

# Report for the DMSC STAP

DMSC Management Team, April 2023

## Versions

Version	Date	Comment
2023-04-19	2023-04-19	- First version shared with STAP
2023-04-21-1	2023-04-21	- Removed all the reference stuff from the IDS job description - Added concerns to Head and IDS sections - Minor changes
2023-04-23-1	2023-04-23	- Text about summer school - Elaborating on SANS and refl IDS role - PaNOSC section added - Extended section about concerns for IDSs - Extended section about concerns for the Head of DMSC - Fixed Figure numbers - Migration of staff to DMSC - Removed section ' Feedback to STAP' because there is a separate report on that.
2023-04-23-2	2023-04-23	- Fixed figures, captions, and figure numbers - Fixed font size
2023-04-25-1	2023-04-25	- Missing text related to IDS job description has been added

## Table of Contents

**Versions** .....1

**Table of Contents**.....1

**Charge**.....2

**Report from Acting Head of DMSC**.....3

    Organization and staffing.....3

    Relocation .....4

    Vision and mission .....4

Dissemination.....	5
PaNOSC .....	6
Other international collaborations/ roles.....	7
Concerns.....	8
<b>Administration and Project Coordination.....</b>	<b>8</b>
Staffing .....	8
Office relocation .....	8
<b>Project management, communication and milestones.....</b>	<b>9</b>
<b>Data Systems and Technologies (DST).....</b>	<b>11</b>
Staffing .....	11
Achievements since last STAP.....	13
Status on the hardware installations .....	13
Plans for 2023.....	15
PaNOSC & HighnESS .....	17
Participation in synergy groups on IT systems.....	18
Concerns.....	19
<b>Scientific Web Applications (SWAP).....</b>	<b>19</b>
Staffing .....	19
Achievements since last STAP.....	20
Plans for 2023 and Milestones .....	21
Concerns.....	22
<b>Data Reduction, Analysis, and Modelling.....</b>	<b>22</b>
Staffing .....	22
Activities since last STAP .....	22
Data analysis.....	24
Data modelling .....	25
Plans for 2023 and Milestones .....	26
Concerns.....	27
<b>Instrument Data Scientists .....</b>	<b>28</b>
Staffing .....	28
Concerns.....	29
Job description.....	29

## Charge

The STAP will be presented with the usual updates from the Acting Head of DMSC and the DMSC (incl. ECDC) group and team leads. Moreover, for this STAP meeting there will also be

updates from the Detector Group Leader and the Instrument Data Scientists as requested by the STAP. The STAP are invited to take note and comment on these updates.

The vision and mission of the DMSC are being updated triggered by the fact that the DMSC will move to other premises at the end of the year or early next year and a general perception that there seems to be a poor understanding of what the DMSC is and does amongst some of its stakeholders. The vision, mission and status of the DMSC is discussed in a "[DMSC—Status and Vision](#)", which is an open and dynamic document that will be used to form the basis for an internal discussion across the ESS, incl. the DMSC, to get to a uniform understanding of what the DMSC is. For the time being everybody at the DMSC has been invited to comment and contribute to the document, but apart from that a systematic approach to discuss its content has not yet been implemented. The STAP are invited to read this document or at least its summary. However, the vision and mission will be presented at the STAP meeting as well as at the subsequent SAC meeting. The STAP and SAC are invited to advise.

## Report from Acting Head of DMSC

### Organization and staffing

The DMSC organization or management situation has not changed since the last STAP meeting and will therefore not be further discussed in this report. However, Giovanna Fragneto has been appointed as Science Director, and the (acting) Head of the DMSC now reports to her.

The Operation and Infrastructure Directorate has been established and the IT Division has moved from the Project Support & Administration Directorate to this new directorate. The Infrastructure Group has moved from the Integrated Control System Division to this new IT division. This indirectly affects DMSC which has a large interface to IT and the Infrastructure Group. It is particularly important that ECDC, SWAP, and the DST have a good working relationship with this division. The Heads of IT and DMSC are looking into ways to make the collaboration more efficient across the Sound.

The Director for Project Support and Administration has stepped down, and ESS is searching for a replacement.

The Associate Director for In Kind has accepted a position as Director for the Advanced Light Source, and will consequently stop working for the ESS.

Project Communication has improved across ESS and monthly updates are now given by the Project Director and the Technical Director. DMSC now also, for the first time, has a single milestone being tracked on the monthly updated ESS-wide dashboard.

One ESS staff member has been migrated from the Detector Group to DRAM. This will strengthen the simulation activities in DRAM, but also means that DRAM as a team now is responsible for this person's simulation responsibilities in the Detector Group.

Another staff member is currently being migrated to the DMSC, but will keep his employment in Sweden but with the Acting Head of the DMSC as his line manager. These staff members are additions to the DMSC staff plans.

More detail about staffing is given in the group reports.

## Relocation

The ESS has received an offer from the Technical University of Denmark to host the DMSC on their Lyngby campus. This is currently being considered by the ESS management and has prompted an update of the DMSC vision and mission. Moreover, a lot of work has gone into identifying different options for locations and associated risks and risk mitigations, which has taken bandwidth from other activities.

## Vision and mission

The DMSC has an existing vision and mission statement, but to the best of our knowledge, this has not been communicated and discussed outside the DMSC. Prompted by the relocation and the general perception that important DMSC stakeholders do not have the full picture of what the DMSC currently is and does, work has begun to communicate and discuss an updated vision and mission statement for the DMSC that can assist with shaping the future DMSC. This work is strongly supported by the Science Director and will be a topic for discussion at the upcoming STAP meeting as well as at the subsequent SAC meeting.

Currently, there is a living document [DMSC---Status and Vision](#). It is primarily written by the Head of the DMSC, but all DMSC staff members have been invited to contribute to (and some have). It is the ambition to follow-up with a more structured approach for dissemination of this document across DMSC and the Science Directorate.

**ESS-DMSC Summer School 2023: Maximising experiments through computation and FAIR data**

Pre-Course Assignment: The students will be invited to complete a short proposal in the ESS User Office software and help define their ambitions for the school. The students will also be enrolled in a basic online Python training course, which will be built on during the school.

	Monday	Tuesday (ESS Lund)	Wednesday	Thursday	Friday
0900-1000	Python Training	DMPs and FAIR data	McStas Simulations	Model-dependent analysis with EasyCore	SciCat and scientific metadata
1000-1100		Reading and manipulation of NeXus			
1100-1200					
1200-1300	Lunch	Lunch	Lunch	Lunch	Lunch
1300-1400	Python Training	Facility Tour	Data reduction and the scipp library	Model-dependent analysis with EasyCore	Reproducibility in data reduction and analysis
1400-1500		Keynote Presentation and Poster Session			
1500-1600					
1600-1700	Beamtime Proposals				
After 1700	Free Time	Free Time	Free Time	Summer School Meal	

Figure 1. Tentative program for upcoming ESS-DMSC summer school.

## Dissemination

We have started to use a more structured approach for dissemination towards our stakeholders, particularly in Denmark and with the purpose of getting more collaboration, e.g. in the form of student projects. We have had a *UCPH meets DMSC* event, where a member (not management) of each of the DMSC technology groups presented their work, followed by presentations of current scientific collaborations. Staff from Technical University of Denmark (DTU) also participated. The Group Leaders have previously had a meeting with staff at ITU to foster more student projects. The *UCPH meets DMSC* event was rather successful and will be followed up with a similar event at DTU where smaller universities in the region (Roskilde and the IT University of Copenhagen) as well as UCPH also will be invited. We are currently also discussing to organize a similar event at Rigshospitalet (DK national research hospital), and for that matter, to organize a similar event on site in Lund.

Moreover, we now, in collaboration with staff in the Science Directorate in Lund, try to make a habit out of having participation from the DMSC when staff from Danish universities have site tours in Lund.

On a more general note, DMSC staff participate in various international workshops and conferences (e.g., ECNS, SciPy, and IUCr) and are frequently invited to give presentations at workshops and conferences.

DMSC is in the process of organizing a summer school to take place in the first week of September. For this school, 12 “students” (defined broadly) will be invited to participate to learn about the role of data in neutron scattering experiments. The programme of the school (shown in Figure 1) follows a user's data from proposal to publication and will allow students to gain familiarity with a broad range of ESS software and an understanding of techniques for data reduction and analysis. This school will take place in Copenhagen (both at the Niels Bohr Institute and DMSC) with a one-day excursion to Lund to see the ESS facility. The event has received financial support from DanScatt (the Danish scattering society) and funding has been applied for from the Novo Nordisk Foundation and the Carlsberg Foundation, which will be used to cover student accommodation, food, as well as other costs. We hope that this one-off will define the model for an ongoing regular DMSC summer school.

## PaNOSC

PaNOSC ended last year in December. This has been a major effort for the involved partners, incl. the ESS-DMSC.

The final review from the European Commission was very positive and also had suggestions for ongoing efforts. For the ESS, the PaNOSC has resulted in:

**Updated data policy** endorsed by Science Management Team, ESS Executive Board and the SAC. However, based on the feedback from SAC, the DMSC management team wants to change a few things in the policy to make it clearer. Ultimately, the updated policy needs to be approved by ESS Council, but knowing that they have many other things on their plate, Council approval has not yet been pushed from our side.

**Improved meta-data catalog.** SciCat functionality has been improved and SciCat has become a community standard with now more than seven institutes using it or even contributing to it. This includes Advanced Light Source in the US, DESY, and even a university, namely Tübingen Universität in Germany. Moreover, the German DAPHNE project has selected SciCat as their preferred solution.

**Federated search.** A major achievement of PaNOSC is the federated search at [data.panosoc.edu](http://data.panosoc.edu) that allows users to search data across multiple photon and neutron sources' data catalogs. Currently data catalogs can be searched through this interface from five PaNOSC partners (CERIC, ESRF, ESS, ILL, XFEL) and two ExPaNDS partners (MAX IV and PSI). The search API is hosted by ESS. The federated search is an important step in making FAIR data findable.

**VISA** for remote access was originally developed by ILL and has since been endorsed by PaNOSC as the platform for remote access. It has been rolled out at all the six partner facilities. At ESS VISA is only accessible for internal users for commissioning purposes. The feedback from users of the ILL VISA is positive. For a use case see <https://www.panosc.eu/use-cases/panosc-use-case-30-visa-data-analysis-in-the-cloud/>

**McStasScript** is a Python interface for McStas developed at ESS. It is a powerful tool for setting up complicated instrument and sample environment features for McStas simulations and is easily accessible for users already familiar with Python. It has been used for training in combination with the e-learning platform below.

**Pan-training.eu / pan-learning.org.** Pan-training.eu (<https://pan-training.eu>) is a photon and neutron community platform for training material. It contains both a library of existing training material with links to its location and an e-learning platform, usually called pan-learning, which is accessible through <https://pan-learning.org> or <https://e-learning.pan-training.eu>. pan-learning is an extension of e-neutrons.org to also host material for photon scattering and with Jupyter integrated so that scripts can be run or developed as part of a course. Jupyter, combined with McStasScript, is the preferred solution moving forward for running McStas simulations in the platform (see <https://www.panosc.eu/use-cases/use-case-20-using-an-e-learning-platform-to-support-presentations/>), but, as it turns out, GUI applications can also be run through Jupyter (see <https://www.panosc.eu/use-cases/use-case-31-seamless-connection-of-jupyter-notebooks-and-gui-applications-for-e-learning-purposes/>)

**Summer school.** The aforementioned e-learning platform was used at PaNOSC summer school (<https://indico.eli-laser.eu/event/13/>), which was hosted by ELI but with ESS as co-organizer. The summer school gave students an introduction to FAIR data in relation to photon and neutron sources. The feedback was very positive and prompted the DMSC to start organizing its own summer school focusing on data and the ESS software suite as discussed previously in the Section Dissemination.

**Extended network and collaborations.** The PaNOSC has resulted in increased knowledge and understanding of each other amongst the six partner ESRIs as well as with some of the ExPaNDS partners. We have for instance collaborated on the pan-training platform with HZDR and SOLEIL, and DST has benefited from knowledge sharing about OpenStack, for instance with DESY.

## Other international collaborations/ roles

The Acting Head of the DMSC is currently co-chair of the LENS WG4 on IT, spokesperson for WG4 sub-group on data analysis, active in LEAPS WG3 on IT (e.g., co-author on paper about LEAPS data strategy), ESS representative in the EOSC Association and in an associated ESRFRI working group, member of an EOSC task force and the newly established SNS/HFIR advisory board for scientific computing.

Due to limited bandwidth, the DMSC has not yet been active in shaping new proposals for Horizon Europe.

## Concerns

The major concerns are all related to staffing, recruitment and retention, partly because of a hot labor market and partly because of the uncertainties related to the future location of the DMSC.

As we approach commissioning, the need for collaborating across groups becomes more pronounced and clarification of roles, project priorities, etc, becomes more important. The increased focus on planning and regular planning meetings is an attempt to proactively manage this (see Section *Project management, communication and milestones*) below.

The perception of the DMSC amongst influential stakeholders may lead to decisions that are damaging for the project, staff motivation, or the future build up of strengths and competences at the DMSC. The development of the Status and vision document is an attempt to mitigate these risks.

The work on relocating the DMSC is currently the most important issue for the DMSC. It concerns staff, project deliverables, risk mitigations, and of course the future of the DMSC. This has meant that other tasks, albeit important, have been down-prioritized, both internally (e.g., project planning and staff management) and externally (e.g. international collaboration).

## Administration and Project Coordination

### Staffing

A new administrative assistant has been on-boarded so that the team is now back to have two much-needed assistants for the DMSC.

A tender is currently out for getting framework agreements, so that consultants can be easily recruited as a means to regain lost effort due to challenges with recruitment.

### Office relocation

The DMSC offices have been relocated internally in the COBIS building. This has been a rather smooth process. As previously mentioned it has not yet been decided where to move to in the future.



# Project management, communication and milestones

ESS uses p6 for Project Planning. This is where the rebaselined project plan is to be accessed. There are project meetings each month informing about the status of the overall project.

For DMSC, the activities (in hours) until the end of 2027 have a set budget and the overall milestones are in p6. Each month, the planner from ESS Lund reaches out to see if DMSC is starting the activities that we planned to start and if we finished the activities we planned to have finished. With delays in hiring, some activities are obviously delayed and tracked through this process. There are also possibilities to make changes to 6 each month via our planner in Lund as long as it does not affect the overall budget. The downside with p6 is that it is not easily accessible and it is not a hands-on tool for us at DMSC to use in our daily planning with interdependencies between our groups on the Danish side of the sound.

Since last STAP, DMSC has focused on getting an easy way to track internal milestones, agree on the definition of completion for the particular milestones and identify the dependencies between the groups. Because the DMSC groups already use JIRA from Atlassian, we have chosen to work with Big Picture from Atlassian where JIRA tasks can be linked to a Gantt-chart. All of these milestones are not necessarily implemented in P6 but are important for internal DMSC project development, priorities and allocation of resources.

Each group, as well as the 6 Instrument Data Scientists have been given the task to come up with milestones for the upcoming 18 months. For each milestone the following have been provided:

- Milestone name
- Due date, either absolute date or relative to another milestone (e.g. 6 months before Hot Commissioning)
- Description of milestone, e.g.:
  - What does this actually mean?
  - How would you demonstrate that the milestone is achieved?
  - What can a user expect to be able to do when a milestone has been achieved?

Once per month we have an internal project meeting with the group leaders and we have started to discuss what is needed from whom and when that is needed for them to achieve completion of the milestone. We also try to link milestones to visualize the dependencies. This way of tracking the DMSC part of the ESS project increases the chance of catching delays and issues and ensuring that sufficient resources are available. It also makes it visual for all group leaders.

Examples of how it looks in BigPicture are shown in Figure 2-4.

>	DMT-50	●	DMSC specific milestones
∨	DMT-66	●	DRAM specific milestones
>	DMT-72	●	DRAM Modeling specific milesto...
>	DMT-73	●	DRAM SWAT specific milestones
>	DMT-82	●	DRAM SCIPP specific milestones
>	DMT-48	●	SWAP specific milestones
∨	DMT-47	●	IDS specific milestones
>	DMT-53	●	IDS ESTIA specific milestones
>	DMT-54	●	IDS DREAM specific milestones
>	DMT-57	●	IDS LoKI specific milestones
>	DMT-62	●	IDS NMX specific milestones
>	DMT-92	●	IDS BIFROST specific milestones
>	DMT-97	●	IDS ODIN specific milestones
>	DMT-46	●	DST specific milestones

Figure 2. High-level view of DMSC milestones in BigPicture.



Figure 3. Specific milestones for SWAP in BigPicture (grouped under DMT-48 in Figure 2).

## SWAP VISA integration with access to experimental data

[Edit](#) [Add comment](#) [Assign](#) [More](#) [Start Progress](#) [Done](#)

**Details**

Type:  Task Status: **TO DO** ([View Workflow](#))  
Priority:  Medium Resolution: Unresolved  
Labels: [created-by-BigPicture#93413](#) [milestone](#)

**Description**  
Dependant on DST

**Attachments** ...

Drop files to attach, or [browse](#).

**Issue Links** +

**has to be done after**

[DMT-33 DST VISA: Auth integration with Useroffice \(p6\)](#)  **IN PROGRESS**

**has to be finished t...**

[DMT-14 SWAP Creation of proposal permissions for data management, includes user permissions f...](#)  **TO DO**

**is required by**

[DMT-48 SWAP Project](#)  **TO DO**

[esss.lu.se/browse/DMT](https://esss.lu.se/browse/DMT)

Figure 4. Detailed view of a specific milestone (DMT-18 in Figure 3).

We are still in the cradle of tracking our milestones like this and hence, there are still improvements to be made - both in terms of what milestones to track but also in terms of defining completion of some of the milestones. We need to find the balance between making this a useful tool and not overdoing it, which risks taking time from the actual work or preventing staff from taking actions that are not planned.

## Data Systems and Technologies (DST)

### Staffing

The DST group currently (Q2 2023) consist of 4 persons working full time including the group leader plus 2 student workers working 10 hours per week. In addition to the current staff level there are 3 vacant positions in the group in accordance with the re-baselined plan for DST. DST is running a recruitment process for those three vacant positions. So far, it has proven very difficult to attract viable candidates and since the 2022 October STAP no full-time positions have been filled. A fourth position starting in August 2023 is also planned for.

So far the effects of lost effort have been mitigated partially by focusing on the bare necessities needed to ensure other groups can progress and by out-sourcing parts of the externally funded work in the PaNOSC and HighNESS projects. As further mitigation steps, DST is looking into whether relevant consultants can be identified from existing framework agreements and DMSC is working on getting framework agreements in place for consultants with DST specific competencies and with place of work in Denmark. DST has also recruited two student workers starting in March and April (2023) working 10 hours per week. These students will be utilized to perform some of the more trivial work and thereby freeing up full time DST staff for more complicated tasks. Furthermore, recruiting students when possible, also serves the purpose of promoting DST as a workplace and a plausible career path for newly graduated candidates. If the concept of student workers turns out to be a success, DST will seek permission to hire a third student worker in Q3 2023 in order to mitigate the effect of lost effort from not being able to fill full-time positions.

As part of the ESS re-baselining process, a new and revised staff ramp-up plan is in place for DST. Figure 5 shows the staff ramp-up plan for DST in the coming years. Since the 2022 rebaselining DST has lost 1.2 FTE in effort from not being able to fill vacant positions.

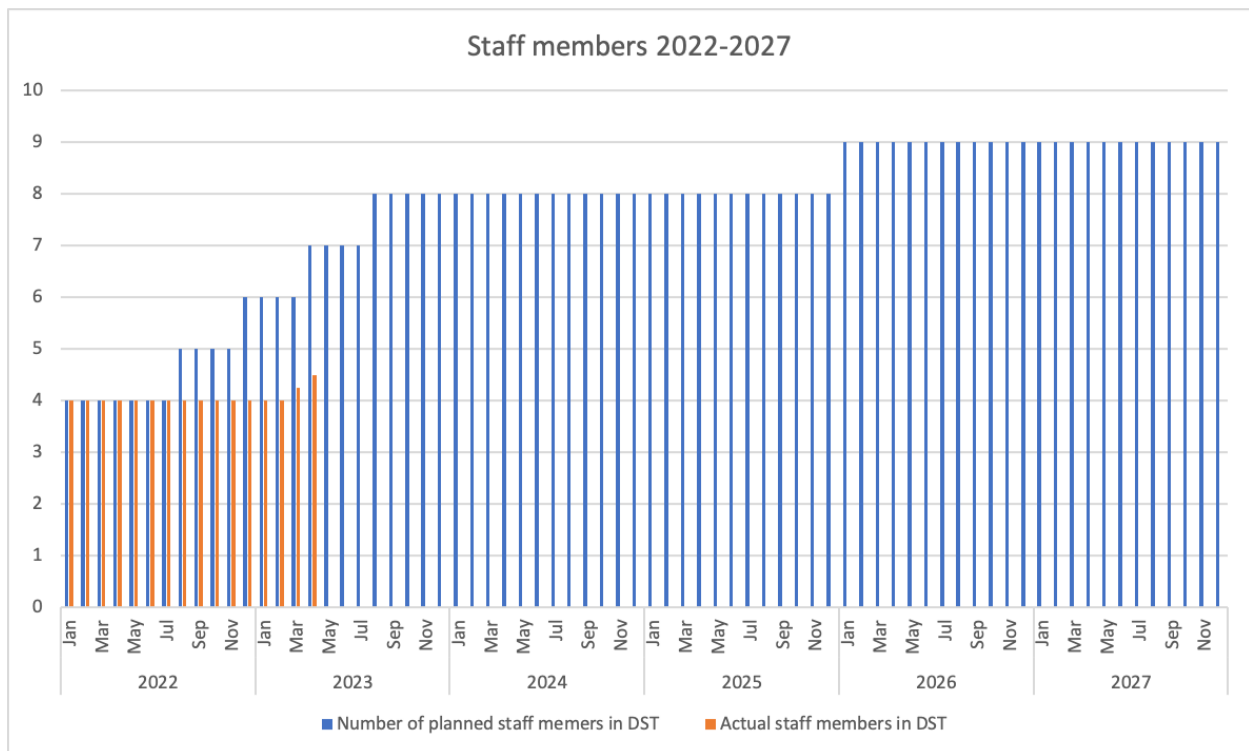


Figure 5: Staff ramp up plan for DST in the period 2022 to 2027

## Achievements since last STAP

Since the STAP in October 2023, DST has made the following major achievements:

- **YMIR test beamline:** DST has progressed the work on setting up the YMIR test instrument as a real production instrument. This work is well underway and in close collaboration with the ECDC group and the ICS division with ECDC having the coordinating unit.
- **Spectrum Scale:** The Spectrum Scale storage system has been fully integrated with DMSC data centers and are used for HPC cluster users and for experimental data alike. An important note here is that DST and SWAP have jointly succeeded in implementing a fully automatic process for setting up a (fileset-based) folder structure based on proposal creation in the Useroffice software and subsequently setting up automatic data transfer from the Lund based storage system to the Copenhagen one.
- **VISA:** DST is progressing on enhancing our VISA installation and every opportunity to obtain experience from the users (IDS's and instrument scientists mostly) is utilized. DST is also active in the continuing collaboration on VISA in a post-PaNOSC environment, and expects to participate in the upcoming inter-facility MoU on VISA-collaboration.
- **OpenStack:** DST has put a lot of effort into improving our OpenStack virtualization environments and we are currently at a stage where a "classic" virtualization cluster can be deployed automatically from scratch using the DST Puppet/Foreman setup. This can be achieved in both data centers alike. While the current deployed setup is very stable and well under configuration change management, there is still work to be done on moving towards a coherent multi-region private cloud setup and it is expected that DST will continue to improve on these installations in the coming years.
- **HPC:** A significant increase of the HPC compute capacity has been procured to budget, installed and deployed into production. This extension consists of 24 nodes of 56 cores and 512GB of memory each. This constitutes a 30% increase of DMSC cluster compute capacity.
- **Beamline and user hardware:** Experimental control workstations (user facing hardware) procured for the first 9 instruments well under budget. Deployment and commissioning is in progress.
- **Control systems networks:** Deployed networks for control systems for the first 9 instruments.

## Status on the hardware installations

Having executed the 2022 procurement plan and subsequent hardware installations, the status of the two DST operated datacenter installations are:

## Installations in Copenhagen

The hardware installations in Copenhagen consists of:

- HPC: 116 nodes in 4 different queues.
- GPU: 2 A100 card server deployed into production.
- InfiniBand HDR fabric: Deployed into production
- InfiniBand QDR fabric: To be sunset.
- 1GBE, 10 GBE and 100GBE ethernet fabric: deployed into production.
- OpenStack VM environment: deployment on-going. Current status is:
  - Prototype running with all relevant componets
  - Production system deployed and migration from legacy Ovirt system is in progress.
- Test and dev setup is in place and being worked on for the next feature roll out and update.
- Legacy Ovirt based virtualization environment: Sunsetting in progress.
- Spectrum Scale Storage system: Deployed into production with AFM-based data transfer from Lund to Copenhagen in place.
- Legacy ZFS systems of a total of 480TB

Moreover, DST has increased the effort on providing various JupyterHub and VISA instances on the available compute infrastructure. We continue to receive a lot of very constructive feedback and are growing more and more confident in the viability of VISA as a main interface to the DMSC computing systems.

The utilization of the HPC cluster continues at a high level of around 85% node utilization.

## Installations in Lund

Currently the hardware installations in Lund consists of:

- Lenovo DSS Spectrum Scale 850TB storage system
- Networking (100Gbe, 10Gbe, 1Gbe and HDR InfiniBand) is in place
- Connection to the Copenhagen data center established and services are provided across that link for
- Deployment and provisioning (Foreman and Puppet)
- Virtualized services.
- 8 Filewriter servers
- 20 Kafka servers
- 18 Event Formation Unit (EFU) servers
- 2 VM hosts (96 cores and 1TB memory each).
- 8 server staging environment hosted for the ECDC group. These are legacy servers from the Utgaard laboratory.

These systems are being used for cold commissioning of the experimental control and data acquisition systems developed by the ECDC group. This work continues to provide value using the Ymir test stand by the ECDC group to test early cold commissioning of devices and IT infrastructure.

## Plans for 2023

In 2022 ESS rebaselined the ESS project and this has resulted in a fairly detailed activity plan for DST. A number of milestones has been put in place with associated activities, all linked to activities and milestones in the other DMSC groups providing transparency and trackability of the work provided by DST. Figure 6 shows the milestone plan for DST from 2023 to mid 2024.

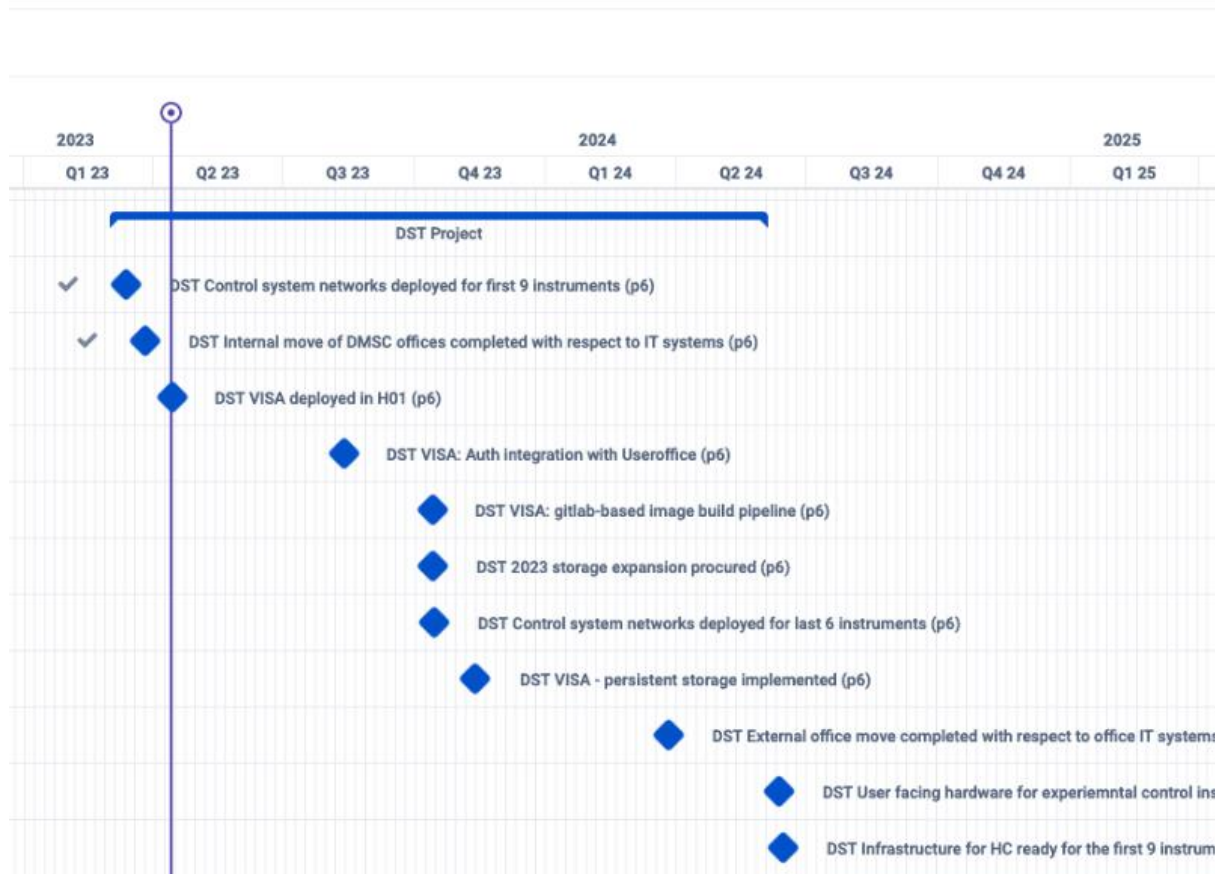


Figure 6 Milestone plan for DST 2023 -> mid 2024

These milestones point both to milestones in P6 but also to concrete JIRA tasks, stories and epics in the DST specific JIRA board. These milestones have resulted in the planning of the activities for 2023 shown in Figure 7.



Figure 7. Activity plan for DST in 2023

Due to the challenges related to recruitment in DST, the uncertainty regarding the housing situation and the fact that the neutron instruments are starting the cold commissioning phase in 2023, DST does expect to adjust plans regularly to be aligned with reality as it unfolds.

## Procurements

In 2023, DST will procure hardware, services and licenses for a total budget of 2797KEUR. These purchases will be distributed on the following categories:



- **HPC:** This is partly to ensure that a replacement capacity for legacy compute capacity is in place but also to scale up the existing HPC capacity to cater for cold commissioning activities and in preparation for BOT and hot commissioning.
- **Network equipment:** Procuring 100GbE equipment required for the general scale up of compute and storage capacity and to replace legacy 10GbE equipment. An important aspect of these purchases is to prepare the scale up of the site-to-site link bandwidth from the existing 10Gb/s to 100Gb/s.
- **Beamline and user equipment:** Workstations, monitors etc. for instrument cabins.
- **Licenses:** Miscellaneous software licenses, predominately for the DRAM group.
- **Services:** Service agreements for operating the COBIS data center, for the site-to-site link and for the connection to the Danish NREN.
- **Miscellaneous equipment:** This covers things like workstations, office network, minor spare parts etc.
- **VM environment:** Scaling up the DST OpenStack virtualization environment to cater for cold commissioning activities and in preparation for hot commissioning. Specific requirements are being developed in collaboration with the other DMSC groups.
- **Storage expansion:** Scaling up the existing Lenovo based DSS Spectrum Scale system to cater for hot commissioning.
- **Server room build-out:** According to plan the Copenhagen server room needs to be build out to 16 racks with increased redundancy for cooling and power. Given the uncertainty regarding the future of the DMSC hosting facility in Copenhagen, it is more than likely that this build out would either be converted into building a new server room or postponed until 2024.

Figure 8 shows the cost distribution for the categories above.

A significant risk recognized by the DST team is the significant supply chain issues which continue to impact most of the IT market. Of special relevance for DST, is that Mellanox (the preferred provider of HPC networking (e.g. switches and adapters) is heavily impacted by this.

These investments add to and replace legacy equipment in the existing installations already in place in the two DST operated datacenters.

DST will also, in 2023, install computer equipment in the control cabins for users to control and monitor their experiments and for reducing and analyzing resulting data on ESS infrastructure. This is done in collaboration with ECDC and the instrument data scientists.

## PaNOSC & HighnESS

DST was involved in WP4 Data Analysis Services, WP6 EOSC integration and WP8 Staff and user training in PaNOSC and WP9 Computing Infrastructure in HighnESS. While PaNOSC ended in November 2022, DST continues to be an active collaborator in the VISA project.

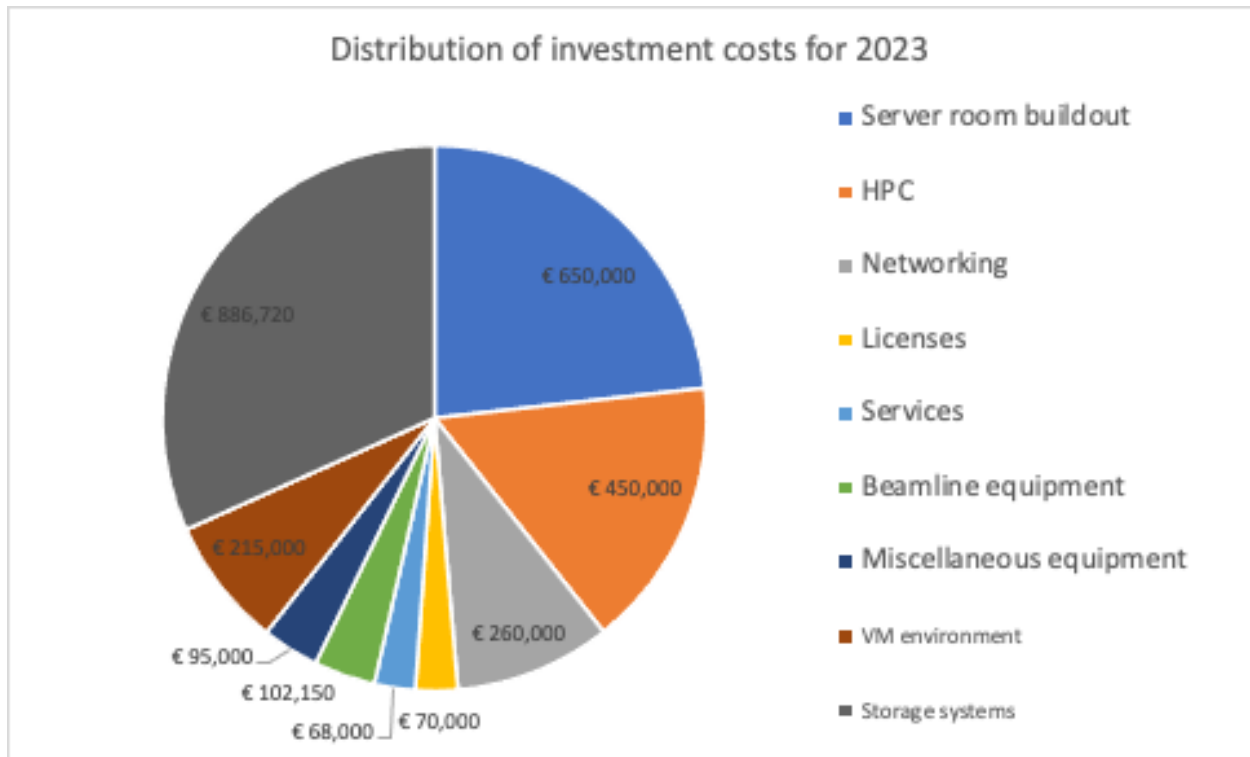


Figure 8 Cost distribution of the 2023 capital investments in DST

Figure 8. Cost distribution of the 2023 capital investments in DST

## HighnESS WP9 Computing Infrastructure

HighnESS is an EU Horizon 2020 project on designing a second moderator for ESS allowing for high-intensive cold neutron experiments to be conducted at ESS. Work package 9 is concerning the compute infrastructure needed for the design work in the project. The work focuses on:

- Providing a cloud service for running McStas simulations
- Providing a cloud service for running Ncrystal simulations
- Ensuring that any (if any) changes to the McStas source code are merged into the McStas production branch.

DST is driving this work by having the work package leader role. While all three bullet points above are progressing, there is still work that needs to be done.

## Participation in synergy groups on IT systems

DST collaborates extensively with the IT Division in the new Operation & Infrastructure Directorate. The IT Division now also contains the ICS Infrastructure group. Besides the aforementioned Common Electrical Project, the areas of collaboration are:

- Information security
- Infrastructure architecture - to ensure as many commonalities as possible between the groups
- Networks
- Hosting infrastructure
- Off-site hosting solutions for other ESS groups (ICS and IT).
- Remote access
- Identity and Access management system for all of ESS (driven by the IT division).
- Disaster recovery for control systems etc.
- Procurement (e.g. tender design, review and common framework agreements).
- Participating in framework agreement on server hardware together with ESS-IT and ICS.

DST will continue and expand this collaboration in the years to come

## Concerns

Despite the progress made since the last STAP, some challenges and concerns will continue to play a role for DST in the remaining period of 2023. These are:

- Not having been able to find candidates to backfill positions since 2020 for various reasons continues to impact the development of the DST deployment stack as well as integrations between DST and other systems.
- The general staffing level in DST prevents work that spans beyond keeping the DST systems running and implementing the absolute necessities in order not to block other groups from working. Considerable risk exists that the latter will not even be possible to avoid despite having onboarded 2 student workers providing a total of 54% of a full time position.  
The very hard prioritization of tasks is also greatly accelerating the accumulation of technical debt on the systems maintained by DST
- The supply chain issues impacting most of the IT market poses significant risks that the DST capital investment plan for hardware and services will not be executed to the full extent in 2023. DST is tracking this very closely and are adjusting plans accordingly.
- Uncertainty regarding the future housing situation as well as the future hosting facility in Denmark does impact both the procurements of general hardware and also the planned 2023 build-out of the Copenhagen server room and the risks associated with potentially having to relocate the Copenhagen server room during hot commissioning.

# Scientific Web Applications (SWAP)

## Staffing

The team consists of five people (including one consultant), in the last six months there has been one resignation and two hires. The plan is to replace the consultancy effort by hiring an additional developer in the coming months.

Finding qualified candidates for the SWAP positions has been difficult, this is due to several factors, main one's being:

- Uncertainty regarding office location
- Strong local IT-market, high salary requests
- Slow internal HR processes

The resignation of the team's senior technology expert in February led to a significant amount of time spent on handover of knowledge and distribution of tasks. At the same time the onboarding of new staff is requiring significant resources from existing staff.

The team continues to work in an agile way, distributing tasks across people and with a high level of knowledge sharing among team members. This has enabled the team to better manage the latest staff turnover.

## Achievements since last STAP

### Integrations

As reported in the last STAP, a considerable amount of work has been focused on integrating SWAP software into the rest of the ESS IT-Landscape. There has been significant progress in this area and there now exist integrations with:

#### **NiCOS**

- Retrieves proposal and sample information from user office
- Send status messages to SciChat

#### **File writer**

- Creates datasets in SciCat via connector service

#### **VISA**

- Instruments created in the user office are also created in VISA

#### **Identity Access Management**

- User creation and profile management

#### **Folder creation**

- Proposals accepted in the user office triggers folder creation in the spectrum scale storage

#### **EAM**

- Shipments from user office visible for the logistics department

Creating a solution for the above-mentioned integrations has meant the development of a new software called the connector. The connector is responsible for listening and acting on messages coming from Kafka and RabbitMQ. The service should be seen as a temporary solution in wait for a centrally hosted IT integration platform.

## Identity Access Management (IAM)

Integration between IAM and User Office has allowed the creation of user accounts in the central repository. Work is ongoing to configure the life cycle management of accounts and the migration of existing user accounts. The tasks associated with IAM continue to be more time consuming than originally envisioned. Additional resources have been contracted by central IT to enable faster resolution to issues and help advance future tasks.

## Pan-learning

The e-learning platform created as part of WP8 for PaNOSC has been migrated over to the SWAP infrastructure for long-term support. Work has included updating the e-learning platform built on Moodle to the latest version and rewriting the theme and user creation to be in a more easily maintainable state.

## User Office

The development of the user office system is continuing with the support from the Scientific Activities Division (SAD) and our STFC partner. The system is currently being tested by CERIC and ELI in preparation for joining the collaboration. In order to ensure a beneficial collaboration, it is necessary to align the different facilities processes, this work is being driven by the SAD team with input from SWAP.

## Data Management Plans

As part of WP2 in PaNOSC a DMP template with an associated technical solution was developed and deployed across facilities. This work will be published in the Data Management Journal as a practice paper.

## Data Curation

The migration of the SciCat backend is now in the final stages, a tagged version has been released and deployed to the staging environment as a last step before moving to production. The suite of auxiliary software developed to support data curation and interactions with SciCat has continued to mature.

## Plans for 2023 and Milestones

- Proposal folder creation

- Finished March 2023
- Proposal folder permissions
  - Ongoing
- User access to VISA
  - Working with DST
- Access card issuance, linked with visit management, e-learning and scheduling
  - Depends on the integration with IAM
- Sample safety review
  - Waiting for sample experiment team

## Concerns

The main concern continues to be staffing and problems delivering tasks that involve groups across divisions/directorates.

The staffing issue has required consultancy to be extended and some tasks to be delayed. We are closely monitoring the situation and considering options such as student workers etc.

The complexity regarding ownership and responsibilities of tasks that span groups has been summarized and presented to the Science Director. With the goal of clarifying responsibilities and highlighting orphan scope in the organization.

## Data Reduction, Analysis, and Modelling

### Staffing

The DRAM group is responsible for providing the data reduction, analysis and modelling software for all instruments at ESS. See last STAP report for more information about the scope covered by the DRAM group.

The data reduction team has four members today and we are currently in the process of recruiting two new developers to the team. The data analysis team has three members today and we are about to start the recruitment for a fourth developer to join the data analysis team. In addition, we have help from two IDS's for SANS and reflectometry data analysis software. The last team is the data modelling team. Here we have one person working as an In-Kind contribution from the Technical University of Denmark, and one post doc funded by the EU projects HighnESS. Finally, from 1<sup>st</sup> of August 2022 a postdoc, who is funded by the Swedish Research Council, works on questions related to both data reduction and analysis for NMX.

## Activities since last STAP

### Data reduction – The Scipp team

#### Scipp: General updates

We have initiated the process of defining more detailed requirements from the IDSeS. The goal is to move from the current high-level lists of things that need to be done to concrete and actionable requirements. To this end, after a kickoff meeting with the IDSeS group, we have defined a template that can be used to describe requirements at a level of detail where we can begin working on them and include them in more detailed planning.

Since then, we have started work on (and partially completed) on a number of such requirements, including SANS beam center finder, coordinate transformations to  $(Q_x, Q_y, Q_z)$  and  $(hkl)$ , source frame unwrapping including unwrapping in pulse-skipping mode, and WFM frame unwrapping for scattering beamlines.

A paper titled "*Systematic underestimation of uncertainties by widespread neutron-scattering data-reduction software*" (by SH, JLW, NV) has been accepted by the Journal of Neutron Research.

#### Scipp: Scipp

We have introduced a more flexible version of `scipp.Dataset` called `scipp.DataGroup`, which does not enforce alignment and supports nesting. It enables a significant variety of usability improvements, e.g., when loading structures files such as NeXus as well as holding data during data-reduction in a more natural manner. Aside from this major addition, we have fixed various bugs, performance, and memory issues.

#### Scipp: ScippNeXus

We have introduced recursive loading of NeXus file structure, and the result is returned as a nested `scipp.DataGroup`. ScippNexus now supports writing a subset of NXcanSAS. We have encountered some performance issues when working with files from YMIR, BIFROST, and DREAM. Specifically, small files with many groups and datasets were affected. This did not come entirely unexpected, since the initial implementation did not focus on performance for small datasets. We are currently implementing a better solution that should be available shortly. Finally, we are working with ECDC to weed out various smaller issues with the NeXus files written by ESS, to ensure that the files are valid and interpretable.

#### Scipp: Plopp

As reported previously, we have re-implemented Scipp's plotting as a separate Python project based on Scipp, called Plopp. Plopp has been consolidated over the past 6 months. As hoped, the removal of the complex dynamic refactoring ability and other improvements compared to the old Scipp built-in plotting solution, the number of reported bugs by users has gone down significantly. Additionally, we have introduced powerful features for custom interactive figures.

Plopp is now the default plotting solution for Scipp, the built-in Scipp solution is deprecated and will be removed later this year.

### Scipp: Scitacean

In collaboration with SWAP, we have begun the development of Scitacean, a Python package for interaction with SciCat. It provides an easy-to-use, high-level interface and handles both metadata and data files. It is already in active use by internal experts. We plan to use Scitacean with Scipp for writing reduction workflows and other data management tasks once a more accessible login system has been deployed to SciCat.

### Scipp: Beamline

We have begun work on our new "live reduction" dashboard, called Beamline (not a typo). We expect to have a working prototype before summer and aim to provide a usable "product" by the end of the year.

### Scipp: Releases

A large number (>45) of software packages has been released since the last STAP. This covers works on the packages: scipp, scippnexus, scippneutron, plopp, ess, mpltoolbox, scitacean.

## Data analysis

The data analysis team is working on in-house projects within the EasyScience framework, but also collaborating with other facilities on e.g. the SasView project for SANS and the PACE project for spectroscopy. Our shared IDS for macromolecular diffractometer is working on using DIALS for the NMX instrument. See last STAP report for more information about the data analysis activities in the DRAM group. Within the EasyScience project we have worked on EasyReflectometry and EasyDiffraction.

For EasyReflectometry, work has been centered around a new release with improvement for file loading, plotting, continuous integration setup as well as distributing snapcast images for installation on Linux platforms. In addition, work has been put into updating the EasyReflectometry roadmap.

Following the STAP advice from last meeting EasyDiffraction is our main focus now, and we want to bring it into a state where it can be presented to a selected group of users. To achieve this goal, we are working on two parallel tracks. One track is related to the usage of the crystallographic libraries CrysPy and CrysFML, while the second track is related to the internals of EasyDiffraction for improving the UX. As mentioned at the last STAP, ILL and ESS have agreed on continuing the collaboration on sustaining CrysFML. In this agreement the ILL scientific computing group takes care of maintaining the CrysFML library jointly with Juan Rodríguez-Carvajal, Elisa Rebolini, and Nebil Katcho whereas ESS focuses on user interfaces including Python bindings for CrysFML. We have established monthly video meetings with the ILL team (ER & NK) and we also had a productive code hackathon with them at DMSC in



December 2022. To continue the collaboration, another code hackathon is scheduled to be held in Grenoble in early summer 2023. The second track is related to the internals of EasyDiffraction UX with improvement for performance, undo/redo, bug fixes, as-well-as a new API for the CrysPy interface, which overall leads to improved performance when using the EasyDiffractionApp (i.e. the desktop installation).

Activities for promoting EasyDiffraction will include an oral presentation at the IUCr-2023 conference and a tutorial session at the Crystallographic Software Fayre (also at IUCr).

Erik Fransson from Chalmers University of Technology is working on a joint Chalmers-MAX IV-ESS post doc project. The project is about statistical learning applied to diffraction data and will make use of easyScience / easyDiffraction. A visit to DMSC is scheduled for May 8-12 2023.

## Data analysis: Releases

New releases of EasyDiffraction, EasyReflectometry and SasView have been made since last STAP.

EasyCore 0.3.0, EasyCrystallography 0.2.1 & 0.3.0, EasyReflectometry 0.0.6, EasyDiffractionApp 0.8.5, and EasyDiffractionLib 0.0.9, and SasView 5.0.6rc1.

## Data modelling

The ray-tracing Monte Carlo instrument simulation tool McStas is celebrating its 25<sup>th</sup> anniversary, yet is still in active development. In the period between October 2022 and March 2023 there were improvements on the process of writing NeXus files, as it previously required a separate installation, but is now part of the installed package. This will make it easy for users to directly make McStas event datafiles for Scipp and Mantid. The development team held a hackathon with a wide focus, including improvements to the installation procedure, performance optimization, bug fixes in the plotting tools and new methods for displaying the geometry.

Improvements and new features are important, but they are of no benefit unless communicated to the user base. To that end a 5-day McStasScript / Union / guide\_bot course was developed in this reporting period and held at University of Copenhagen for the group of Kim Lefmann. The course is fully cloud based, requiring no local installation, and consists of quizzes in Jupyter notebooks where the students submit code answers. The code is checked and students receive instant feedback

The EU project PaNOSC ended during the reporting period and had its closing event where WP 5 on simulations were highlighted as a major success of the project due to the harmonization of simulation interfaces for X-ray and neutron instrument simulations. There is also ongoing involvement in HighNESS WP 7 where instruments are being simulated in order to benchmark different moderator options for the second moderator slot at ESS.

IDS Søren Schmidt led a beamtime with the purpose of testing the new timepix-3 detector at SENJU, J-PARC. A simulation team member joined the beamtime and provided support for all three experiments performed, demonstrating that simulations can aid researchers in decision-making during an experiment. IDS Gregg Tucker and IDS Celine Durniak were supported in their endeavors to prepare simulations that will provide data for testing of analysis workflows. Instrument Scientist for FREIA requested help for building a model of the FREIA slits system using Union components. And in general, we have provided ESS-oriented McStas community services to ECDC, polarization-group, TBL, NMX, DREAM, C-SPEC, and the SAGA proposal.

Throughout the period we have also assisted in modernization efforts for the pan-learning.org platform, end-user support to courses running on the platform (UVA Amsterdam + TU Delft).

As part of Peter Willendrups DTU affiliation, he will be hosting José Robledo as GNeuS (<https://gneus.eu/>) Post Doc for the next couple of months here at DMSC. José will be working on virtual SANS experiments with McStas with side-projects going in the direction of GPU-computation, AI/ML etc.

## Data modelling: Releases

New releases of McStas and McStasScript have been made since last STAP.

McStas 2.7.2, 3.2, 3.3. McStasScript 0.0.52, 0.0.53, 0.0.54, 0.0.55

## Plans for 2023 and Milestones

We are working on making our milestones more transparent and visible to the organization. As mentioned at previous STAP meetings we are using a mixture of GitHub, Zenhub, Jira, Confluence, and P6 in order to accommodate the different needs as seen from a developer perspective as well as for the ESS project organization point of view. At DSMC, all groups are now using Jira's BigPicture to keep track of milestones.

Below is listed a few of the DRAM milestones.

BigPicture ID	Name	Date	DRAM Team
DMT-83	Prototype of Live Dashboard ready	2023-09-31	Scipp

DMT-86	IDS data reduction requirements gathered for Phase HC-1 (2023)	2023-12-31	Scipp
DMT-74	EasyDiffraction release ready for friendly users to use & test	2023-09-31	Data Analysis
DMT-76	Calculation engines used in EasyDiffraction, CryPy & CrysFML, updated to latest versions	2023-12-31	Data Analysis
	McStas intr files for instrument 1-9 updated and validated	2023-12-31	Modelling
DMT-70	Most relevant McStas scattering kernels for instrument 1-9 identified	2023-12-31	Modelling

## Concerns

Below we list some of the concerns, comments and aspirations we have discussed within the teams.

- Scope, Staff, Time problem: Do we have enough staff to get the job done?

## Scipp

While work on initial requirements has started, we experience challenges in gathering concrete requirements. Among other reasons, the lack of actual ESS data files is a contribution to this. We will keep working with the IDS to ensure we can support them in the process of defining concrete requirements.

Some dependencies used in our software stack can lead to SPOF (single point of failure). We try to minimize and/or eliminate these dependencies.

## Data analysis

Supporting several calculation engines in EasyScience does come with some overhead, wrt. getting all dependencies to work together.

How do we keep up with new trends like machine learning and Bayesian data analysis?

## Data modelling

The data modelling team is in a transition phase, going from a more individual project-based setup over to a more ESS-DMSC-DRAM common project-based setup. Here we need to develop best practice for working together as a team.

Our staff profile is smaller when compared to the other two DRAM teams. We have to use the resources we have wisely, so that we can support the organization without creating technical debt.

# Instrument Data Scientists

Each of the Instrument Data Scientists (IDSs) have uploaded their own report and will present their work on the second day of the STAP meeting. The IDSs also participate in their associated instrument STAPs. In some cases, they provide a written report and in other cases an oral presentation, and in some cases both. For the DMSC STAP the IDSs were asked to upload the reports for the instrument STAPs in case they have one, and otherwise to write a brief report specifically for the DMSC STAP. These reports are not merged into this document, but are separate documents that can be found on the indico page.

The VR funded post doc for NMX is in this context considered an IDS and will also present at the DMSC STAP.

Because there are individual reports and because the IDSs are given the opportunity to present their activities themselves, this section will only concern some general considerations and a job description.

## Staffing

There has been no new recruitments for the IDS group since the last STAP. We follow the staffing plan given below with nominally ½ IDS per instrument.

<b>Instrument Data Scientist</b>	<b>Technique</b>	<b>Instruments</b>
IDS1 (WP)	SANS	LoKI, +Skadi
IDS2 (AM)	Reflectometry	ESTIA, +FREIA
IDS3 (SS)	Imaging & Engineering	ODIN, BEER
IDS4 (CD)	Diffraction	DREAM, MAGIC
IDS5 (GT)	Spectroscopy	BIFROST, CSPEC
IDS6 (NN1 1 year before HC for additional diffractometers)	Diffraction	+NMX, +HEIMDAL
IDS7 (NN2 1 year before HC for additional spectrometers)	Spectroscopy	+MIRACLES, +TREX, +VESPA

As mentioned, we currently have a postdoc working on data processing for NMX, who formally is a member of the DRAM group, but in practice can be considered an IDS for NMX. Currently, the IDSs for reflectometry and SANS are only tasked with supporting the ESTIA and LoKI instruments.

In practice, the two IDSs for diffraction and the two IDSs for spectroscopy should jointly manage all the diffractometers and spectrometers, respectively. For the second IDS for a given technique we will therefore be looking for somebody with complementary skills to the other IDS. Members of the IDS group are expected to support each other and coordinate with each other across instrument techniques.

## Concerns

There are generally concerns about the legal and practical limitations on how much the IDSs can work in Sweden, which to some extent is also linked to the future location of the DMSC. The need for their presence on the Lund site also depends on other resources and skills available to the instrument teams, and hence is directly linked to the staffing of instrument teams on the Lund side. These questions are currently being considered by the Science Director.

Related to the job description below, is the question about their role in relation to the instrument teams but also in relation to the technology groups, primarily at the DMSC. This again links to the discussion above. The balance we have tried to strike so far is that of a Product Owner /

Manager, i.e. a person who is not responsible for implementation of solutions, but who is responsible for overseeing and validating delivered solutions.

## Job description

The job description for the IDs have been discussed on previous occasions. A long overdue task is to formulate a formal job description. The following job description (R2A2) has been formulated for the Instrument Data Scientists, but given the cross-functional role of the IDs interfacing groups should be consulted as well as the Science Director.

**Name:** lamada Taex Pert

**Job title:** Instrument Data Scientists for *technique or instruments*

**Job type and job level:** Scientist D

### Role

Manage DMSC deliverables to instrument and user program and support these programs with expertise in scientific computing solutions, especially those supported by the DMSC

### Responsibilities

- Liaise with the relevant instrument teams and technology groups to establish and maintain an effective interface with DMSC activities
- Act as expert in software, data processing, data science, simulations and scientific computing for relevant instrument teams
- Work with the user community to find suitable solutions for their scientific computing needs
- Represent DMSC at relevant instrument STAP meetings
- Coordinate and share knowledge with other instrument data scientists and DMSC technology groups to facilitate sustainable solutions, avoid duplication of work, and foster a collaborative spirit
- Contribute to the development of open-source software where needed and appropriate
- Oversee, review, evaluate and take ownership of DMSC deliverables to relevant instruments, specifically the full data processing pipeline
- Act as the user community's advocate and assume an active approach
- Provide constructive feedback to DMSC development teams where relevant
- Maintain and update roadmap and / or project schedule for DMSC deliverables to relevant instruments

- Conduct scientific or technical research and author scientific or technical publications in the field of scientific computing and neutron scattering
- Take part in peer review of scientific publications, grant proposals, research, and facilities
- Represent ESS at appropriate conferences and workshops
- Establish and maintain constructive contacts with the research and innovation community
- Present scientific and technical research at relevant conferences and workshops, and in seminars and lectures
- Supervise postdocs and students and participate in teaching, as appropriate
- Keep relevant instrument models up to date at an appropriate level of fidelity
- Spend up till 20% of time on local contacting for user experiments in alignment with instrument team
- Support users with treating, analyzing, modelling and simulating data from experiments at ESS in alignment with instrument team
- Provide training on the DMSC software suite to instrument teams and users

## Accountabilities

- To Line Manager
- To Project Manager for reporting on deliverables
- To Instrument Teams for providing expert advice on scientific computing
- To fellow staff, for effective communication, safe and professional conduct

## Authorities

- Initiate and request resource for activities as needed to meet the responsibilities;
- Manage assigned resources to complete assigned tasks and responsibilities;
- Identifying personal training and development needs to stay current with state of the art technology.
- Delegated responsibilities from Line Manager