



# The ESS Data Management and Software Centre

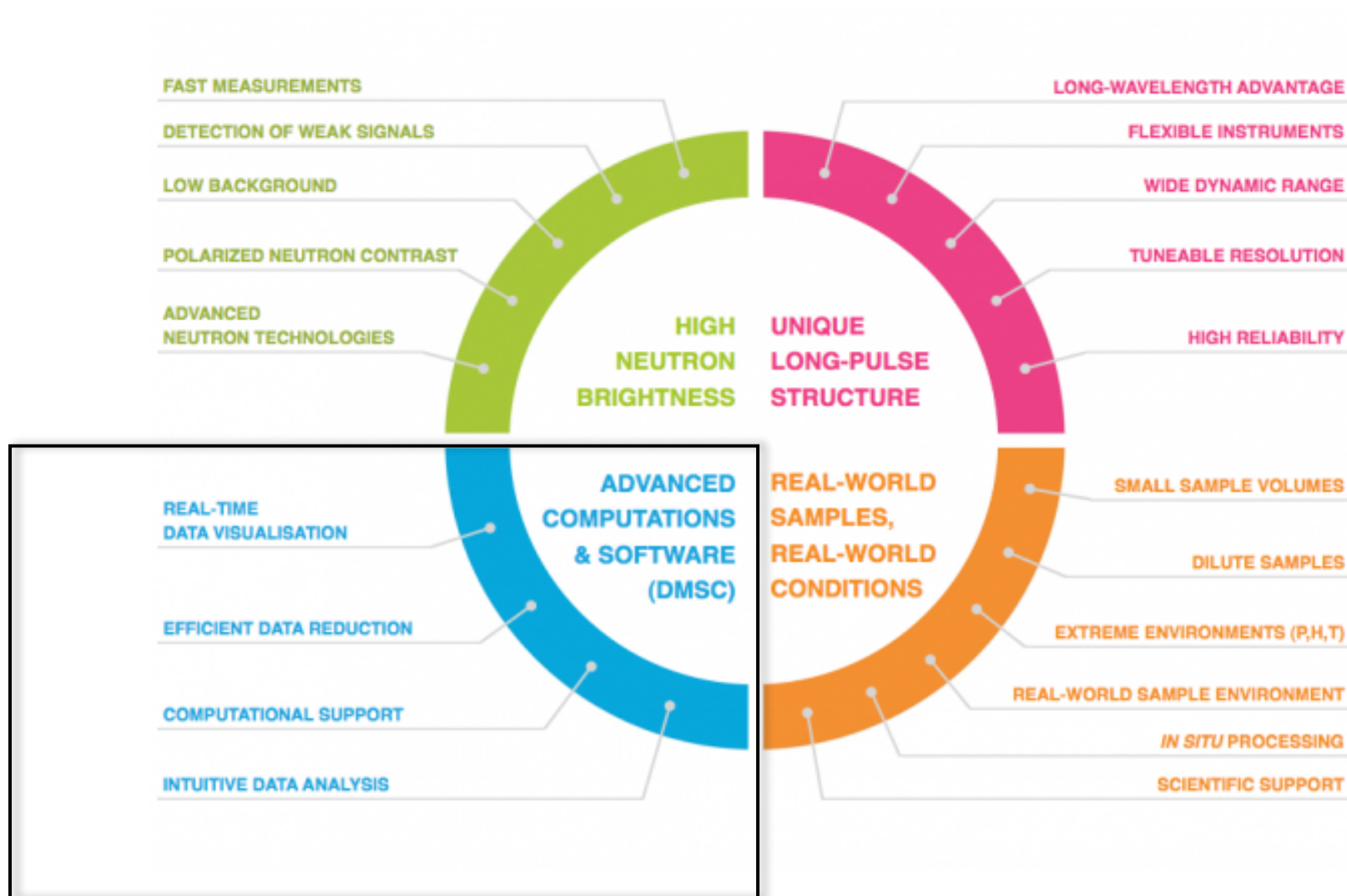
*Getting the most out of data*

FREDRIK BOLMSTEN, GROUP LEADER SCIENTIFIC INFORMATION MANAGEMENT SYSTEMS

# Technical Design Report (2013)

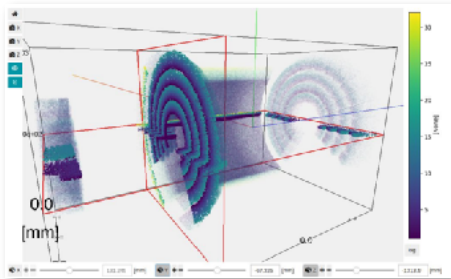
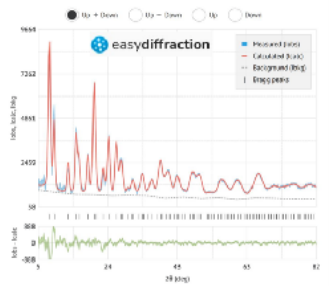
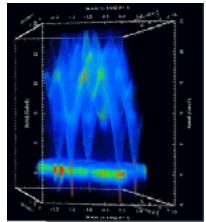
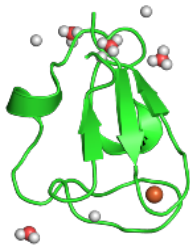
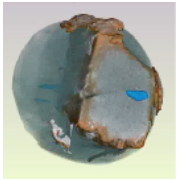
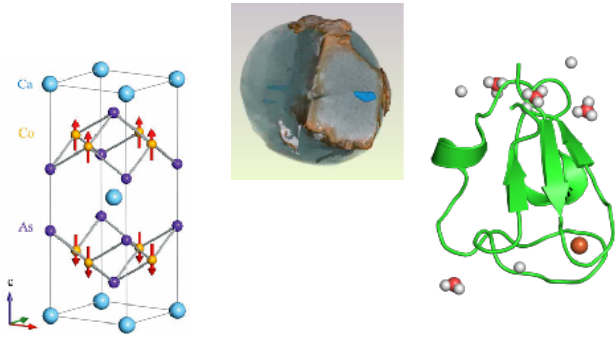


Importance of computing was emphasized already in design phase

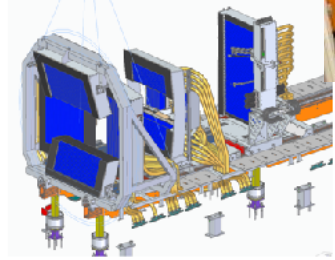
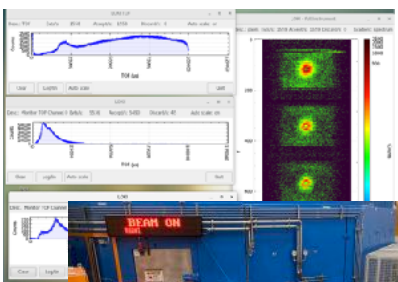
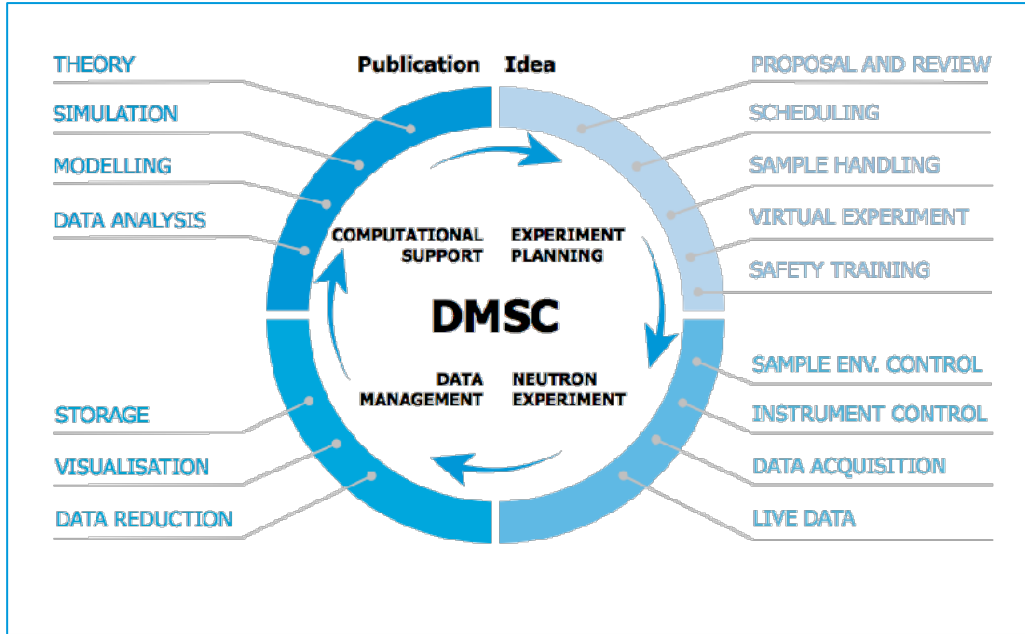


# User journey

Support user from idea to publication with scientific computing tools & services



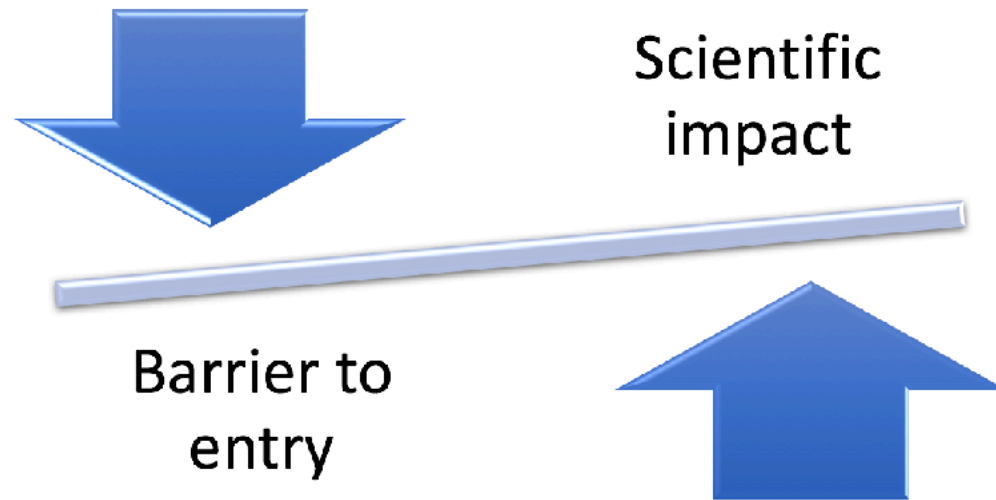
Search	Index	Resolution	Space Group	Unit Cell	Refinement	Validation
1	001	0.100	P21	a=0.356, b=0.356, c=0.356	R=0.010	Good
2	002	0.050	P21	a=0.356, b=0.356, c=0.356	R=0.010	Good
3	003	0.033	P21	a=0.356, b=0.356, c=0.356	R=0.010	Good



# DMSC objective



Minimize the time it takes to analyze and interpret experimental data



This is particular important for neutron sources due to the cost of producing neutrons

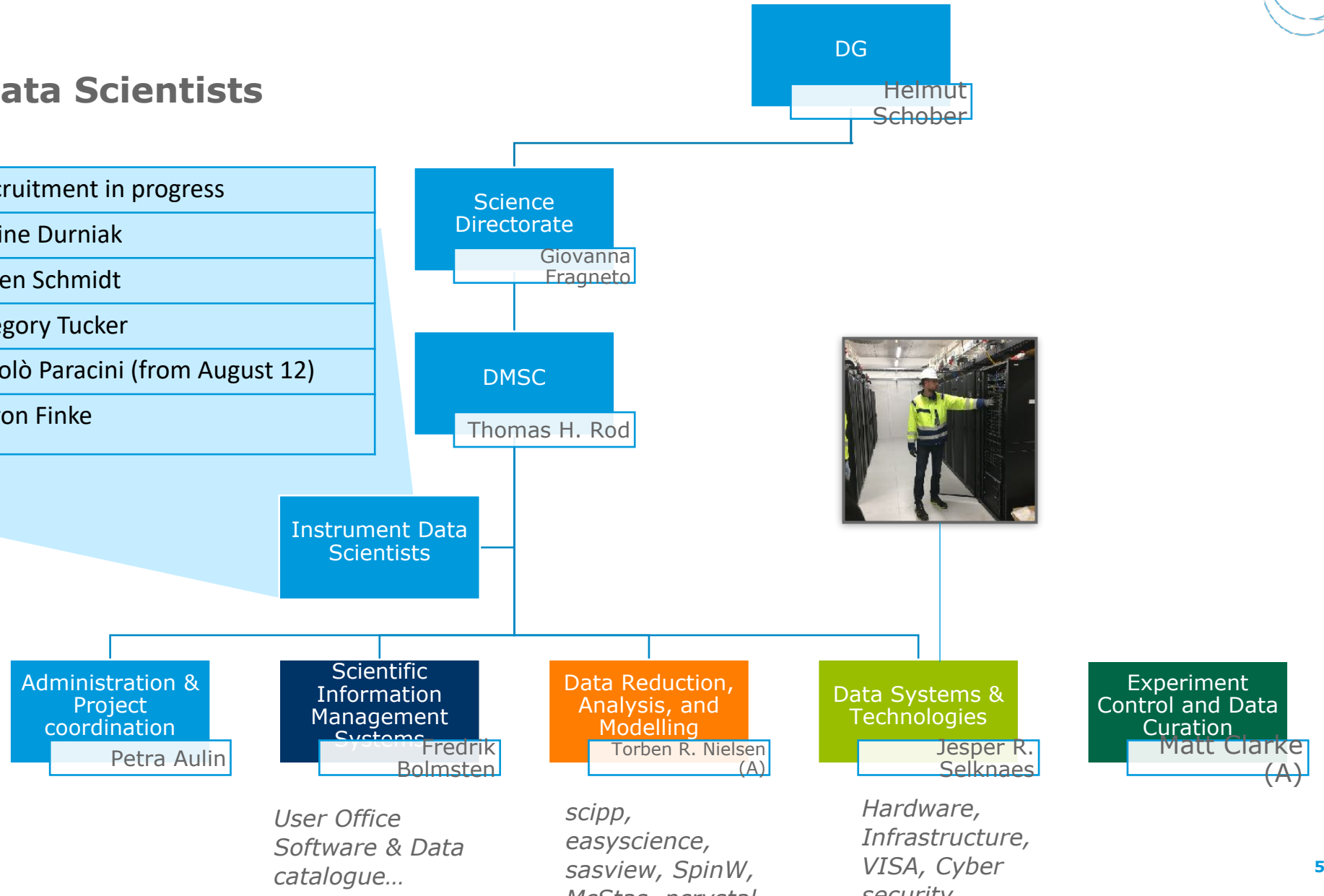
Maximise the scientific impact and success of ESS by serving the needs of both non-expert and advanced users

# DMSC organisation



## Instrument Data Scientists

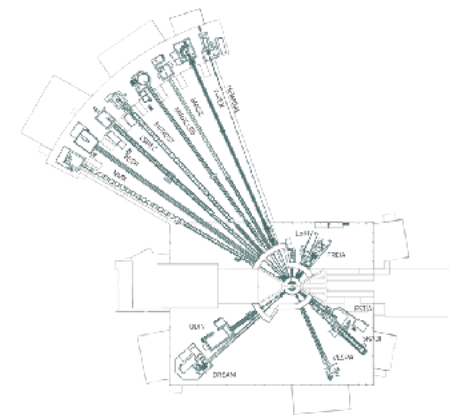
<b>LOKI (&amp; SKADI)</b>	Recruitment in progress
<b>DREAM (&amp; MAGIC)</b>	Céline Durniak
<b>ODIN (&amp; BEER)</b>	Søren Schmidt
<b>BIFROST (&amp; CSPEC)</b>	Gregory Tucker
<b>ESTIA (&amp; FREIA)</b>	Nicolò Paracini (from August 12)
<b>NMX</b>	Aaron Finke



# Integrated data pipeline

for each instrument

- Interface to instrument teams
- Customization to instruments



Data Systems & Technologies:

# User Office



User Office / Template Editor

Logged in as: fredrik.beimston@ess.eu (User Office)

Name: 2022 DEMAX template

Description: Based on DEMAX original template

Reorder topics mode

New proposal	General information	Biological deuteration	Chemical deuteration	Expression of Interest	Proposal text
proposal_base Title/Abstract/Intro-Proposers	id_general General Information	establishment_163871061114 Rbdeuteration test	establishment_163871061114 Chemical deuteration text	establishment_163871061114 Ei text	text_input_157132316388 Summary
	establishment_1572345693496 Science	slide_select_deuteration Is biological deuteration applicable?	chem_deu_enabled Is chemical deuteration applicable?	boolean_163769923695 Is biological deuteration applicable?	text_input_157132316388 Background
	selection_from_options_1572345693496 What is the primary science discipline for research with this sample?	all_select_deuteration_type Select deuteration type(s)	chem_deu_molecule_name Molecule to be deuterated (name):	has_crystallice Is crystal lattice applicable?	text_input_157132316388 Proposed Experiment
	selection_from_options_157117600052 Which is the main sub discipline of chemistry?	is_biomax E. coli cell paste (protein expression strains a g Tuner DE3, BL21 DE3)	chem_deu_amount Amount of material required (mass):		file_upload_157132316388 If you prefer to upload proposal text as a PDF please do so here. Please upload your figures (max. 4 as a single page PDF, including legends)
	selection_from_options_157117600052 Which is the main sub discipline of earth and environmental science?	boolean_1575451912001 Algal cell paste (Boryobococcus strain II)	chem_deu_amount_justification Justify the amount requested		text_input_157132316388 References
	selection_from_options_157117600052 Which is the main sub discipline of engineering?	boolean_1575451912001 Yeast cell paste (Pichia pastoris)	chem_deu_d_percentage Indicate percentage and location of deuteration		file_upload_157132316388 Additional Materials

**DYNAMIC**

Virtually nothing is hard coded. Templating tools are simple to use by non-software developers giving flexibility to changes.



Science and Technology Facilities Council



## Proposal Folder and File Structure

Proposed structure to be mounted in the same place on all relevant machines. The view of this structure should be consistent across machines regardless of location.

**TODO:** Permissions are for the embargo period. What is opened after, if subject to checks with the policy.

There will be one proposal folder per instrument per year.

The top level path for the proposal is set to:

**/ess/data/<instrument>/<year>/<proposal\_id>**

In this folder, the following sub folders will be created:

- **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.

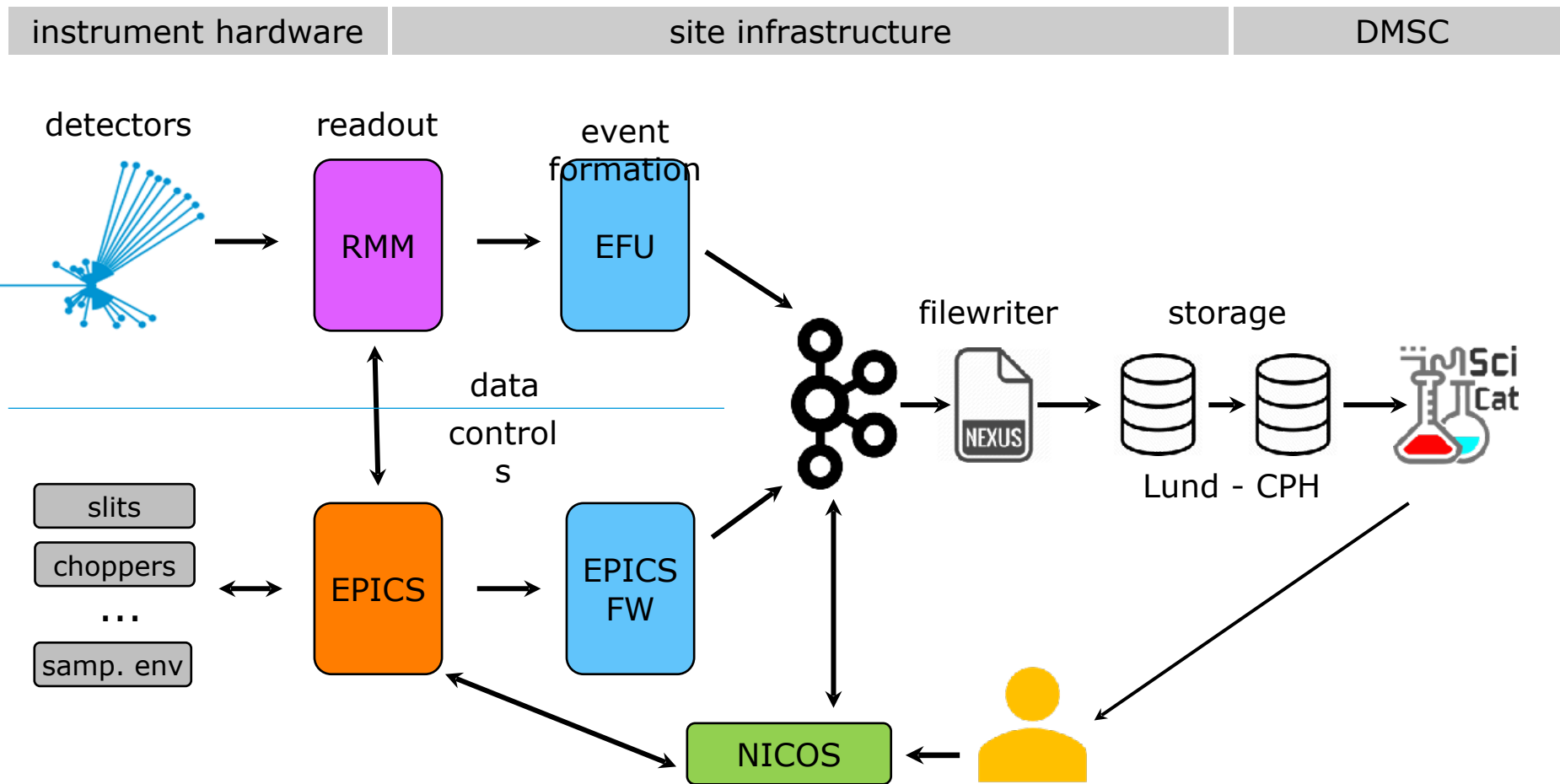
- **reduced**

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

```
[root@sftpserver2 ~]# sudo -u nexus ls -l /sftp/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
```



# Experiment Control



# NICOS

## With User Office integration



NICOS - admin at localhost:3389

Application Output Script control Tools Help

NICOS Instrument: YMIR Experiment: Experiment at YMIR Status: IDLE

Attention: connected as admin user.

Experiment

Setup

Instrument interaction

Batch file generation

History

Logs

### Proposal information

Find my proposal

Proposal ID: 471120

Proposal title: I want beamtime 3

Local contact:

	name	email	affiliation
1			

Users:

	name	email	affiliation
1	Matt Clarke		Other
2	Fredrik Bolmsten		Other

Sample Information:

		1
name	(CH3)(CH2) <sub>x</sub> SO <sub>4</sub> Na	
Formula	(CH3)(CH2) <sub>x</sub> SO <sub>4</sub> Na	
number of	1	
mass/volume	0	
density	0 g/cm <sup>3</sup>	

Notifications (one email address per line):

Do not continue scripts after fatal errors

Apply

Changes to the experiment proposal have not been applied


Discard Changes

# 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**

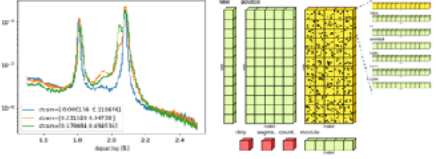
- Representations and Tables
- Plotting Overview

Related projects ▼ 23.05.0 (latest) [refresh] [download]

- scipp
- plopp
- scippnexus
- scippneutron
- ess

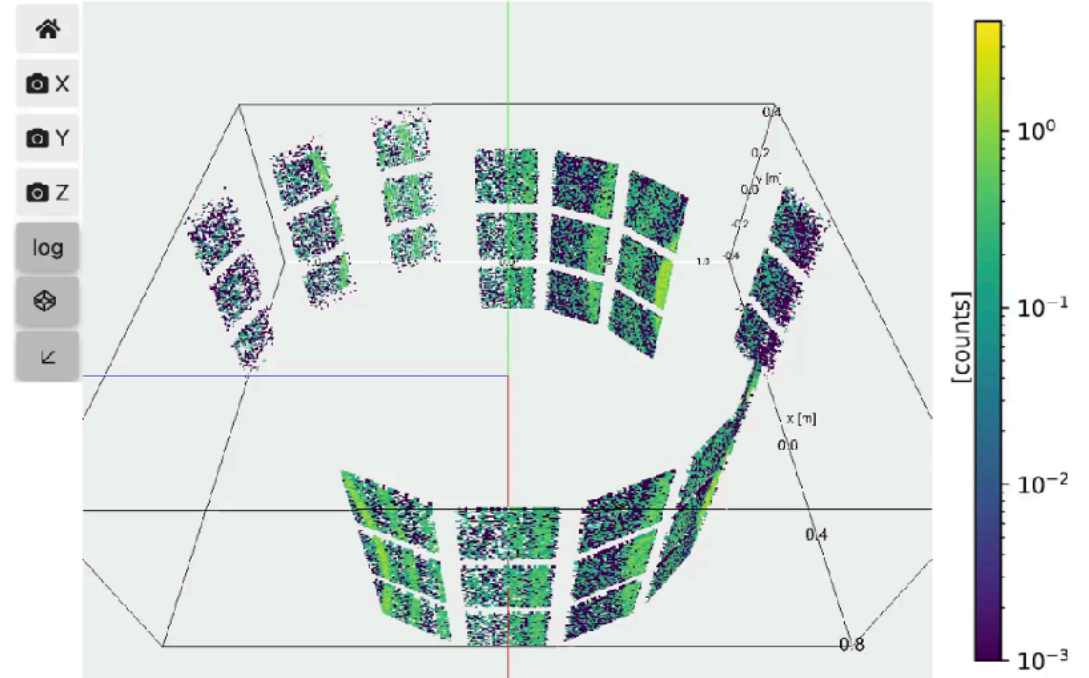
## Scipp – Multidimensional data arrays with dimensions

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 analysis software

<https://easyscience.software>



Home Projects Features Contact

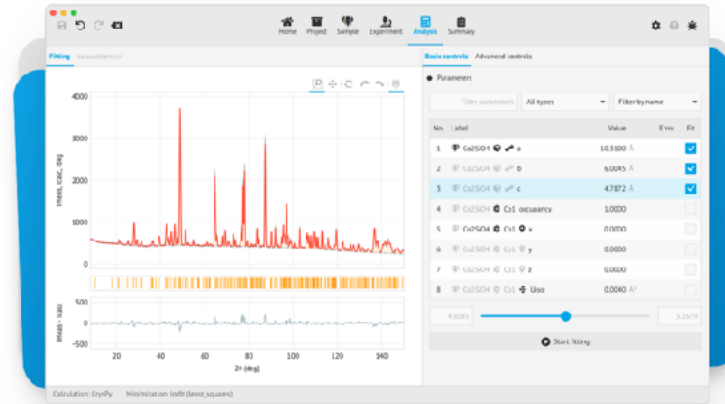


## easydiffraction

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) →

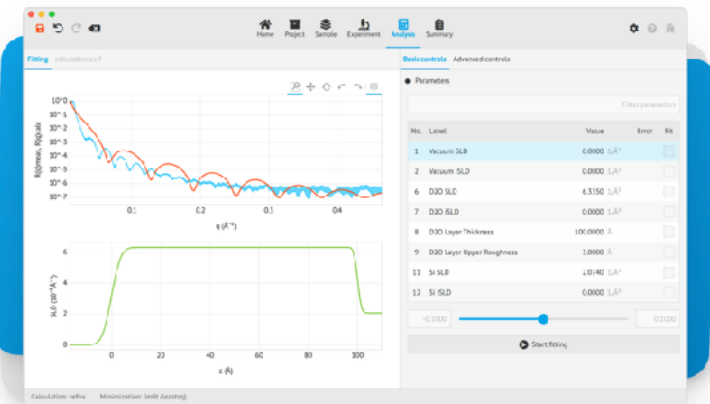


User interfaces:

- Python
- Jupyter
- GUI



Multiple libraries (Calculators)



## 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](#).

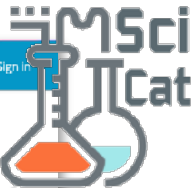
Visit [easyreflectometry.org](http://easyreflectometry.org) →

Other requests are:

- Data analysis for QENS
- Data analysis for TOF imaging

# FAIR data management

SciCat is used as data catalogue by several facilities



SciCat

Items per page: 25 1 - 25 of 2054

Search	Clear	Icon	Icon	Run No.	Size	Start Time	Type	Image	Proposal ID
PID		20.500.12259/4f8c991e-a879-4e00-9095-5bb113	DPC data		212 MB	2020-02-21 Fri 10:01	derived		
Text Search		20.500.12259/74d45720-ab24-bdc9f4c0eb30	Fe3O4		14 KB	2020-02-20 Thu 10:30	raw		
Location		20.500.12259/2d5af6ef-V20 sample data			0 B	2020-02-06 Thu 13:02	derived		
Group		20.500.12259/2511nicosLast Neutrons Liver at ILLZ.		2511	203 MB	2019-12-11 Wed 12:48	raw		YC75Z5
Type		20.500.12259/2510nicosWFM Low Res messing with chopper 1 and 2, slit2 pinhole, slit3 fully open.		2510	81 MB	2019-12-11 Wed 12:37	raw		YC75Z5
Keywords		20.500.12259/2509nicosslitscan, slit2 scanning, slit3 fully open		2509	116 MB	2019-12-11 Wed 11:54	raw		YC75Z5
Start Date - End Date		20.500.12259/2508nicosslitscan, slit2 scanning, slit3 fully open		2508	38 MB	2019-12-11 Wed 11:49	raw		YC75Z5
+ Add Condition		20.500.12259/2507nicosslitscan, slit2 scanning, slit3 fully open		2507	2 KB	2019-12-11 Wed 11:48	raw		YC75Z5
		20.500.12259/2506nicosslitscan, slit2 scanning, slit3 fully open		2506	37 MB	2019-12-11 Wed 11:47	raw		YC75Z5
		20.500.12259/2505nicosslitscan, slit2 scanning, slit3 fully open		2505	20 MB	2019-12-11 Wed 11:45	raw		
		20.500.12259/2504nicosslitscan, slit2 scanning, slit3 fully open		2504	2 KB	2019-12-11 Wed 11:42	raw		
		20.500.12259/2503nicosslitscan, slit2 scanning, slit3 fully open		2503	21 MB	2019-12-11 Wed 11:39	raw		

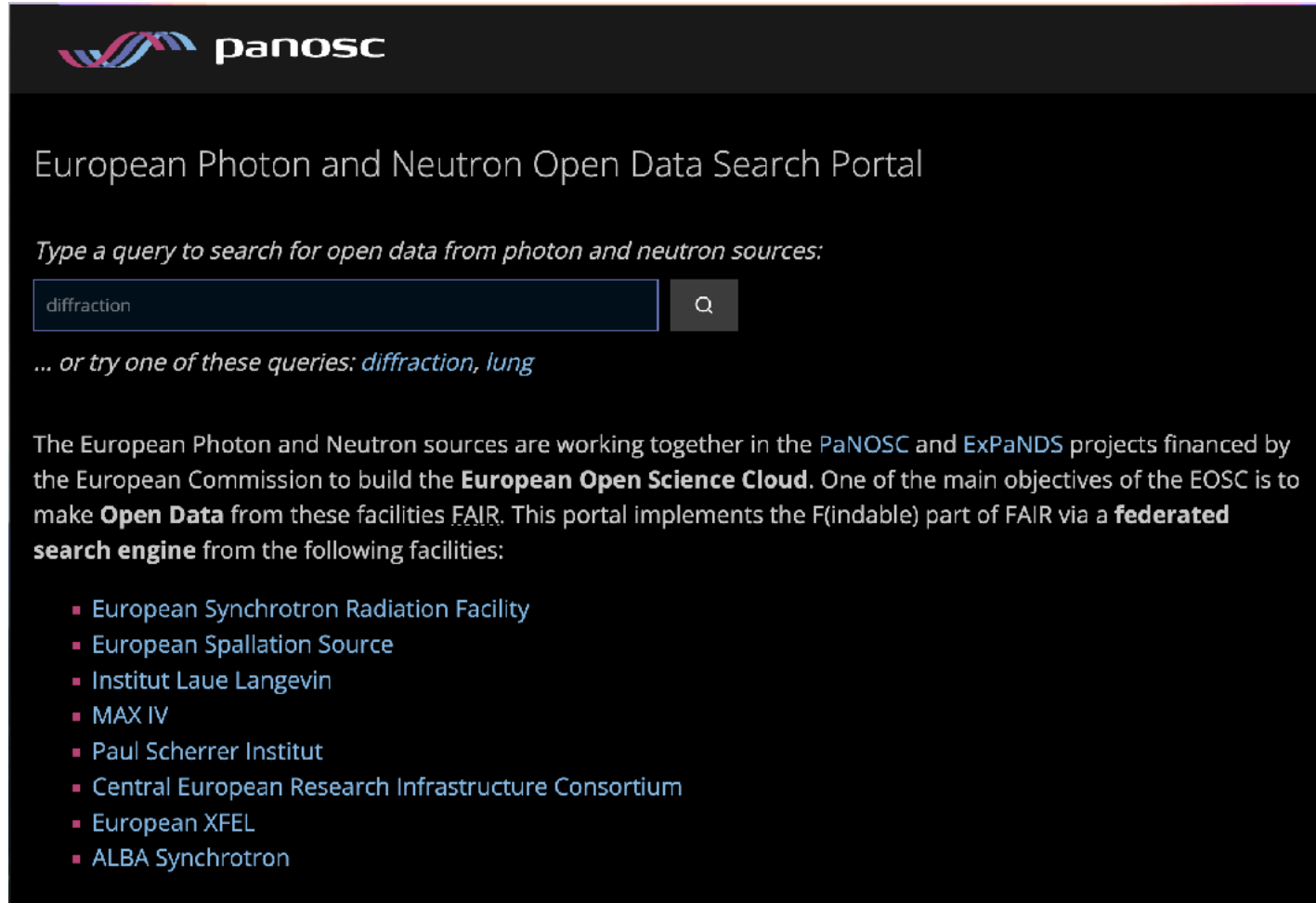
Data will be made public after 3 years



# Federated data portal

You can find and access data from multiple photon and neutron sources from one place

<https://data.panosc.eu>



The screenshot shows the panosc website interface. At the top left is the panosc logo. Below it is the title "European Photon and Neutron Open Data Search Portal". A search bar contains the text "diffraction" and a search button. Below the search bar, it says "... or try one of these queries: diffraction, lung". A paragraph of text explains the project's goals: "The European Photon and Neutron sources are working together in the PaNOSC and ExPaNDS projects financed by the European Commission to build the European Open Science Cloud. One of the main objectives of the EOSC is to make Open Data from these facilities FAIR. This portal implements the F(indable) part of FAIR via a federated search engine from the following facilities:". A list of facilities follows:

- European Synchrotron Radiation Facility
- European Spallation Source
- Institut Laue Langevin
- MAX IV
- Paul Scherrer Institut
- Central European Research Infrastructure Consortium
- European XFEL
- ALBA Synchrotron



# Current status



Organisation | Our Workplace | People | Management | Support

Science Directorate

Blog / 2024 / March / 5

ROAD TO SCIENCE

## The full ESS experiment pipeline has been demonstrated

Created by Carina Lobley, last modified by Ulrika Hammarlund on Mar 05, 2024

The Experiment Controls and Data Curation group (ECDC) and the Scientific Information Management Systems group (SIMS), along with their colleagues at the Data Management and Software Centre (DMSC), have demonstrated the full ESS experiment pipeline.

The ESS architecture for experiment control, data acquisition, and data curation requires orchestration of many software components. \*Project CODA\* (Continuous Data) has recently been completed by ECDC at ESS. Using a mixture of real data and previously recorded experiment data, continuous data collection has been run on ODIN, NMX, BIFROST, and the Test Beamline. The intention is to keep Project CODA running in preparation for commissioning activities and the user programmes.

ESS controls and data path architecture (simplified). The action starts with the user at the NICOS workstation and data ends up in SciCat.

This marks a significant milestone for the ESS experiment data pipeline – congratulations to all those involved!

Please [click here](#) for a full report of project CODA.

Science Directorate | Feature story | General

BIFROST	1	31 Oct 2024	30 Jun 2025	09 Oct 2025	YES
DREAM	1	31 Oct 2024	30 Jun 2025	09 Oct 2025	YES
LOKI	1	31 Oct 2024	30 Jun 2025	09 Oct 2025	YES
ODIN	1	01 Nov 2024	30 Sep 2025	09 Oct 2025	YES
NMX	1	16 Oct 2025	30 Sep 2025	09 Oct 2025	YES

ESTIA	2	31 Mar 2025	31 Dec 2025	same	YES
FREIA	2	01 Jul 2027	30 Jun 2027	same	NO
MAGIC*	2	30 Jun 2025	31 Mar 2027	same	NO
BEER	2	30 Sep 2025	30 Sep 2026	same	NO
SKADI	2	23 Dec 2026	30 Sep 2026	same	NO

HEIMDAL*	3	01 Jul 2027	30 Jun 2027	same	NO
CSPEC*	3	20 Dec 2024	30 Jun 2027	same	NO
MIRACLES*	3	02 Jul 2027	30 Jun 2027	same	NO
TREX	3	02 Jul 2027	30 Sep 2027	same	NO
VESPA	3	01 Jul 2027	30 Sep 2027	same	NO



**Questions?**