New developments and scientific case for offline data analysis

Alex de Maria Antolinos and Gianluca Santoni
ESRF

New developments ~~and scientific case for offline data analysis~~

Alex de Maria Antolinos ~~and Gianluca Santoni~~
ESRF

Overview

- Data Model Meeting at Soleil (September)
- User Interface
  - Roadmap
  - EXI
  - EXI2
- Offline data analysis
- Upgrade of JBoss Server
- Single-Sign On

Data Model
Screening Tables

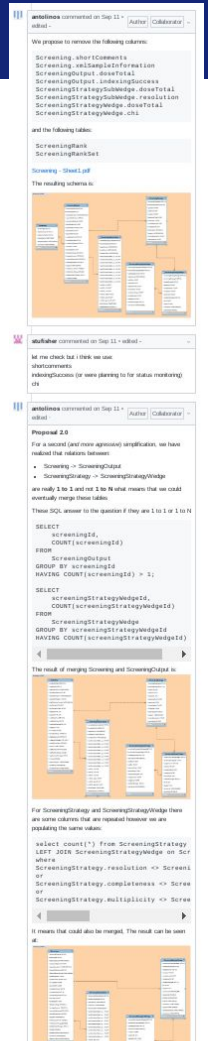
Screening Tables

- Github issue #46
  - https://github.com/ispyb/ispyb-database-modeling/issues/46

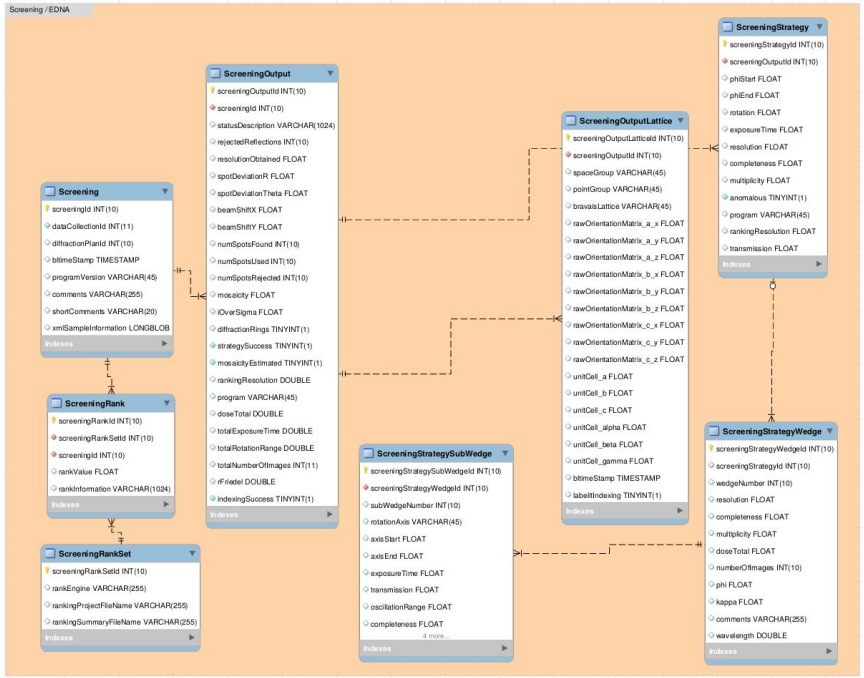
- Clean up old screening tables
  - Unused columns
  - Unused tables
  - Refactor some tables

- Side effects:
  - Changes might break compatibility with the old user interface !!
  - Changes to be done on MxCube

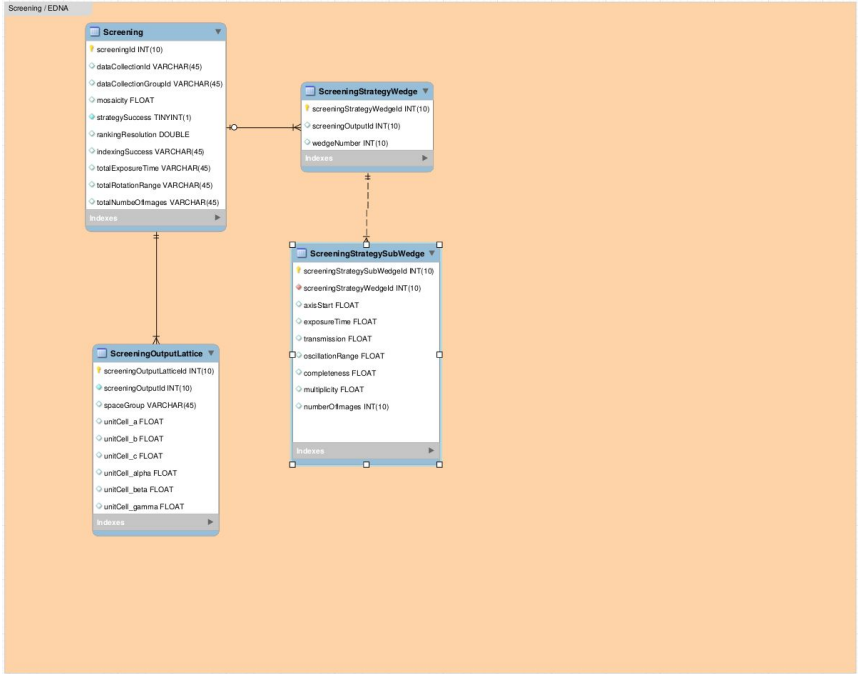
- No deadline defined for this (waiting input for collaborators)



Screening tables



Current data model

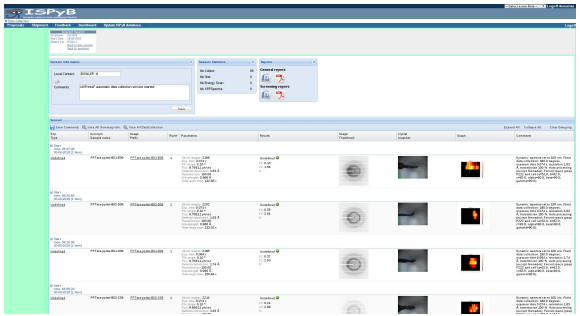


Proposed data model

User Interface

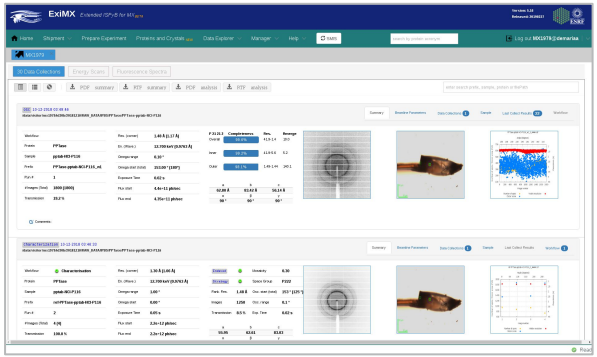
Roadmap

- EXI will be the **official ISPyB UI** at the startup 2020 **at the ESRF**
- Old User interface will be deprecated with **no more support and maintenance**
- EXI2 is being developed since April 2019



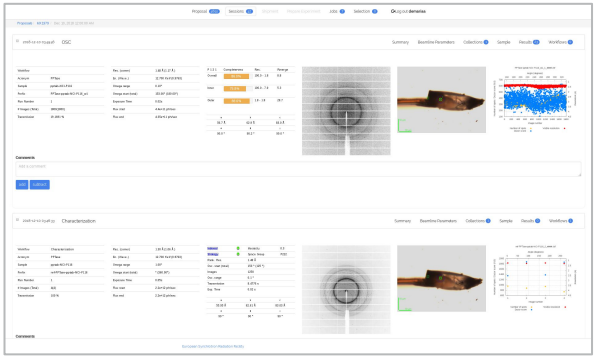
ISPyB (2004 -2020)

https://ispyb.esrf.fr



EXI (2014 - ?)

https://exi.esrf.fr

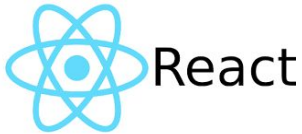


EXI2 (2019 - ?)

https://exi2.esrf.fr

Why do we need EXI2?

- It makes easier to developers to join the project
  - It uses the same technologies like MxCube



  - Components can be reused between MxCube-EXI2
- Responsive (works in mobile devices)
- It is modular

Features:

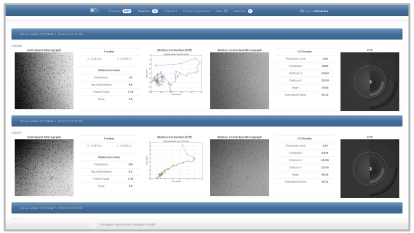
- No major changes about how data is displayed (copy of EXI with some improvements)
- It includes offline data analysis

Do you want to collaborate?

- Suggestions and ideas
- Specific requirements
- Follow-up of the project
- Active developments
- Testing

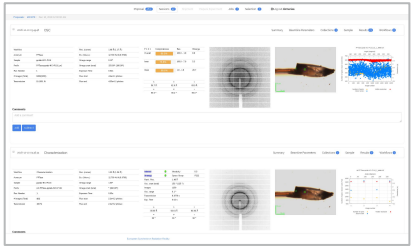
https://exi2.esrf.fr is pointing to the ISPyB instance of the ESRF but it could also point to your public ISPyB instances for you to test

Migration Status from EXI to EXI2



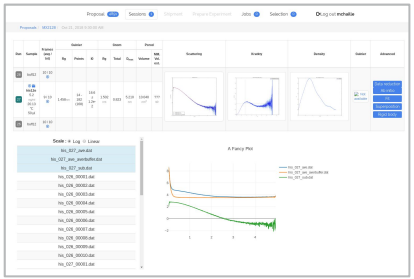
EM

90% Progress



MX

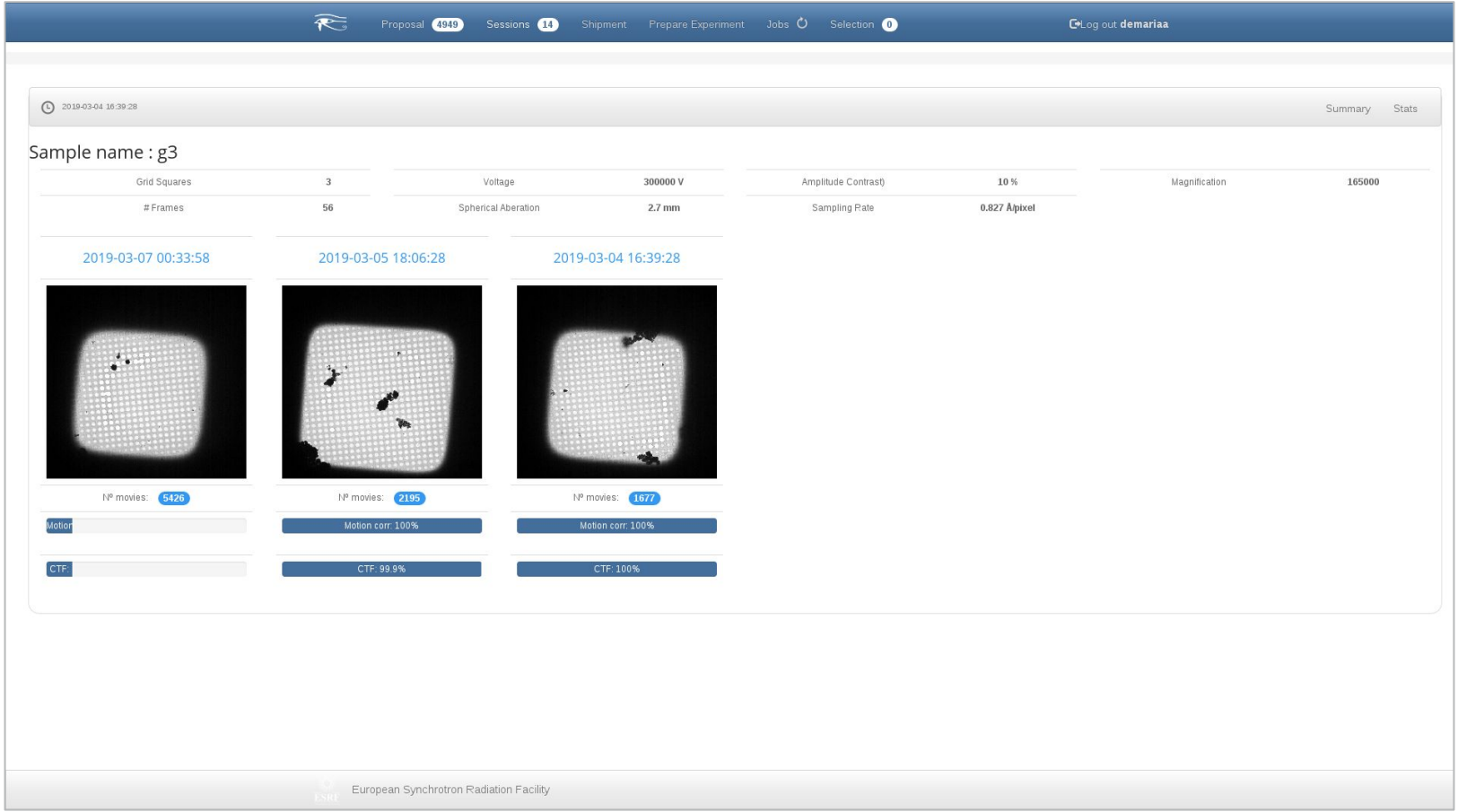
50% Progress



BioSAXS

70% Progress

Status for EM (90%)



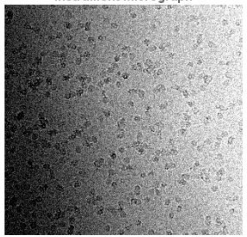
Status for EM (90%)

Proposal 4949 | Sessions 14 | Shipment | Prepare Experiment | Jobs | Selection 0 | Log out **demariaa**

Movie number 15179Mar 7, 2019 12:31:25 AM

338348

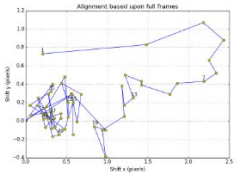
Instrument Micrograph



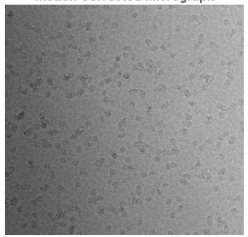
Position	
X: -3.197e-4	Y: -3.197e-4

Motion Correction	
Total Motion:	52
Avg. Motion/frame	0.9
Frame Range	1-56
Dose	1.0

Motion Correction Drift

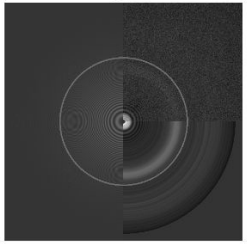


Motion Corrected Micrograph



CTF Results	
Resolution Limit:	3.09
Correlation:	0.062
Defocus U:	21554
Defocus V:	21224
Angle:	19.82
Estimated B factor:	65.43

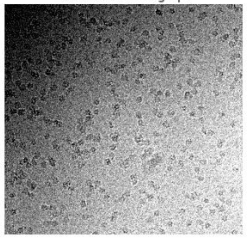
CTF



Movie number 15178Mar 7, 2019 12:31:25 AM

338347

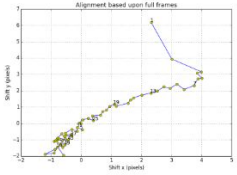
Instrument Micrograph



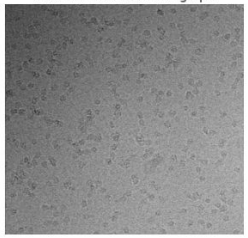
Position	
X: -3.197e-4	Y: -3.197e-4

Motion Correction	
Total Motion:	345
Avg. Motion/frame	6.2
Frame Range	1-56
Dose	1.0

Motion Correction Drift

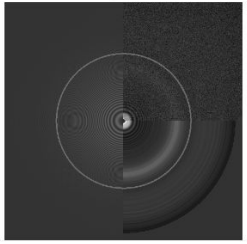


Motion Corrected Micrograph



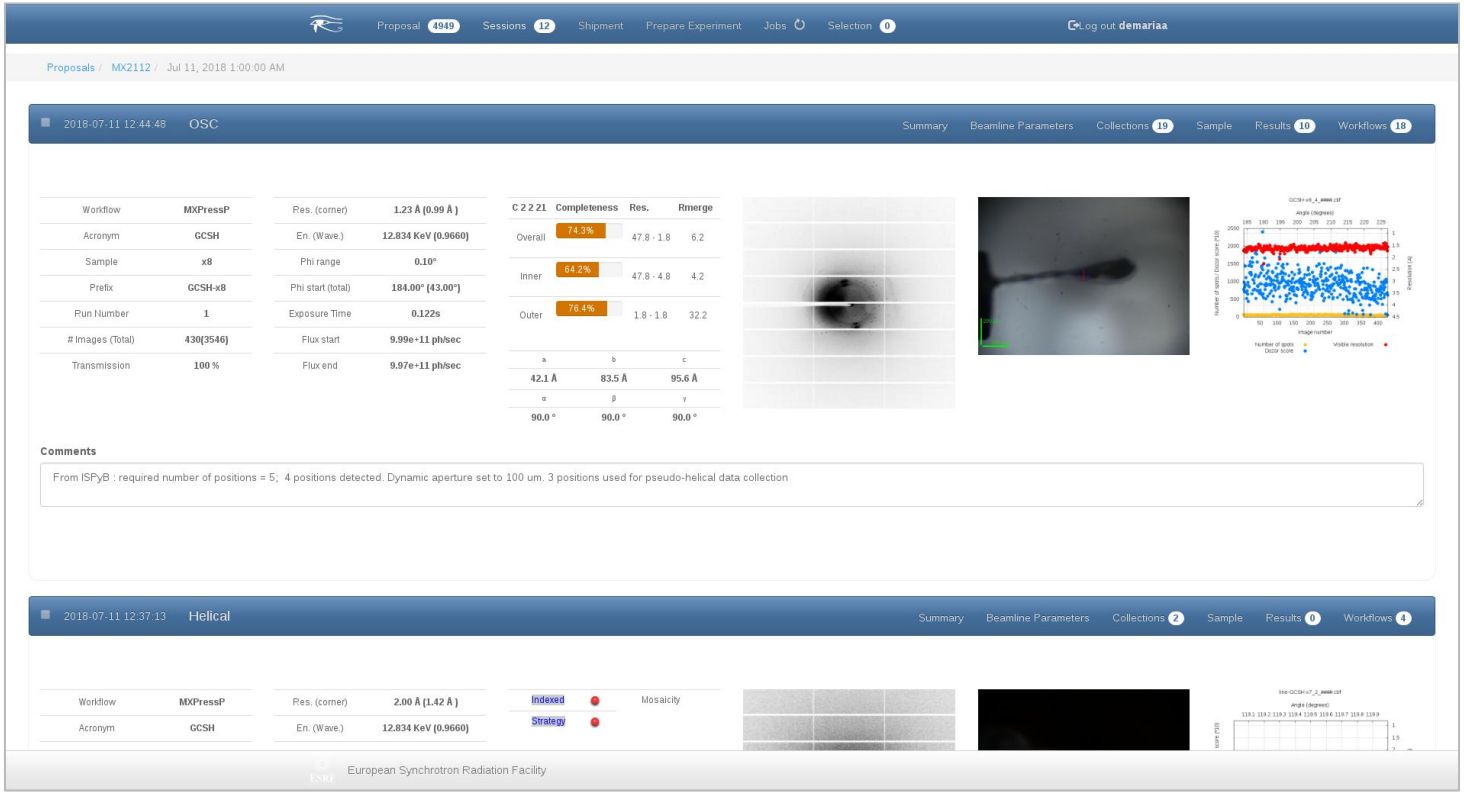
CTF Results	
Resolution Limit:	2.93
Correlation:	0.074
Defocus U:	21238
Defocus V:	21210
Angle:	68.16
Estimated B factor:	65.31

CTF



Movie number 15177Mar 7, 2019 12:28:56 AM

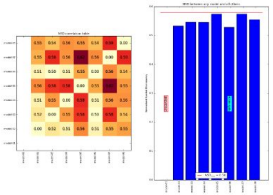
Status for MX (50%)

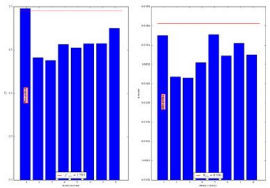


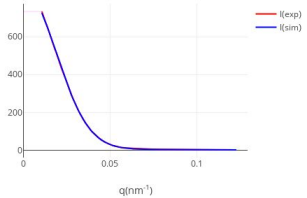
Status for BioSAXS (70% done)

Proposal 4949 | Sessions 10 | Shipment | Prepare Experiment | Jobs | Selection 0 | Log out mchaille

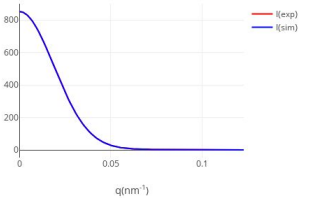
Proposals / MX2085 / Nov 29, 2018 9:30:00 AM



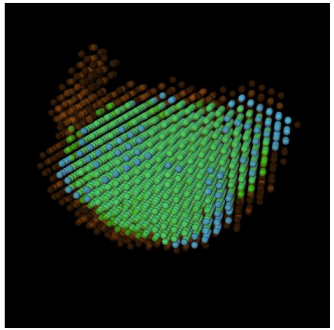




Fit of the simulated scattering curve versus a smoothed experimental data (spline interpolation)

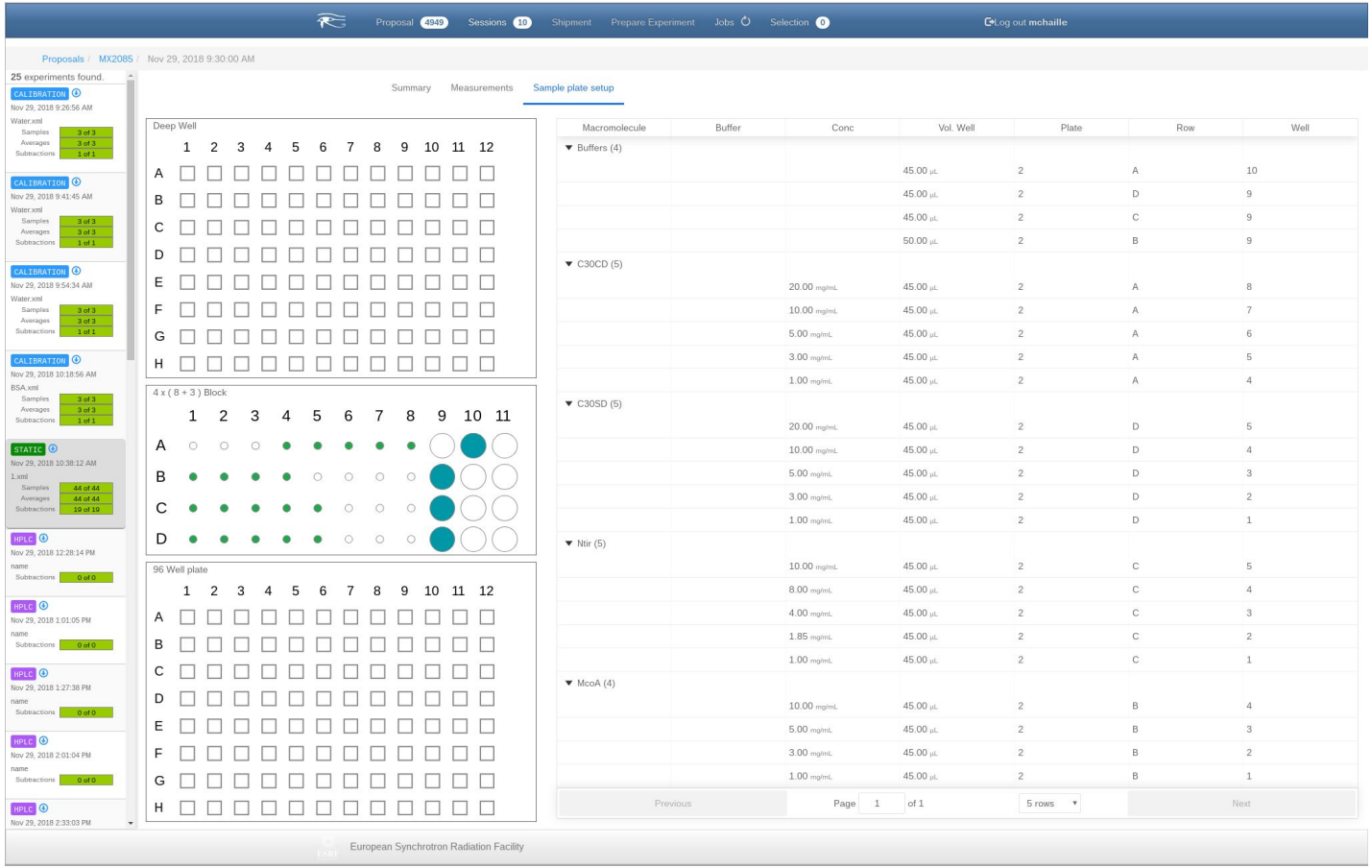


Fit of the simulated scattering curve versus the experimental data.



Type	chiSqrt	rFactor	Rg	PDB	Fir	Log
Reference				damaver		
Refined	1.143	0.00075	64.49	dammin	dammin	dammin
Model	1.41	0.00107	64.498	model_04	model_04	model_04
Model	1.565	0.00122	64.504	model_03	model_03	model_03
Model	1.524	0.00151	64.494	model_01	model_01	model_01
Model	1.574	0.00142	64.498	model_06	model_06	model_06
Model	1.751	0.0013	64.502	model_07	model_07	model_07
Model	1.381	0.00106	64.511	model_00	model_00	model_00
Model	1.977	0.0015	64.512	model_02	model_02	model_02
Model	1.572	0.00129	64.504	model_05	model_05	model_05

Status for BioSAXS (70% done)



Conclusion

- EXI is being migrated to React smoothly
- Fully migration expected by the end of 2020
  - No official deadline
- EXI and EXI2 are compatible with your installed version of ISPyB
- It is a good time to:
  - Provide feedback
  - Help with the developments
    - Coding
    - Testing

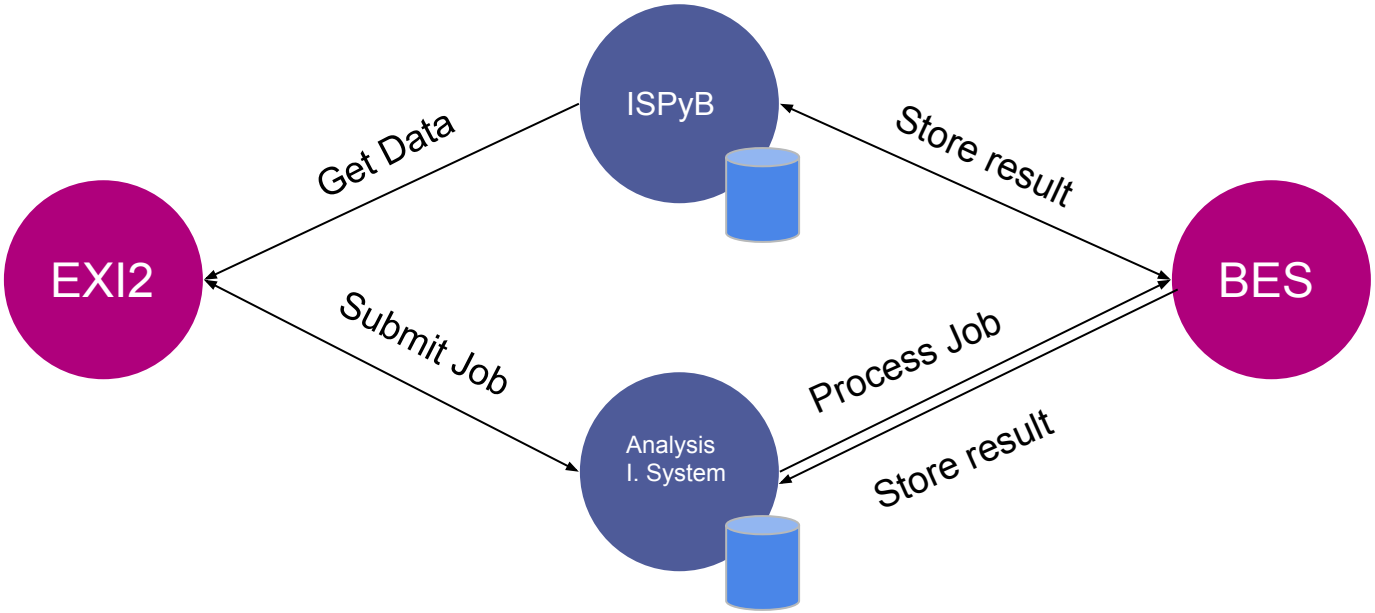
Offline Data Analysis

Offline Data Analysis: Goal

Goal: Users can launch predefined jobs from UI

- Decoupled architecture
  - components to execute independently while still interfacing with each other
- Flexible
  - Easy to add new type of jobs
  - Easy to maintain
- Allows interactive jobs
- Not specific to ISPyB

Offline Data Analysis: architecture



Presentation Layer

Services Layer

Processing Layer

Offline Data Analysis: Implementation

- **Information System for Analysis** (ISA)
  - NodeJS application (https://gitlab.esrf.fr/icat/is4a)
  - Queueing system

- It does:
  - Expose an API to
    - Store a job, input, output and status
    - Publishes a catalogue of tools for a given entity e.g: data collection
  - Use a MongoDB for storing both data and metadata

- It does not
  - run the jobs

Offline Data Analysis: why mongo

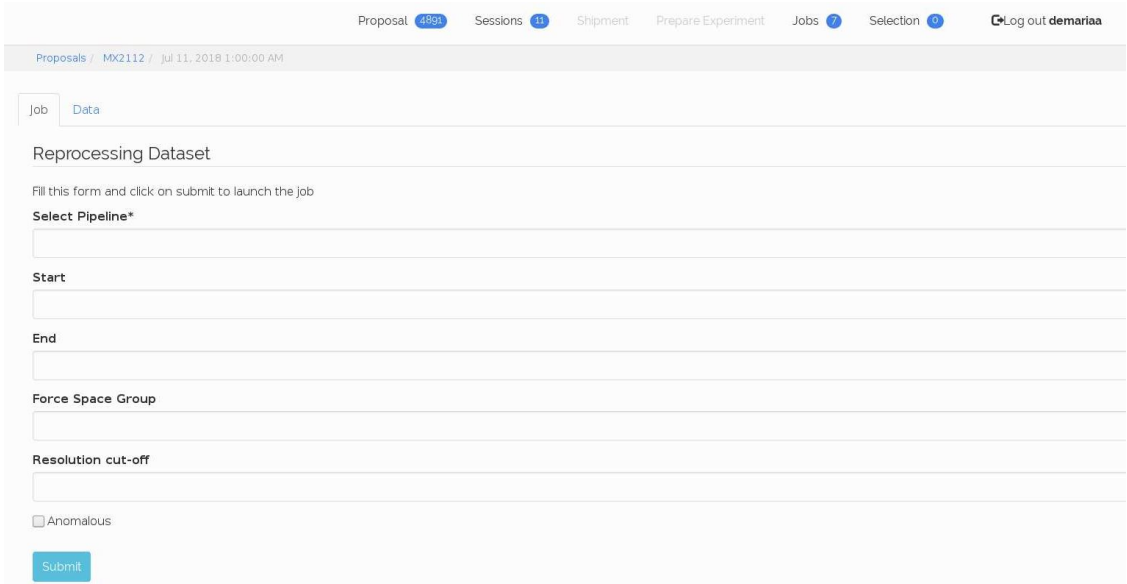
- More versatile
  - No schema constraints
- Good for prototyping
- GridFS
  - storing large files may be more efficient in a MongoDB database than on a system-level filesystem.
- Apply a different data policy
  - Data acquisition and online data analysis to be stored forever
  - Offline data analysis might be removed after some time
- The use of a MongoDB does not prevent to store data on ISPyB if needed

Job catalogue

```
datacollection: {
    reprocess: {
        "title": "Reprocessing Dataset",
        "description": "Fill this form and click on submit
        type : "object",
        "required": [
            "type"
        ],
        "properties": {
            "type": {
                "title": "Select Pipeline",
                "type": "string",
                "enum": [
                    "EDNA_proc",
                    "autoPROC",
                    "XIA2_dials",
                    "grenades_fastproc",
                    "grenades_parallelproc"
                ],
                "enumNames": [
                    "EDNA_proc",
                    "autoPROC",
                    "XIA2_dials",
                    "grenades_fastproc",
                    "grenades_parallelproc"
                ]
            },

            "Start": {
                "type": "number"
            },
            "End": {
                "type": "number"
            },
            "Space Group": {
                "title": "Force Space Group",
                                ring"
            },
            "R
                "type": "string"
```

IS4A



React-JSON-Schema

Conclusion

- A mockup for offline data analysis has been developed
- It allows store data and metadata from processing jobs
- It is supposed to be versatile and easy to maintain
- We are ready to test it with real use cases
  - Feedback is appreciated

- If anyone interested please contact us!

Wildfly

FROM 

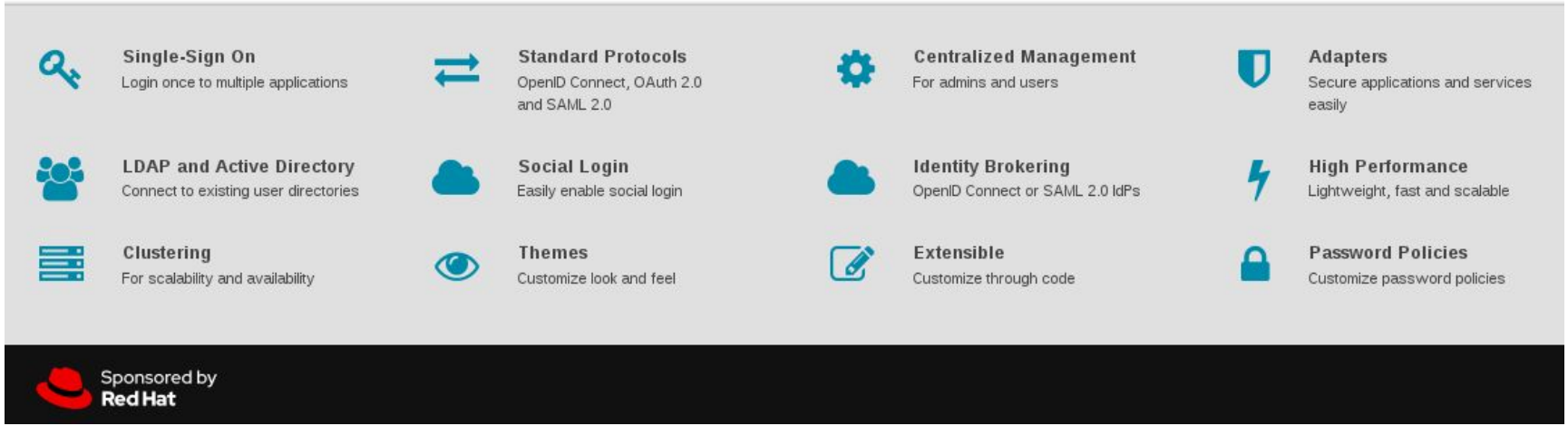
8.2

TO 

18.0

- Upgrade to Wildfly version 18
  - No big deal but some dependencies might change in the pom.xml
- Aiming to be backward compatible
- Aiming to upgrade to a recent version of Wildfly more frequently

Single-Sign on



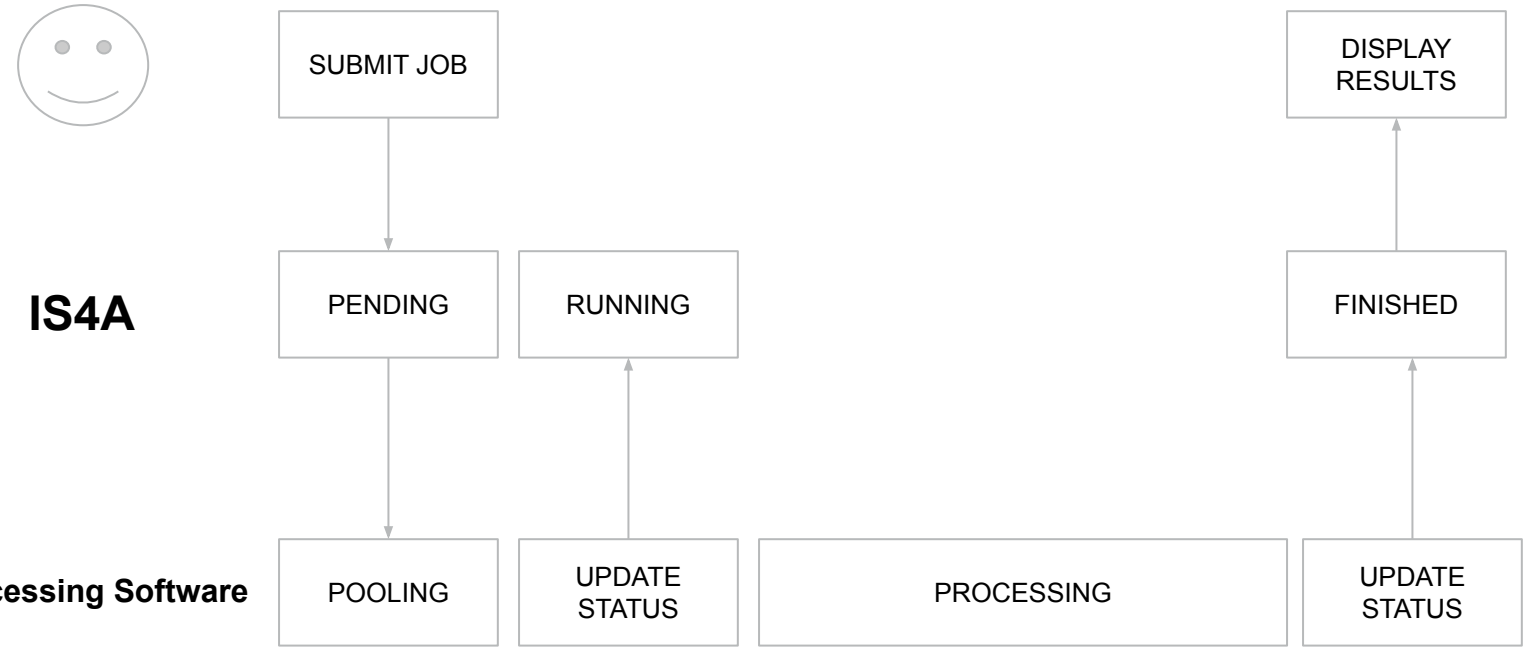
Single-Sign on

- **Support**
  - standalone.xml to be modified
  -
- **Advantages**
  - No specific code in ISPyB
  - MxCube-ISPyB single-sign on
  - Allows more ids
    - orcID
    - UmbrellaId

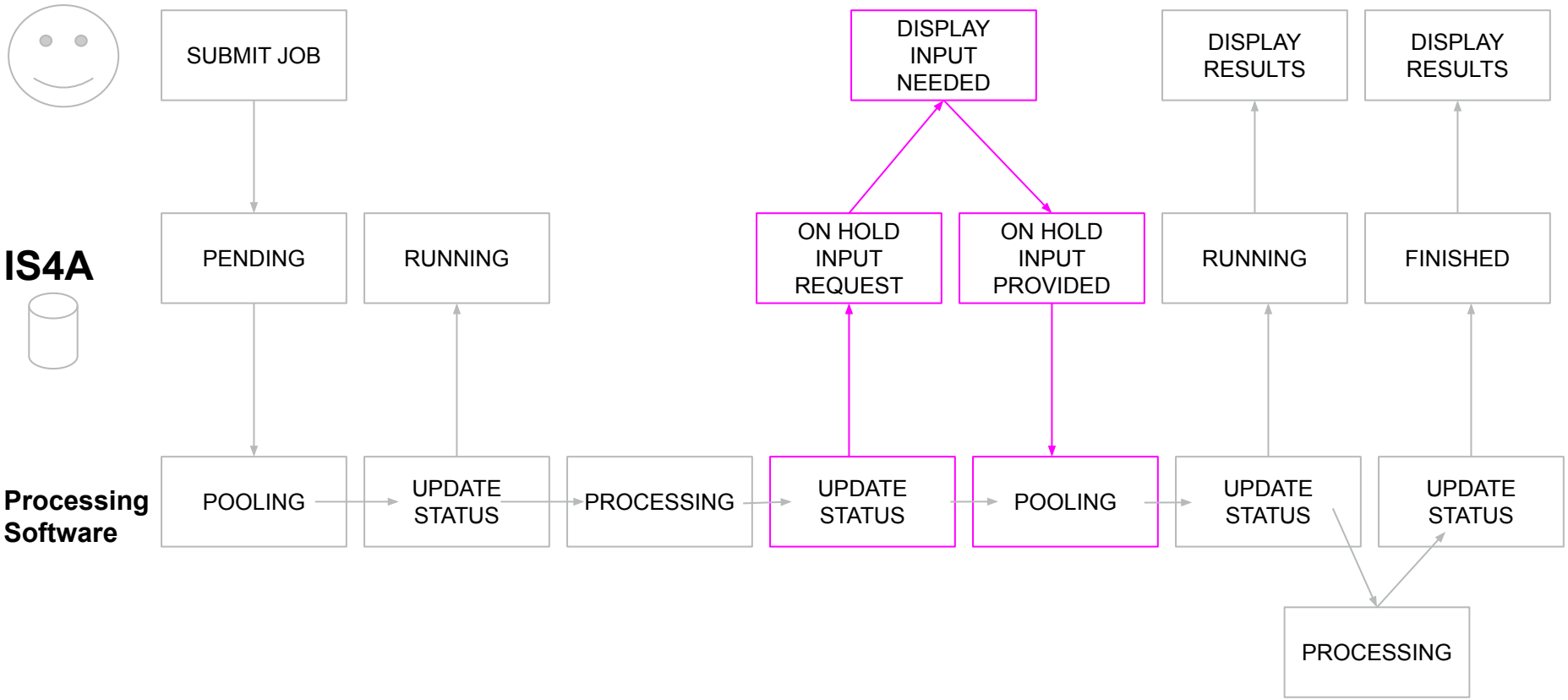
Thanks!

Backup Slides

Use case non-interactive



Use case interactive



IS4A Service

- NodeJS server
- Mongo DB
- Simple Restful API

JOB

GET	`/jobs`	Returns all jobs
POST	`/jobs`	Creates a new job
GET	`/jobs/type/datacollection`	Returns json schema of jobs that can be run for a data collection
GET	`/jobs/type/datacollectiongroup`	Returns json schema of jobs that can be run for a data collection
GET	`/jobs/{id}/status/{status}`	Returns allowed types of job
POST	`/jobs/{id}/output`	Add output to the job
GET	`/jobs/{username}`	Gets all jobs from user
POST	`/jobs/upload`	Uploads a file

Datacollection placeholder

Proposal 4891 | Sessions 11 | Shipment | Prepare Experiment | Jobs 7 | Selection 0 | Log out **demariaa**

Proposals / MX2112 / Jul 11, 2018 1:00:00 AM

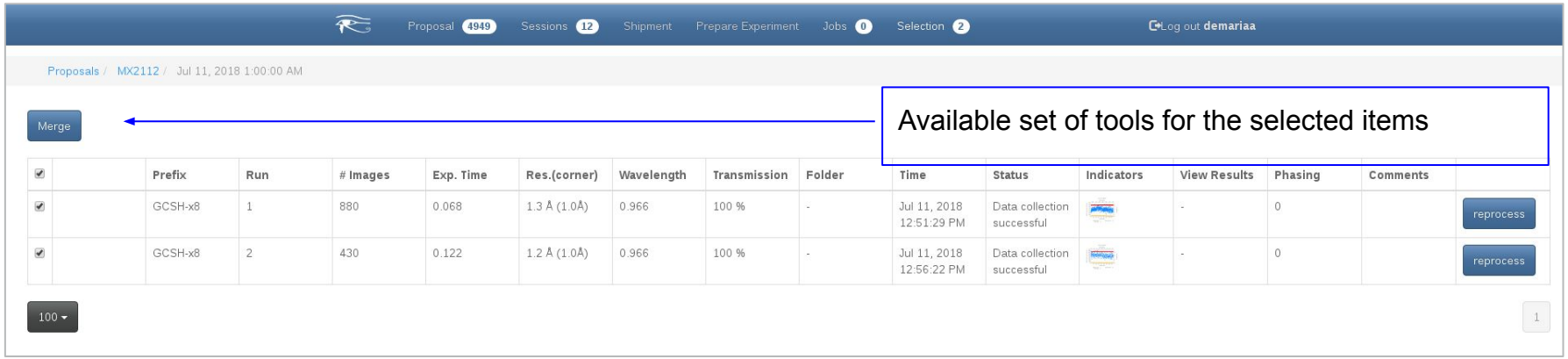
2018-07-11 12:44:48 OSC

Summary | Beamline Parameters | Collections 19 | Sample | Results 22 | Workflows 18

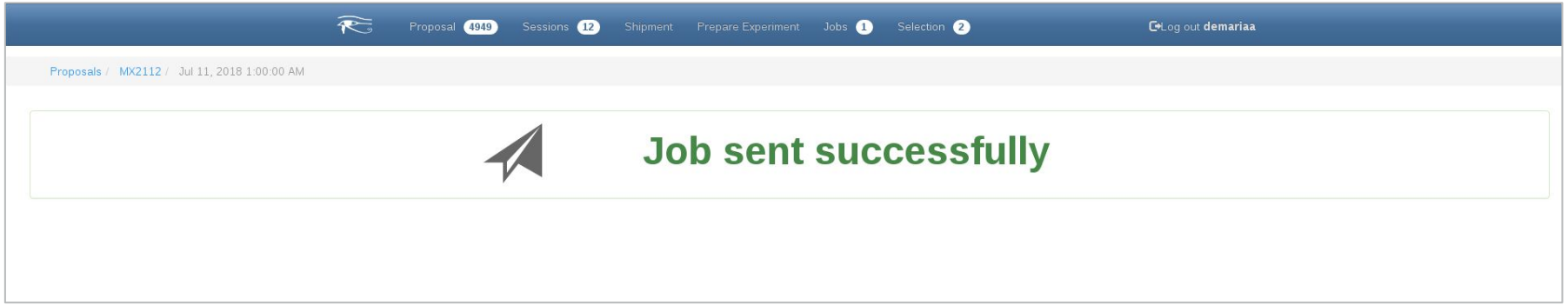
Prefix	Run	#Images	Exp. Time	Res. (corner)	Wavelength	Transmission	Directory and image template	Time	Run status	Indicators	View Results	Phasing	Comments	
mesh-GCSH-x8	1	168	0.1 s	2.0 Å (1.4 Å)	0.966 Å	100		Jul 11, 2018 12:44:48 PM	Data collection successful			0		reprocess
line-GCSH-x8	2	100	0.1 s	2.0 Å (1.4 Å)	0.966 Å	100		Jul 11, 2018 12:46:35 PM	Data collection successful			0		reprocess
line-GCSH-x8	3	100	0.1 s	2.0 Å (1.4 Å)	0.966 Å	100		Jul 11, 2018 12:47:28 PM	Data collection successful			0		reprocess
line-GCSH-x8	4	100	0.1 s	2.0 Å (1.4 Å)	0.966 Å	100		Jul 11, 2018 12:48:22 PM	Data collection successful			0		reprocess
ref-GCSH-x8	4	4	0.1 s	1.6 Å (1.2 Å)	0.966 Å	100		Jul 11, 2018 12:49:13 PM	Data collection successful			0		reprocess
ref-GCSH-x8	5	4	0.1 s	1.1 Å (0.9 Å)	0.966 Å	100		Jul 11, 2018 12:50:20 PM	Data collection successful			0		reprocess
GCSH-x8	1	880	0.068 s	1.3 Å (1.0 Å)	0.966 Å	100		Jul 11, 2018 12:51:29 PM	Data collection successful			0		reprocess
line-GCSH-x8	6	100	0.1 s	1.1 Å (0.9 Å)	0.966 Å	100		Jul 11, 2018 12:53:38 PM	Data collection successful			0		reprocess
line-GCSH-x8	7	100	0.1 s	1.1 Å (0.9 Å)	0.966 Å	100		Jul 11, 2018 12:54:32 PM	Data collection successful			0		reprocess
line-GCSH-x8	8	100	0.1 s	1.1 Å (0.9 Å)	0.966 Å	100		Jul 11, 2018 12:55:27 PM	Data collection successful			0		reprocess

Launch a job

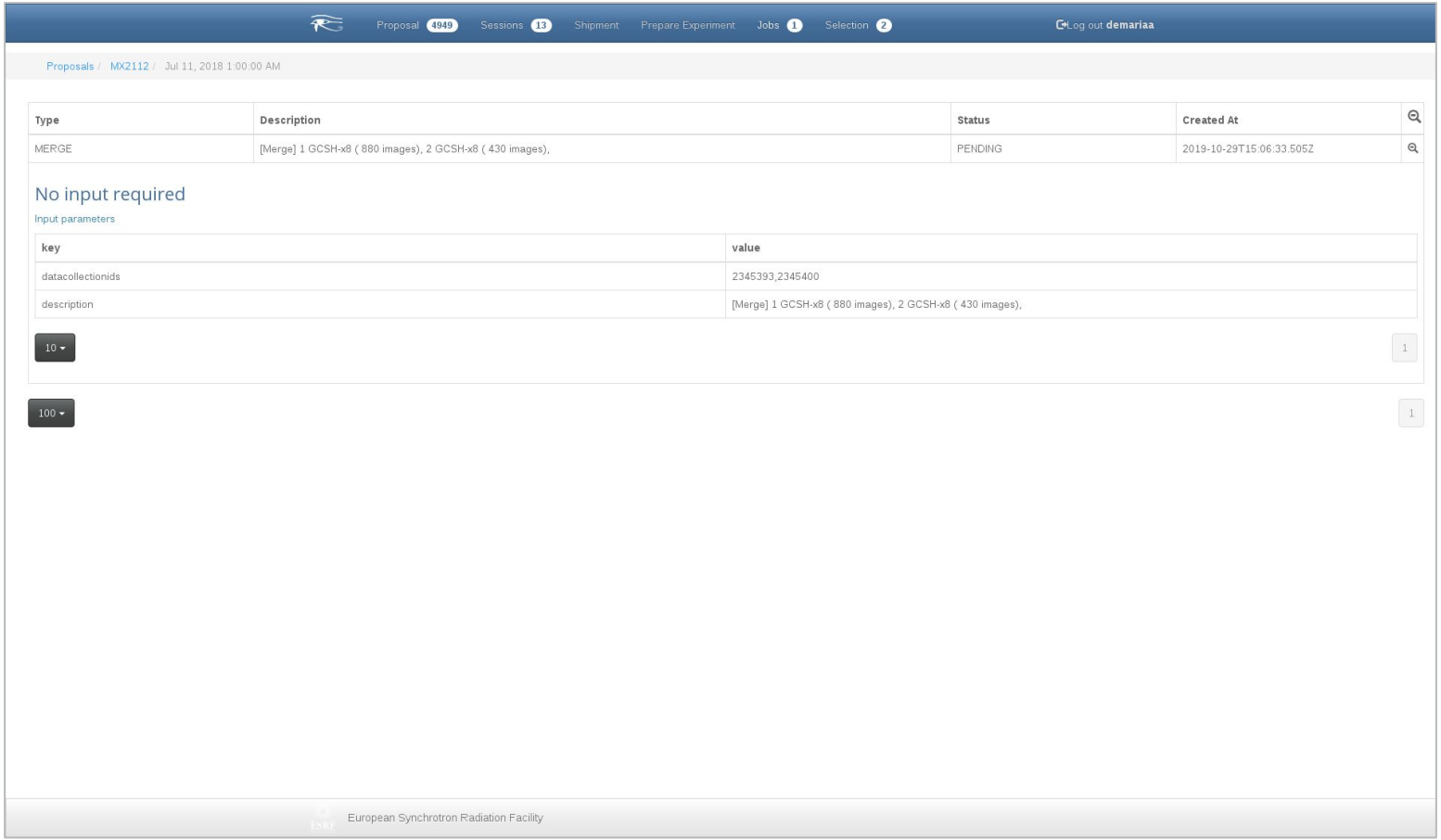
1)Select items to process

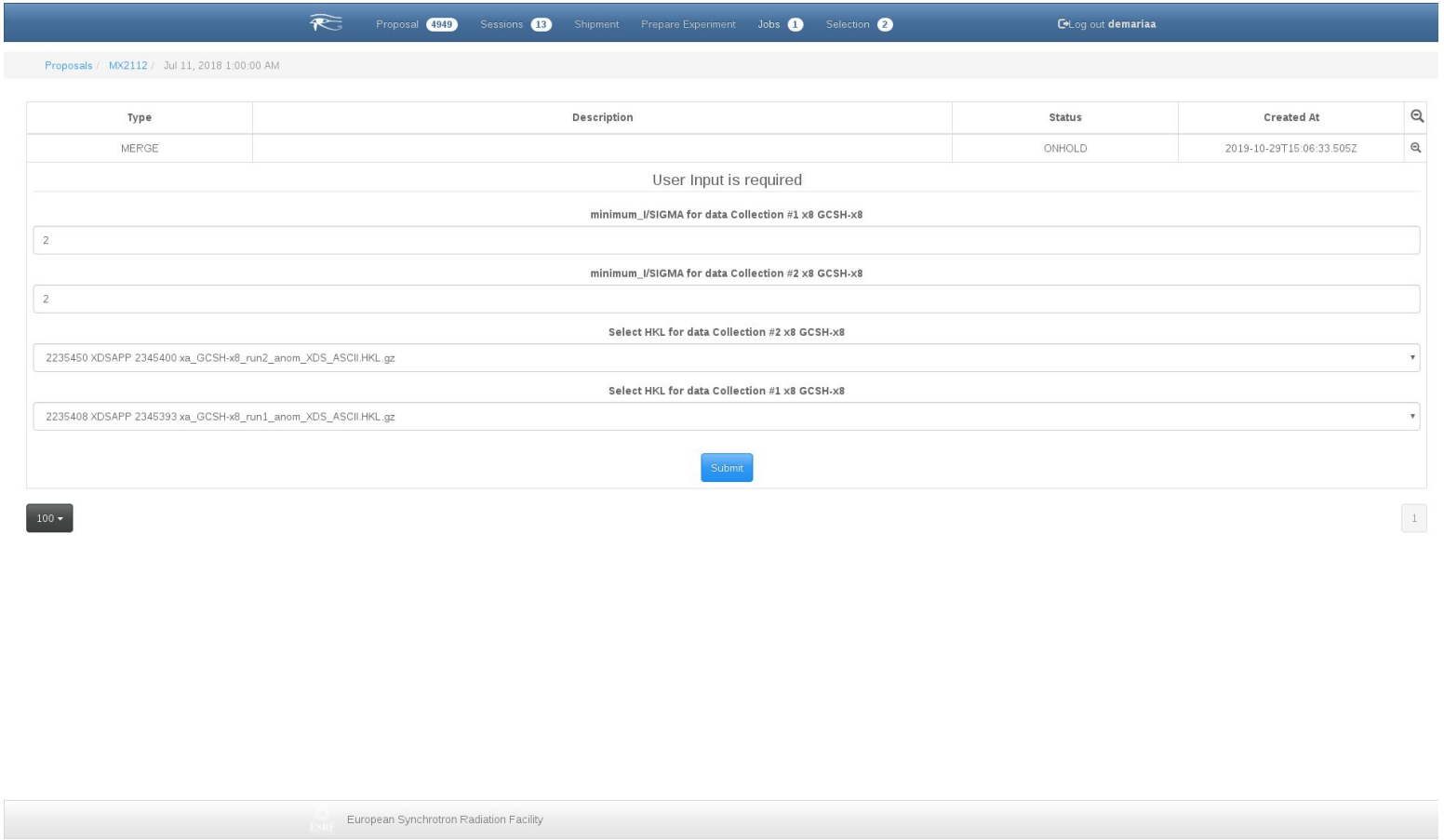


2) Launch the job



Run a job


Proposal 4949
Sessions 13
Shipment
Prepare Experiment
Jobs 1
Selection 2
Log out demariaa
Proposals / MX2112 / Jul 11, 2018 1:00:00 AM
Type
Description
Status
Created At
MERGE
[Merge] 1 GCSH-x8 (880 images), 2 GCSH-x8 (430 images),
PENDING
2019-10-29T15:06:33.505Z
No input required
Input parameters
key
value
datacollectionids
2345393,2345400
description
[Merge] 1 GCSH-x8 (880 images), 2 GCSH-x8 (430 images),
10
1
100
1
European Synchrotron Radiation Facility


Proposal 4949
Sessions 13
Shipment
Prepare Experiment
Jobs 1
Selection 2
Log out demariaa
Proposals / MX2112 / Jul 11, 2018 1:00:00 AM
Type
Description
Status
Created At
MERGE
ONHOLD
2019-10-29T15:06:33.505Z
User Input is required
minimum_I/SIGMA for data Collection #1 x8 GCSH-x8
2
minimum_I/SIGMA for data Collection #2 x8 GCSH-x8
2
Select HKL for data Collection #2 x8 GCSH-x8
2235450 XDSAPP 2345400 xa_GCSH-x8_run2_anom_XDS_ASCII.HKL.gz
Select HKL for data Collection #1 x8 GCSH-x8
2235408 XDSAPP 2345393 xa_GCSH-x8_run1_anom_XDS_ASCII.HKL.gz
Submit
100
1
European Synchrotron Radiation Facility

Run a job

Type	Description	Status	Created At
MERGE	[Merge] 1 GCSH-x8 (880 images), 2 GCSH-x8 (430 images),	RUNNING	2019-10-29T15:06:33.505Z