Py-ISPyB developers' meeting
6 Oct 2022

DRAFT

Participants:

- Alessandro de Maria, Mael Gaonach, Olof Svensson, Marjolain Bodin(ESRF)
- Karl Levik, James Hall, (Diamond)
- Rasmus Fogh, Gerard Bricogne, (Global Phasing)
- Alberto Nardella (MAX IV)
- Idrissou Chado (SOLEIL)
- Annie Heroux (Elettra)

Minutes: Rasmus Fogh

SSX Updates

Discussion of DB needs for SSX. GB pointed out that SSX had more and different processing needs, and required more care than rotation data. The situation was more similar to XFEL. ESRF ID29 has done some work on how best to do this. GB suggested a need to rethink the database structure to support SSX. Now that SSX was being rapidly developed would be a very good opportunity. A PDB mmCIF working groups has recently done considerable modelling work to support SSX and (especially) XFEL, which is already done and prototyped. It would be a pity to ignore the opportunity to establish an ISPyB model and vocabulary that was synchronised with this work. Global Phasing (Clemens Vonrhein, Peter Keller, GB) volunteered to serve as go-betweens; the people to contact on the wwPDB side were particularly Aaron Brewster and Tom White.

AdM proposed a talk on this work at the next half-yearly ISPyB meeting, and GB promised to take this further (**ACTION**: GB). It would require setting the time of the talk at late afternoon / evening to allow participation of US-based people, ca. 1700 CET.

A further discussion covered the use and limits of the DataCollectionGroup for multiple sweeps and multiple samples. The original use of the DC Group was to group multiple data collections on the same sample. For SSX it is now used to make groups of, say, 10000 images from different crystals. OS noted that the DC Group had several problems in practice. For workflow work it was used to combine all data collected on a given sample, including characterisation images, and this made it unsuitable to use for multi-sweep processing. In a recent discussion with Stuart Fischer (unfortunately absent from this

meeting) it had been proposed to start using some tables from Diamond for this purpose. ProcessingStates? But the main limit is in the GUI. GUIs had until now been organised with a result for every sweep rather than for multi-sweep processing, and the overriding division of data had been per sample. This did not match either multi-sweep work or multi-sample data combinations, and it had been considered to reorganise the GUI to work by Project. Some work in this direction has been done by MG at ID29, but these changes are urgent for the Global Phasing workflow, which is currently working on MASSIF-1. The use of the Project table for this purpose was not without problems, since 1) the obvious meaning of 'Project' was 'everything related to a given macromolecule', and 2) Diamond was using the table for something different. There had been discussion at Diamond about adding a ProcessingGroup table to combine data collections that were processed together, but no tables had been written yet.

It was decided to interrupt the discussion and arrange a single-topic meeting specifically to deal with this problem. **ACTION**: AdM.

MySQL Python connector and MariaDB

KL raised the issue that the MySQL python connector used so far was no longer compatible with MariaDB, starting from version 8.0.30. MariaDB started as a fork of MySQL, but the two have now diverged so the same connector cannot be used. There is no prospect of remedying this, as MySQL does not see MariaDB compatibility as a priority. KL proposed changing over to the MariaDB connector, even though this would require a certain amount of work. The question was raised whether a MariaDB connector would work with MySQL. AN note that at MAX IV they used MySQL, as he had not been able to get MariaDB to work. MAX IV had absolutely no programmer resources to use for making such a move. It was agreed to pin the MySQL python connector to version 8.0.29, in the expectation that MAX IV would become able to make a move before the situation became untenable.

Latest developments, by MG

- One-time tokens have been introduced to allow downloading of large files in a single click without requiring separate authorisation setting. They are single-use only, and expire after a very short time, on the order of ten seconds.
- A BeamlineGroup has been introduced to allow setting permission to access all sessions on a particular beamline (e.g. for synchrotron personnel).

Nobugs feedback

People from LNLS Brazil had been discussing which of the various ISPyB interfaces to adopt, and MG had recommended PyISPyB as the most future-proof. The Brazilians had been visiting widely, and other alternatives had also been presented to them.

People at Nobugs had been surprised at the number of alternative GUIs for ISPyB (Original,

SynchWeb, EXI, EXI2, now PyISPyB). How might this number be reduced? OS considered this to be unfair, since the GUI had not been part of the collaboration. He saw it as an argument for the proposition that the GUI **should** be part of the collaboration, building the shared GUI in PyISPyB.

JH noted that the problem had been a very tight coupling between front end and back end in the past. He had been impressed by Stuart Fischers talk about Daiquiri, at Nobugs, and was strongly in favour of a modular GUI, ideally with new panels configurable by YAML.

AdM noted that the GUIs do not need to **look** the same at different sites, as long as the underlying code and technology was shared, and agreed that the GUI should be(come) part of the collaboration. JH was in favour, but noted that changes were hard. It was very difficult to get qualified personnel (one co-worker had recently left for the private sector at twice the salary). Even though the current code base was very opaque and hard to induce people to there was still a temptation to avoid the work of changing and to keep extending the existing system. Given a modular and extensible alternative there would be a good case for change. One area that might be ripe for a PyISPyB GUI was tomography, which was new but prioritised at several synchrotrons.

Feedback from Digital LEAPS-STAR:

AdM had been at the meeting, and reportedting ideas around creating a DoI permanent identifier for samples, so thar a sample could be tracked through its life history and eventually referenced also in publications. CRIMS had already done work that could be used in this direction. AH noted as a complication that pins with bar codes had been available for quite a while, but that people tended to reuse the barcodes for different samples, and generally not keep track of the codes.

Next half-yearly meeting

The next meeting is set for December 6-7. Some topics for the meeting have already been raised (above). In addition it was proposed to have a hands-on session on installing ISPyB, and a brainstorming session on the kind of UI one would want for the future.

Any Other Business.

None

Next Meeting

ACTION: AdM