# ESRF Experiment Division Tuesday Events
## 15/09/2009


# From your sample to your data analysis: how to track every step of your experiment in a database. An example with ISPyB for MX experiments

**Patrice Brenchereau ESRF/CS/MIS**

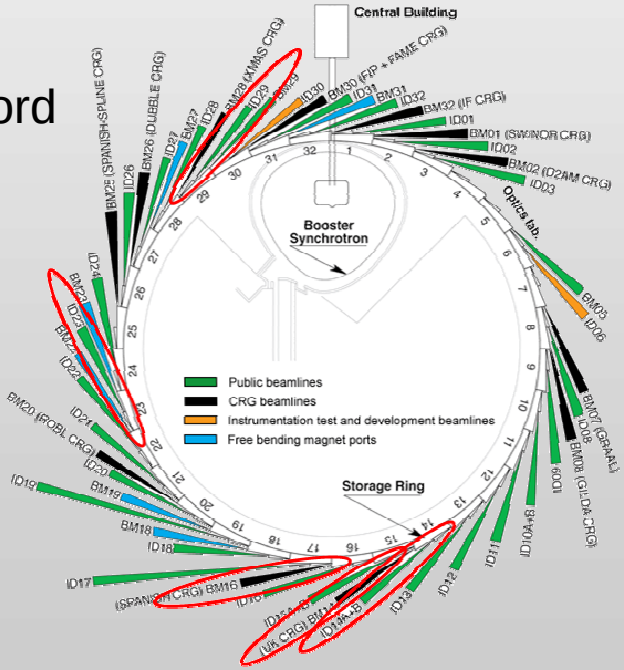
# Contents

- Tracking, experiments, samples, data analysis…
- What is ISPyB?
- Features
- Architecture
- What's next?

# From your sample to your data analysis: how to track every step of your experiment in a database…

# What is ISPyB?

- ISPyB: **I**nformation **S**ystem for **P**rotein Crystallograph**Y** **B**eamlines
  - Laboratory Information Management System (LIMS) for protein crystallography experiments on synchrotron beamlines
  - Web based interface
- Access
  - http://ispyb.esrf.fr
  - To log-in: ESRF experiment number and password
- Where is it used at the ESRF?
  - ID14-1, ID14-2, ID14-4
  - ID14-3 soon (BioSAXS)
  - ID23-1, ID23-2, ID29
  - BM14, BM16



# Proposals and Users

- Type of Proposals:
  - IX: Industrial Users who come to the ESRF
  - FX: Industrial Users who don't come to the ESRF (MXPress Service)
  - MX: Academic Users (mainly BAG / Block Allocation Group)
- ISPyB Users:
  - End user (users on site or remote, Esrf local contact)
  - Local Contacts (dewar tracking)
  - Fx Managers (managing Industrials)
  - Bloms (managing beamlines)
  - Stores (dewar tracking)

# History

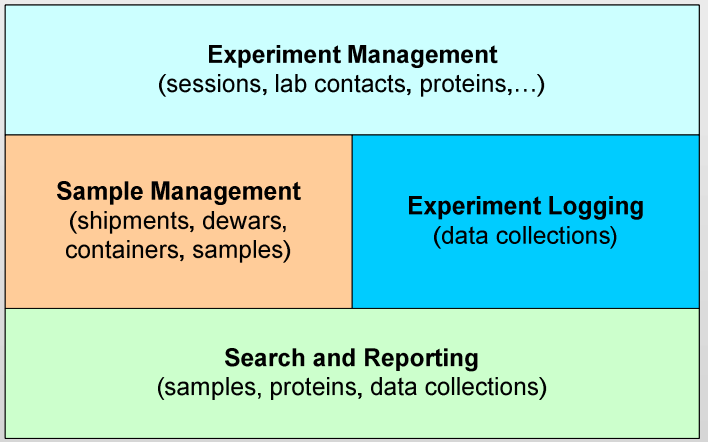
- 2001-2005: Pxweb
  - Python/Zope, MySql
  - JSBG Project (ESRF/EMBL)
  - Experiment logging for MXPress users (FX users)
- 2005-now: ISPyB
  - Java/Struts/Jboss, Mysql
  - ESRF/Spine/Bioxhit and BM14/MRC/eHTPX collaboration
  - Pxweb features + new features + Dewar Tracking
  - 2009: Collaboration with Diamond (code sharing)

# ISPyB Project Management

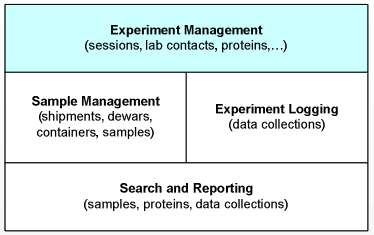
- Management Structure
  - Steering Committee
  - Executive Committee
  - Scientific and User Committee
  - Project Manager
  - Developers
- Meetings
  - ESRF / DLS once a month (Marratech)
  - Synchrotrons: every 6 months / 1 year (last was April 2009)
- Source Code
  - Shared on SourceForge / SVN
  - Bug & New Feature tracking
  - Site specific statements : is_ESRF(), is_DLS()

# ISPyB Features & Modules

- Experiment Management
- Sample Management
- Experiment Recording
- Search and Reporting



# Experiment Management




- Items:
  - Proposals
  - Sessions (date, beamline, local contact)
  - Lab Contacts (email, address,..)
  - Proteins (approved sample sheets)
- Features:
  - Data retrieved from the User Portal (SMIS)
  - Edit Lab Contact Card

# Lab Contacts

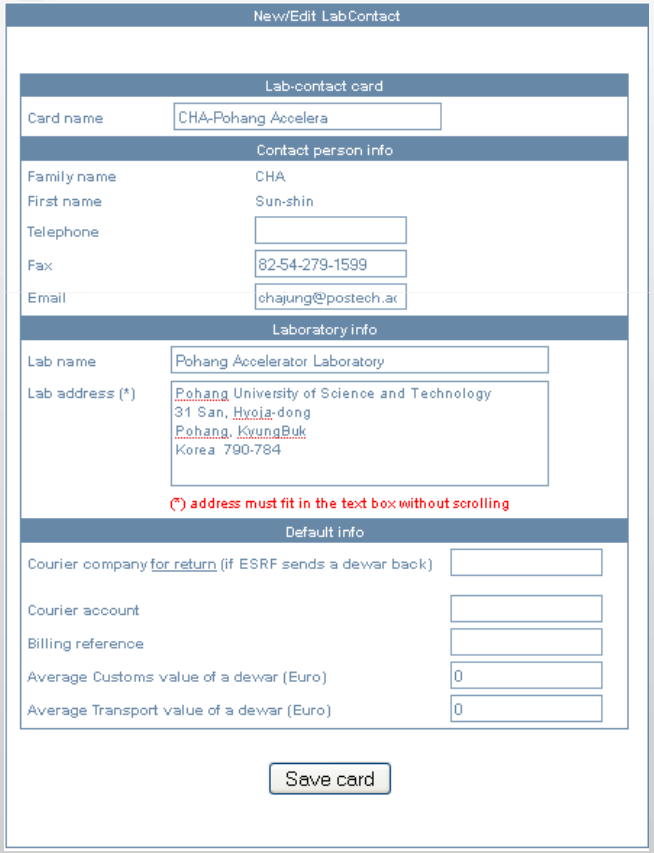
## Search


## Select

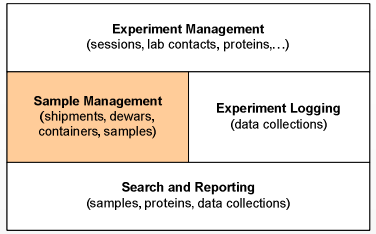
Select the scientist contact

| Scientist name | Scientist first name | Lab name | Action |
|---|---|---|---|
| TURKENBURG | J.P. | York Structural Biology Laboratory | Select |
| LEONARD | Gordon | E.S.R.F. | Select |
| MONACO | Stéphanie | E.S.R.F. | Select |
| TERRADOT-PIOT | Laurent | E.S.R.F. | Select |
| THIBAULT | Xavier | E.S.R.F. | Select |
| SPRUCE | Darren | E.S.R.F. | Select |
| GUIJARRO | Matias | E.S.R.F. | Select |
| PIPELINE | Data | E.S.R.F. | Select |
| CHA | Sun-shin | Pohang Accelerator Laboratory | Select |
| KIM | In-kwon | Department of Chemistry | Select |
| KIM | Min-Kyu | Laboratory of Biophysics | Select |
| AN | Young jun | Korea Ocean Research & Development Institute | Select |
| SEO | Koung hye | Department of Chemistry | Select |
| LEE | Kon ho | Gyeongsang National University | Select |
| KOSKI | Kristian | Department of Biochemistry | Select |
| HAAPALAINEN | Antti | Department of Biochemistry | Select |
| BUJACZ | Grzegorz | Institute of Bioorganic Chemistry | Select |

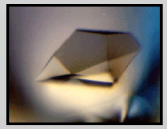
## Edit



# Sample Management

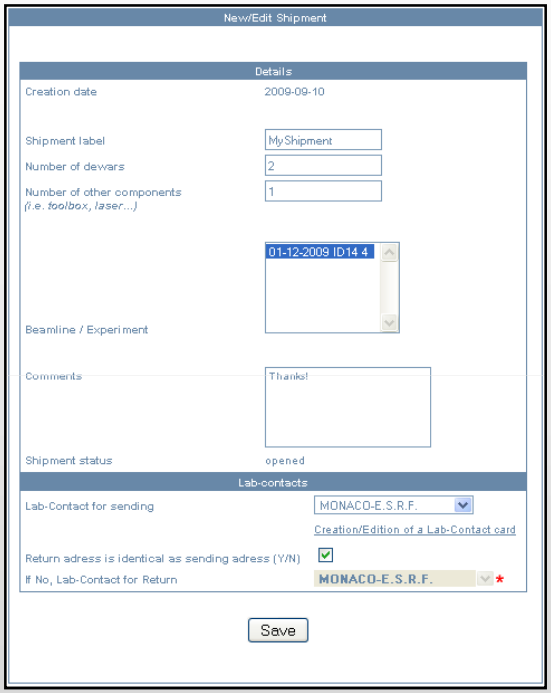



- Items:
  - Shipments (set of dewars)
  - Dewars (barcode, courier tracking number)
  - Containers (barcode)
  - Samples (barcode, protein acronym, crystal form,…)
- Features:
  - Online shipment description
  - Shipment description upload (Excel file)
  - Dewar Tracking

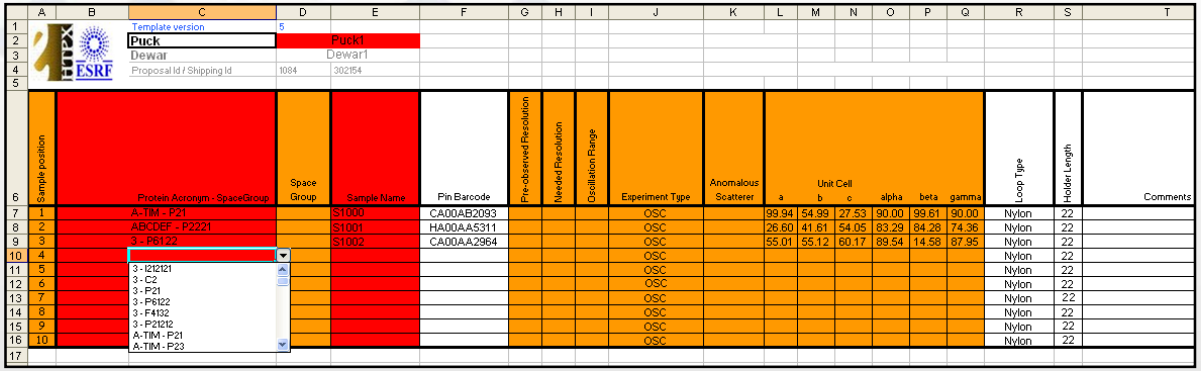


# Creating Shipments

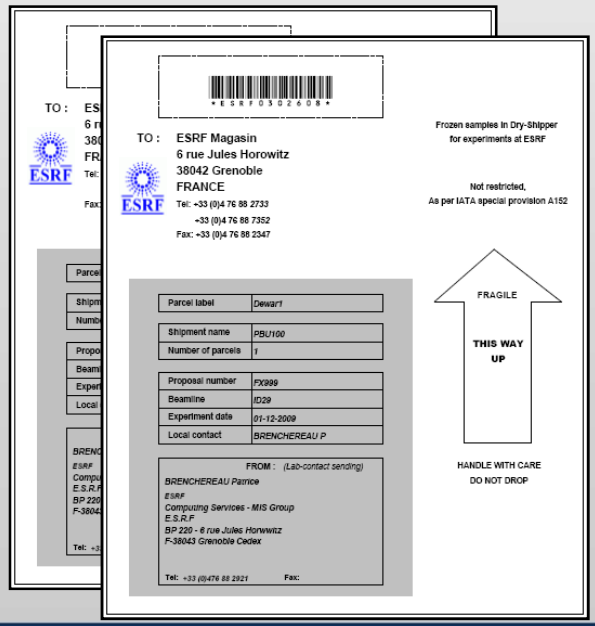
## Shipment Creation



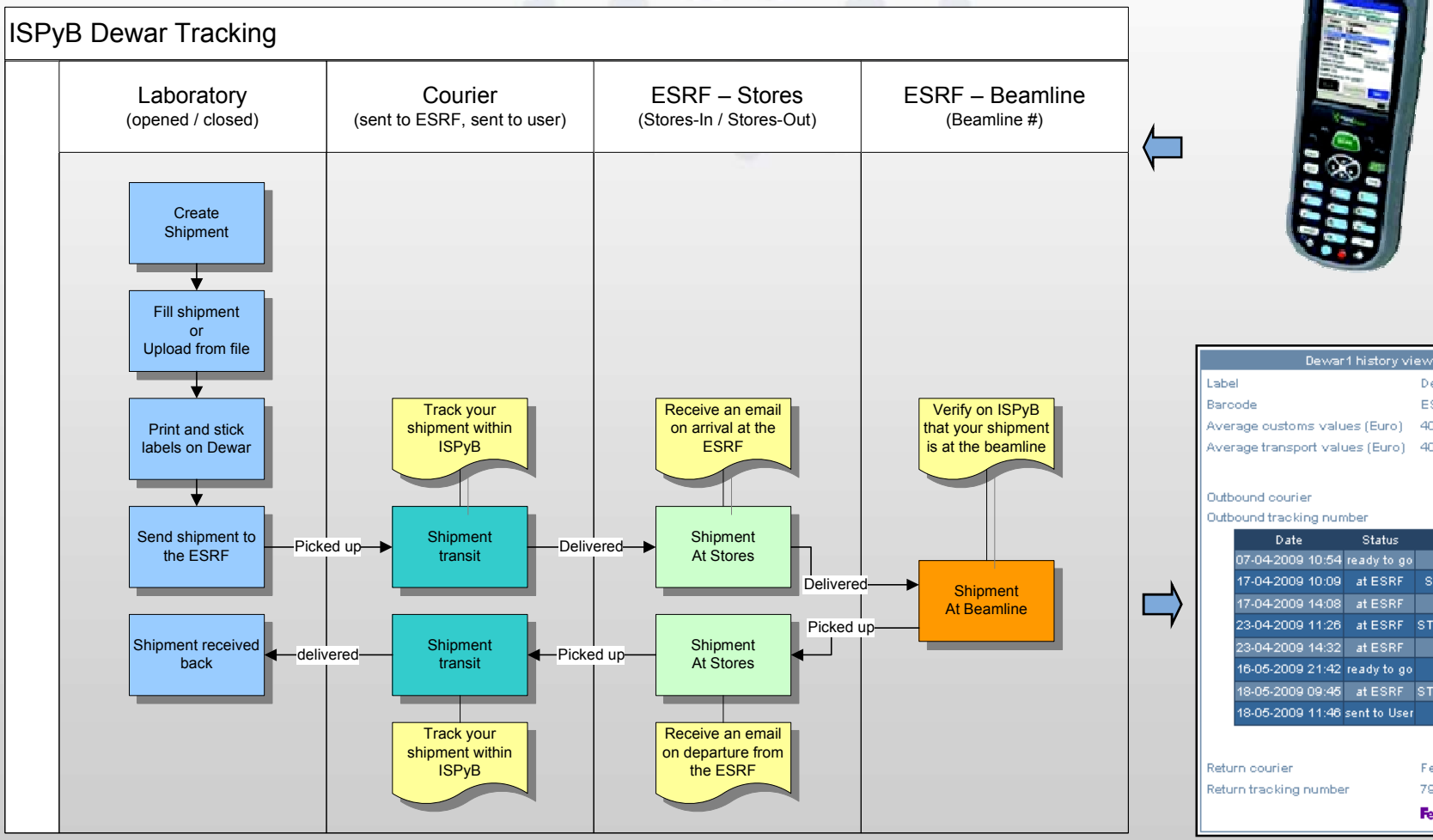

## Shipment Description (Excel)




## Dewar Labels (sending, return)



# Dewar Tracking






Dewar1 history view

| | |
|---|---|
| Label | Dewar1 |
| Barcode | ESRF0302349 |
| Average customs values (Euro) | 4000 |
| Average transport values (Euro) | 4000 |

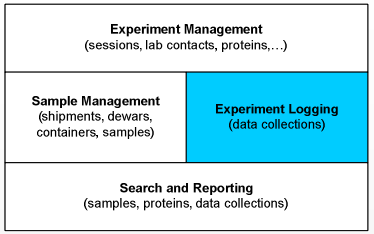
Outbound courier

Outbound tracking number

| Date | Status | Location |
|---|---|---|
| 07-04-2009 10:54 | ready to go | |
| 17-04-2009 10:09 | at ESRF | STORES-IN |
| 17-04-2009 14:08 | at ESRF | ID14 |
| 23-04-2009 11:26 | at ESRF | STORES-OUT |
| 23-04-2009 14:32 | at ESRF | ID14 |
| 16-05-2009 21:42 | ready to go | |
| 18-05-2009 09:45 | at ESRF | STORES-OUT |
| 18-05-2009 11:46 | sent to User | |

| | |
|---|---|
| Return courier | Fedex |
| Return tracking number | 796614000970 |

# Experiment Recording




- Items:
  - Session info (date, beamline name, operator,…)
  - Experiment parameters (wavelength, energy, detector distance,…)
  - Beamline parameters (name, beam transmission, beam size,…)
  - Crystal snapshots
  - Images (thumbnails, link to real images)
  - Data Analysis / EDNA results (characterisation & strategy, mosflm.log, pointless.log,…)

# Data Collections

**Data Collections of a Session**

Data Collections | Energy Scans

| Image Prefix | Run No | Protein Acronym | Start Time | # images | Wavelength | Transm. | Ex. Time | Phi start | Phi range | Detector Resolution | Status | Sample Ranking | Skip | Comments |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | All | | | | | | | | | | Rank | | Save |
| ins | 2 | | 08-07-2009 17:15:58 | 14 | 0,976 | 100 | 0,34 | 115,95 | 0,3 | 1,54 | | | | |
| ins | 1 | | 08-07-2009 17:14:03 | 57 | 0,976 | 100 | 0,1 | 96 | 0,35 | 1,54 | | | | |
| ref-ins | 2 | | 08-07-2009 17:05:01 | 2 | 0,976 | 100 | 1 | 0 | 1 | 1,5 | | | | Collecting 2 reference images |
| ref-ins | 1 | | 08-07-2009 16:54:21 | 2 | 0,976 | 100 | 1 | 0 | 1 | 1,7 | | | | Collecting 2 reference images |
| FAE-X35 | 4 | | 08-07-2009 14:42:52 | 21 | 0,976 | 100 | 0,65 | 112,95 | 0,5 | 1,7 | | | | |
| FAE-X35 | 3 | | 08-07-2009 14:41:33 | 31 | 0,976 | 100 | 0,4 | 92,8 | 0,65 | 1,7 | | | | |
| FAE-X35 | 2 | | 08-07-2009 14:40:39 | 22 | 0,976 | 100 | 0,31 | 63 | 0,9 | 1,7 | | | | |

This collection has a crystal snapshot.
collection successful
DNA indexing failed.
No samples for the data collection.

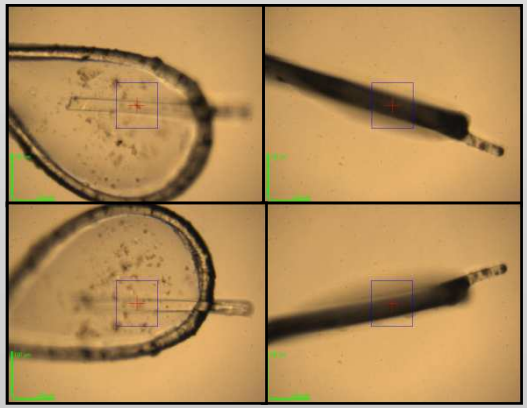
# Data Collections

## Parameters & Results

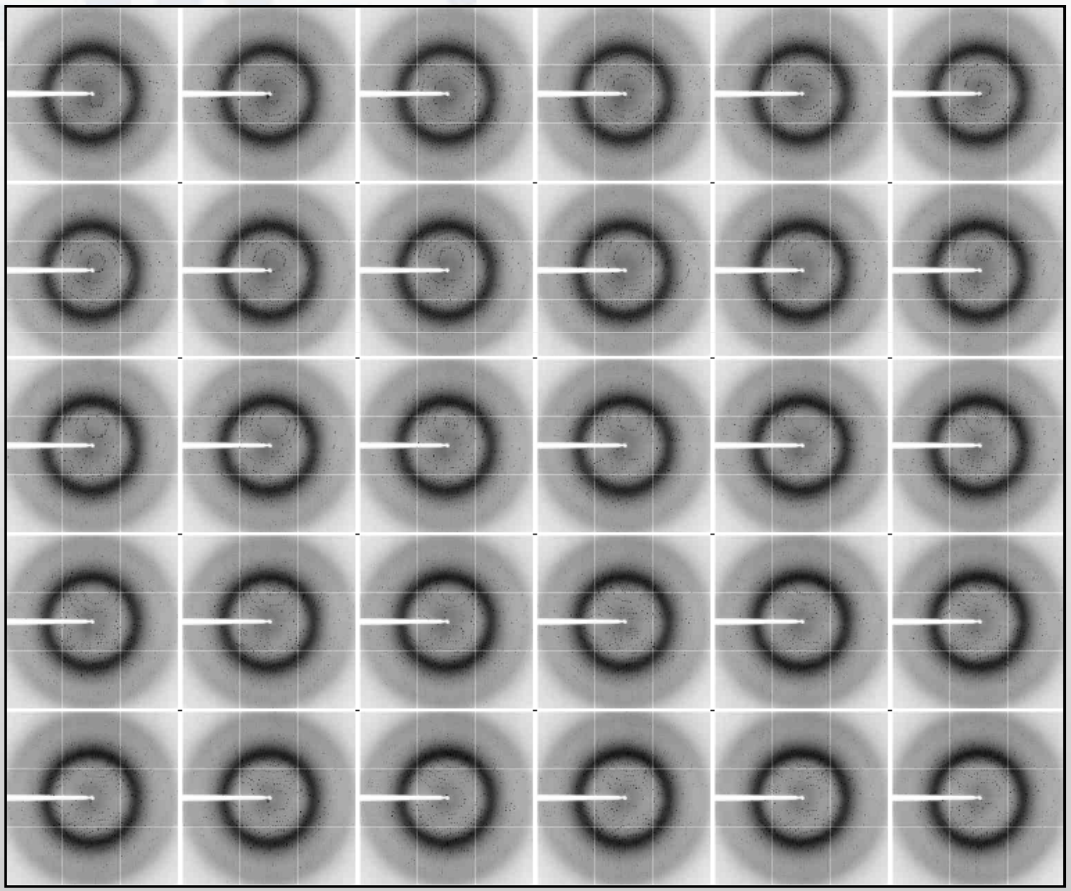
Experiment parameters | Beamline parameters | DNA results

| | |
|---|---|
| Img directory | /data/visitor/mx415/id14eh4/20090909/SAD |
| Img prefix | stand-fae |
| Run no | 3 |
| Start Time | 09-09-2009 11:27:53 |
| End Time | 09-09-2009 11:39:01 |
| Type of experiment | |
| Wavelength | 0.9795 Å |
| Energy | 12.658 keV |
| Phi start | 167.0° |
| Oscillation range | 0.75° |
| Overlap | 0.0° |
| Exposure Time | 0.5 s |
| Number of passes | 1 |
| Detector Distance | 287.9 mm |
| Resolution at edge | 1.99 Å |
| Resolution at corner | 1.52 Å |
| Xbeam | 157.72 mm |
| Ybeam | 156.91 mm |
| Detector 2theta | 0.0 |
| Kappa | N/A |
| Phi | N/A |
| Experiment comment | DNA data collection |

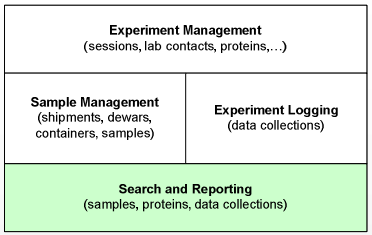
## Crystal Snapshots



## Image thumbnails (image wall)



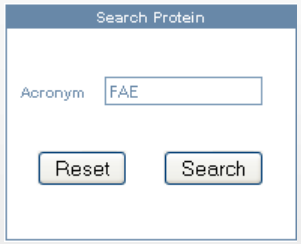
# Search and Reporting

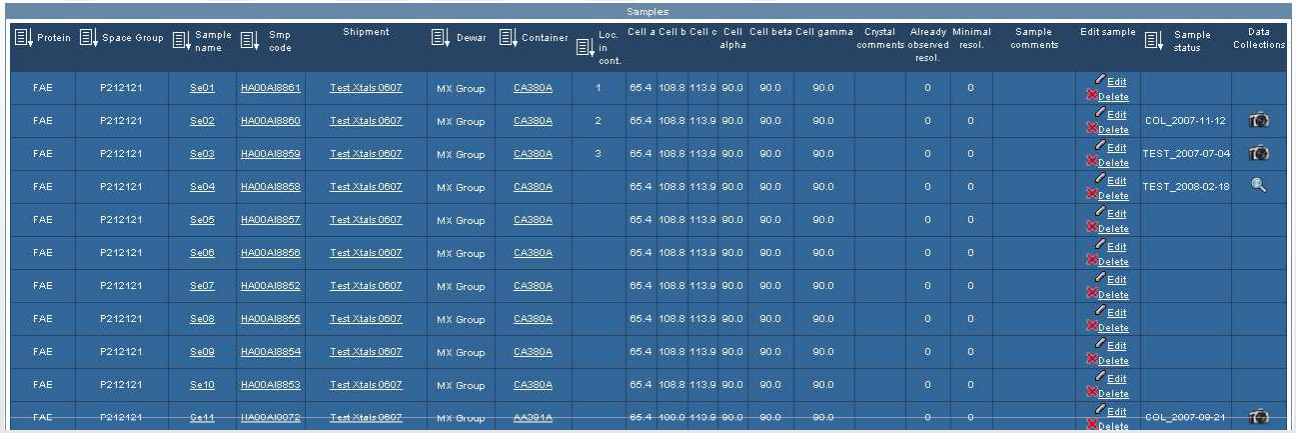



- Search
  - Data Collection (protein, sample, experiment date,…)
  - Proteins
  - Samples
- Reports
  - Pdf, Word, Csv

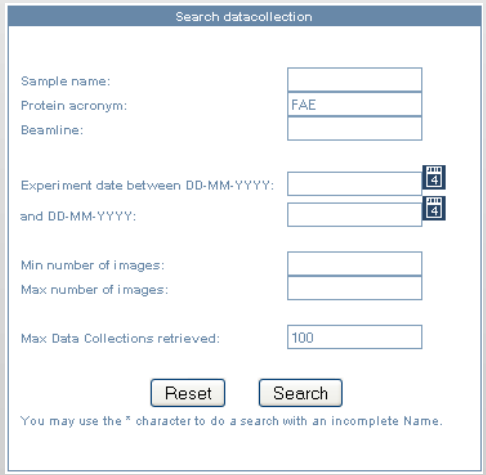
# Search

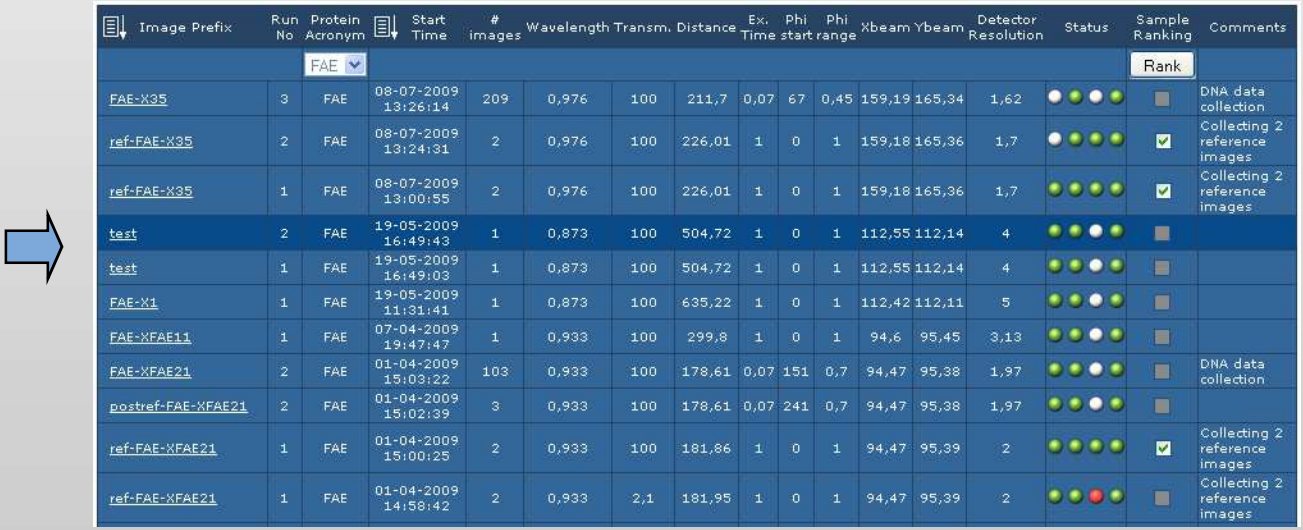
## Protein or Sample



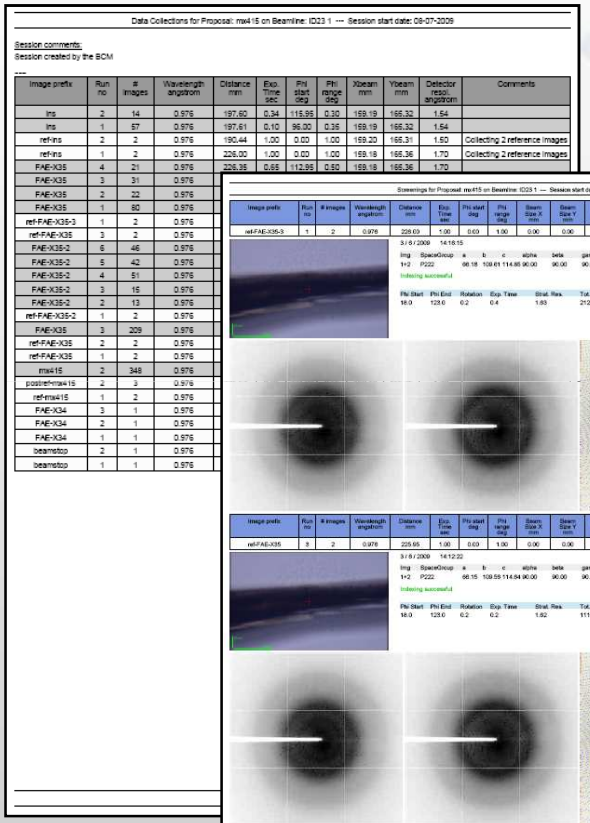

## Data Collection



# Reporting

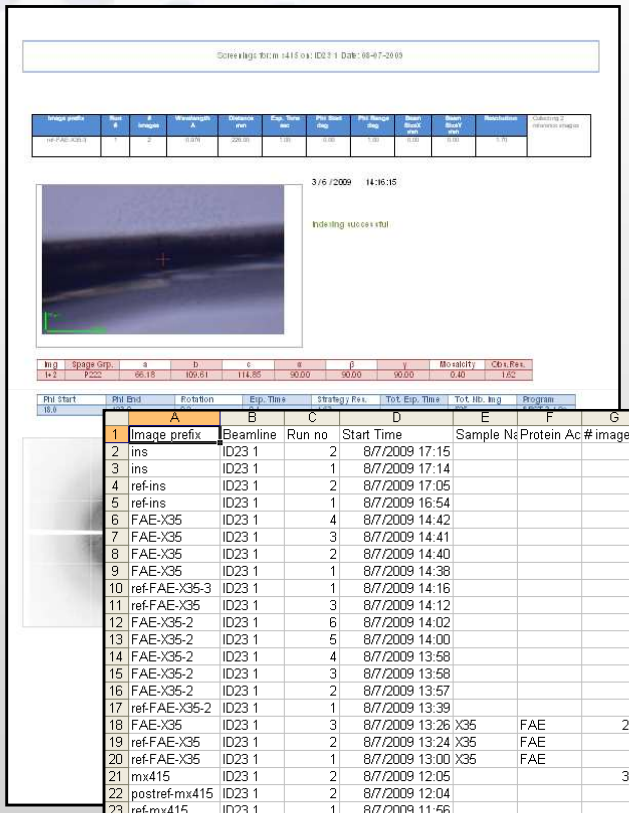
Pdf




Pdf



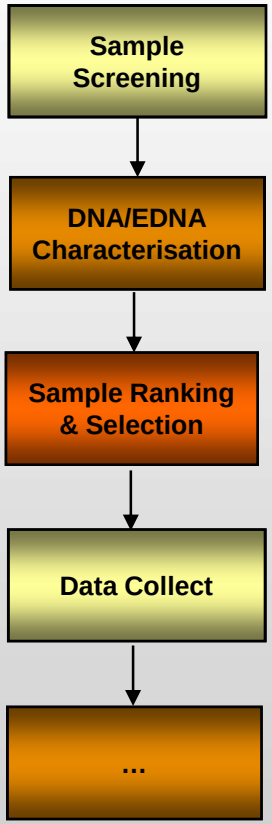
Word



Excel / Csv

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Image_prefix | Beamline | Run no | Start Time | Sample Na | Protein Ac | # images | Wavelengt | Distance ( | Exp. Time | Phi start (d | Phi range |
| 2 | ins | ID23 1 | 2 | 8/7/2009 17:15 | | | 14 | 0.976 | 197.6 | 0.34 | 115.95 | 0.3 |
| 3 | ins | ID23 1 | 1 | 8/7/2009 17:14 | | | 57 | 0.976 | 197.61 | 0.1 | 96 | 0.35 |
| 4 | ref-ins | ID23 1 | 2 | 8/7/2009 17:05 | | | 2 | 0.976 | 190.44 | 1 | 0 | 1 |
| 5 | ref-ins | ID23 1 | 1 | 8/7/2009 16:54 | | | 2 | 0.976 | 226 | 1 | 0 | 1 |
| 6 | FAE-X35 | ID23 1 | 4 | 8/7/2009 14:42 | | | 21 | 0.976 | 226.35 | 0.65 | 112.95 | 0.5 |
| 7 | FAE-X35 | ID23 1 | 3 | 8/7/2009 14:41 | | | 31 | 0.976 | 226.35 | 0.4 | 92.8 | 0.65 |
| 8 | FAE-X35 | ID23 1 | 2 | 8/7/2009 14:40 | | | 22 | 0.976 | 226.35 | 0.31 | 63 | 0.9 |
| 9 | FAE-X35 | ID23 1 | 1 | 8/7/2009 14:38 | | | 80 | 0.976 | 226.35 | 0.1 | 33 | 0.5 |
| 10 | ref-FAE-X35-3 | ID23 1 | 1 | 8/7/2009 14:16 | | | 2 | 0.976 | 226 | 1 | 0 | 1 |
| 11 | ref-FAE-X35 | ID23 1 | 3 | 8/7/2009 14:12 | | | 2 | 0.976 | 225.95 | 1 | 0 | 1 |
| 12 | FAE-X35-2 | ID23 1 | 6 | 8/7/2009 14:02 | | | 46 | 0.976 | 251.3 | 0.55 | 103.75 | 0.45 |
| 13 | FAE-X35-2 | ID23 1 | 5 | 8/7/2009 14:00 | | | 42 | 0.976 | 251.3 | 0.37 | 74.35 | 0.7 |
| 14 | FAE-X35-2 | ID23 1 | 4 | 8/7/2009 13:58 | | | 51 | 0.976 | 251.3 | 0.13 | 53.95 | 0.4 |
| 15 | FAE-X35-2 | ID23 1 | 3 | 8/7/2009 13:58 | | | 15 | 0.976 | 251.3 | 0.18 | 44.2 | 0.65 |
| 16 | FAE-X35-2 | ID23 1 | 2 | 8/7/2009 13:57 | | | 13 | 0.976 | 251.37 | 0.1 | 39 | 0.4 |
| 17 | ref-FAE-X35-2 | ID23 1 | 1 | 8/7/2009 13:39 | | | 2 | 0.976 | 226.04 | 1 | 0 | 1 |
| 18 | FAE-X35 | ID23 1 | 3 | 8/7/2009 13:26 | X35 | FAE | 209 | 0.976 | 211.7 | 0.07 | 67 | 0.45 |
| 19 | ref-FAE-X35 | ID23 1 | 2 | 8/7/2009 13:24 | X35 | FAE | 2 | 0.976 | 226.01 | 1 | 0 | 1 |
| 20 | ref-FAE-X35 | ID23 1 | 1 | 8/7/2009 13:00 | X35 | FAE | 2 | 0.976 | 226.01 | 1 | 0 | 1 |
| 21 | mx415 | ID23 1 | 2 | 8/7/2009 12:05 | | | 348 | 0.976 | 225.66 | 0.05 | 131 | 0.55 |
| 22 | postref-mx415 | ID23 1 | 2 | 8/7/2009 12:04 | | | 3 | 0.976 | 225.62 | 0.05 | 221 | 0.55 |
| 23 | ref-mx415 | ID23 1 | 1 | 8/7/2009 11:56 | | | 2 | 0.976 | 242.39 | 0.5 | 0 | 1 |
| 24 | FAE-X34 | ID23 1 | 3 | 8/7/2009 11:50 | | | 1 | 0.976 | 276.56 | 1 | 0 | 5 |
| 25 | FAE-X34 | ID23 1 | 2 | 8/7/2009 11:43 | | | 1 | 0.976 | 276.56 | 1 | 0 | 0.1 |
| 26 | FAE-X34 | ID23 1 | 1 | 8/7/2009 11:33 | | | 1 | 0.976 | 276.56 | 1 | 0 | 1 |
| 27 | beamstop | ID23 1 | 2 | 8/7/2009 10:09 | | | 1 | 0.976 | 242.39 | 1 | 0 | 1 |
| 28 | beamstop | ID23 1 | 1 | 8/7/2009 10:08 | | | 1 | 0.976 | 242.39 | 1 | 0 | 1 |
| 29 | | | | | | | | | | | | |
| 30 | | | | | | | | | | | | |

# Sample Ranking

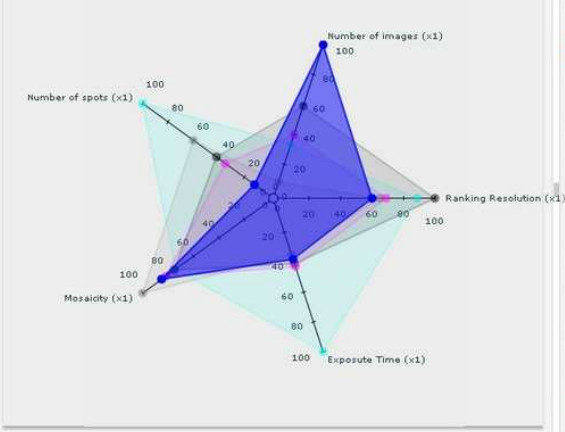



**Ranked Samples**

Sample Ranking (Rank - Value)

| Select Sample | Image Prefix | Run No | Start time | Space Group | UC a | UC b | UC c | Ranking resol. Å | Exposure time s | Mosaicity ° | Number of spots | Number of images | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Wt: 1 | Wt: 1 | Wt: 1 | Wt: 1 | Wt: 1 | Rank |
| ✓ | ref-t2-35b1-2 | 1 | 10:15:50 | C2221 | 70 | 206 | 107 | #2 2,24 | #1 223,0 | #3 0,54 | #1 2606 | #8 107 | #1 85 % |
| ✓ | ref-p2_7_dna | 1 | 15:31:14 | P4 | 66 | 66 | 101 | #1 1,99 | #4 547,5 | #2 0,53 | #6 1107 | #2 61 | #2 70 % |
| ✓ | ref-dna-HA00AQ0066 | 1 | 16:09:47 | P222 | 65 | 126 | 230 | #8 3,08 | #9 604,0 | #8 0,72 | #13 510 | #1 48 | #3 58 % |
| ✓ | ref-t2-35d6-1 | 4 | 11:25:50 | P2 | 74 | 107 | 102 | #10 3,34 | #10 618,3 | #1 0,48 | #10 775 | #5 80 | #4 57 % |
| ✓ | ref-mfe1fl1b | 1 | 20:16:36 | P222 | 66 | 126 | 229 | #5 2,85 | #3 470,8 | #2 0,53 | #11 755 | #7 99 | #5 57 % |
| | ref-t2-35d6-1 | 1 | 10:19:09 | P2 | 74 | 107 | 102 | #7 2,99 | #6 585,1 | #9 0,74 | #7 1100 | #3 70 | #6 56 % |
| | ref-mfe1fl9b | 1 | 20:34:00 | P222 | 65 | 126 | 230 | #3 2,81 | #2 386,0 | #5 0,62 | #4 1259 | #13 830 | #7 52 % |
| | ref-t2-35d6-1 | 2 | 10:38:23 | P2 | 74 | 107 | 102 | #9 3,12 | #11 631,9 | #10 0,82 | #9 849 | #4 74 | #8 51 % |
| | ref-t2-35d5 | 1 | 10:26:32 | P2 | 74 | 107 | 102 | #10 3,34 | #13 647,5 | #7 0,70 | #5 1129 | #6 98 | #9 51 % |
| | ref-mfe1fl9b | 2 | 20:35:55 | P222 | 65 | 126 | 230 | #4 2,84 | #8 597,4 | #5 0,62 | #3 1558 | #11 760 | #10 50 % |
| | ref-dna-HA00AC1704 | 1 | 16:06:19 | P222 | 65 | 126 | 230 | | | | | | |
| | ref-mfe1-rpmfe1fl-11 | 1 | 16:13:15 | P222 | 66 | 126 | 229 | | | | | | |
| | ref-t2-35d6-2 | 1 | 10:23:13 | P2 | 75 | 107 | 102 | | | | | | |

Ranking Chart

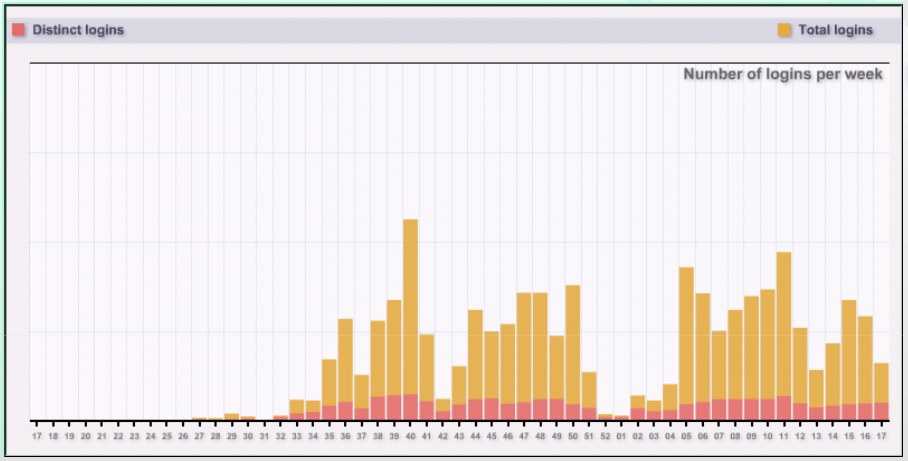


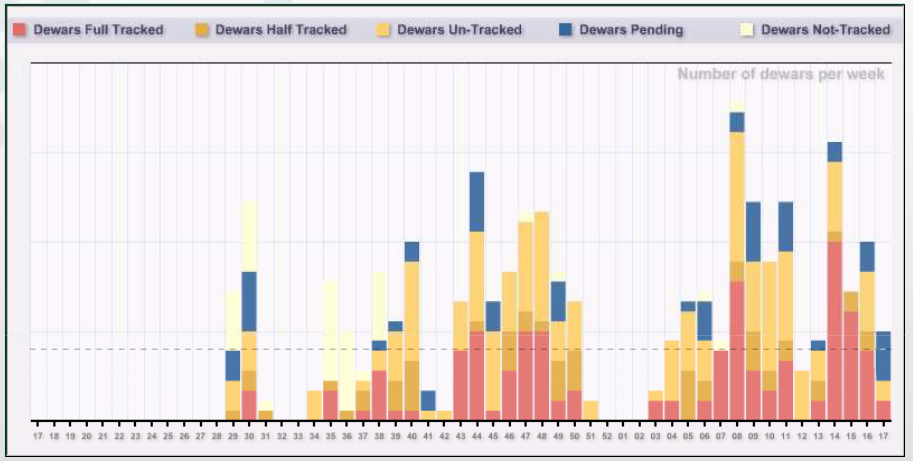

Sample List

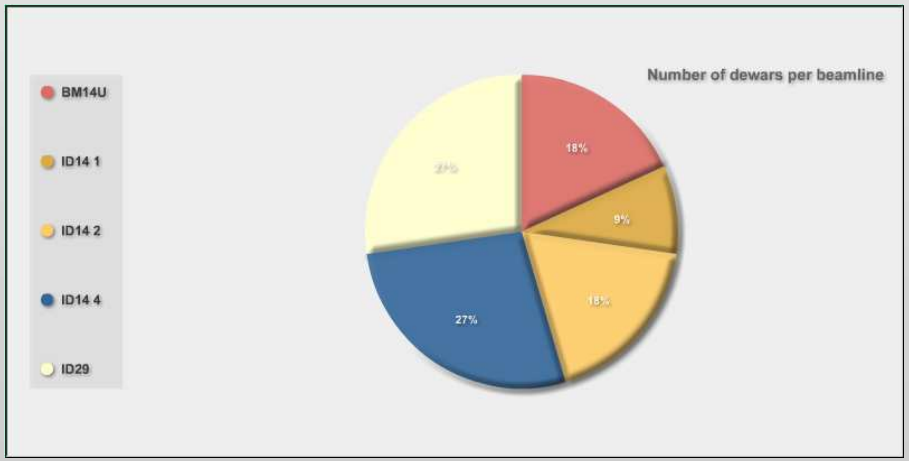
| Colour | Select | Rank | Score | Image prefix |
|---|---|---|---|---|
| | ✓ | 1 | 79 | ref-t2-35b1-2 |
| | ✓ | 2 | 64 | ref-p2_7_dna |
| | ✓ | 3 | 59 | ref-mfe1fl7 |
| | ✓ | 4 | 54 | ref-mfe1fl6 |
| | ✓ | 5 | 54 | ref-mfe1-rpmf... |
| | | 6 | 53 | ref-mfe1fl6b |
| | | 7 | 52 | ref-mfe1-rpmf... |
| | | 8 | 52 | ref-mfe1-rpmf... |
| | | 9 | 52 | ref-mfe1fl1b |
| | | 10 | 51 | ref-t2-35d6-1 |
| | | 11 | 51 | ref-mfe1-rpmf... |
| | | 12 | 50 | ref-t2-35d6-1 |
| | | 13 | 50 | ref-dna-HA00A... |
| | | 14 | 50 | ref-mfe1fl1a |
| | | 15 | 49 | ref-mfe1fl9b |
| | | 16 | 47 | ref-mfe1fl9b |
| | | 17 | 47 | ref-mfe1fl6a |
| | | 18 | 47 | ref-dna-HA00A... |
| | | 19 | 46 | ref-t2-35d5 |

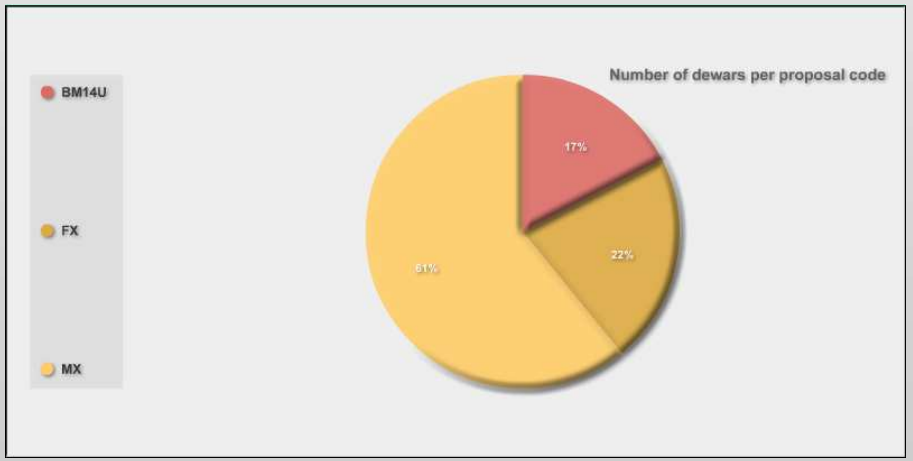
Submit Cancel

# Reporting for Managers Dashboard

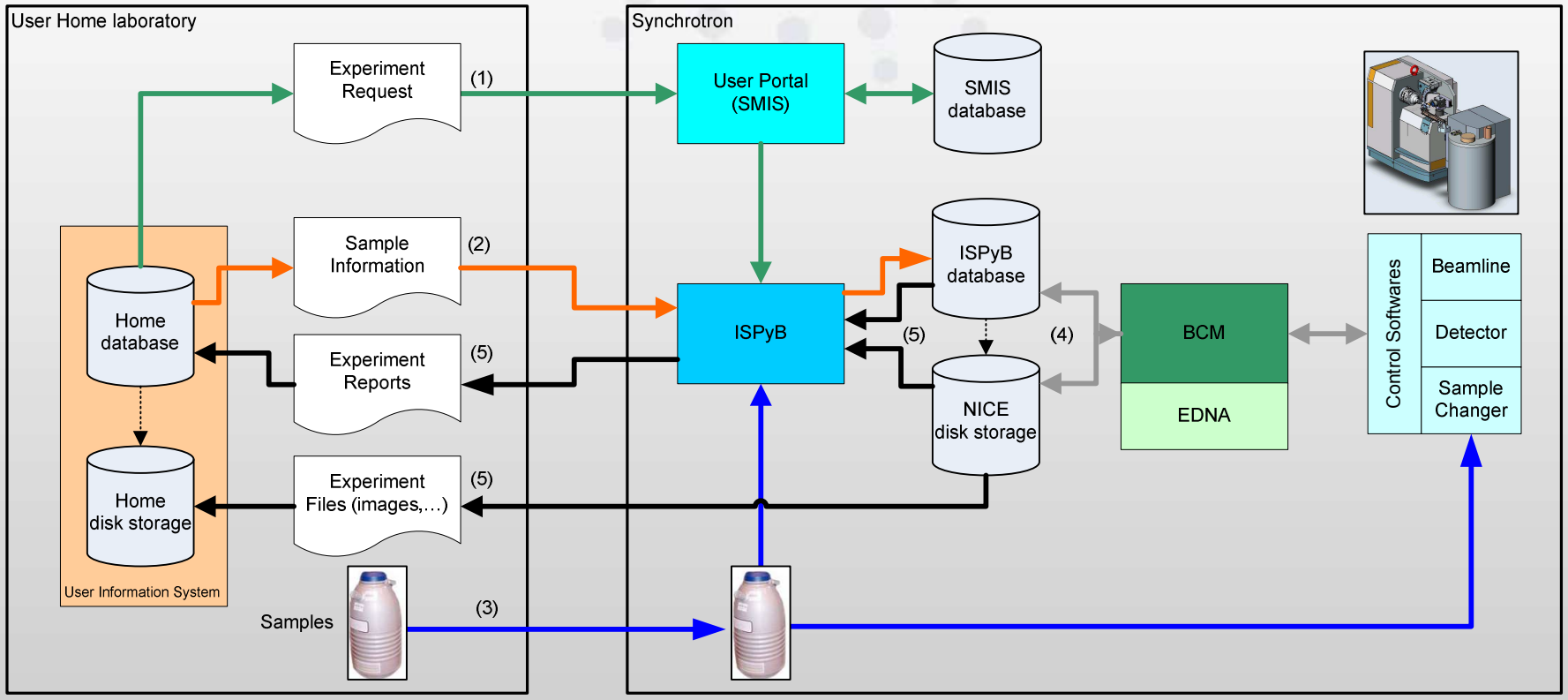


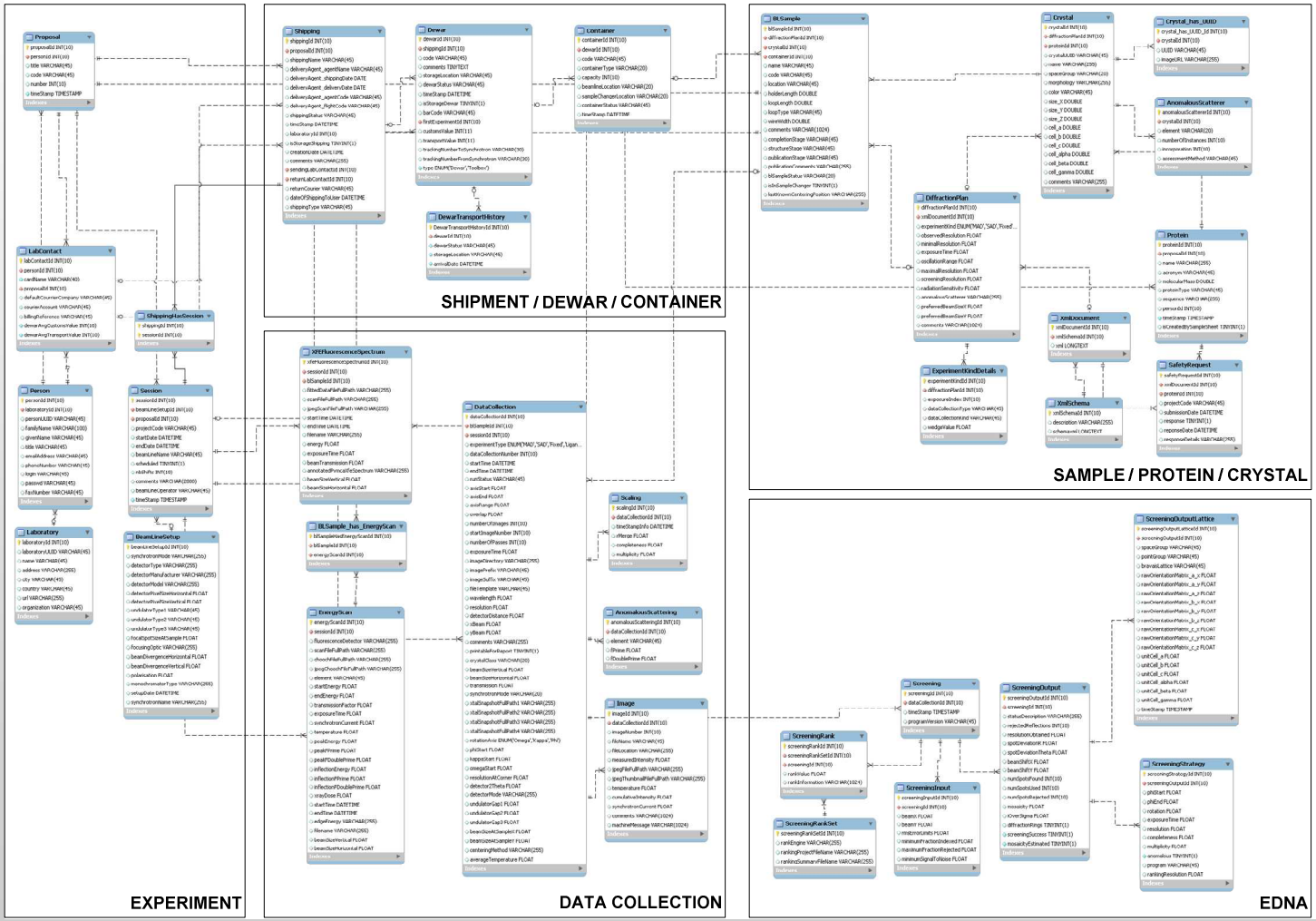



# Application Environment and Data Flow



# ISPyB Data Model



# ISPyB Technology

- Pure Web Interface
- Java
- Jboss Application Server (Ejb)
- Struts (MVC Web Application Framework)
- MySql / Oracle Database
- Communication with User Portal (SMIS) via Web Services
- Communication with BCM and EDNA
  - Through the Data Base
  - …soon via Web Services



Struts


ORACLE

# How does it help?

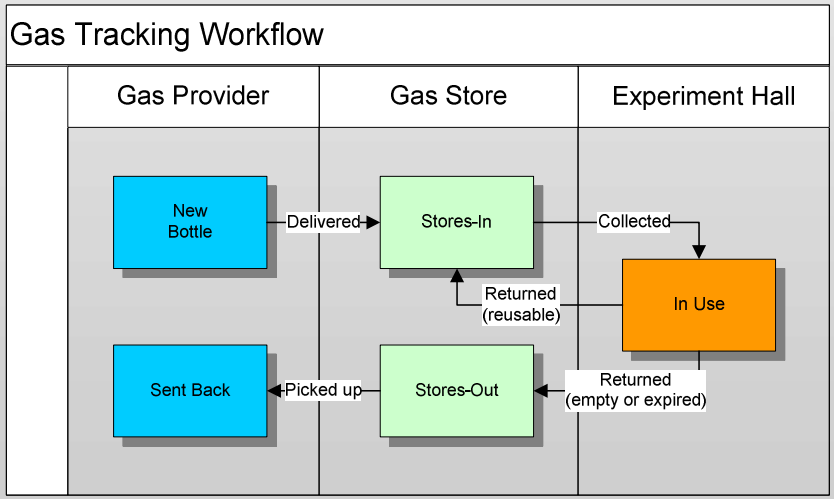
- Users
  - Prepare Experiment
  - Remote tracking by colleagues at home while the experiment is performed
  - Process experiment results back into the Home LIMS
  - Long term tracking of experiments (stored in the database)
- Communication between Users and ESRF Operations
  - Dewar description (beamline, local contact, user address)
  - Sample description (description, bar-code, location, known unit cell dimensions, space group, diffraction plans, comments)
  - Operator comments on data collections
- Safety (Dewar Tracking)

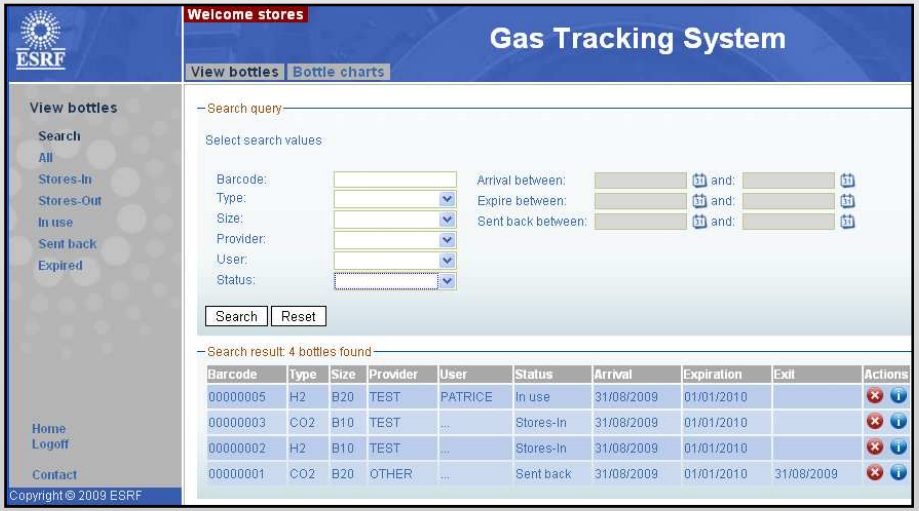
# Next Steps…

- ISPyB new features (new EDNA data, AutoProcessing,…)
- Object Tracking
- ISPyB for BioSAXS (ID14-3)
- ISPyB for other beamlines
- ISPyB for other MX facilities
- ISPyB and UPBL10/MASSIF

# Gas Tracking

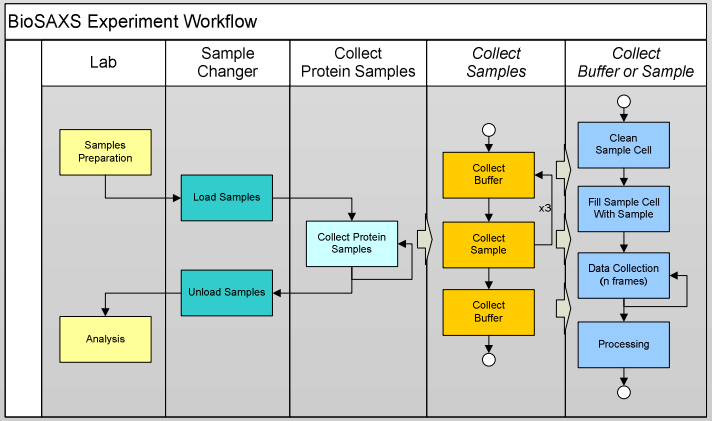
- Why Tracking bottles of gas?
  - Gas available at storage -> Users
  - Location of old bottles -> Safety, Finance
  - Location of ‘dangerous’ bottles ->Safety
- Uses Dewar Tracking architecture and technology



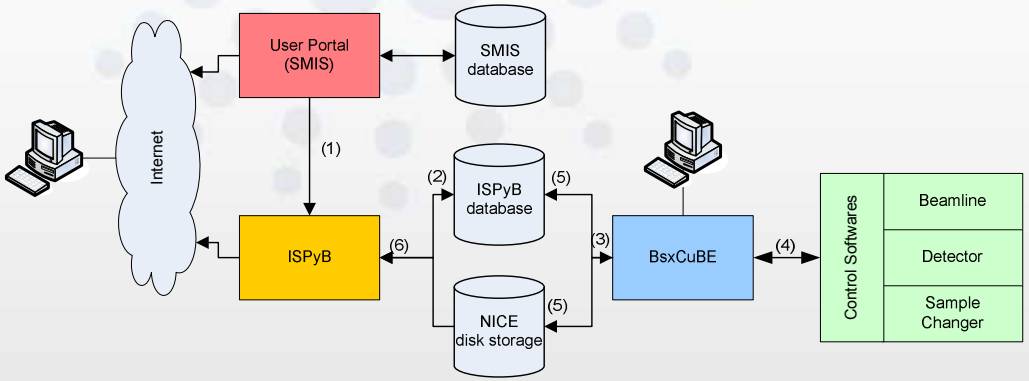
# ISPyB For BioSAXS (ID14-3)

- BioSAXS needs
  - Sample Tracking
  - Experiment Management (sample preparation, …)
  - Experiment Recording (images, data files, beamline parameters,…)
  - Search and statistics on past experiments
  - Web interface (remote access)
  - Link with MX experiments and SANS experiments…




# ISPyB For BioSAXS (ID14-3)

- Architecture



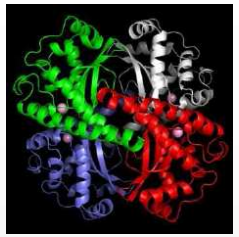

- Developments
  - Experiment Management: as it is in ISPyB
  - Sample Management: little customization (dewar/box, container/plate)
  - Experiment Management:
    - Data model extensions for BioSAXS data
    - GUI for BioSAXS data
  - Search and Report: customization for BioSAXS data
  - BCM/BsxCuBE evolutions

- An opportunity to see how we can use the ISPyB 'framework' and data model on other beamlines

# ISPyB for Other Beamlines

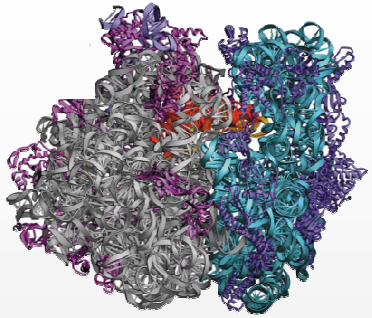
- Features
  - Sample Tracking
  - Experiment Logging
- Developments
  - Experiment Management: as it is in ISPyB (link with User Portal/SMIS)
  - Sample Management: beamline sample description (generic ?)
  - Experiment Recording
    - Experiment data model (one per experiment type … or generic data model)
    - Gui to browse & view data collections
  - Search and Reports: depends on experiment data model
  - BCM (Beamline Control Management)
    - MX: MxCuBE
    - BioSAXS: BsxCuBE
    - …

# ISPyB Light for Other MX Facilities



- Facilities interested in using ISPyB
  - Max-Lab, Bessy, Soleil, APS,… (ISPyB meeting, April 2009)
- Features
  - Sample Management: as it is in ISPyB
  - Experiment Recording: as it is in ISPyB
  - Search and Report: as it is in ISPyB
  - Experiment Management: link with Proposal Submission System
    - Proposals, Sessions, Lab Contacts, Sample sheets,…
    - Current ISPyB: data filled in by User Portal / SMIS
    - ISPyB Light: API or Gui to fill in / manage this information

# UPBL10 / MASSIF



- Many more samples…
  - Instead of a few samples, hundreds will be evaluated to select the best ones for data collection
  - An order of magnitude increase in samples and data
- …and new Architecture / Workflow
  - Automated end stations for sample screening
  - End station for data collection
- Needs
  - As automation increased – more reliant on information from user (crystal mounting, size, composition,…)
  - Automated links with Home Lab Systems
  - Tracking by individual Samples
  - Sample Ranking (over hundreds of samples)
  - New data arrangement: not by session, by project ?
  - Sample Management: sorting facility associated with beamlines

# Thanks to…


| | | | |
|---|---|---|---|
| Elspeth Gordon | ESRF | Alun Ashton | DLS |
| Stéphanie Monaco | ESRF | Martin Walsh | DLS/BM14 |
| Gordon Leonard | ESRF | Karl Levik | DLS |
| Gianluca Cioci | ESRF | Graeme Winter | DLS |
| Pernot Petra | ESRF | | |
| Adam Round | EMBL | Solange Delagenière | ESRF |
| Matthew Bowler | ESRF | Ludovic Launer | MRC/BM14 |
| | | Emmanuel Eyer | ESRF |
| Martine Moroni | ESRF | Christian Rolland | ESRF |
| Serges Ronseaux | ESRF | Samuel Jones | ESRF |
| Thierry Caruana | ESRF | CS/MIS Group | ESRF |
| Eric Fertin | ESRF | | |
| | | Oleg Konovalov | ESRF |
| Olof Svensson | ESRF | | |
| Darren Spruce | ESRF | … people I don't know but who have contributed to ISPyB | |
| Matias Guijarro | ESRF | … people I forgot to mention (sorry) | |
| Bernard Lavault | EMBL | … and all users for their constructive feedbacks | |


# Questions…