# **Developers’ discussions**

Discussion were spread throughout the developers’ meeting, more general discussions, and the Code Camp. They are all summarised together.

## **Centring**

Martin Savko’s Murko software is being deployed outside SOLEIL. Program functionality and transferability appears to be good, and there is a consensus to recommend the program. Murko produces a 3D representation of the crystal, as either a point cloud or a 3D mesh, and MXCuBE should make space to capture and store this information.

## **Remote operation**

Remote operation is dominant. SLS is 100% remote, and BioMAX, P14 and Elettra are 70-80% remote, with only local groups working in person. At P14 in-person operation is limited to TREXX and imaging users. ANSTO plans to offer remote operation.

## **Queues and processing of multiple data collections**

This topic was raised by Jacob Oldfield, who suggested the need for some kind of (data) structure to support this. This led to a more general discussion of the queue system. The current queue has the DataCollectionGroup as a unifying object, but (RF notes) it is not clear what the restrictions are on the use, interpretation, or storage (e.g. in ISPyB) of these.

The ESRF/Web version is moving to a Pydantic-defined queue, which would give a welcome definition to the data necessary. RF raises the question of the relationship between the queue-defining data structures in mxcubeweb and mxcubecore, the exact actions that control the content of the queue, and how precisely each of these are documented, as this is an area that is very complex to figure out from analysing the code.

RF raised the possibility of adopting the MXLIMS experiment description for also specifying the queue, but such a move was seen as premature.

## **Interleaving and HDF5 files**

Martin Savko raised the problem of interleaved experiments in the context of HDF5 files.

With complex interleaved experiments, HDF5 leads to name clashes or file overwriting, with successive wedges of an interleaved sweep being prone to name clashes. The problems are (1) that HDF5 master files assume homogenous images (e.g. same energy for all), (2) that there is no good mechanism for going from overall image number (equivalent to omega value) to identify a particular image, and (3) that processing programs like e.g. XDS requires data given as a single monotonic sweep. This can be handled without too much trouble using mini-CBF files, but a complete and precise description of the experiment in current HDF5 is not possible without storing some information outside HDF5. Processing this kind of data will therefore require some kind of repackaging, e.g. to one master file per sweep, that will unavoidably be missing some information (as a minimum, the original acquisition order).

MXCuBE does have an interleaving mechanism, either through the DataCollection dialogue or the Web version, but it has not been used for a long time. In practice, interleaving is done by separate collections per scan (MS), and MXCuBE does not know that they go together. MAX IV has produced a script for dealing with interleaved data in an HDF5 context, but it is old, not often used, and cannot handle all possible experiments.

It is suggested that MXCuBE itself should be doing the repackaging of master files, probably by generating additional master files to point to the stored data files. This could be done either prior to the experiment (on the assumption that everything will work), or after the fact.

## **Cyber security**

MAX IV bases their security on VPNs, while several other sites use two-factor authentication instead, and ANSTO explicitly wants to avoid VPNs.

The main precaution that MXCuBE can take in this regard is to make sure that all dependencies are kept updated. The partners in the MXCuBE Collaboration should continue to share experiences about Internet exposure.

Sites generally seem less worried about security than at the last meeting.

ANSTO is planning an external security review, complete with penetration testing, once their systems are ready. ESRF is also considering an external review. Results from these tests would be of great interest to other sites.

## **Developer organisation**

It was proposed to encourage developers to join into subgroups working on particular topics, and to consider tools that would make this easier, such as Slack. The meeting was in favour of such cooperation but recommended that communication should continue through the centralised channels of GitHub Issues and Discussions, the regular mxcubecore and mxcubeweb developers’ meetings, and the mailing list. This reflected a concern that explicit small-group communication entailed too high a risk of information fragmentation, and that more real-time tools like Slack were not well suited to a collaboration with members widely spread across time zones.

## **Additional targets**

To be discussed at a round table at the next developers’ meeting:

* Each facility should provide conda environments and version restrictions.
* We should make PRs for documentation.
* Better test coverage