



**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

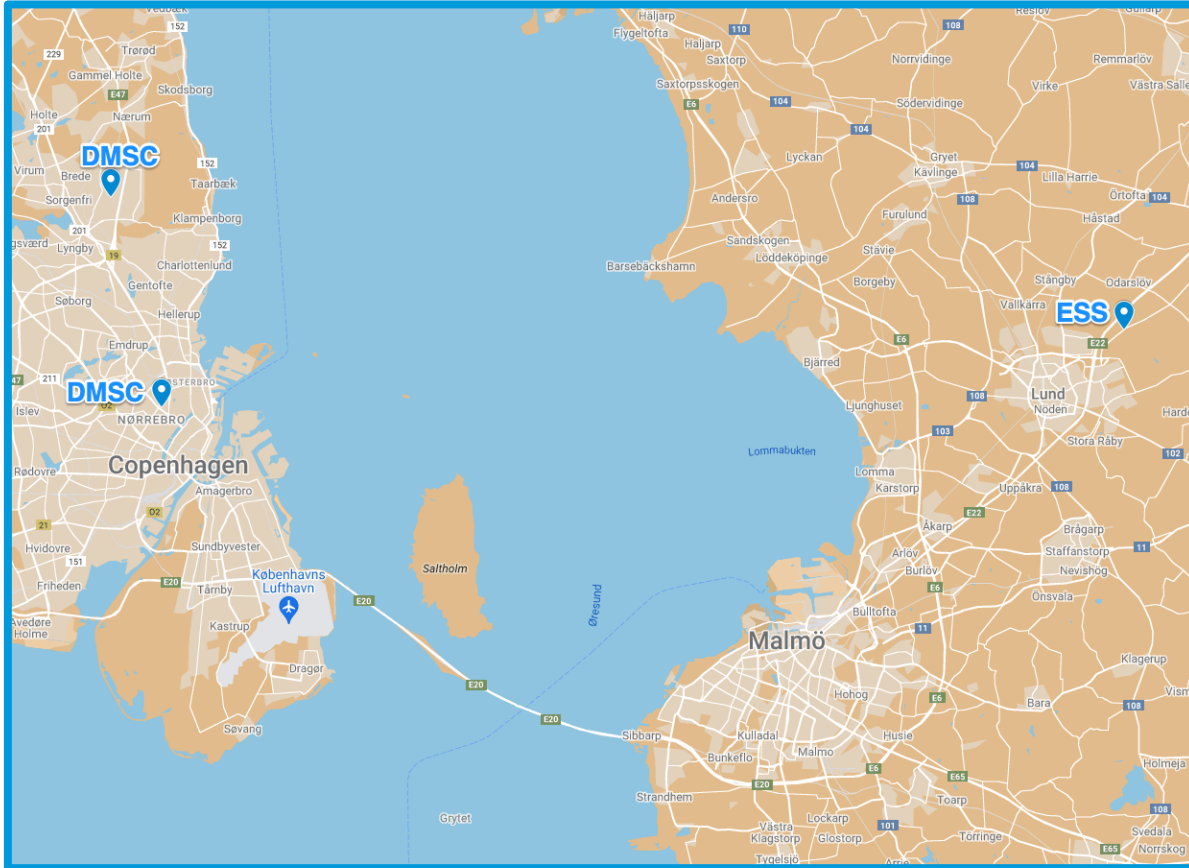
DMSC



# Data Management and Software Centre (DMSC)



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



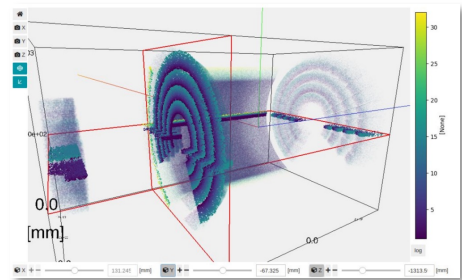
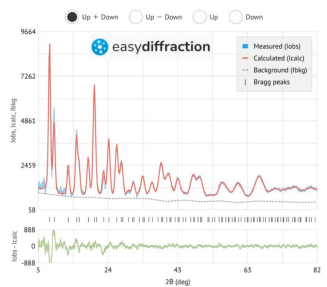
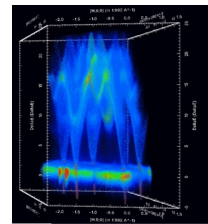
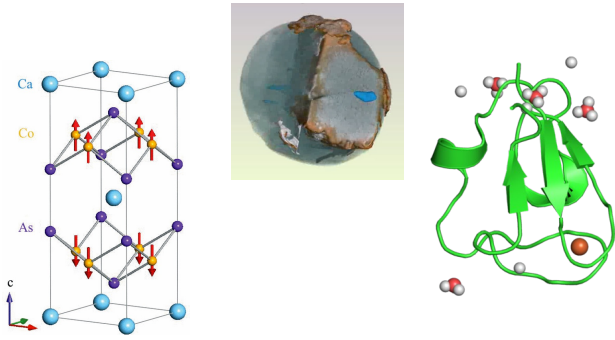
ESS - Lund

DMSC - Copenhagen area

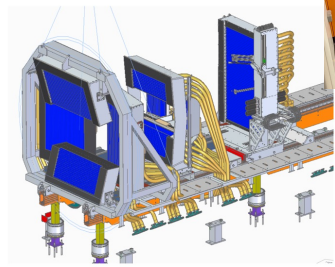
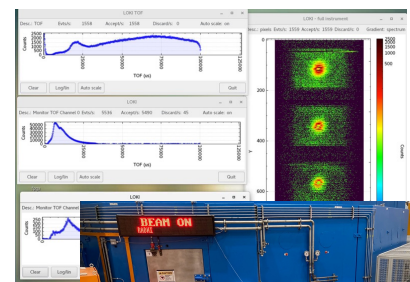
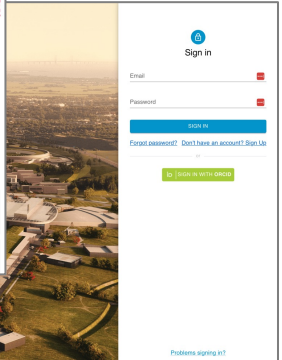
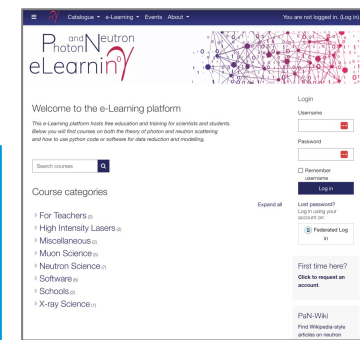
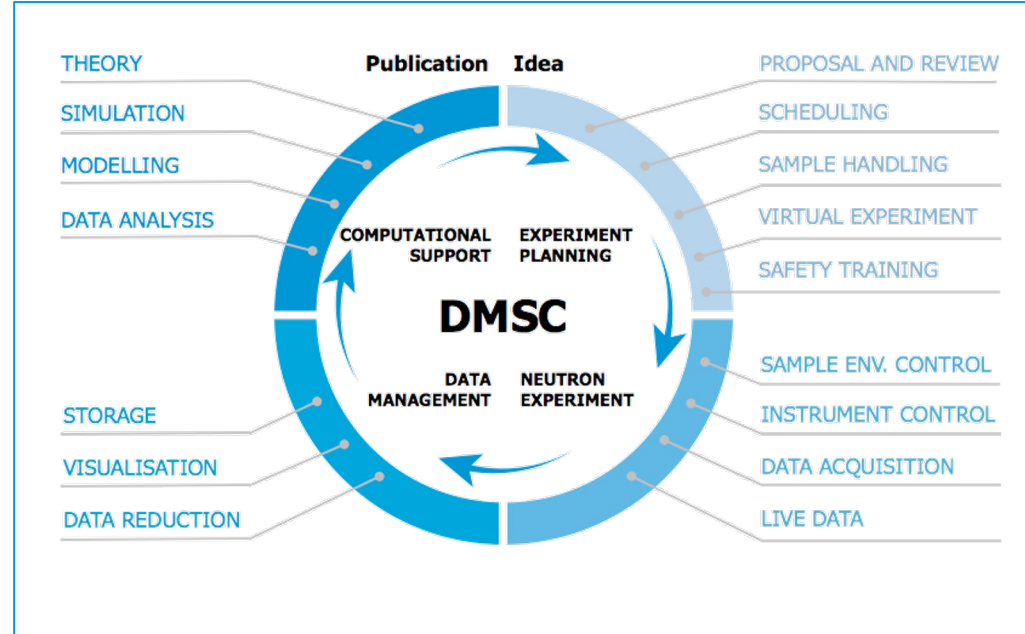
# DMSC & Scientific computing



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

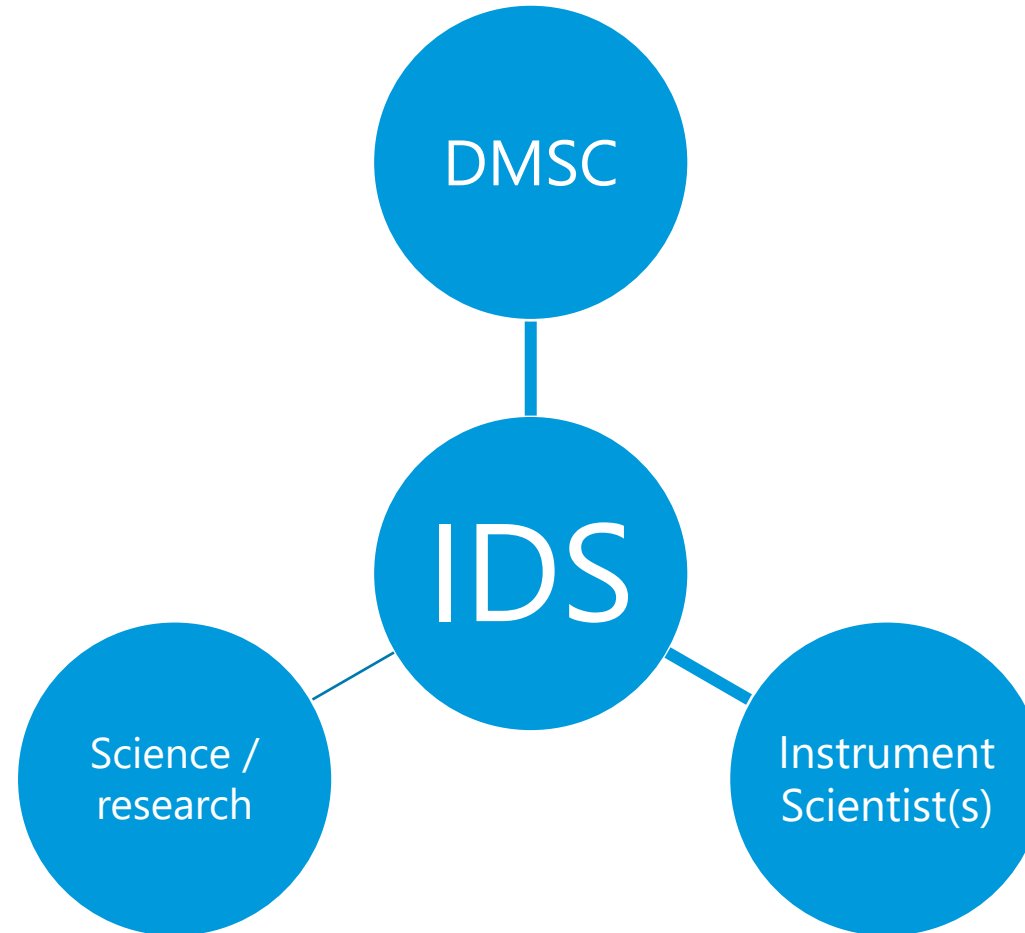


Search	Results
Search	1-31
1	1-31
2	1-31
3	1-31
4	1-31
5	1-31
6	1-31
7	1-31
8	1-31
9	1-31
10	1-31
11	1-31
12	1-31
13	1-31
14	1-31
15	1-31
16	1-31
17	1-31
18	1-31
19	1-31
20	1-31
21	1-31
22	1-31
23	1-31
24	1-31
25	1-31
26	1-31
27	1-31
28	1-31
29	1-31
30	1-31
31	1-31

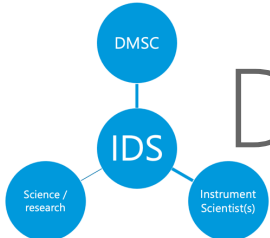


# Instrument Data Scientists (IDS)

DMSC interface to instrument teams





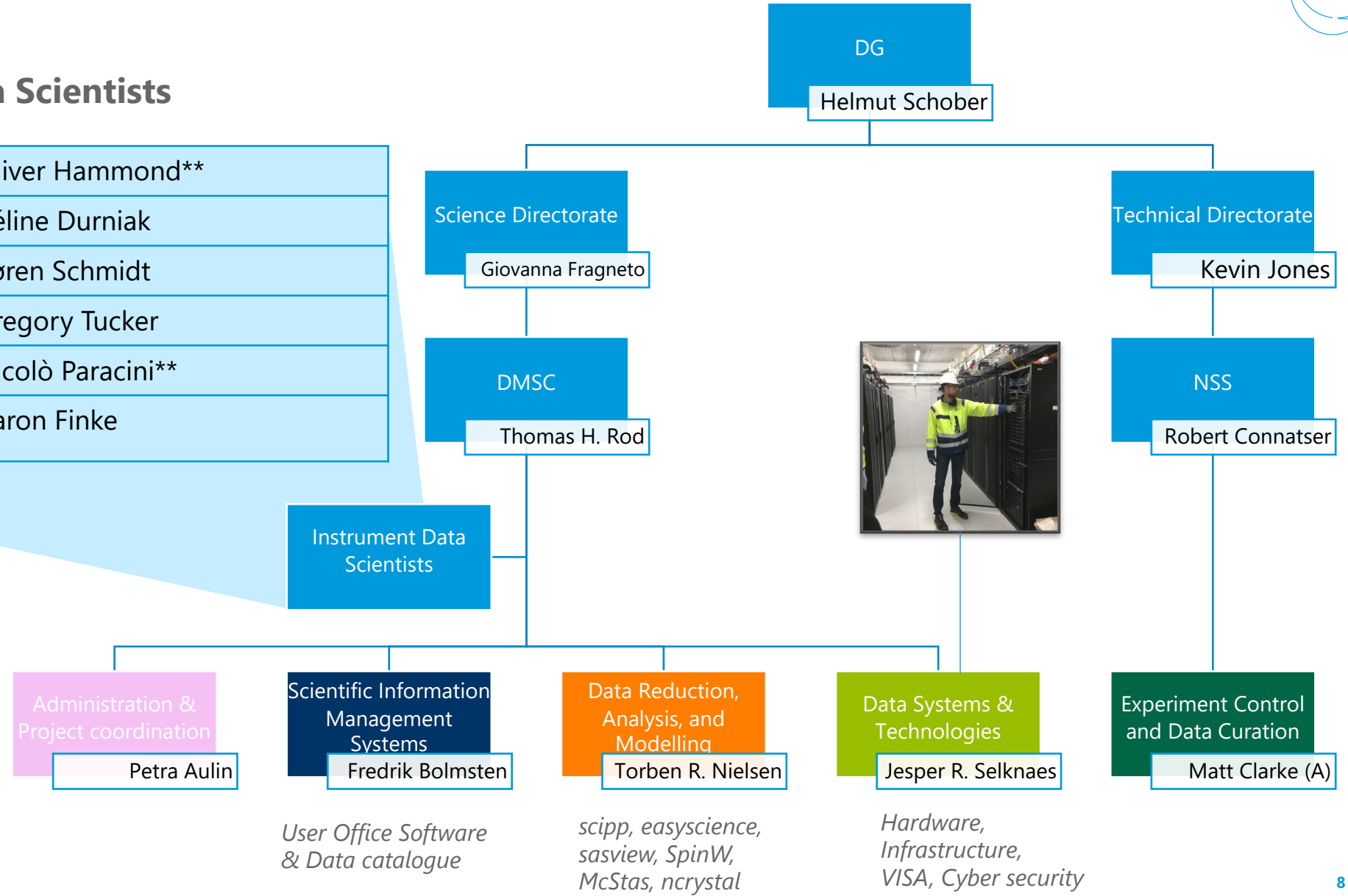


# DMSC organisation



## Instrument Data Scientists

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

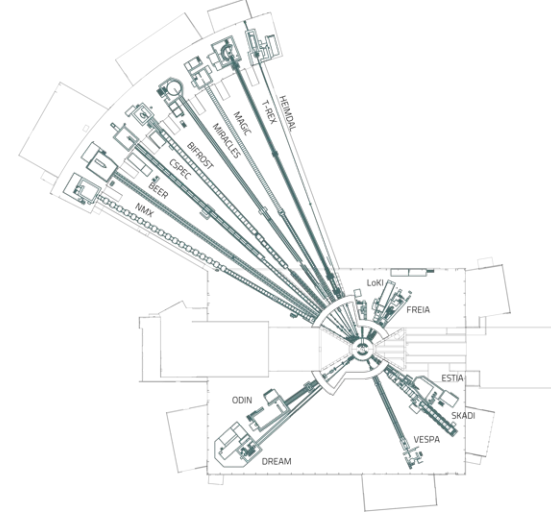


*User Office Software & Data catalogue*

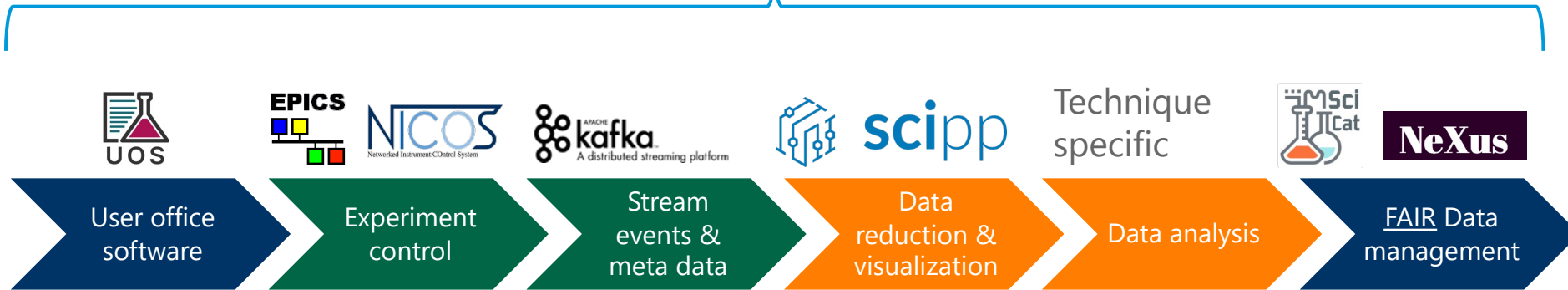
*scipp, easyscience, sasview, SpinW, McStas, ncrystal*

*Hardware, Infrastructure, VISA, Cyber security*

# Integrated data pipeline



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



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.

```

jesperrudeselknaes — root@a1node02:/expdat
[root@a1node02 100110]# pwd
/expdata01/YMIR/2023/100110
[root@a1node02 100110]# ls -lhn
total 3.0K
drwxr-xr-x 2 10278 20318 4.0K Mar 24 2023 aux
drwxr-xr-x 2 10278 20318 4.0K Mar 24 2023 derived
drwxr-xr-x 4 10278 20318 4.0K Jan 3 17:27 raw
drwxr-xr-x 2 10278 20318 4.0K Mar 24 2023 reduced
drwxr-xr-x 2 10278 20318 4.0K Mar 24 2023 tmp
drwxr-xr-x 2 10278 20318 4.0K Mar 24 2023 upload
[root@a1node02 100110]#

```

```

[1]: import scipp as sc
import scippneutron as scn
import numpy as np
import h5py as h5
from shutil import copyfile
import sys
from ipynb_launcher import FileChooser

[2]: def fix_nexus_scipp(infile, outfile):
    ...
    Currently there are some tweaks to make file working in scipp. This function won't be necessary in the future
    copyfile(infile, outfile)
    with h5.File(outfile, 'a') as f:
        group = f['entry/instrument']
        for monitor_name in filter(lambda k: k.startswith('monitor'), group):
            monitor_group = group[monitor_name]
            monitor_event_group = monitor_group['(monitor_name)events']
            for key in list(monitor_event_group):
                monitor_group[key] = monitor_event_group.pop(key)
            del monitor_group['(monitor_name)events']

[3]: def fix_nexus_mantid(infile, outfile, dds):
    ...
    Additional tweaks for loading into Mantid
    ...


```



# Updates from SIMS

User Office & Data Catalogue

# Scientific Information Management Systems



EUROPEAN SPALLATION SOURCE

USERNAME OR EMAIL

PASSWORD

Sign In

VISA

Home

Help



## 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	Name	Size	Type	Date Modified
Computer	aux	4.0 KiB	folder	Yesterday
oliverhammond	derived	4.0 KiB	folder	Yesterday
Desktop	ingestor	4.0 KiB	link to /ess/ingestor/odin/2024/325723/ingestor	Yesterday
Trash	raw	4.0 KiB	link to /ess/raw/odin/2024/325723/raw	Yesterday
Devices	scicat	4.0 KiB	link to /ess/scicat/odin/2024/325723/scicat	Yesterday
File System	tmp	4.0 KiB	folder	Yesterday
	upload	4.0 KiB	folder	Yesterday

New proposal

New proposal

Proposal ID

Title

Abstract

Principal Investigator


Co-Proposers

Pick your Instrument.

BACK SUBMITTED DOWNLOAD PDF

[PRIVACY STATEMENT](#) [FAQ](#)

Welcome to SarView



EUROPEAN SPALLATION SOURCE

7 folders, Free space: 398.3 TiB

16

# 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



ESS EUROPEAN SPALLATION SOURCE User Office / Edit Proposal Logged in as fredrik.bolmsten@ess.eu (User)

Dashboard  
New Proposal  
Experiment Times  
Help

I want beamtime 3 Proposal ID: 471120

✓ New proposal ✓ Samples & Info ✓ Review

New proposal

Proposal ID 471120

Filter 36 K

4 9

Home

People

Henrik Johansson  
matt.clarke (he/him)

ESS EUROPEAN SPALLATION SOURCE

Proposals / 471120 /

Details Datasets Logbook

General Information

Title I want beamtime 3

Abstract

Proposal ID: 471120



# Updates from DRAM

Scipp, EasyScience, McStas


# Data Reduction, Analysis & Modelling

# Data reduction & visualization software



<https://scipp.github.io>

We are hiring! Come work with us as a [Software Engineer \(Python\)](#) at the European Spallation Source



Search the docs ...

**GETTING STARTED**

- What is Scipp?
- Installation
- Quick start
- Frequently Asked Questions

**USER GUIDE**

- Data Structures
- Indexing and Selecting
- Computation
- Masking
- Binned Data
- GroupBy
- Coordinate transformations
- Reading and Writing Files
- Tips, tricks, and anti-patterns

**VISUALIZATION**

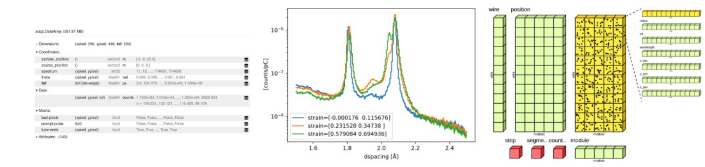
- Representations and Tables
- Plotting Overview

Related projects 23.05.0 (latest)

- scipp
- plopp
- scippnexus
- scippneutron
- ess

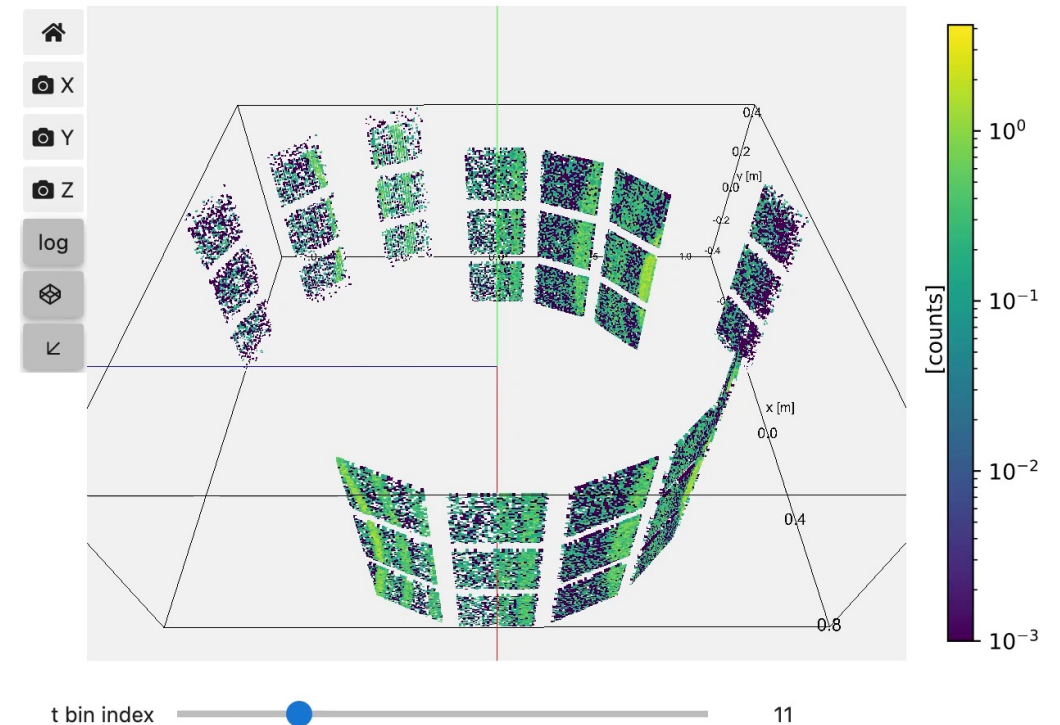
## Scipp – Multidimensional data arrays with named dimensions

A Python library enabling a modern and intuitive way of working with scientific data in Jupyter notebooks



Scipp is heavily inspired by [Xarray](#). It enriches raw NumPy-like multi-dimensional arrays of data by adding named dimensions and associated coordinates. Multiple arrays can be combined into datasets. While for many applications Xarray is more suitable and matured than Scipp, there is a number of features missing in other situations. If your use case requires one or several of the items on the following list, using Scipp may be worth considering:

- **Physical units** are stored with each data or coord array and are handled in arithmetic operations.
- **Histograms**, i.e., **bin-edge axes**, which are by 1 longer than the data extent.
- Support for non-regular or scattered data and **non-destructive binning**.

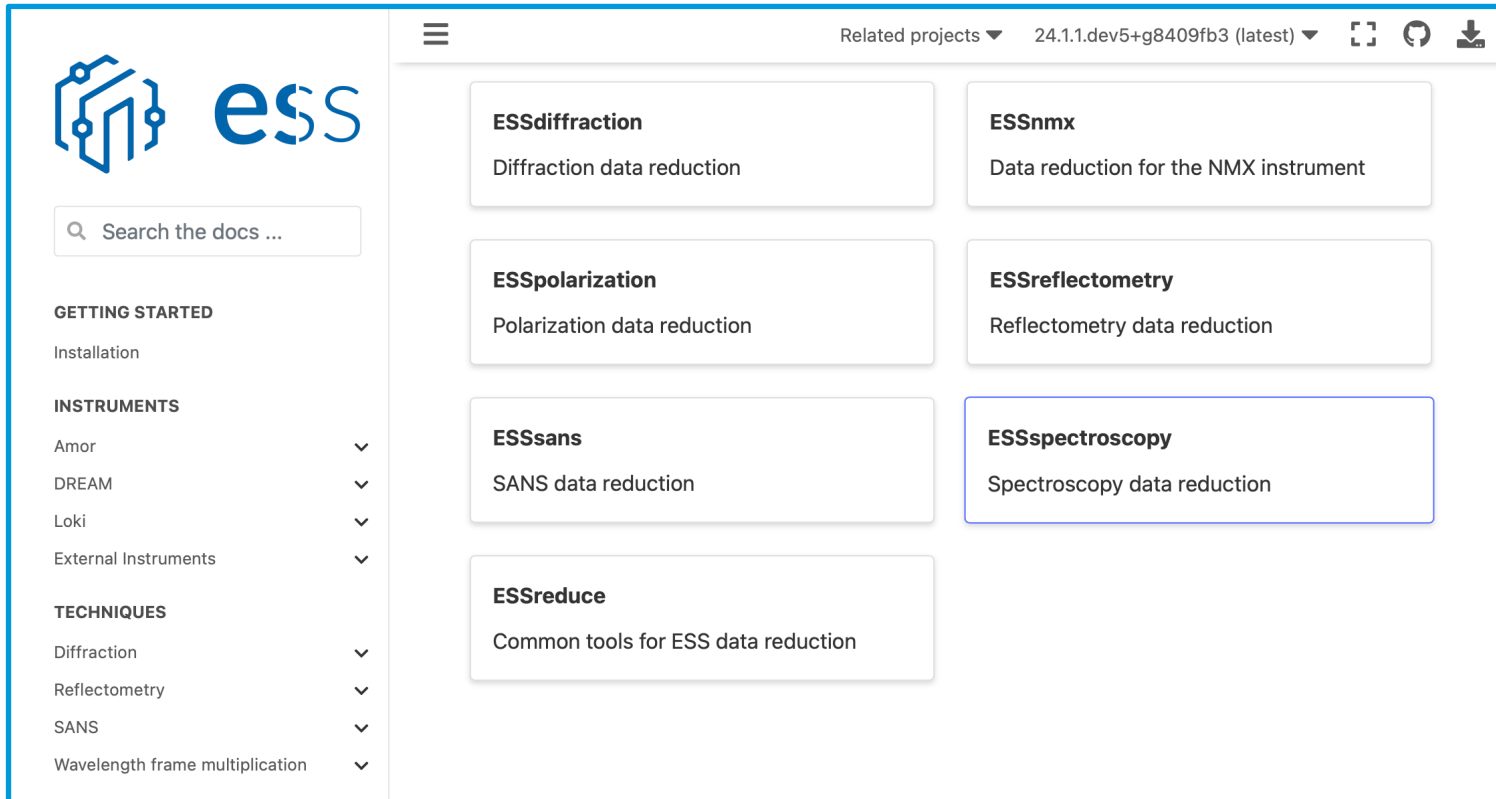




# Data reduction workflows for ESS



*On-line documentation // Getting started*



❑ [scipp.github.io/ess](https://scipp.github.io/ess)



- ❑ **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 ESSans



[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
			σ

### 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/ess/loki/2/60339-2

Filename[EmptyBeamRun]: /home/runner/.cache/ess/loki/2/60339-2

angstrom start: 2 stop: 12 nbins: 300

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

None  DirectBeam

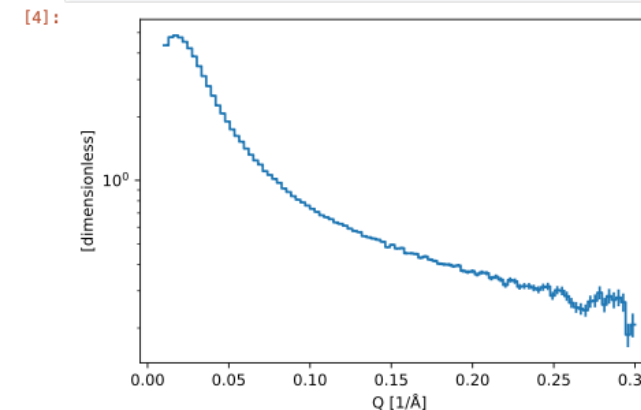
x = -0.02914868 y = -0.01816138 unit: m

## 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")
```



# Data analysis software

<https://easyscience.software>



**es easyscience**

Home Projects Features Contact

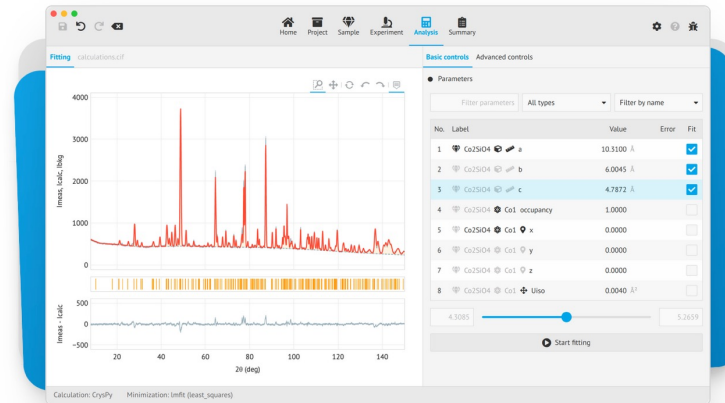


## easydiffraction

Simulation of diffraction patterns based on structural models and refinement against experimental data.

Integrates such crystallographic data analysis libraries as [CrysPy](#) and [CrysFML](#).

Visit [easydiffraction.org](http://easydiffraction.org) →



User interfaces:

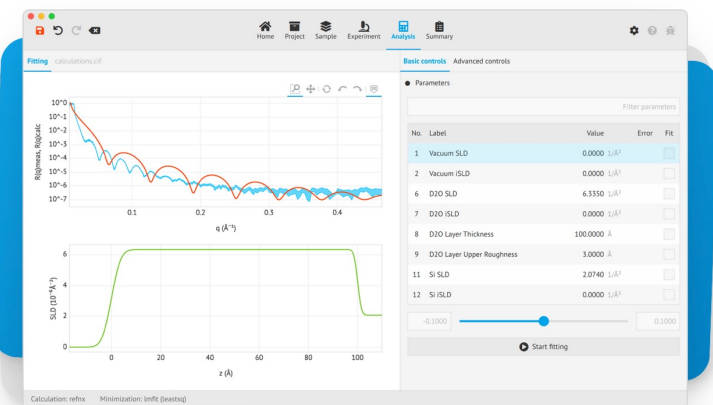
- Python
- Jupyter
- GUI



Multiple libraries (Calculators)

Other requests are:

- Data analysis for QENS
- Data analysis for TOF imaging



## easyreflectometry

Simulation of reflectometry profiles based on layered structures and refinement against experimental data.

Integrates such reflectometry data analysis libraries such as [refnx](#) and [refl1d](#).

Visit [easyreflectometry.org](http://easyreflectometry.org) →

# Updates from integration test

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



# ECDC Data pipeline

--> (DMSC) Integration testing

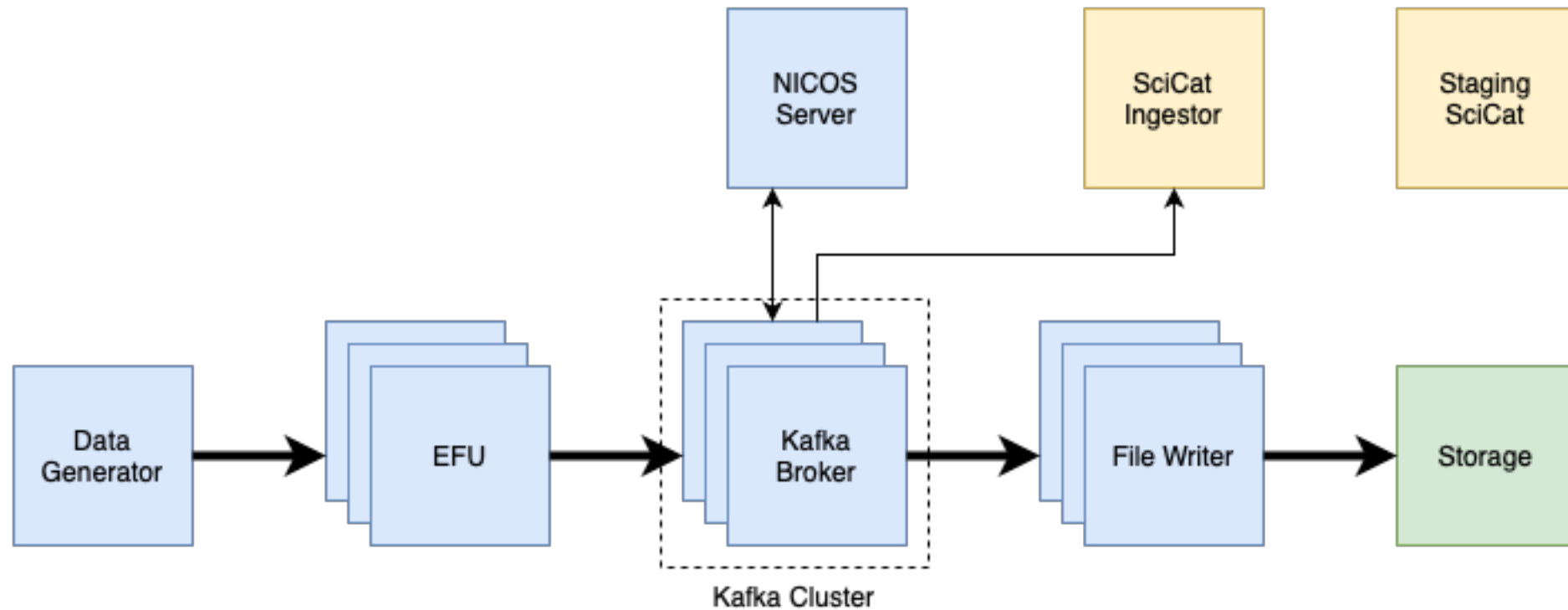
- **CODA** (COntinuous Data Acquisition) add for various instruments
  - Writes multiple datafiles per hour by **ECDC**
  - Datafiles are stored on **DST** hardware
  - Data catalogued in SciCat by **SIMS**
  - Data reduced and analysed by **DRAM**
  - Has been the driver for solving many issues
- Established monthly data surgeries
  - DRAM, ECDC, IDSs, SIMS
  - Discuss/resolve issues with data and the data pipeline

# CODA



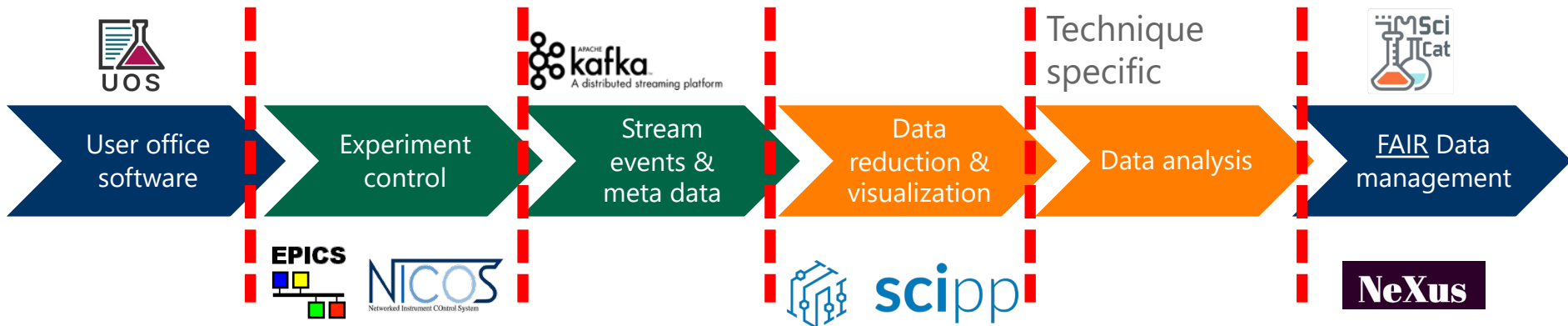
<https://confluence.esss.lu.se/display/ECDC/CODA>

- ❑ 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)

### Chexus

❌ chexus-tests: [bifrost, manual]	🔄
❌ chexus-tests: [dream, manual]	🔄
❌ chexus-tests: [loki, manual]	🔄
✅ chexus-tests: [nmx, manual]	🔄
❌ chexus-tests: [odin, manual]	🔄
✅ chexus-tests: [tbl, manual]	🔄

### Ingestor

✅ ingestor-tests: [bifrost]	🔄
✅ ingestor-tests: [dream]	🔄
✅ ingestor-tests: [loki]	🔄
✅ ingestor-tests: [nmx]	🔄
✅ ingestor-tests: [odin]	🔄
✅ ingestor-tests: [tbl]	🔄

### McStas-Scipp

❌ mcstas-scipp-tests: [bifrost]	🔄
❌ mcstas-scipp-tests: [dream]	🔄
❌ mcstas-scipp-tests: [loki]	🔄
❌ mcstas-scipp-tests: [nmx]	🔄
❌ mcstas-scipp-tests: [odin]	🔄
❌ mcstas-scipp-tests: [tbl]	🔄

### NeXus-Scipp

❌ nexusfiles-scipp-tests: [bifrost, manual]	🔄
❌ nexusfiles-scipp-tests: [dream, manual]	🔄
❌ nexusfiles-scipp-tests: [loki, manual]	🔄
❌ nexusfiles-scipp-tests: [nmx, manual]	🔄
❌ nexusfiles-scipp-tests: [odin, manual]	🔄
❌ nexusfiles-scipp-tests: [tbl, manual]	🔄

### Scipp-Analysis

❌ scipp-analysis-tests: [bifrost]	🔄
❌ scipp-analysis-tests: [dream]	🔄
✅ scipp-analysis-tests: [loki]	🔄
❌ scipp-analysis-tests: [nmx]	🔄
❌ scipp-analysis-tests: [odin]	🔄
✅ scipp-analysis-tests: [tbl]	🔄



# What are we testing?

## Chexus:

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

## File ingestor:

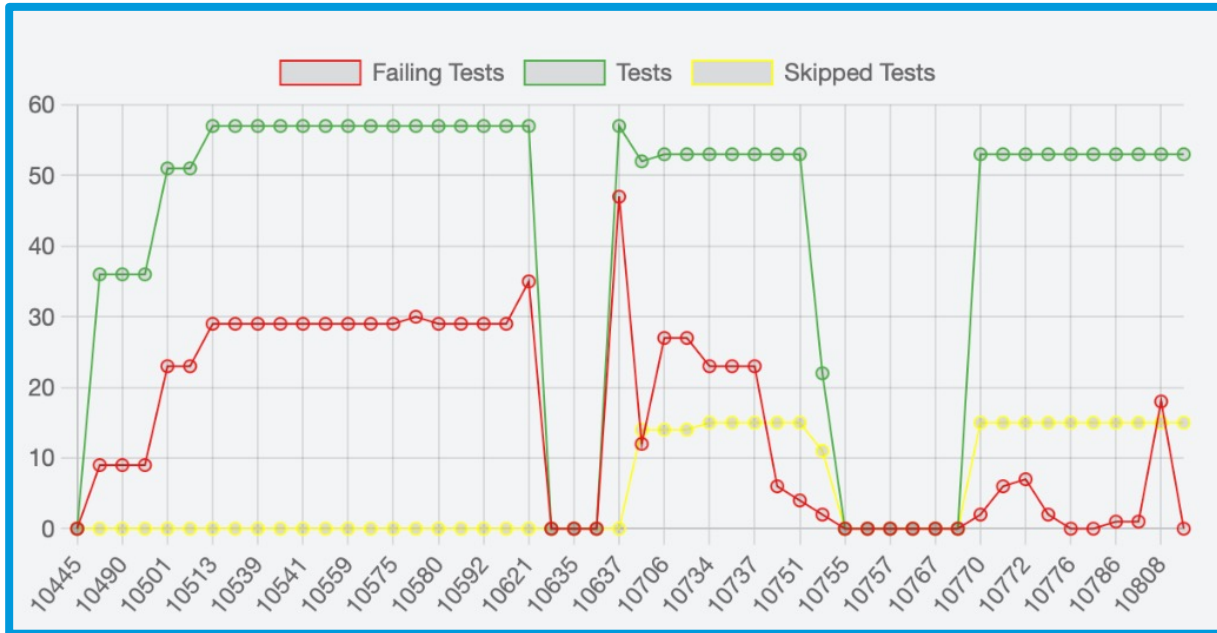
- Check that the latest file found in the CODA folder is less than 24h old
- Check that a Scicat query for the latest file in the CODA project is less than 24h old
- Check that the file returned by the Scicat query is the same as the latest file found on disk

## 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

Pipeline Needs Jobs 22 Tests 53

**setup**

- ✓ inspect-coda

**test**

- ✓ chexus-tests 6
- ✓ ingestor-tests 6
- ✓ mcstas-scipp-tests: [nmx] 6
- ✓ nexusfiles-scipp-tests 6
- ✓ scipp-analysis-tests 2

Dropdown for nexusfiles-scipp-tests:

- ✓ nexusfiles-scipp-tests: [bifrost, manual]
- ✓ nexusfiles-scipp-tests: [dream, manual]
- ✓ nexusfiles-scipp-tests: [loki, manual]
- ✓ nexusfiles-scipp-tests: [nmx, manual]
- ✓ nexusfiles-scipp-tests: [odin, manual]
- ✓ nexusfiles-scipp-tests: [tbl, manual]

# Updates from Summer School



- 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>

ESS Data Management and Software Centre Summer School

Search

Welcome

Python

- Introduction to JupyterLab
- Python basics
- Using external libraries

Proposals, DMPs and FAIR

- The ESS Viewpoint

McStas

- Simulation
- SANS exercise
- QENS exercise

Reduction with scipp

- Introduction to Scipp

ESS Data Management and Software Centre Summer School

The purpose of this online book is as a companion to the workshop materials from the ESS Data Management and Software Centre Summer School. This material can be treated in a self-guided fashion.

## Welcome

The aim of the summer school is to cover *in-silico* some the ESS data pipeline.

User Office Software → Experimental Control → Stream Events & Metadata → Data Reduction & Visualisation → Data Analysis → FAIR Data Management

UOS, kafka, EPICS, NICOS, scipp, NeXus, MSci, JCat, Technique Specific

Reduction with scipp

- Introduction to Scipp
- Coordinate transformations
- SANS data reduction
- QENS data reduction

Analysis with EasyScience

- Model-dependent analysis
- Thinking about data probabilistically
- The EasyScience framework
- Fitting data with easyscience

## Introduction to Scipp

Multi-dimensional arrays with labeled dimensions and physical units

Reduction with scipp

- Introduction to Scipp
- Coordinate transformations
- SANS data reduction
- QENS data reduction

Analysis with EasyScience

- Model-dependent analysis
- Thinking about data probabilistically
- The EasyScience framework
- Fitting data with easyscience
- Uniform priors in easyscience
- Fitting SANS data
- Fitting QENS data
- Markov chain Monte Carlo
- Bayesian Model Selection
- Bayesian analysis of SANS data
- Bayesian analysis of QENS data

Scicat

- Dataset
- Data Curation

## 3. Binned data

Scipp distinguishes **histogrammed** data from **binned** data:

- Histogrammed data refers to regular dense arrays of, e.g., floating-point values with an associated bin-edge coordinate.
- Binned data refers to the precursor of histogrammed data, i.e., each bin contains a "list" of contributing events or values. Binned data can be converted into a histogram by computing the sum over all events or values in a bin.

Binned data

Histogrammed data

count records inside bins

x	count
0.0 - 0.5	1
0.5 - 1.0	2
1.0 - 1.5	3
1.5 - 2.0	4
2.0 - 2.5	2
2.5 - 3.0	1
3.0 - 3.5	1

# 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