

ISPyB Collaboration Meeting @ ALBA

30/11/2023



| The European Synchrotron

**Andy Götz on behalf of
Data Automation Unit +
SB Group @ ESRF**

- 1. State of the collaboration**
- 2. Good news (for everyone?)**
- 3. What c/should we collaborate on?**

HAVE WE REACHED A DEAD END?



WE REACHED A FORK – WE NEED TO CHOOSE



STATE OF THE ISPYB COLLABORATION

- 1. Good attendance at meetings but ...**
- 2. A lack of developers to contribute code**
- 3. No common software core shared amongst all partners**
- 4. The collaboration partners do not have the means to address the needs of a critical scientific domain**

STATE OF THE ISPYB COLLABORATION

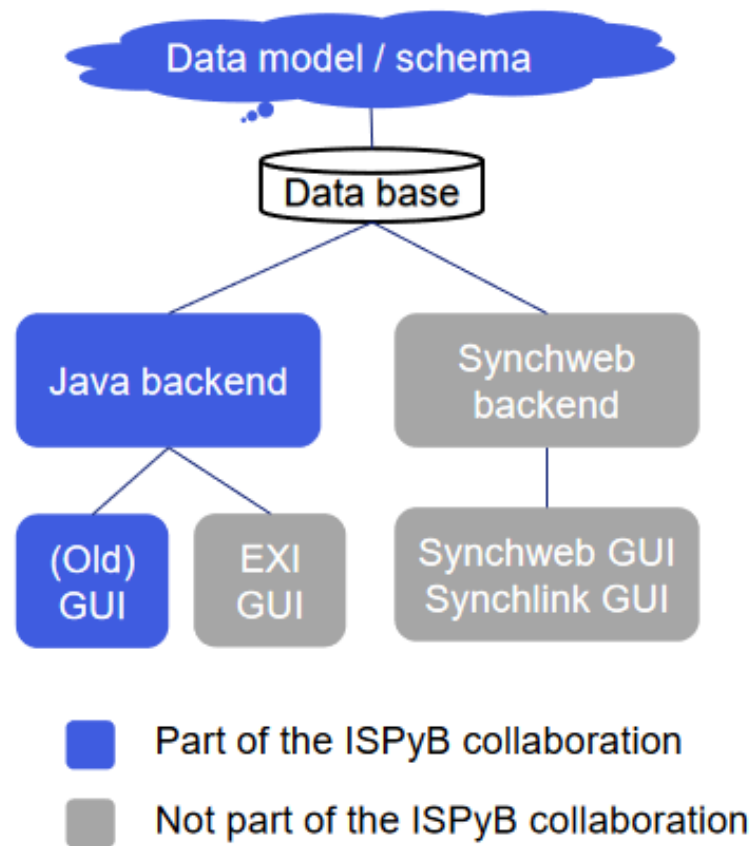
Olof's presentation in Lund 03/2019

- **The ISPyB collaboration is about:**

- The database schema / data model
- The Java backend / API of web services

- **ISPyB GUIs :**

- **The official GUI** is part of the Java backend, however, it is in the process of being phased out
- The DLS GUIs **Synchweb** and **Synchlink** are not using the Java backend and are not part of the collaboration
- The **EXI** GUI developed at the ESRF and used by several facilities is using the Java backend and is not part of the collaboration



STATE OF THE ISPYB COLLABORATION

EMBL Hamburg February 2020

Extra-ordinary ISPyB strategy meeting @ EMBL-HH
(Hamburg, Germany), 12 February 2020

Overview presentations

- Introduction - Gleb Bourenkov
- ISPyB strategy - Gianluca Santoni
- SynchWeb for MX Diamond - Dave Hall
- ISPyB @ EMBL HH MX - Gleb Bourenkov
- Global Phasing's viewpoint on the ISPyB Collaboration - Gerard Bricogne
- ISPyB status at SOLEIL
- ISPyB at MAX IV - Jie Nan

Technical presentations

- ISPyB Collaboration - Alejandro de Maria Antolinos
- ISPyB Back end evolution - Neil Smith

<https://ispyb.github.io/ISPyB/>

- *All participants agree to go forward together and to try and make the collaboration work.*
- *The Collaboration intends to work towards a common framework, with as much shared code as possible, and try to modularise.*
- *The immediate target is to develop a shared backend based on the existing database.*

... a pilot proposal should be made on an entirely new, future application for both sites to use. In order to work, this would require:

- *Precise up-front agreement on the scope of the project, major technical choices, and decisions about the adoption of existing standards (such as ICAT)*
- *Inclusion of the full code stack, including the user interface, within the collaboration.*
- *Empowering the developers to take decisions on technical questions*
- *Sufficient resources*

Martin Walsh *comments that different sites will unavoidably have different priorities, pressures, and requirements, and that any agreement must allow each site to cater for those.*

https://ispyb.github.io/ISPyB/webpages/Hamburg_20200212/Minutes_20200212.pdf

- *The two main development centres will stay with their separate implementations and work towards increased sharing; there are no plans to abandon either implementation or to start a completely new alternative for now.*
- *One view of the upcoming work, promoted by **Martin Walsh**, emphasised producing useful results that would add value at both ESRF and DLS*

REASONS WHY ESRF STARTED AFRESH

- 1. Sustainability**
- 2. Lack of developers**
- 3. Technical debt**
- 4. New start in Python**
- 5. Duplication of metadata**
- 6. Common solution**
- 7. Easier to add new features**
- 8. Idempotent for ISPyB collaborators**

<https://cloud.esrf.fr/s/DJo9n8oG2ncZeMQ>

RESOURCES FOR EVALUATING ICAT @ ESRF

- **Evaluation took 9 months**
- **2 Data Managers (Alex, Marjolaine) (fullstack developers)**
- **2 Data Engineers (Olof, Mael) (fullstack developers)**
- **EWOKS workflow team for processing data**
- **MXCuBE team (Marcus, Antonia)**
- **ICAT metadata catalogue (STFC) + ICATplus (ESRF)**
- **H5Web for Nexus/HDF5 file viewing (Axel, Loic)**
- **Team of SB Scientists expert in MX (Max, Matthew, Didier, Romain) and EM (Isai, Gregory)**
- **Global Phasing expertise (Gerard, Rasmus, Clemens)**

OUTCOME OF THE EVALUATION OF ICAT

- **Scientific Use Cases**
 - CryoET + MX successfully implemented
- **Developers**
 - Easier to implement due to [Key,Value] pattern for storing metadata
 - Faster development time
 - Less technical debt
- **Scientists**
 - Improved user interfaces
 - New features implemented
 - Unblocks the potential of data processing for structural biology
- **Sustainable solution**

*The proof of the pudding
(is in the eating)*



NEW IDEAS FROM SCIENTISTS – MAX, MATT, ROMAIN

- *Merging multiple datasets into a single dataset*
 - *Aimless is one of the available tools, but the way it works is that each dataset is indexed and integrated (also, the indexing needs to be checked for consistency in some cases) and then aimless works at the level of integrated data -- intensities.*
 - *In this case they belong to the same crystal, but there are other cases like mesh and collect in which they will come from different crystals, and in other cases even different sample holders (this latter might be a good example for you since I think it would require major changes to ispyb to implement).*
- *for the GPL workflows the workflow figures out whether to collect multiple oscillations with different kappa angles, then merges them (with aimless?). In this case, we need to make it easy to ingest this data, keep track of this relationship in the database AND display it*
- *Link data from different sites (a user would connect to icat and would be able to fetch data from, e.g. different synchrotron sources)*

- We should be discussing what features the experts are asking for, but ...
- Instead we are discussing which software stack to use
- We can't continue like this ...

ADOPTING MX-ICAT

- 1. ESRF to package ICAT + ICATplus (containers + documentation)**
- 2. ESRF to provide frontend for MX (and SSX, EM, ...)**
- 3. Partners to install MX-ICAT and push metadata to ICAT from MXCuBE and processing programs**
- 4. Similar to today but collaboration includes full software stack**
- 5. Partners to contribute to (1) scientific data model (scientists) (2) scientific processing (scientists+engineers) and (3) backend+frontend software (engineers)**

ALL members of the collaboration

- Otherwise it is not a collaboration but a Users forum
- How to contribute
 - **Software:** Hire a frontend or backend developer to help develop the software
 - **Scientist:** Contribute to the scientific use cases

1. Share the full software stack

- MX on ICAT / MX on PY-ISPYPB / SynchWeb

2. Share only the scientific data model

- Specified as a CIF Dictionaries e.g. mmCIF
See <https://mmcif.wwpdb.org/>

OPTIONS FOR THE FUTURE

- 1. Stay with Java backend + EXI frontend – no new features**
- 2. Adopt the py-ispyb backend + frontend – develop MX backend**
- 3. Adopt ICAT backend + frontend - profit from new features**
- 4. Adopt/Stay with SynchWeb – profit from refactoring**
- 5. Adopt SciCat – develop MX frontend**
- 6. Other – develop / adopt another backend + frontend**

THE FUTURE DEPENDS ON YOU + US

- **The Collaboration needs to make a choice what it wants to collaborate on:**
 - 1. Scientific Data model**
 - 2. Single features e.g. shipping**
 - 3. Full software stack per technique**
- **If full software stack then multiple collaborations can/will be formed**
- **Collaborate on definition of scientific metadata**



NAME FOR THE NEW ONTOLOGY

mmCAT



ESRF Data Portal



SynchWeb