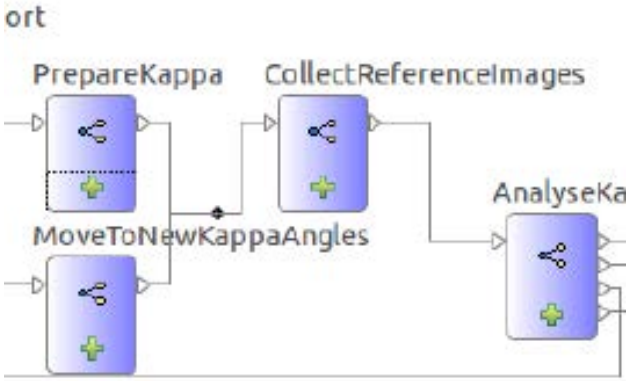
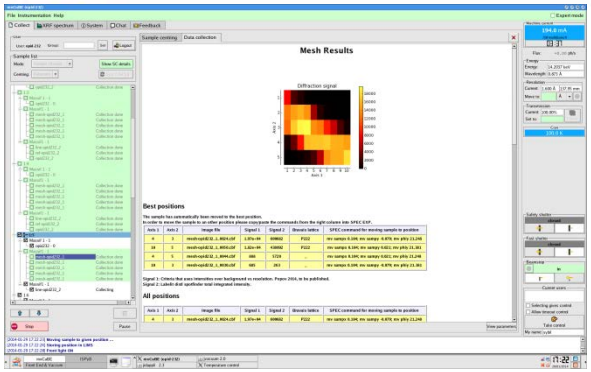
ESRF
The European Synchrotron





Workflows for Structural Biology

Olof Svensson
Data Analysis Unit
ESRF

WORKFLOWS ⇨ WORKFLOW TOOL

Wikipedia workflow definition:

"A workflow consists of a sequence of connected steps where each step follows without delay or gap and ends just before the subsequent step may begin"

Therefore, any experiment (manual or automated) and most computer programs can be described as a workflow!

The work described in this presentation uses a workflow tool for designing and executing the workflows for complex structural biology experiments at ESRF MX beamlines

WHY USE A WORKFLOW TOOL?

They offer a visual, higher level programming language than traditional programming languages like C, Python, Fortran etc

The goal is not to replace these languages but to complement them!

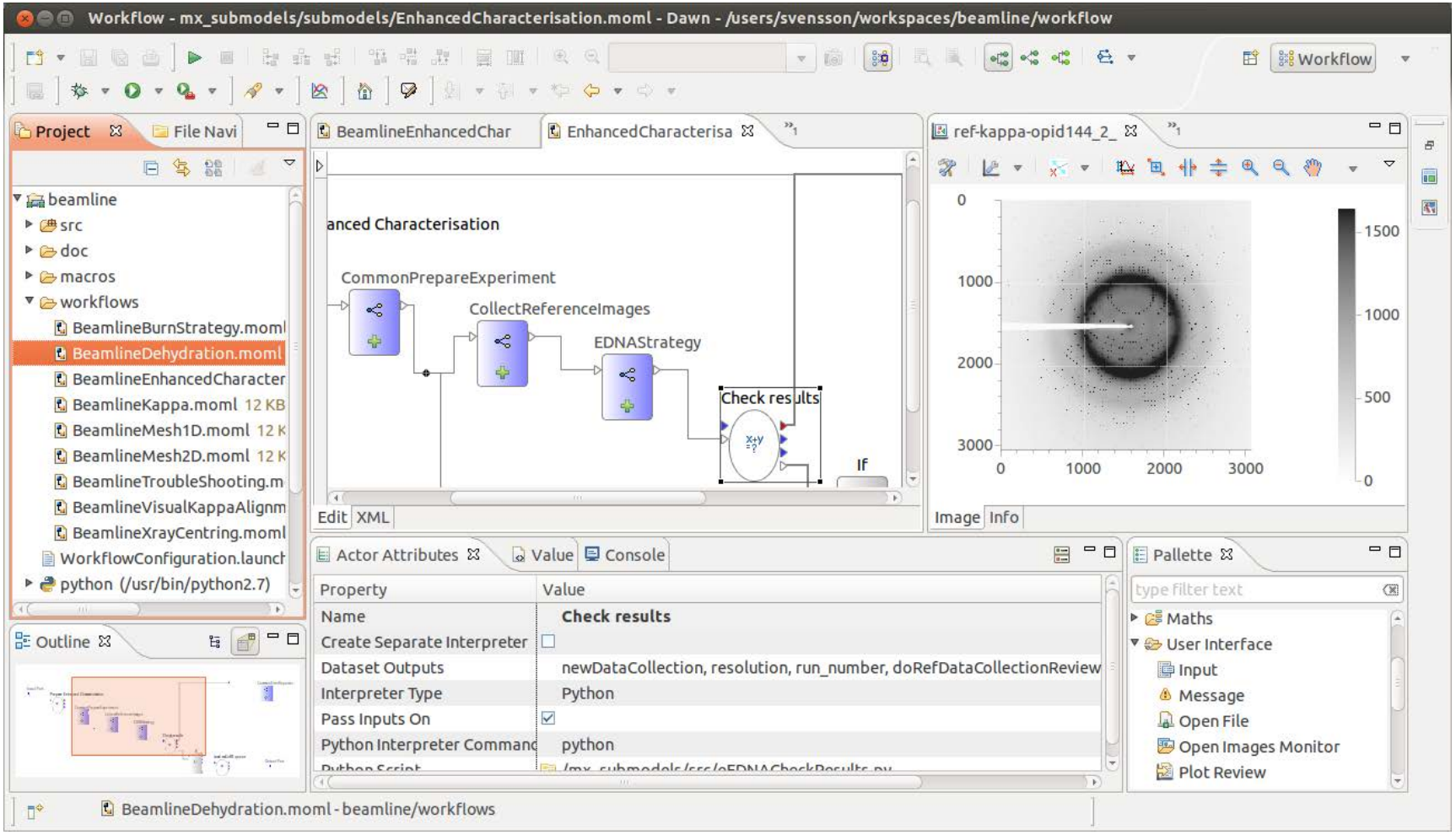
Workflows facilitates development of the high level analysis:

- Visual programming → (beamline) scientists can participate in the design and make modifications
- Easy to implement parallelism, error handling, LIMS connection etc.
- Documentation by design!

Workflow tools for data analysis :

- Widely used in many scientific fields e.g. biology
- New for synchrotron radiation facilities

THE WORKFLOW TOOL : DAWN AND PASSERELLE



http://www.dawnsci.org

WORKFLOWS DEPLOYED AND UNDER DEVELOPMENT

The following workflows have been deployed on ESRF MX beamlines:

Sample positioning:

- Mesh scan
- X-ray centring

Kappa goinostat re-orientation:

- Cell re-orientation using STAC (align lattice vector with rotation axis, optimise anomalous signal etc)
- Visual sample re-orientation

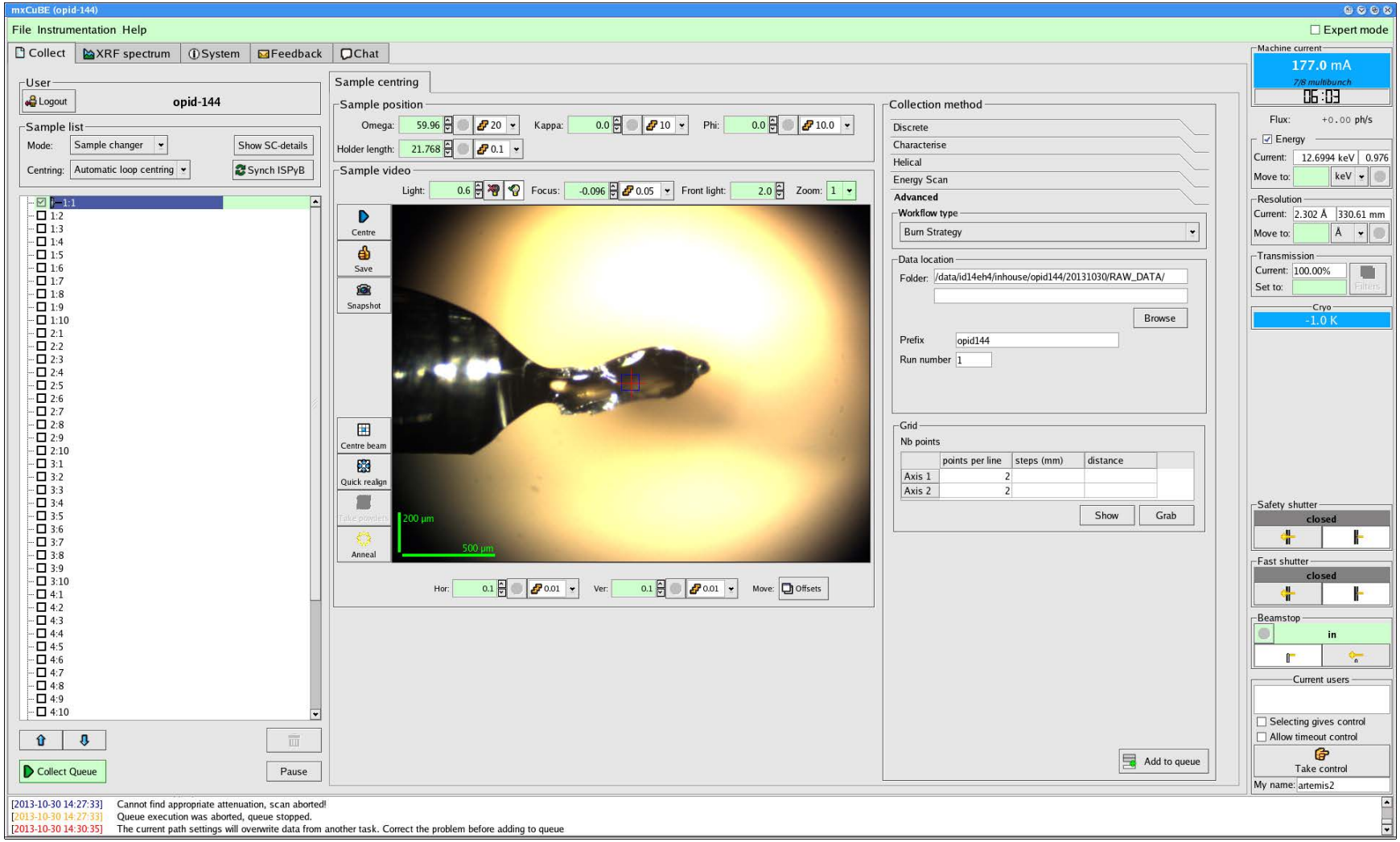
Sample characterisation:

- Enhanced EDNA characterisation
- Helical EDNA characterisation

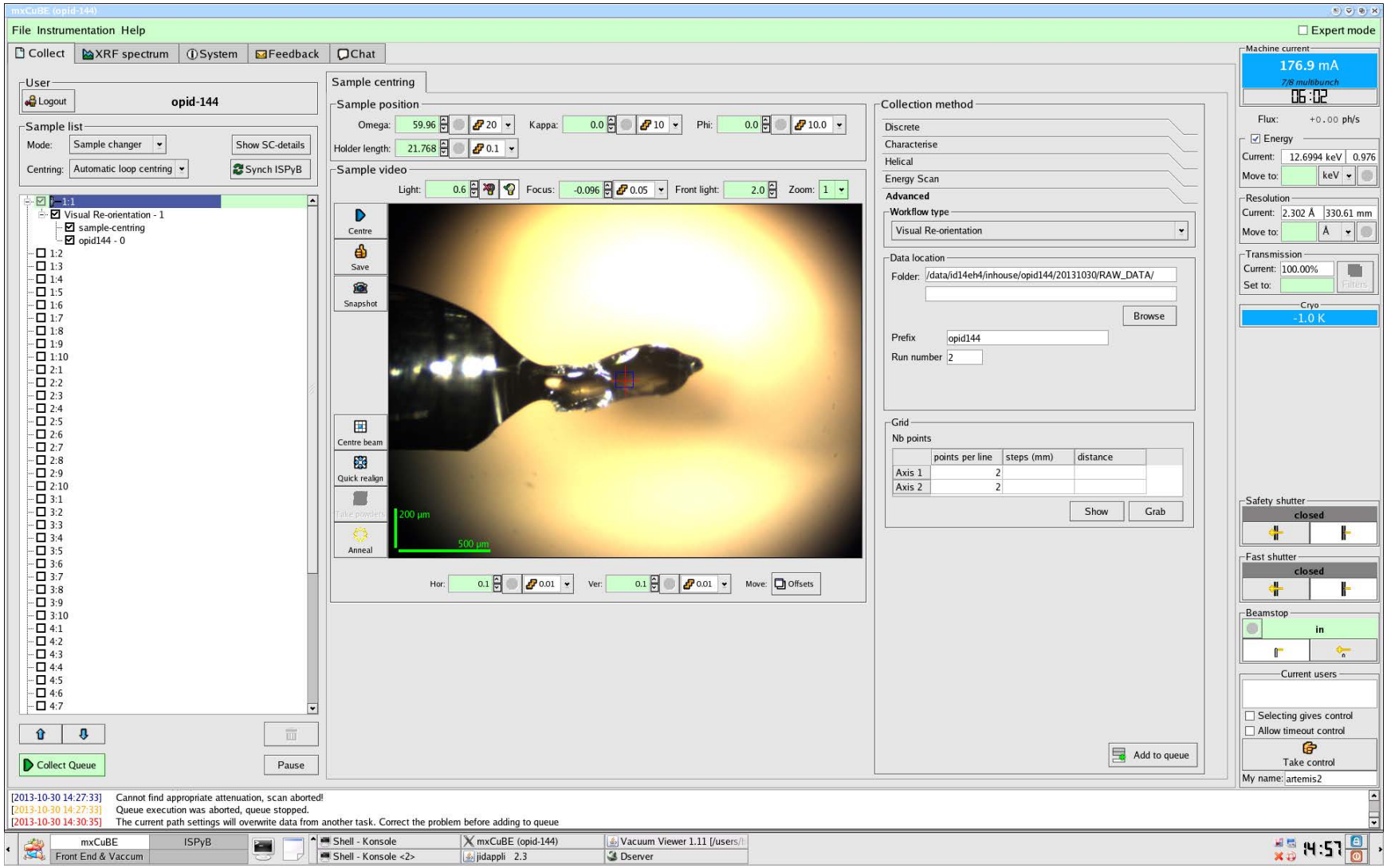
Automatic control of dehydration device

Automatic estimate of crystal radiation damage susceptibility (burn strategy)

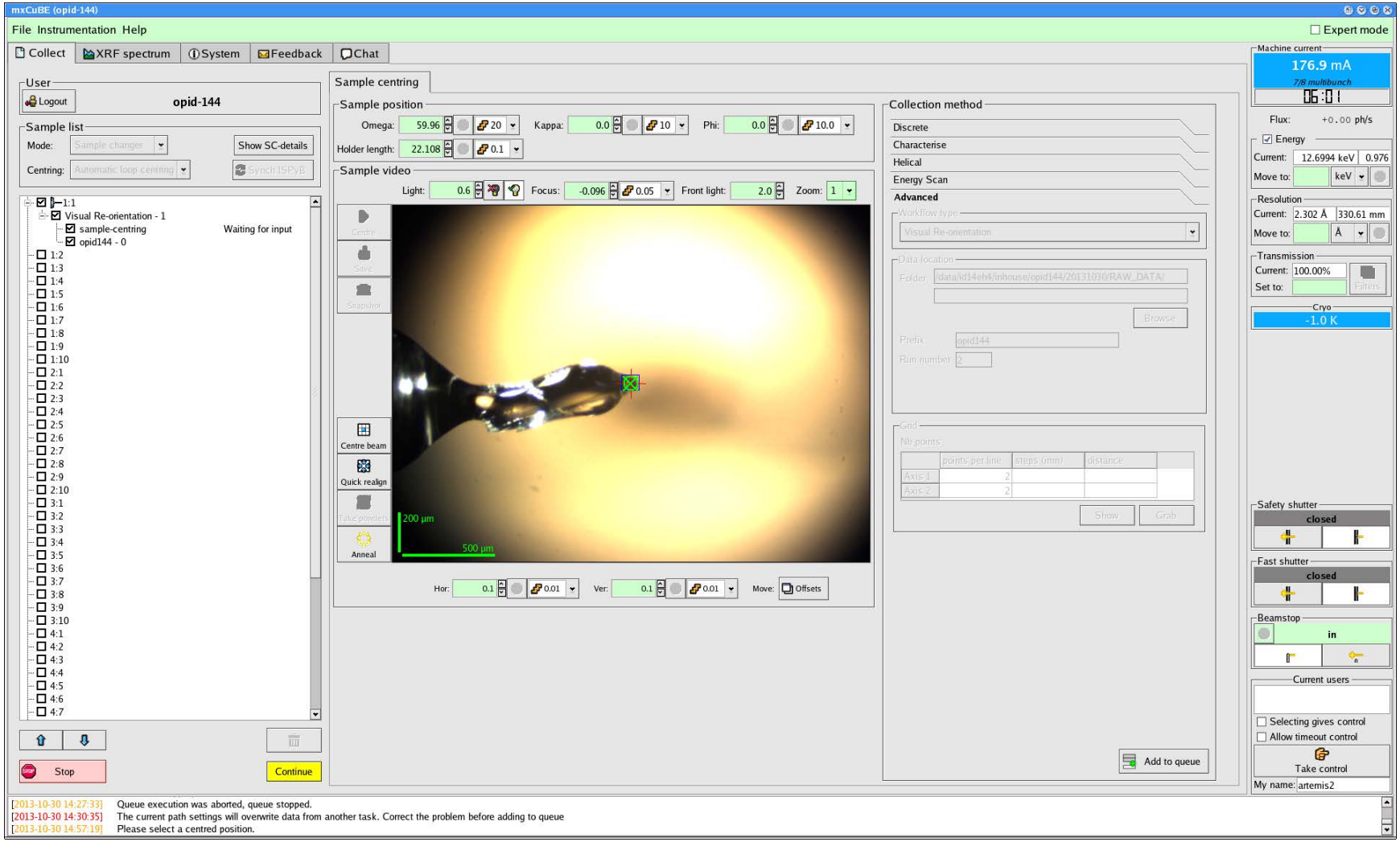
VISUAL KAPPA GONIOSTAT RE-ORIENTATION



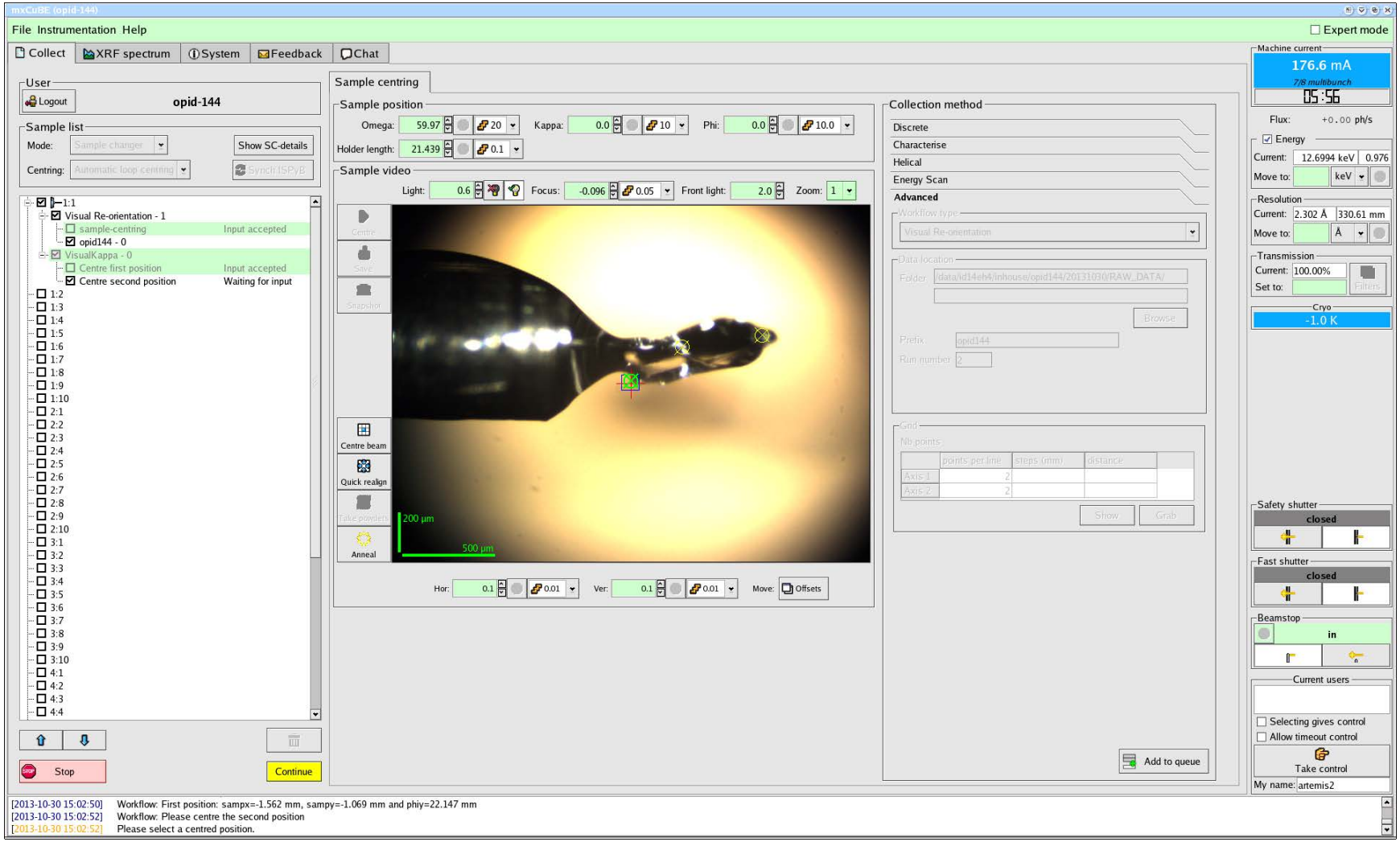
VISUAL KAPPA GONIOSTAT RE-ORIENTATION



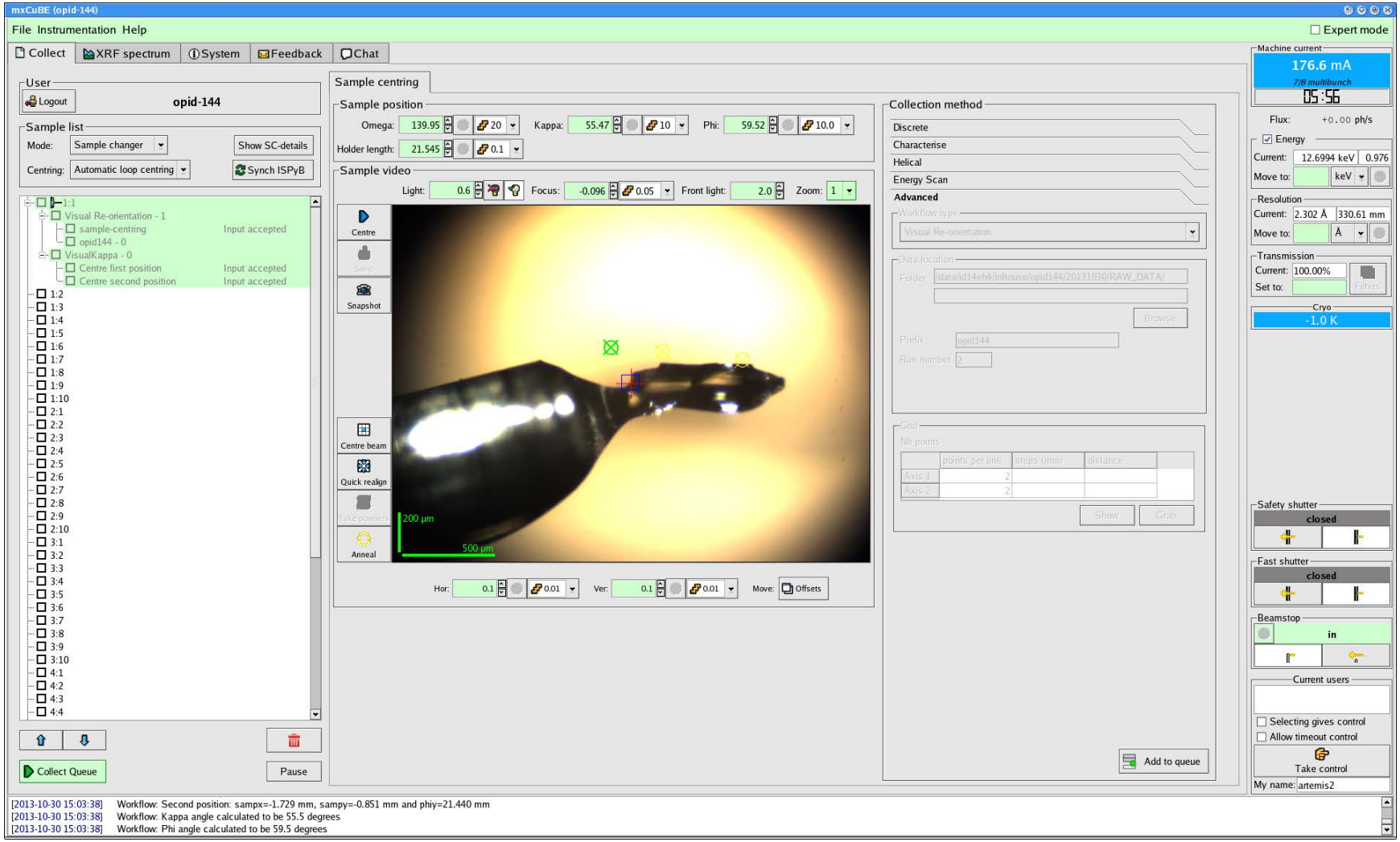
VISUAL KAPPA GONIOSTAT RE-ORIENTATION



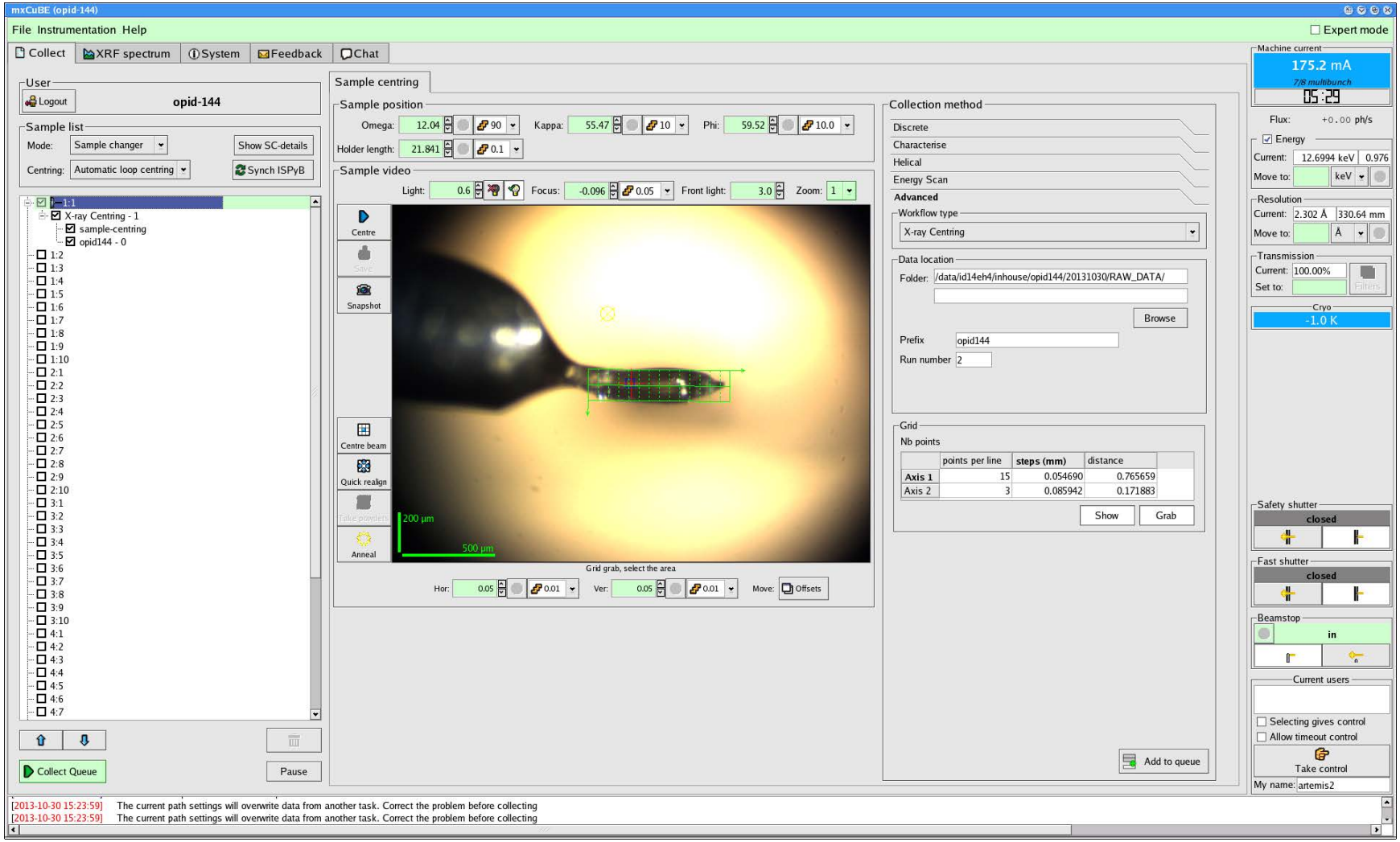
VISUAL KAPPA GONIOSTAT RE-ORIENTATION



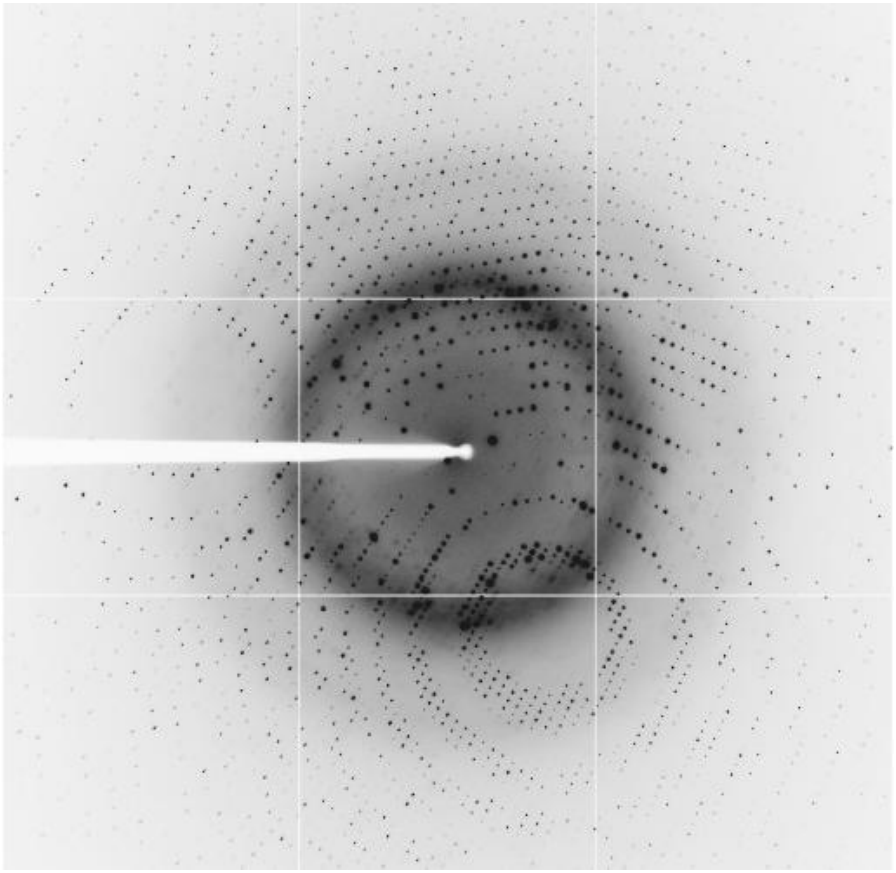
VISUAL KAPPA GONIOSTAT RE-ORIENTATION

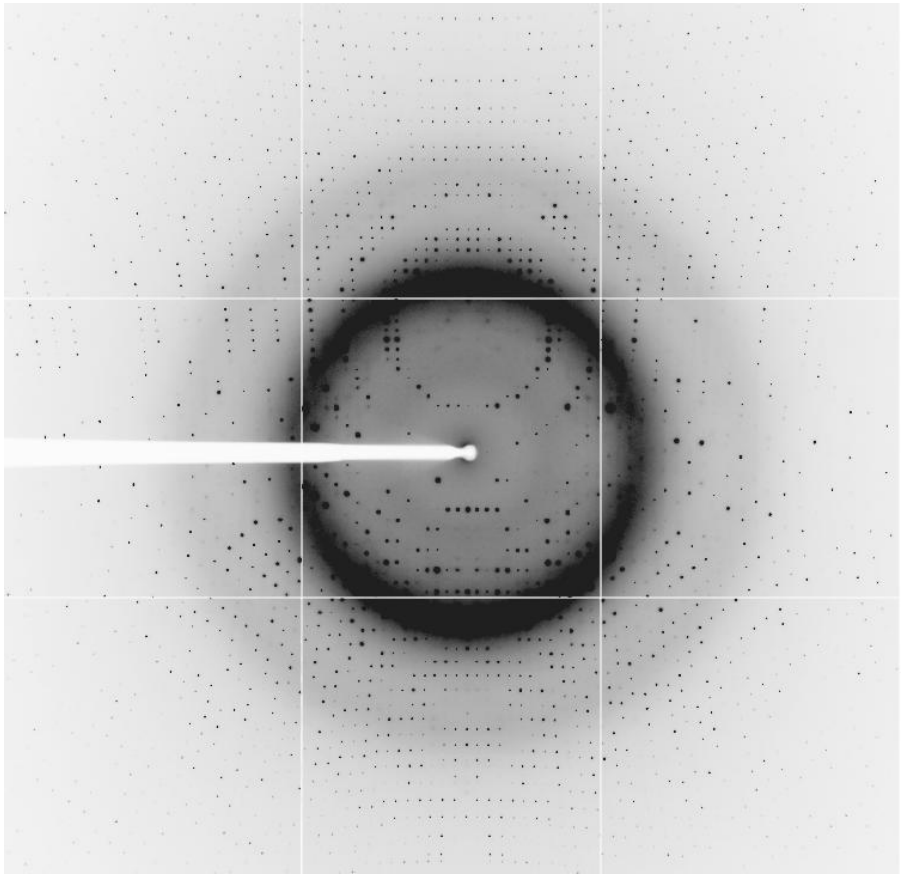


VISUAL KAPPA GONIOSTAT RE-ORIENTATION

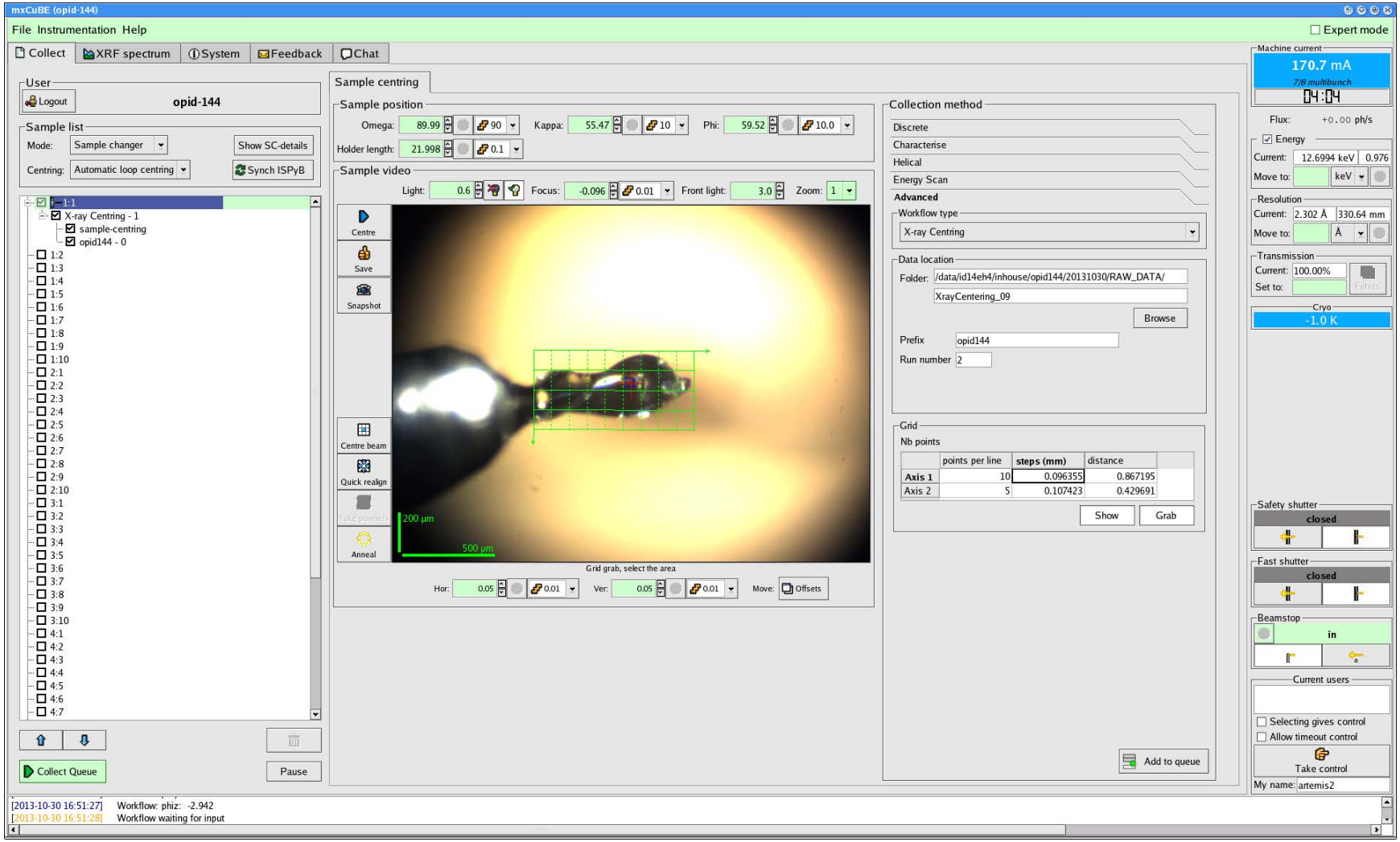


KAPPA GONIOSTAT CELL RE-ORIENTATION

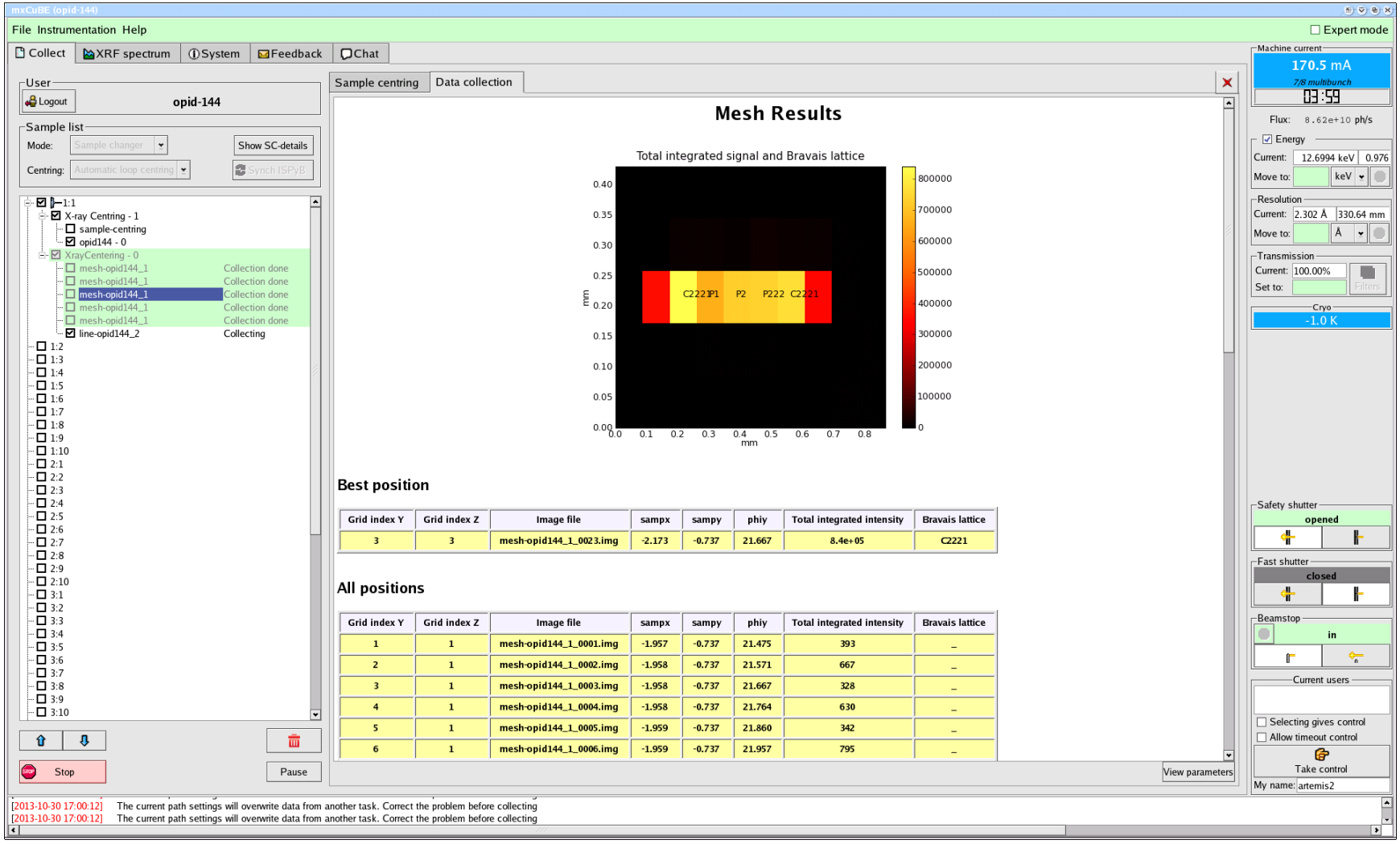




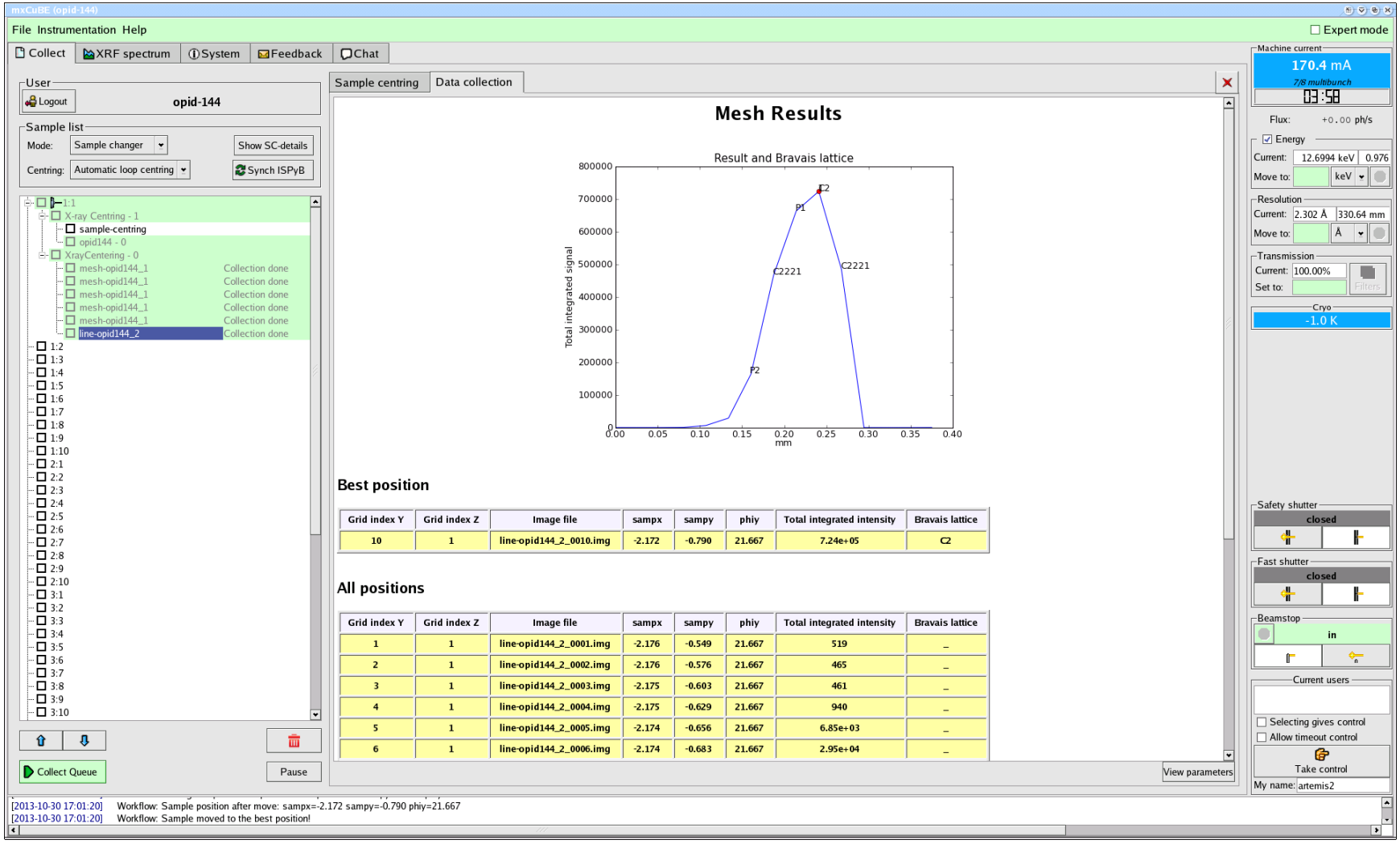
X-RAY CENTRING



X-RAY CENTRING



X-RAY CENTRING

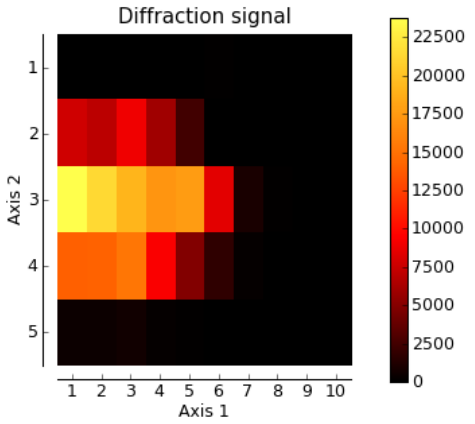


MESH SCAN DATA ANALYSIS

Two programs are run in parallel for each mesh scan image:

- Background 3D (written by Sasha Popov, ESRF)
- Labelit / Phenix spot finder (http://www.phenix-online.org)

For the five mesh scan images with the highest score MOSFLM indexing is executed:



Best position

The sample has automatically been moved to the best position.
In order to move the sample to an other position please copy/paste the commands from the right column into SPEC EXP.

Axis 1	Axis 2	Image file	Signal 1	Signal 2	Bravais lattice	SPEC command for moving sample to position
1	3	mesh-opid232_1_0021.cbf	2.38e+04	1016817	P222	mv sampx 0.182; mv sampy -0.807; mv phiy 21.380

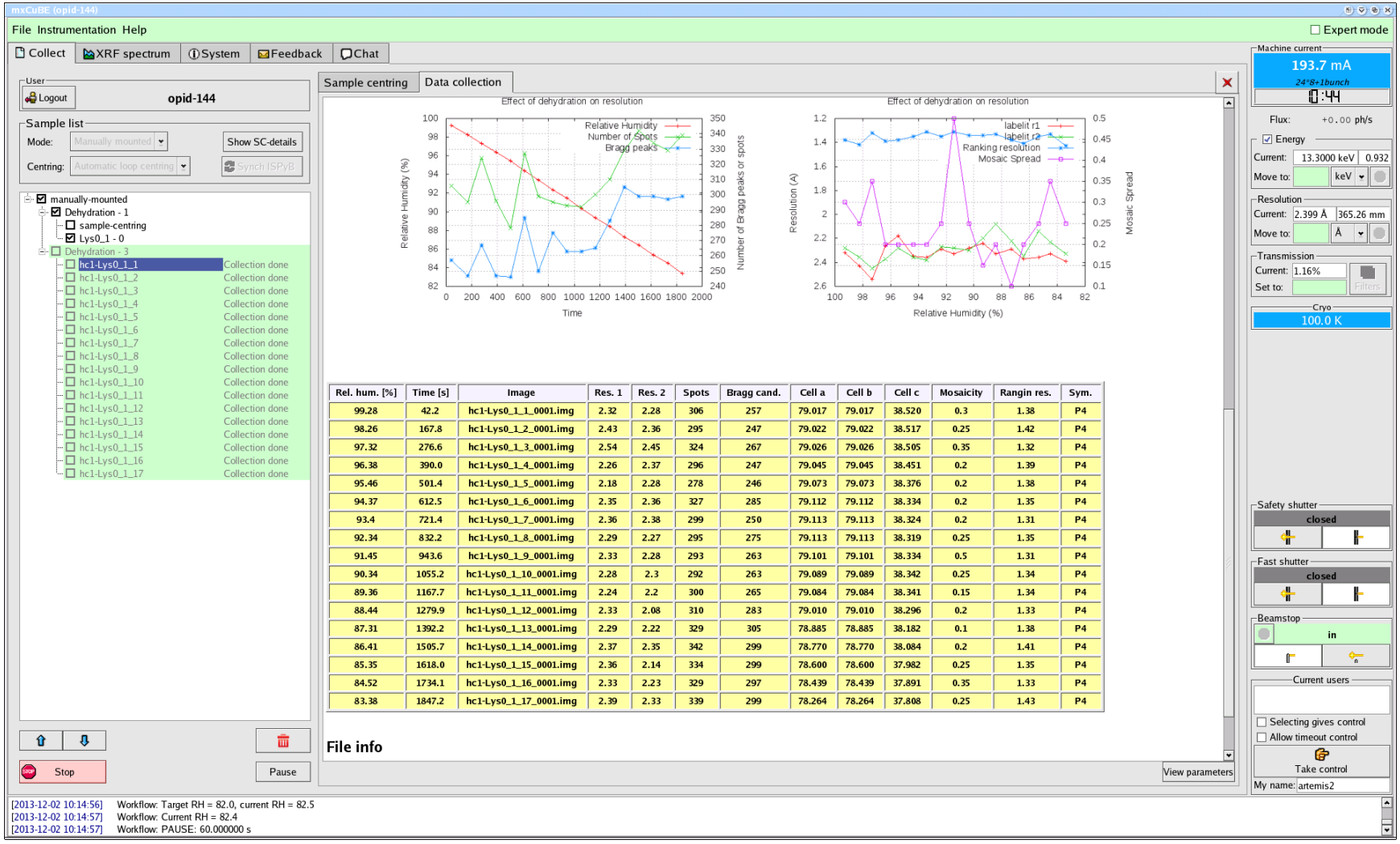
Signal 1: Criteria that uses intensities over background vs resolution. Popov 2014, to be published.
Signal 2: Labelit distl spotfinder total integrated intensity.

All positions

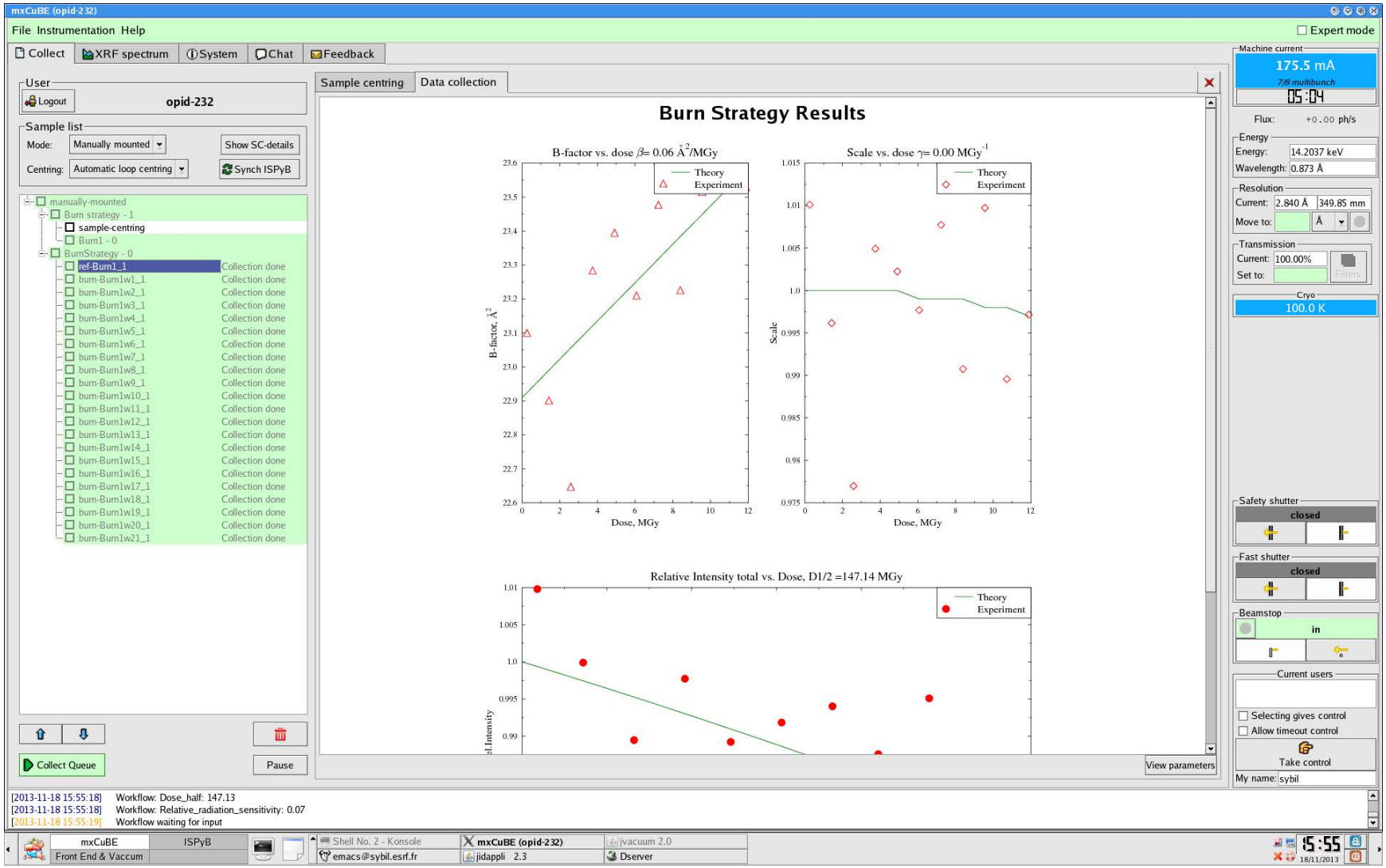
Axis 1	Axis 2	Image file	Signal 1	Signal 2	Bravais lattice	SPEC command for moving sample to position
1	3	mesh-opid232_1_0021.cbf	2.38e+04	1016817	P222	mv sampx 0.182; mv sampy -0.807; mv phiy 21.380
2	3	mesh-opid232_1_0022.cbf	2.14e+04	714677	P222	mv sampx 0.182; mv sampy -0.807; mv phiy 21.402
3	3	mesh-opid232_1_0023.cbf	1.92e+04	726267	P222	mv sampx 0.182; mv sampy -0.807; mv phiy 21.424
5	3	mesh-opid232_1_0025.cbf	1.75e+04	537506	P222	mv sampx 0.182; mv sampy -0.807; mv phiy 21.469
4	3	mesh-opid232_1_0024.cbf	1.72e+04	541746	P1	mv sampx 0.182; mv sampy -0.807; mv phiy 21.446
3	4	mesh-opid232_1_0033.cbf	1.53e+04	423657	_	mv sampx 0.182; mv sampy -0.757; mv phiy 21.424
2	4	mesh-opid232_1_0032.cbf	1.4e+04	521400	_	mv sampx 0.182; mv sampy -0.757; mv phiy 21.402
1	4	mesh-opid232_1_0031.cbf	1.4e+04	688343		mv sampx 0.182; mv sampy -0.757; mv phiy 21.380

DEHYDRATION

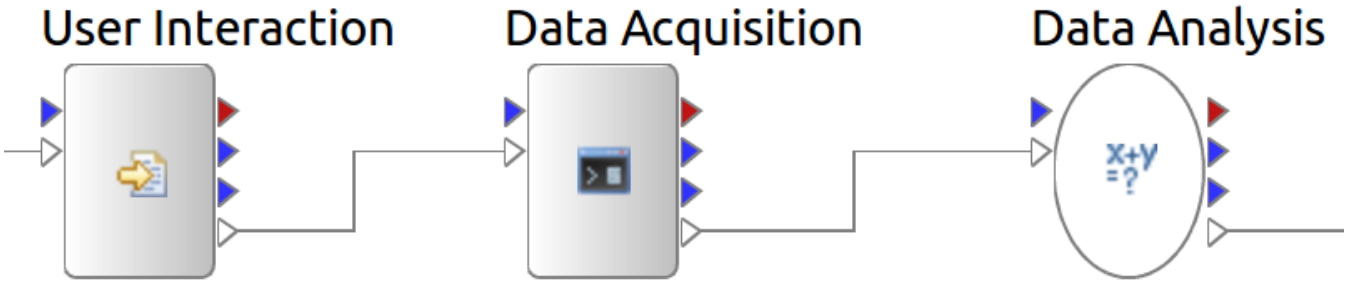
DEHYDRATION



AUTOMATIC ESTIMATE OF CRYSTAL RADIATION DAMAGE SUSCEPTIBILITY



GENERAL ASPECTS OF ESRF MX WORKFLOWS



- Executed from mxCuBE
- Log and working directories created in PROCESSED_DATA
- Result HTML page displayed in mxCuBE
- Connected to ISPyB:
  - *Status and workflow meta data uploaded automatically*
  - *Log and result HTML pages uploaded to pyarch (long time archive)*
- Robust error catching / handling

FUTURE DEVELOPMENTS

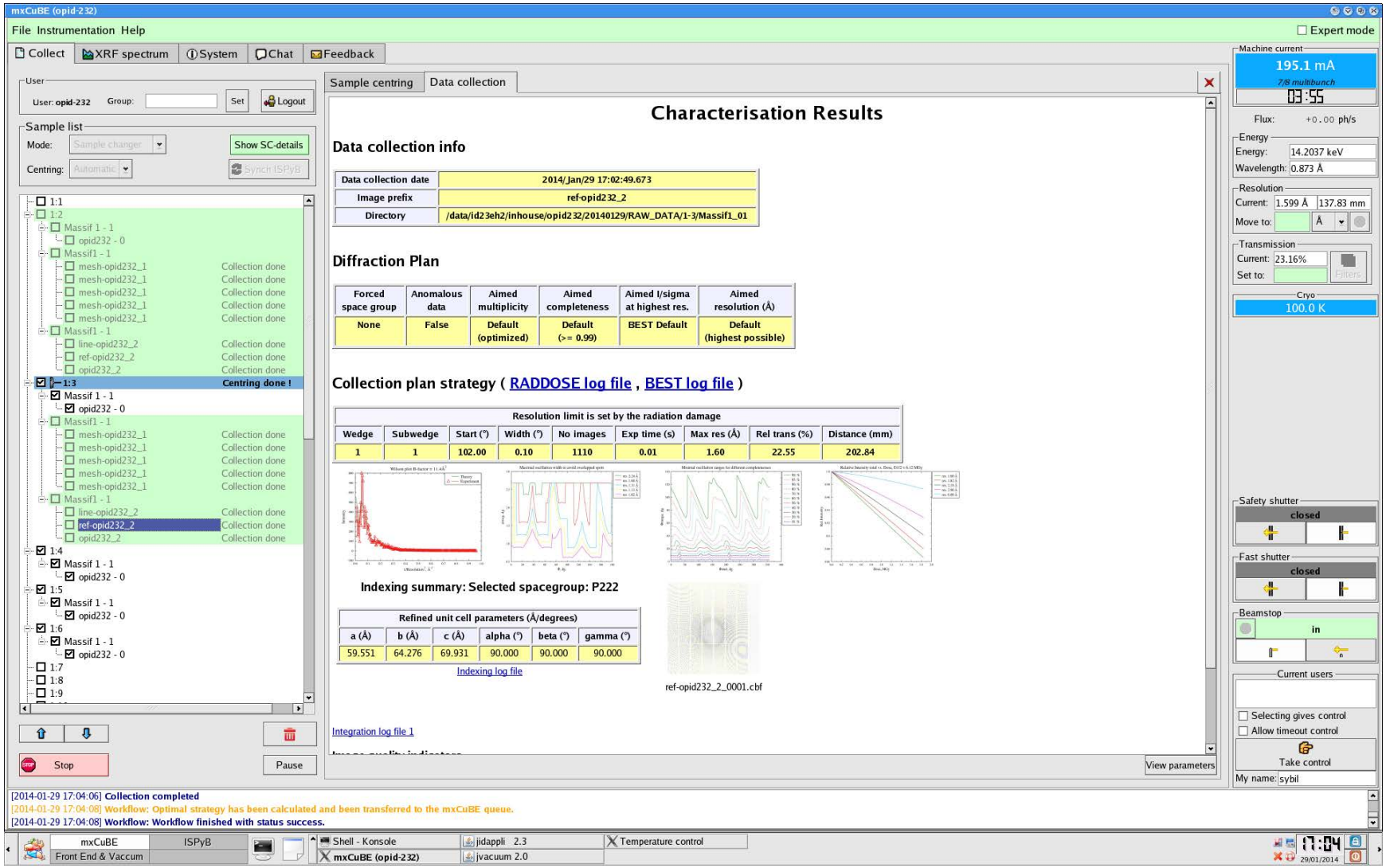
Improvements of existing workflows:

- Interaction with mxCuBE
- Error handling / recovery
- Mesh scan data analysis

The following workflows are currently under development:

- Diffraction tomography
- Fully automatic X-ray centring, enhanced characterisation and data collection
- Inter-leaved MAD data collection

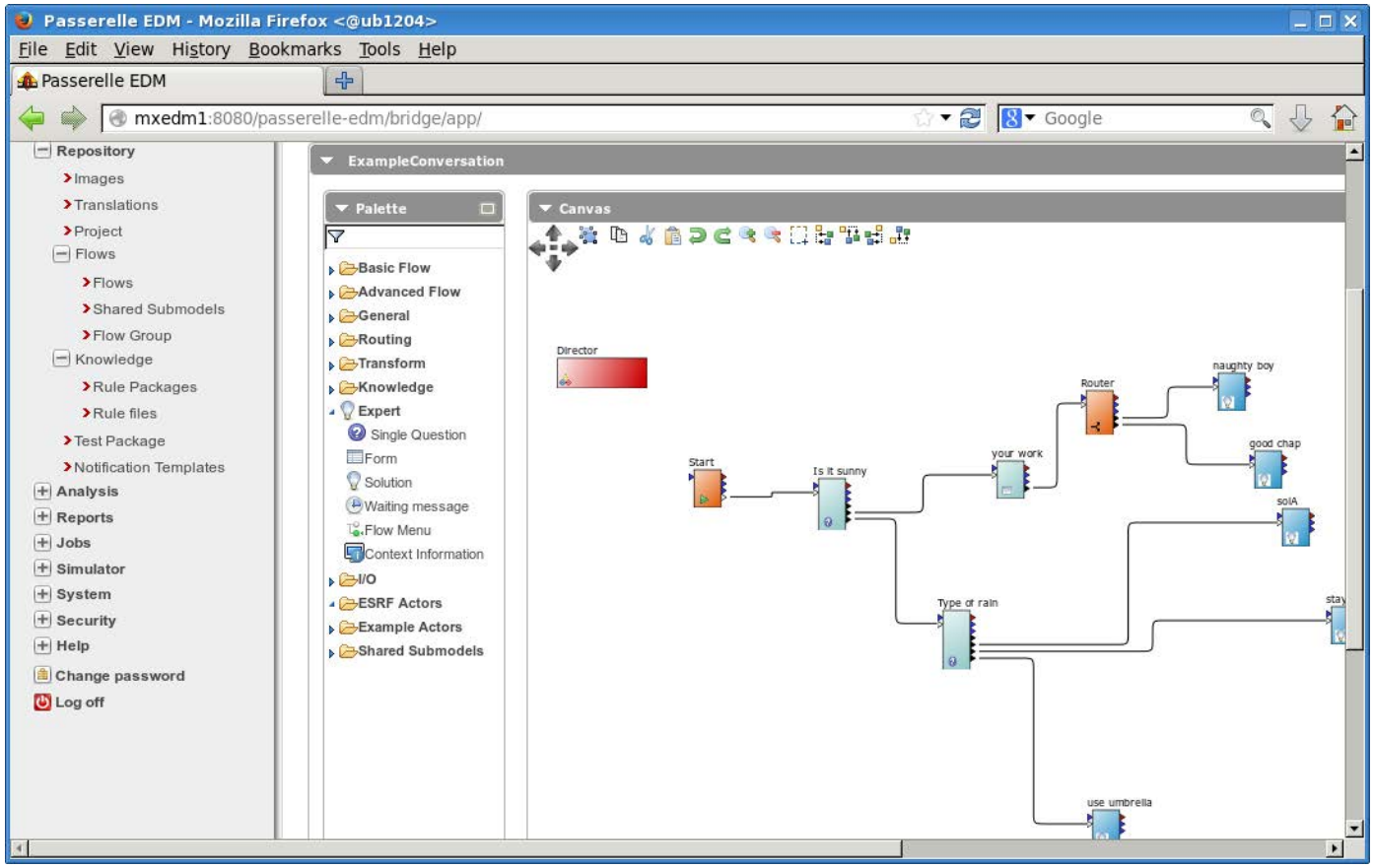
AUTOMATIC X-RAY CENTRING, CHARACTERISATION AND DATA COLLECTION



FUTURE DEVELOPMENTS – PASSERELLE EDM

Workflows for downstream processing – Passerelle Enterprise Decision Manager:

- Web based / collaborative workflow management
- Connection with ISPyB
- Goal: allow remote re-processing via ISPyB web interface



FUTURE DEVELOPMENTS – WORKFLOWS FOR BIOSAXS

The following workflows have been proposed to be implemented for the ESRF bioSAXS beamline BM29:

- Burn strategies to assess radiation sensitivity and optimize acquisition parameters.
- Comparison of separate concentrations and extrapolation to zero concentration
- Merging of regions not affected by radiation damage
- Comparison of separate concentrations and taking actions to produce additional dilutions and triggering data collections
- Advanced HPLC analysis
- Automated water calibration

ACKNOWLEDGEMENTS

- Sandor Brockhauser, Matthew Bowler and Max Nanao (EMBL Grenoble)
- Andy Götz, Elspeth Gordon and Sasha Popov (ESRF)
- Erwin de Ley and Koen Heunick (Isencia, Belgium)
- Matthew Gerring and the DLS Scisoft Team
- The ESRF Data Analysis Unit, Beamline Control Unit and Structural Biology group

Further reading:

- S. Brockhauser, O. Svensson, M. W. Bowler, M. Nanao, E. Gordon, R. M. F. Leal, A. Popov, M. Gerring, A. A. McCarthy and A. Gotz, The use of workflows in the design and implementation of complex experiments in macromolecular crystallography, Acta Cryst. (2012). D68, 975-984