

## DMSC activities related to DREAM for First Science

**Author** Céline Durniak, Instrument Data Scientist (IDS) for Diffraction (DREAM, MAGIC)

**Date:** April 2024

### Table of Contents

1	Introduction	1
2	Data processing pipeline	1
2.1	Instrument simulations	2
2.2	NeXus files	2
2.3	Data reduction in Scipp	3
2.4	Data analysis with EasyDiffraction	5
2.5	Data archiving	6
3	Tools for user experience	7
3.1	VISA	7
3.2	Scichat	7
3.3	User office software	7
3.4	Preparing the community	8
4	Project management	8
5	Development for cold and hot commissioning	8
6	Challenges	9
7	References	9

## 1 Introduction

This document briefly describes the status of the development activities by the Data Management and Software Centre (DMSC) related to Diffraction. It is split into two parts; the first one focussing on the data processing pipeline and the second one on user experience. At this stage, these activities are mostly focused on First Science for the DREAM instrument and Powder Diffraction. The work reported in this document has involved close collaborations between DMSC, the instrument team, the ESS Detector group and other institutes.

## 2 Data processing pipeline

Figure 1 shows the data processing workflow used in this section to report on the development of the related software or frameworks:

- instrument simulations with McStas [1] and GEANT4 [2]
- data reduction with Scipp [3]
- data analysis with EasyDiffraction [4]
- data archiving with NeXus files [5] and Scicat [6]



Figure 1: The data processing pipeline. Each chevron corresponds to a class of services needed for processing data. The software or frameworks described in Section 2 are listed above the chevrons.

**Note:** In order to test the processing pipeline, we also use data from existing instruments from other facilities, like POWGEN at SNS.

## 2.1 Instrument simulations

Because of the 3D geometry of DREAM detectors, we have to model them in GEANT4 [2]. The procedure to generate simulated data for DREAM is first to use McStas for the processes from the moderator to the sample, and then to feed this output in a GEANT4 model of the detectors. This GEANT4 model was initially developed by Irina Stefanescu from the ESS detector group. All detectors have now been modelled in their latest designs (see Figure 2 for Day 1 configuration).

The model of DREAM in the high-intensity and high-resolution configurations has been updated to run in the latest version of McStas. It is under version-control (gitlab repository).

A library of samples for McStas simulations has been compiled: these are standard samples for Powder Diffraction, Single Crystal Diffraction, PDF, SANS or samples with texture and incommensurate properties.

Future plans for instrument simulations include the addition of components related to polarisation by Wai Tung Lee (Hal) *et al.*

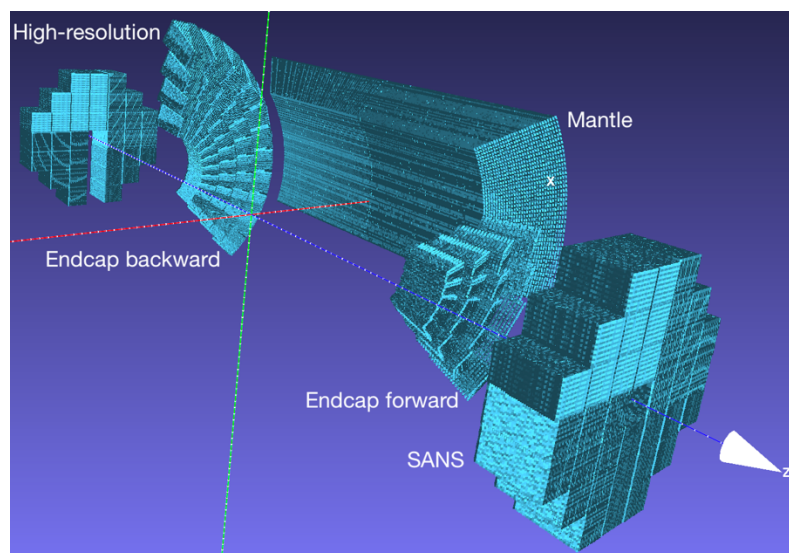


Figure 2: 3D view of Day 1 configuration of DREAM detectors (done in Meshlab).

## 2.2 NeXus files

The ESS NeXus file will contain detected neutrons as events by recording the position of the voxel and a time stamp at which the detection occurred.

It also contains the instrument geometry in addition to other metadata (proposal, users, varying parameters, monitors' counts...). Figure 3 shows the existing tree structure of the NeXus file for

DREAM. Neither McStas nor GEANT4 offer an output with this format. Therefore, an additional step is required to add simulated data to DREAM Nexus files.

**Note:** The final design for the NeXus file still has to be decided. The level of details of metadata to be stored in the file needs to be a compromise between a sufficient amount of information for data processing and some possible computing limits (bandwidths, data rate, file size...).



Figure 3: Tree structure of DREAM NeXus file showing the storage of different components of the instrument: slits, monitors, detectors, choppers. Note that the structure is displayed in 3 columns because of size constraints; but it should be read as a single entity. The detectors are split by banks: mantle, endcap backward and forward, high-resolution and sans detectors. The 2<sup>nd</sup> column details the information stored for the mantle detector.

## 2.3 Data reduction in Scipp

### 2.3.1 Description of Scipp & co

Figure 4 shows the software stack for reduction at ESS defined as a series of levels, with Scipp sitting at the bottom of the stack, providing generic and scientific functionality, such as labelled data arrays and uncertainty propagation. The ScippNeutron module builds on Scipp to provide neutron scattering-specific, facility-independent functionality, including unit conversions. ScippNeXus provides a link between Scipp and the NeXus data format. Plopp is the visualization library, offering different ways to plot in 1, 2 or 3 dimensions with different and customizable levels of interactivity. Above these packages, are different ess instrument- and technique-specific submodules. Sciline helps building scientific pipelines. Jupyter notebooks and Python scripts are the interfaces to reduce data with the Scipp stack of frameworks.

**Note:** In addition to the frameworks listed in Figure 4, the Scipp team has also developed [beamtime](#) for live data reduction visualization framework for ESS [7].

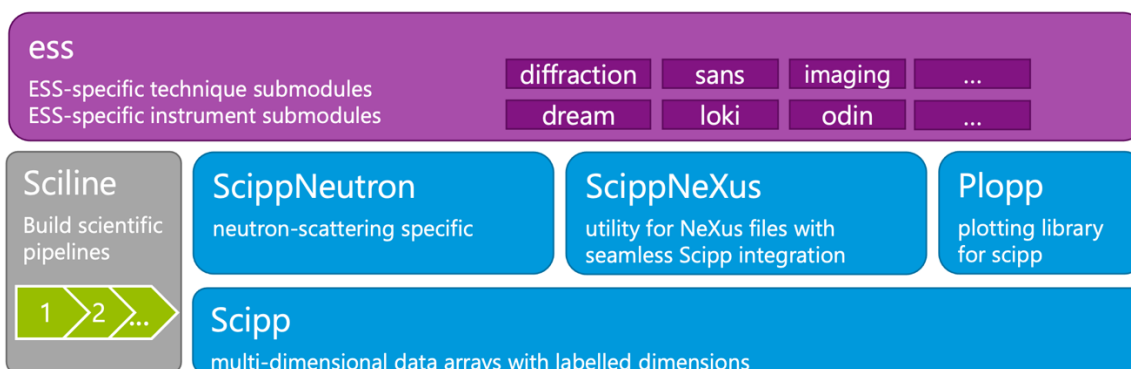


Figure 4: The software stack for reduction at ESS.

### 2.3.2 Reduction of Powder Diffraction data

As listed in Table 1, users of Scipp can load DREAM NeXus or ASCII files, convert to different units, such as time-of-flight, wavelength, d-spacing or momentum transfer, mask or group detectors and slice, bin, sum, divide or subtract datasets in order to go from events to a 1D plot that could be used by analysis software. At all stages of the reduction, depending on its structure, data can be displayed as 1 or 2D plots or in 3D via the instrument view.

Loading	Event NeXus file, tabular data → data structures with labelled dimensions
Unit transformation	ToF ↔ $\lambda$ ↔ d-spacing ↔ Q & user-defined
Plotting	1D/2D & 3D instrument view, built-in & user-defined widgets
Selecting	Non-destructive masking, filtering, cropping
Grouping	Binning, histogramming, merging
Correcting	Normalization
Saving	XYE, CIF files

Table 1: Summary of reduction steps for Powder Diffraction available in Scipp.

#### Example: Workflow with Sciline

Figure 5 a) shows the task diagram for computing the final result of a simple data reduction workflow. The pipeline can be used to compute results or to visualize the structure of the workflow: b) and c) shows 2D plot of an intermediate dataset and a 1D plot of the result, respectively.

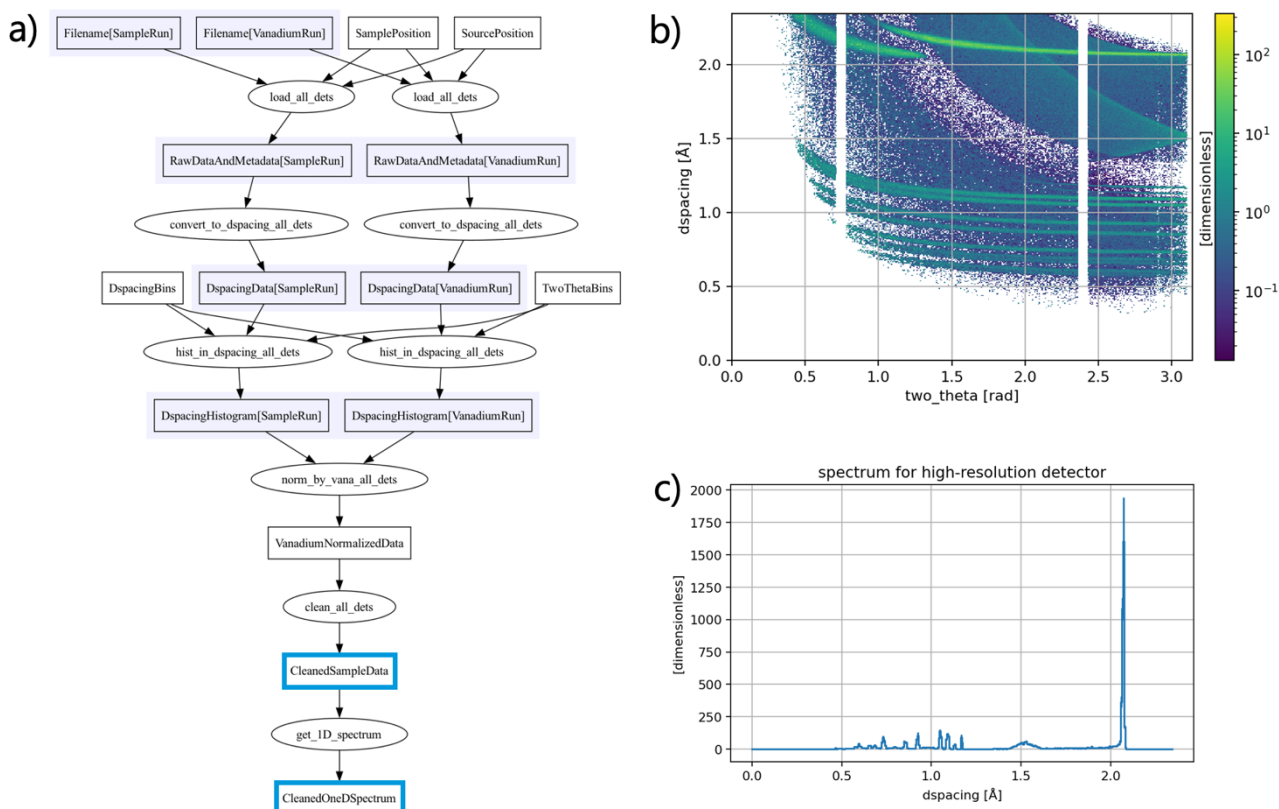


Figure 5: a) Graph of the workflow for DREAM (data are marked by rectangles and operations on the data by ovals) b) plot of an intermediate result c) 1D plot of final result for the high-resolution detector. Data plotted in b) and c) are highlighted by blue rectangles in the workflow (a).

#### Notes:

- As input files of this workflow, we used ASCII files from McStas/GEANT4 simulations for DREAM because of the absence of interface with GEANT4 and the non-final design of the ESS NeXus files. Any required metadata were coded by hand in the workflow.
- There is another example of a workflow for Powder Diffraction using POWGEN data and Sciline in the online documentation of essdiffraction [8].

Some of the future plans for reduction of Powder diffraction are to implement Wavelength Frame Multiplication stitching and to add mask to the workflow.

### 2.3.3 Other scattering techniques

For other scattering techniques covered by DREAM, we will adapt what has been done for other ESS instruments: for example, LOKI in the case of SANS.

#### Work in progress

- Implementation of a reduction workflow for (SANS) polarised data by Annika Stellhorn (post-doc in SKADI team) and the Scipp team
- Investigation of existing solutions for reduction of Single Crystal data (DAWN used by NMX@ESS, Mantid by TOPAZ@SNS)

## 2.4 Data analysis with EasyDiffraction

These last few months, the EasyDiffraction developers had focussed on improving its performance with a limited number of features (see Table 2). Now ToF support has to be reimplemented in this faster version. For crystallographic libraries, they still collaborate with the Institut Laue Langevin on CrysFML [9], CrysPy [10].

Number of refined parameters	Computing time version 0.8.5-beta → version 0.9.0-alpha9	Factor of improvement
12	16s → 3s	x 3
21	44s → 3s	x 15
34	7m20s → 12s	x 36

Table 2: Comparison of performance between 2 versions of EasyDiffraction as function of the number of refined parameters.

PDF data analysis with EasyDiffraction and pddfitt2 [11] library has been investigated in Jupyter notebooks (not available in the GUI yet).

Figure 6 shows a), b) the Graphical User Interface at different stages of the analysis and c) the layout of a Jupyter notebook using EasyDiffraction as a module.



Figure 6: Screenshots of a) and b) EasyDiffraction Graphical User Interface (GUI) of the Model and Analysis panels, c) Jupyter notebook using EasyDiffraction as an imported library.

## 2.5 Data archiving

Datasets from POWGEN and NOMAD @ SNS have been stored on Scicat [6] as test data for the interfaces with Scicat in the data processing pipeline. Figure 7 shows the table of POWGEN test data on the Scicat web interface.

pid	Name	Size	Start Time	Type	Image	Proposal ID
20.500.12269/2b0d9e7f-a98f-462e-be0c-4ecf5a89840b	2020B Calibration Empty PAC Instrument HR July13rd	672 MiB	2023-02-21 Tue 14:07	raw		276591
20.500.12269/317cb406-29aa-49f3-b20f-e55893d37b56	2021B Calibration V rod in PAC HR July13rd	4 GiB	2023-02-21 Tue 11:50	raw		276591
20.500.12269/938a6ce1-1a23-4b9b-d0f1bbfba20	2021B Calibration Si HR July13rd	975 MiB	2023-02-21 Tue 13:46	raw		276591
20.500.12269/95abbf62-0c40-44f5-b6c9e9d628bd	2021B Long Calibration diamond HR on July13rd	12 GiB	2023-02-21 Tue 14:35	raw		276591
20.500.12269/cf6fe1cb-52ba-439f-8ad4-319e51c7	MT PAC06 new at 300K CWL 0.8A	911 MiB	2023-02-21 Tue 14:32	raw		276591

Figure 7: Screenshot of Scicat webpage showing POWGEN data stored in the catalogue.

There are also on-going discussions with the data curation scientist to list searchable metadata relevant for Diffraction.

### 3 Tools for user experience

#### 3.1 VISA

DMSC is adapting Virtual Infrastructure for Scientific Analysis (VISA), initially developed at the Institut Laue Langevin, to provide a standard user interface to access instrument control, Scicat and live reduction during beamtime as well as for post-beamtime data processing (reduction, analysis). With VISA, one can create compute instances to access a remote desktop or JupyterLab to process experimental data without having to neither install software nor download large files as shown in Figure 8.

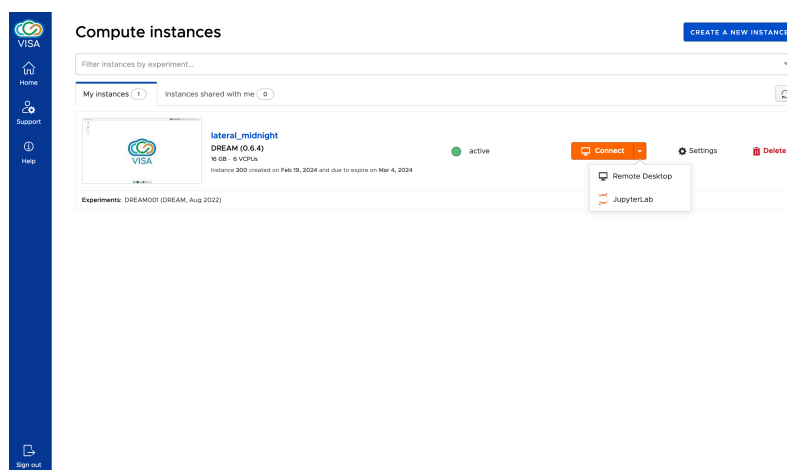


Figure 8: Example of a DREAM instance on VISA showing possible access to a remote desktop or JupyterLab.

#### 3.2 Scichat

SciChat is the chat tool developed by DMSC to allow team members to interact during beamtime. Users can expect to have a chatroom for their proposal and see selected actions from NiCOS (see Figure 9).

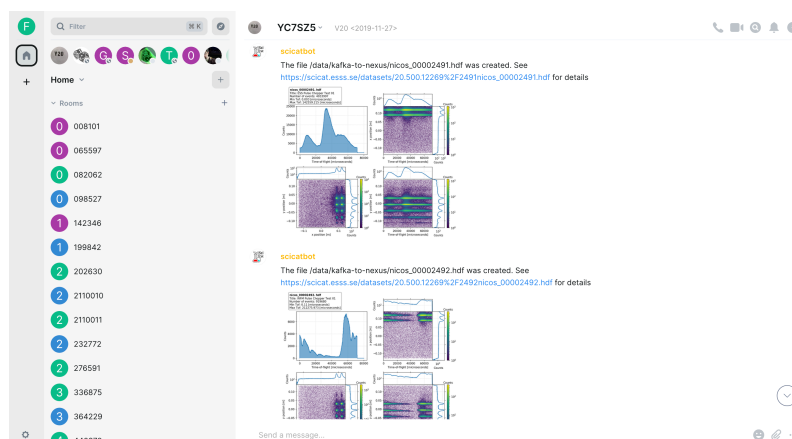


Figure 9: Automated graphs posted to Scichat during an experiment.

#### 3.3 User office software

The User Office is a web platform for proposals, handling admission, evaluation, administration and scheduling. Figure 10 shows the typical ESS user journey from the "registration" in order to submit a proposal (creation of an account, access to webpages, handling of personal data) to "wrap-up and

harvesting” (feedback form, access to experimental data, refund of expenses, experiment report, publications).



Figure 10: The ESS user journey.

Figure 11 shows the website, which is already available to access the deuteration & macromolecular crystallization platform (DEMAX).

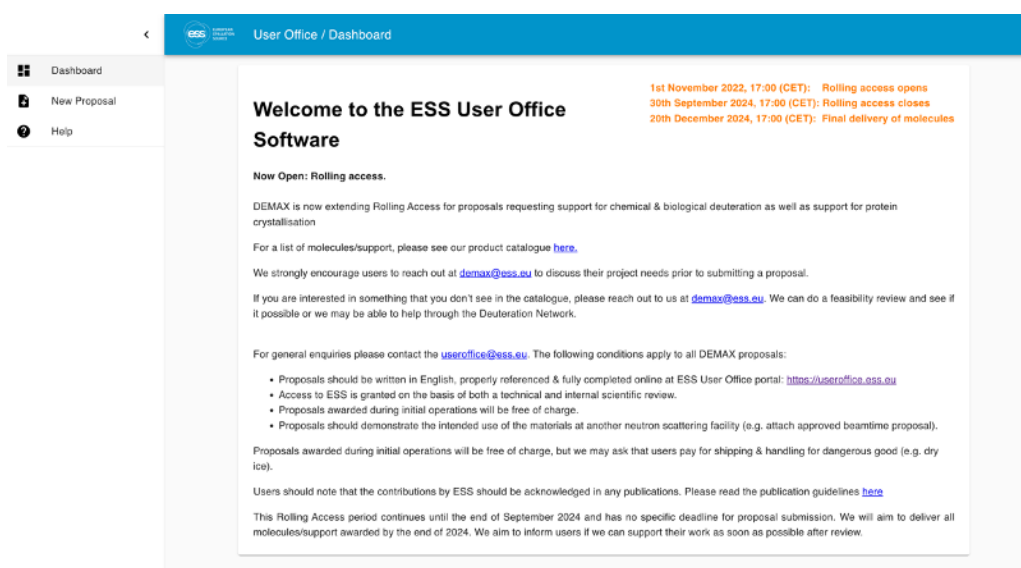


Figure 11: Webpage of the ESS User office software.

### 3.4 Preparing the community

Last year, DSMC hosted its first DMSC Summer School (4-8 September 2023) [12]. The programme covered a tutorial on Python programming and an in-depth introduction to the data pipeline for two scattering techniques (SANS and QENS) covering experiment simulation (McStas), data reduction (Scipp), analysis (EasyScience), and cataloguing (Scicat) with a focus on findable, accessible, interoperable, and reusable (FAIR) data practices throughout.

A second edition is in preparation for September 2024.

## 4 Project management

In order to plan and track the development of the software and frameworks required for Cold Commissioning, Hot Commissioning and First Science for DREAM, the required tasks related to instrument simulation, data reduction, analysis, Scicat, Scicat, User Office, NeXus, instrument control and VISA have been compiled into a “DMSC-DREAM” JIRA dashboard.

## 5 Development for cold and hot commissioning



Earlier this year, the ESS Detector group tested some of DREAM detectors at Utgård. The data collected using cosmic rays will be used to develop and benchmark plotting features for commissioning.

## 6 Challenges

- The relocation of DMSC to the campus of the Danish Technical University and the related restructuration (Experiment Control Data Curation group) make the retention of staff and of the associated knowledge uncertain. Moreover, ESS staff based in Lund is even less likely to visit DMSC on a regular basis
- The future role of IDS might also change: the time to be spent in Copenhagen and Lund is unclear. The solution of remote support during user operation might not be ideal to collaborate with users
- The communication between the teams involved in a project could be improved (ESS, DMSC, external collaborators)
- The small amount of test data from DREAM's detectors (only those collected at Utgård earlier this year) makes the development of adequate tools for commissioning more uncertain

## 7 References

- [1] P. Willendrup, et al., McStas: Past, present and future, *Journal of Neutron Research* **17**, 35 (2014).
- [2] J. Allison et al., Recent developments in GEANT4, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835**, 186 (2016).
- [3] S. Heybrock et al., Scipp: Scientific data handling with labeled multi-dimensional arrays for C++ and Python, *Journal of Neutron Research* **22**, 169 (2020).
- [4] easyDiffraction, <https://easydiffraction.org/>.
- [5] M. Könnecke et al., The NeXus data format, *J. Appl. Cryst.* **48**, 301 (2015)
- [6] Scicat <https://scicat.ess.eu/>.
- [7] Beamline <https://scipp.github.io/beamline/>.
- [8] Reduction workflow for POWGEN with Sciline, [https://scipp.github.io/essdiffraction/examples/POWGEN\\_data\\_reduction.html](https://scipp.github.io/essdiffraction/examples/POWGEN_data_reduction.html).
- [9] CrysFML <https://code.ill.fr/scientific-software/crysfml>.
- [10] cryspy, <https://github.com/ikibalin/cryspy>.
- [11] Pdffit2, <https://github.com/diffpy/diffpy.pdffit2>.
- [12] First DMSC Summer School, September 2023. <https://indico.ess.lu.se/event/3267/page/648-programme>.