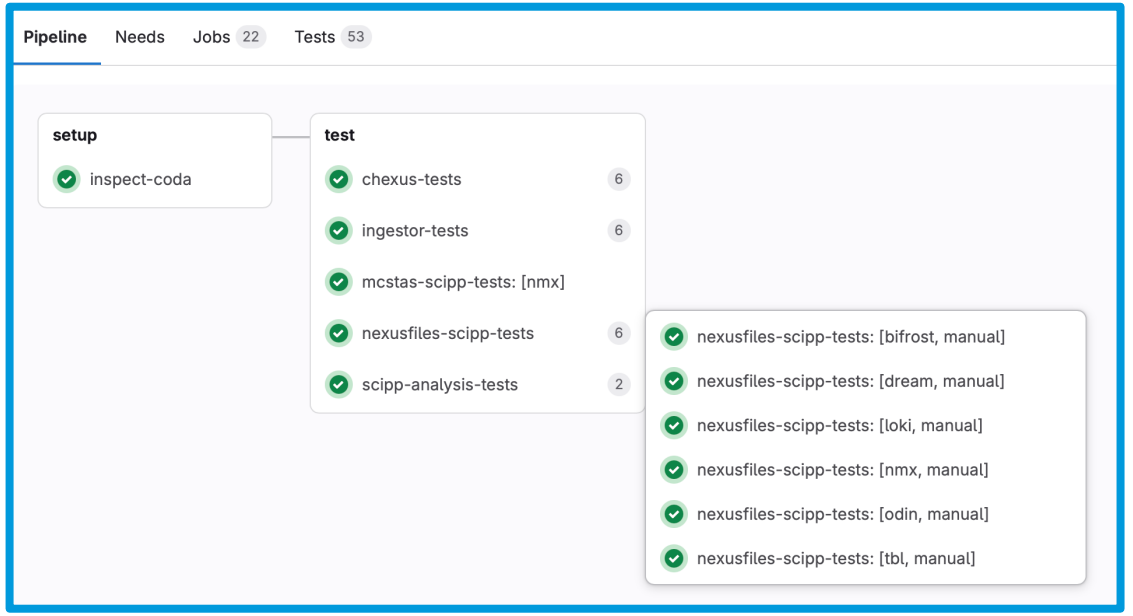
- **Nexus-Scipp:**
  - Verify that files written by ECDC **can be loaded by Scipp** without any errors
  - Attempt to compute/unwrap the time-of-flight event coordinate in the files
  - Attempt to perform a reduction workflow (e.g. compute I(Q))
  - Can reduced result be uploaded to Scicat?

DMSC integration test

Monitoring over time



21-Sep-2024



Updates from Summer School

2 - 6 September 2024

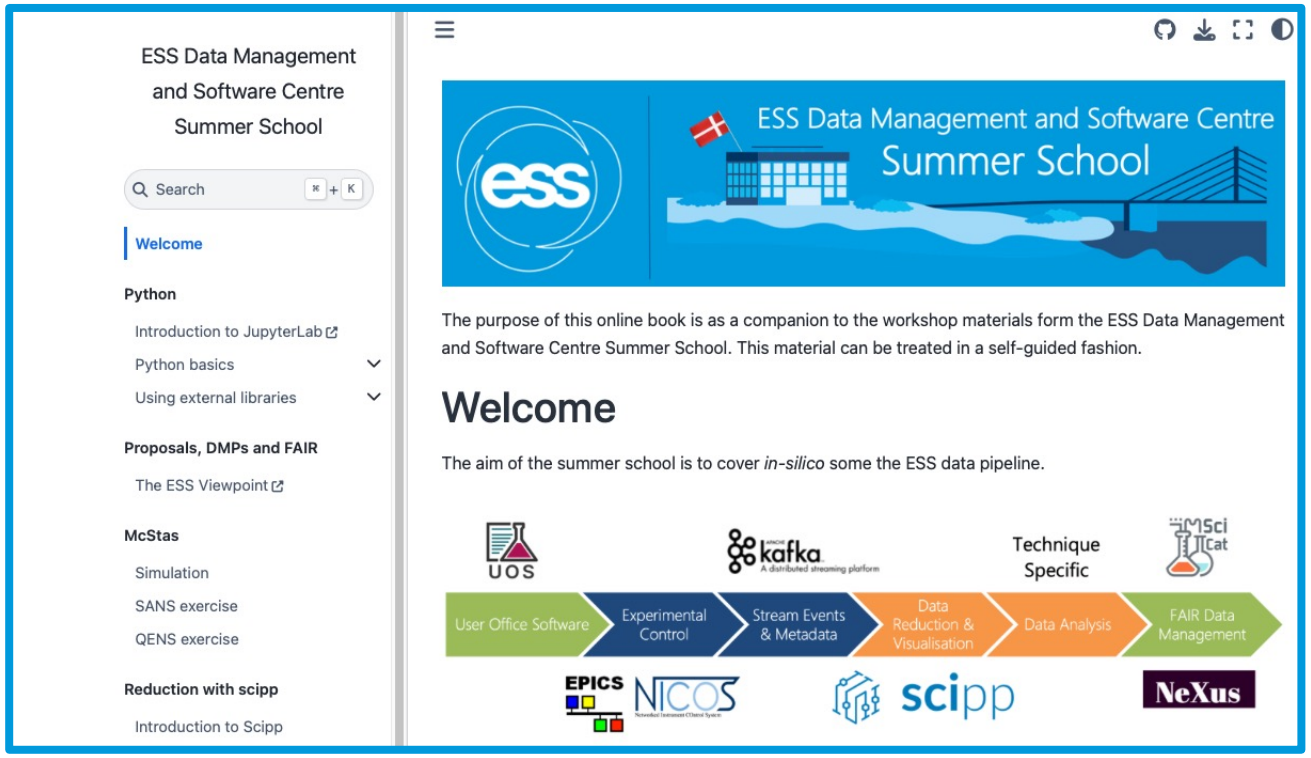
- Around 20 people participated

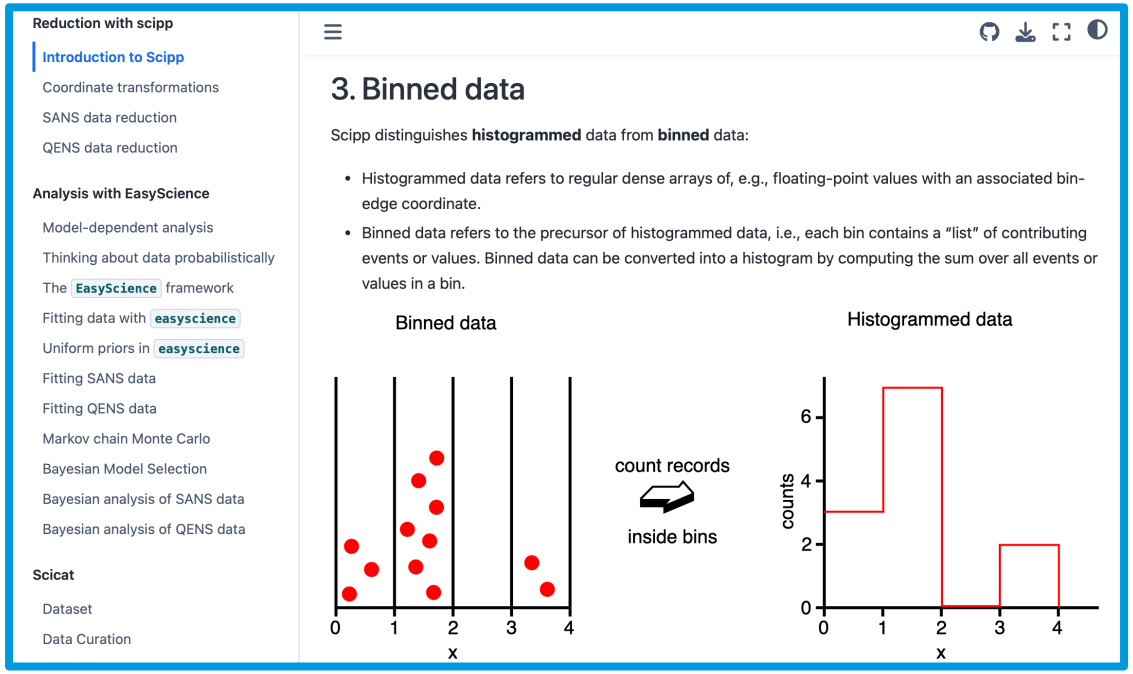
With a mix of lectures & workshops on subjects including:

- Using Python for scientific computing
- Data management and the role of data in the beamtime proposal
- Simulating neutron scattering experiments
- Modern reduction and analysis of data
- The importance of scientific metadata and reproducibility

DMSC summer school 2024

https://ess-dmsc-dram.github.io/dmsc-school





Concluding remarks

Summary

Overall good progress

- ❑ Staff recruited (IDS, DST)
- ❑ Testing DMSC data pipeline (Joint projects across DMSC):
  - ❑ Test systems in place (CODA) data files written by the ESS infrastructure
  - ❑ VISA & User office integration
  - ❑ Data reduction for first instruments
  - ❑ DMSC Summer School

Thank you