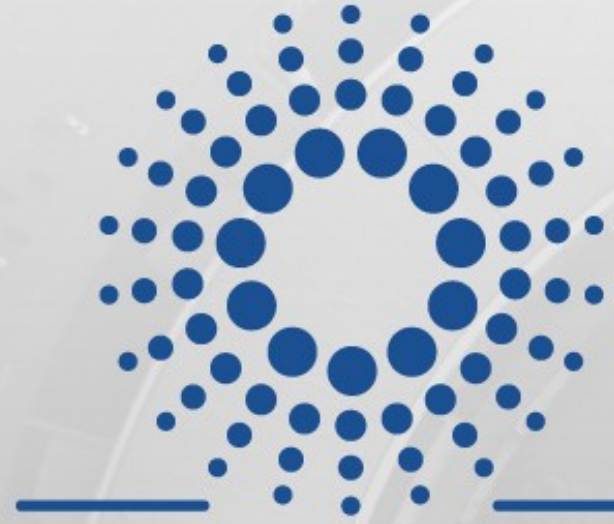
A Light for Science



ESRF

European Synchrotron Radiation Facility

Data Analysis Workbench

- Project started in June 2010
- 'Workbench for online data analyis'
- Based on and contributing to existing products
- Delivered using RCP
- ESRF management tracking project using Jira, http://jira:8080/browse/DAWB

  (Internal to ESRF)

Project Goals

1. Ability to view various file formats:
   - srs, dat, png, jpg, jpeg, tif, tiff, cbf, img, ciff, mccd, edf, pgm, cor, bruker, h5, nxs, pdb, (gz, bz2, zip)
   - Others?
2. Ability to visualize and run workflow pipelines.
3. Ability to interact with data and plotting using CPython.
4. Ability to configure data analysis in 'DAD' XML based system at ESRF.
5. Provide a workbench which can run other analysis tools.

Project Running

- Collabortations / APIs where possible, for example:
  - GDA/SDA, Passerelle, Fable, Eclipse/RCP, jmol etc. Other small APIs
- Supported on Windows, Mac and Linux
- Packaged to work 'out of the box'
- Unit tests, regression tests and nightly builds
- Agile project management

Why RCP?

- It uses SWT and JFace an MVC layer for SWT
- Programming at a higher level
- Reusable components available on web
  - CSV editor, GEF, XY Graphing
- Modular plugins and parts increase productivity
- Growing and large user community with significant inward investment.

Current Visualization Status

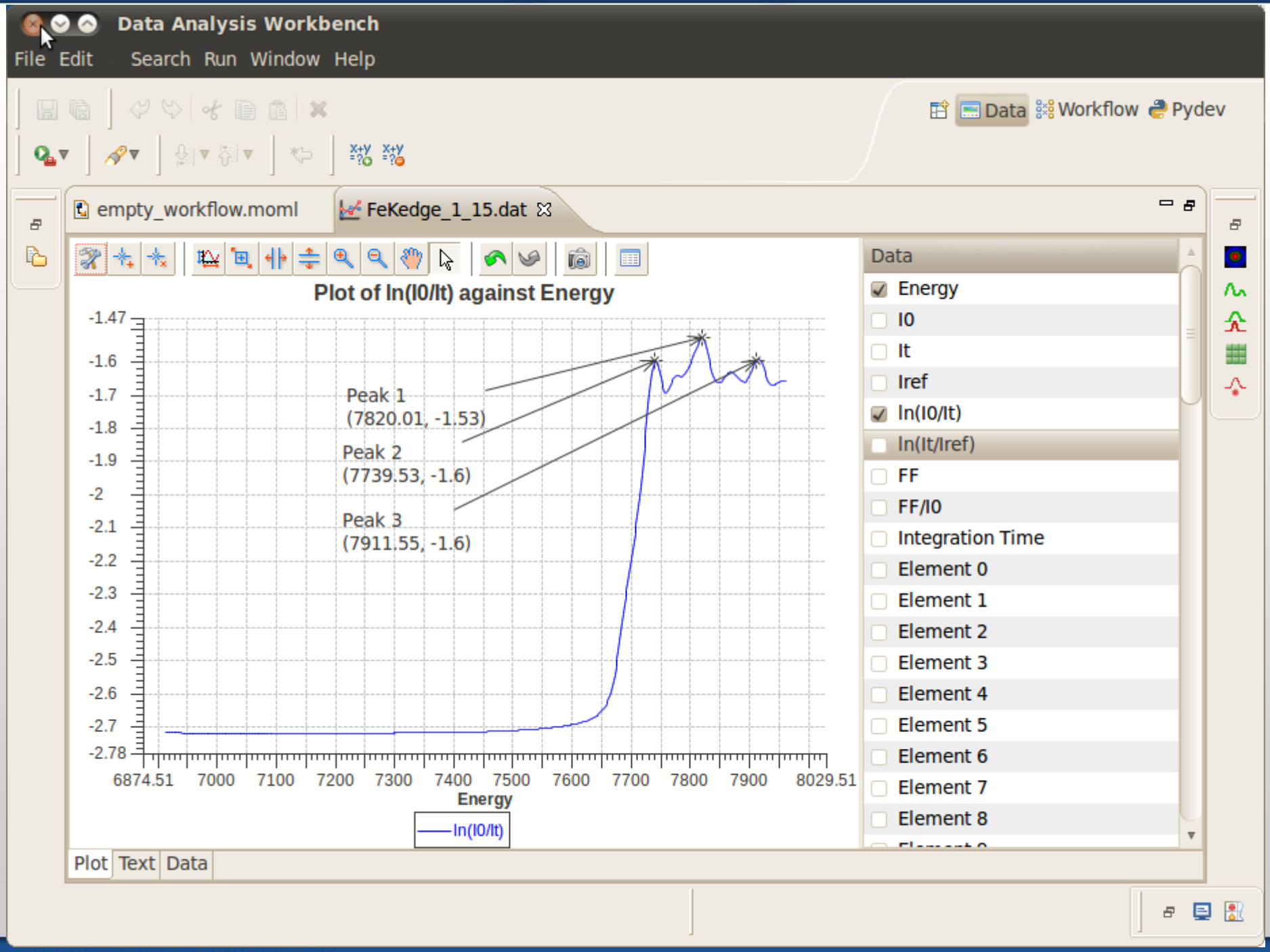
- Diamond SDA plotting
  - + Hardware acclerated, Same interface for 1D, 2D and 3D
  - - Specific OS and hardware requirements. Does not integrate to RCP properly. Crashes. Does not pass through remote desktop properly at ESRF.
  - > Does not work well at ESRF but is available in workbench just in case
- JLPlotting used by Fable for 1D
  - + Easy to use and used widely at ESRF
  - - Uses swing so not really an option
  - > Being phased out of fable 1D views
- Fable Image Viewer
  - + Works well on any system as uses raw SWT layer. Clean integration with RCP.
  - - Only supports images
  - > Chosen as the default image viewer in the workbench
- SWT XY Graph
  - + Works well on any system as useds raw SWT layer. Based on GEF.
  - - Only supports X/Y graphs
  - > Chosen as the defalut XY graph
- **Conclusion: SWT based graphics are the best choice for 1D and 2D**

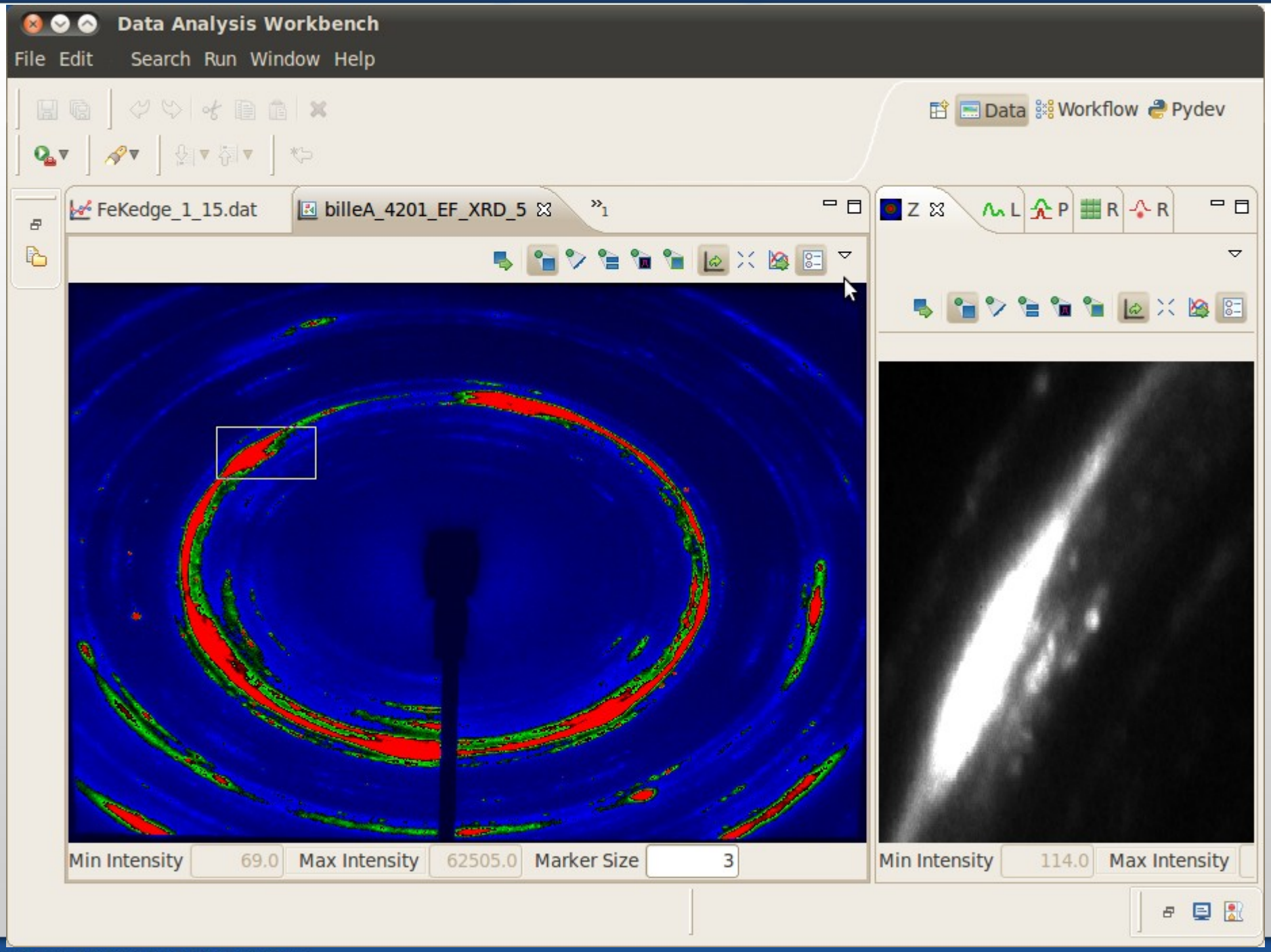
Current HDF5 Status

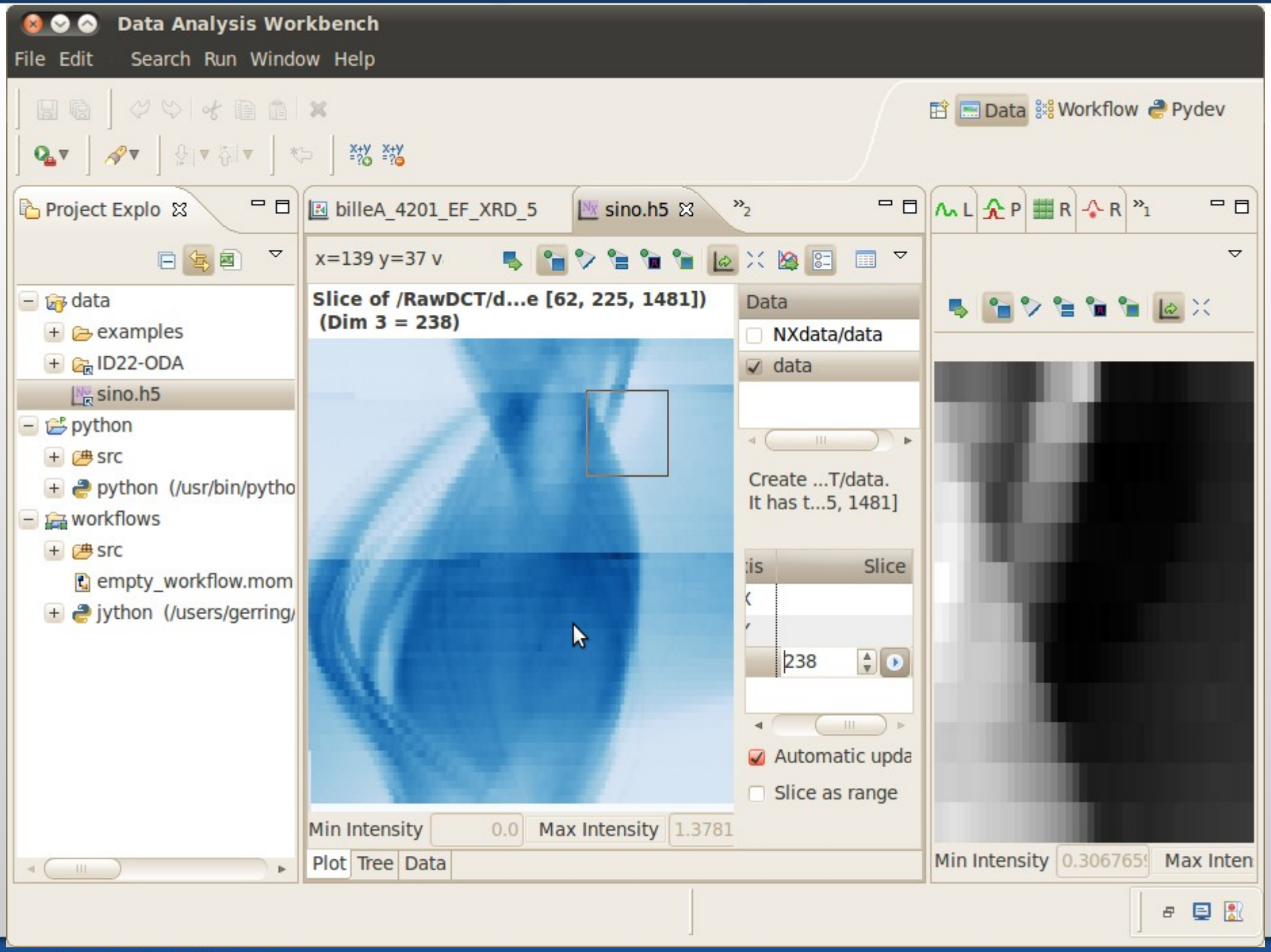
- Nexus plugin from Diamond
  - Poor stability in multi-treaded environment
  - Synchronization used to combat this.
  - Does not support most of old h5 files at ESRF
- HD5 plugin
  - Supports multi-threaded reading
  - Does not cause JVM exit with same tests as nexus
  - No need to set LD_LIBRARY_PATH
  - Works with larger files better based on testing at ESRF.
- Conclusion
  - Use raw hdf5 for reading files at ESRF.

Other SDA Experiences

- Complex dependency tree, highly coupled.
- Abscence of clean and reusable API's.
- Side plot system poor RCP design.
- Plotting causing OS crash, application hanging and application exits when tested at ESRF.
- Design principles such as single use objects, DRY and design patterns not currently used.
- Poor unit testing and low quality.

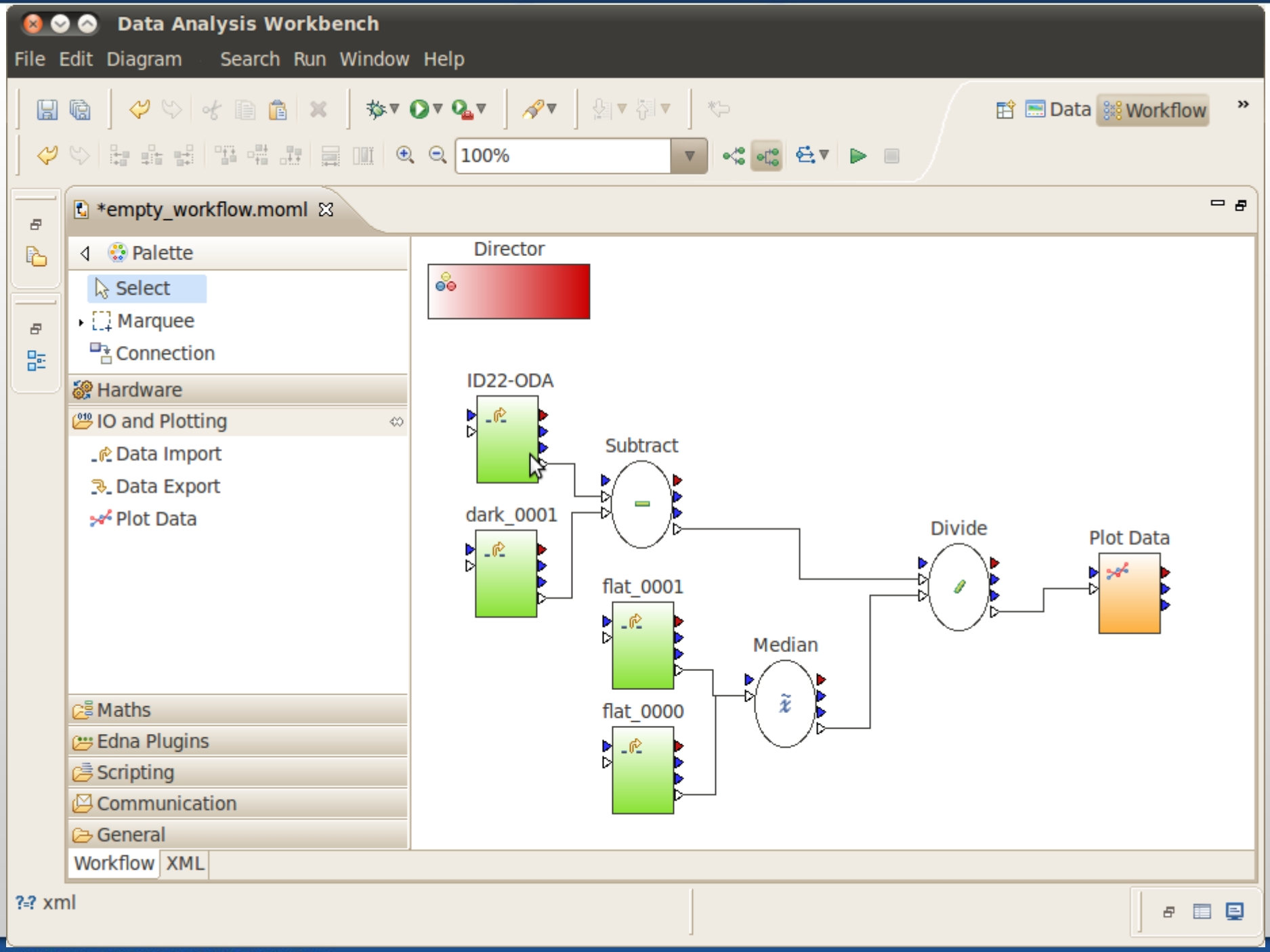

Data Analysis Workbench
File Edit Search Run Window Help
Data Workflow Pydev
empty_workflow.moml
FeKedge_1_15.dat
Plot of ln(I0/It) against Energy
Peak 1
(7820.01, -1.53)
Peak 2
(7739.53, -1.6)
Peak 3
(7911.55, -1.6)
-1.47
-1.6
-1.7
-1.8
-1.9
-2
-2.1
-2.2
-2.3
-2.4
-2.5
-2.6
-2.7
-2.78
6874.51 7000 7100 7200 7300 7400 7500 7600 7700 7800 7900 8029.51
Energy
ln(I0/It)
Data
Energy
I0
It
Iref
ln(I0/It)
ln(It/Iref)
FF
FF/I0
Integration Time
Element 0
Element 1
Element 2
Element 3
Element 4
Element 5
Element 6
Element 7
Element 8
Plot Text Data


Data Analysis Workbench
File Edit Search Run Window Help
Data
Workflow
Pydev
FeKedge_1_15.dat
billeA_4201_EF_XRD_5
Min Intensity 69.0 Max Intensity 62505.0 Marker Size 3
Z
L
P
R
R
Min Intensity 114.0 Max Intensity


Data Analysis Workbench
File Edit Search Run Window Help
Data
Workflow
Pydev
Project Explo
billeA_4201_EF_XRD_5
sino.h5
x=139 y=37 v
Slice of /RawDCT/d...e [62, 225, 1481])
(Dim 3 = 238)
data
examples
ID22-ODA
sino.h5
python
src
python (/usr/bin/pytho
workflows
src
empty_workflow.mom
jython (/users/gerring/
Data
NXdata/data
data
Create ...T/data.
It has t...5, 1481]
Slice
238
Automatic upda
Slice as range
Min Intensity 0.0
Max Intensity 1.3781
Plot
Tree
Data
L
P
R
R
Min Intensity 0.306765
Max Inten

Workflow for Results Analysis

- 'Simple to use subset of Python'
- A pipeline which runs as a service
- Workbench can author and deploy pipeline
- Workbench connects to and starts / stops pipeline
- Able to run nodes from:
  - CPython projects, KNIME extension points, SDA data analysis


Data Analysis Workbench
File Edit Diagram Search Run Window Help
Data
Workflow
100%
*empty_workflow.moml
Palette
Select
Marquee
Connection
Hardware
IO and Plotting
Data Import
Data Export
Plot Data
Maths
Edna Plugins
Scripting
Communication
General
Workflow
XML
?-? xml
Director
ID22-ODA
Subtract
dark_0001
flat_0001
flat_0000
Median
Divide
Plot Data

CPython in Workflow Integration

- Support of Numpy nodes.
- Support of EDNA nodes.
- Support of Fable diffraction nodes.
- Interfaces to transform data to and from workflow.
- Interfaces to support plotting.
- Expectation to use same memory space as Java process.

Extensions

- Connection of Data Collection and Data Analysis in integrated system.
- More Fable tools
- Bioclipse views/editors
- Domain Specific Language (DSL) for SPEC
- Beamline control with Passerelle

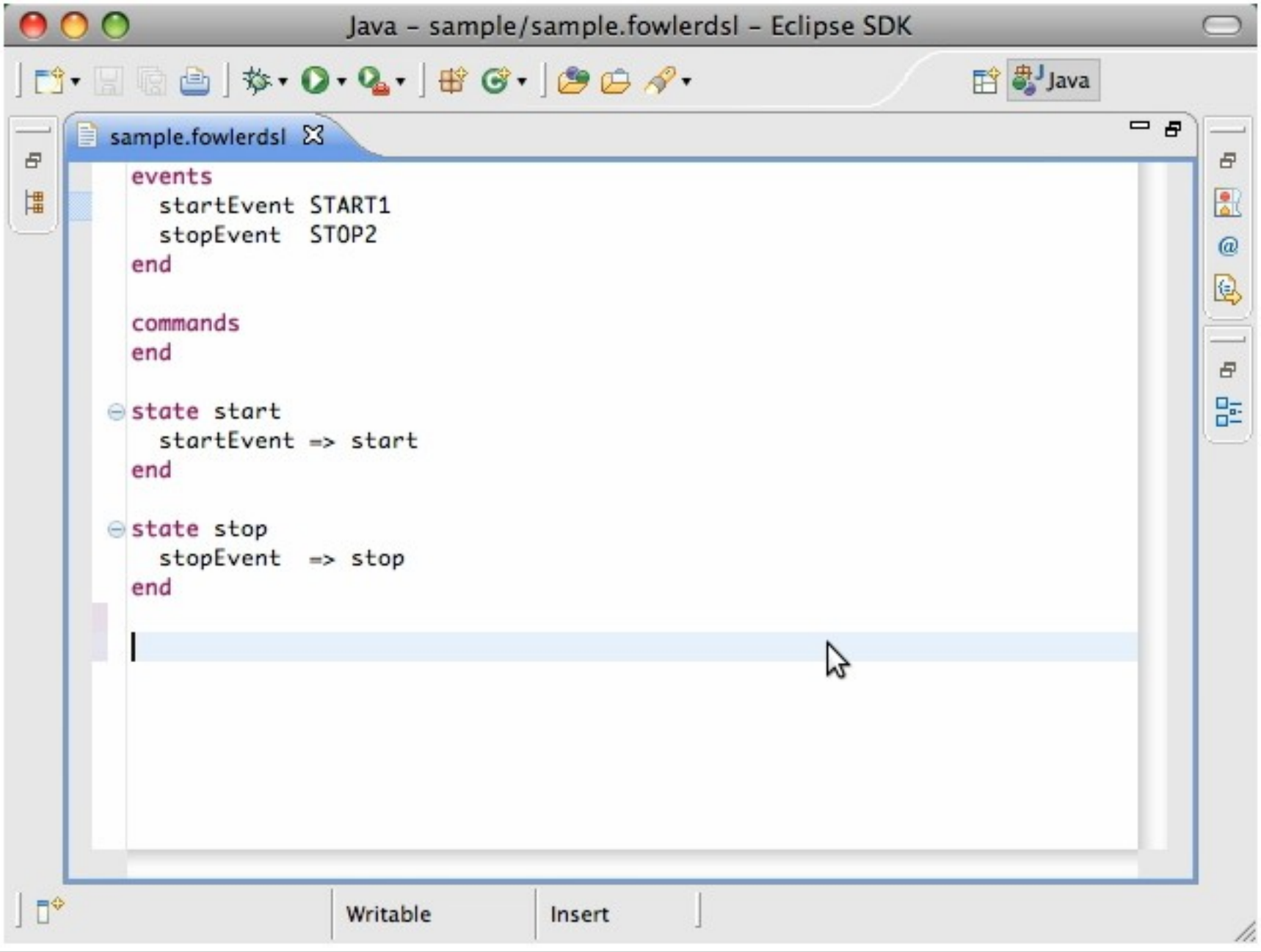
Fable Tools

- Some Fable tools being packaged with project.
- Being moved incrementally.
- Packaged in standard distribution.
- Heavy use of CPython in Fable a challenge.
- Diffraction tools most likely candiate to be moved next.

SPEC DSL

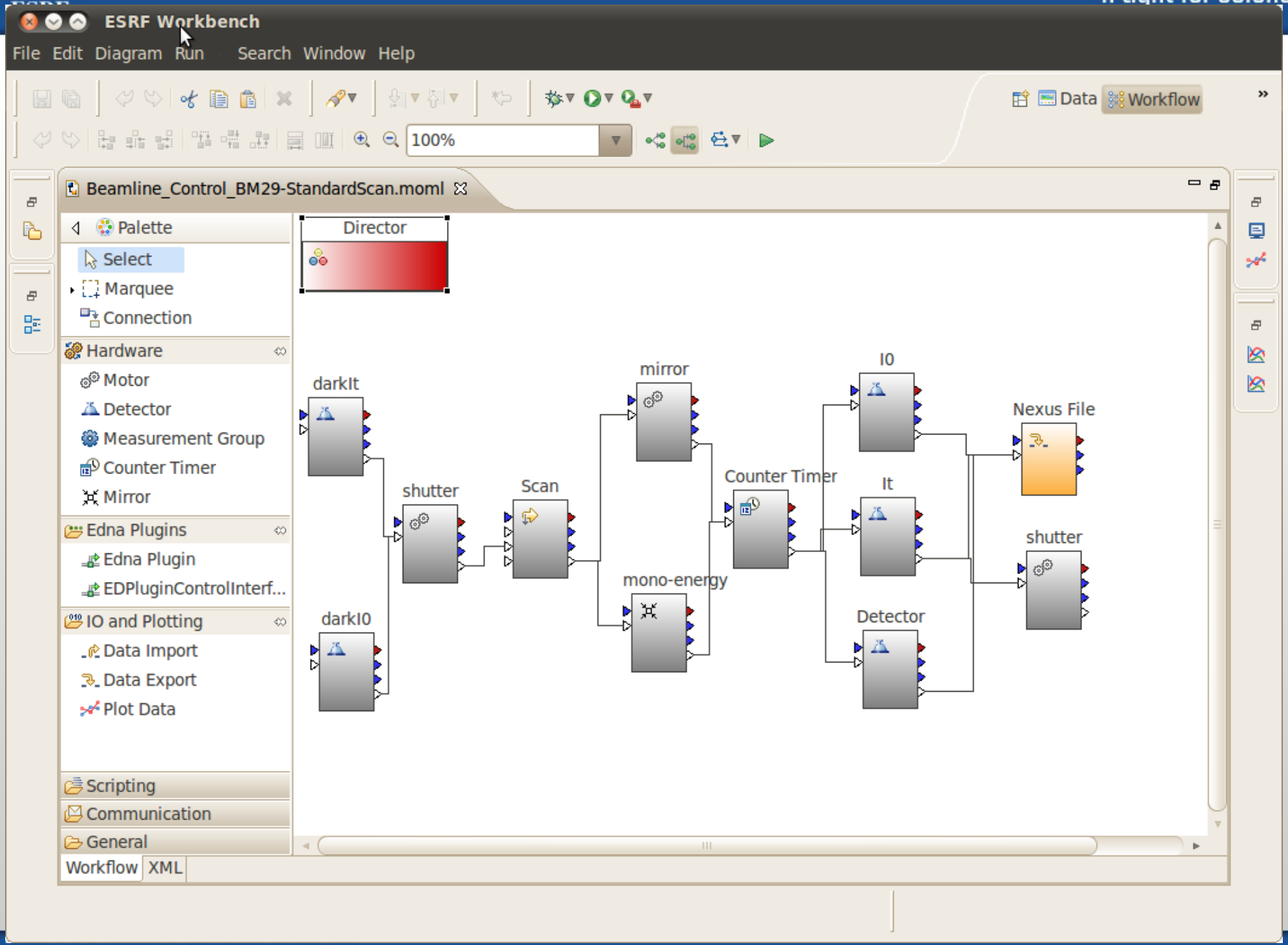
- Syntax checking, highlighting, search, refactoring of SPEC scripts.
- Scripts orgnanised into projects.
- Better integration with plotting and other UI operations.
- Ability to upgrade beamlines incrementally as SPEC is supported.
- Better ability to analysis and visualize script structure.
- Better sharing between beamlines.
- SPEC console integrated to rest of GUI.

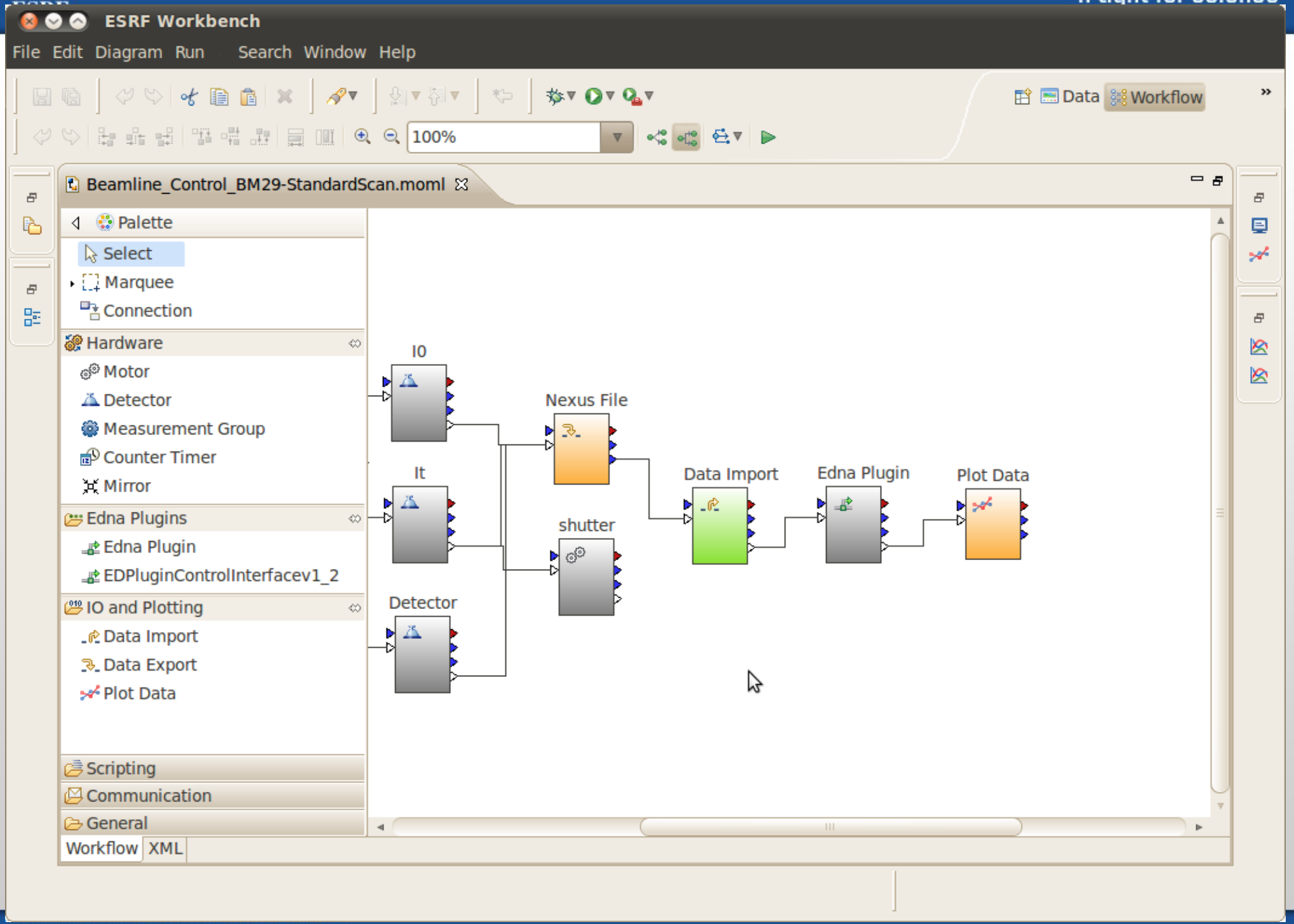
http://www.eclipse.org/Xtext/


Java – sample/sample.fowlerdsl – Eclipse SDK
Java
sample.fowlerdsl
events
startEvent START1
stopEvent STOP2
end
commands
end
state start
startEvent => start
end
state stop
stopEvent => stop
end
Writable
Insert

Beamline Control using Passerelle Workflows

- Ability to easily change experiment.
- Ability to add in new hardware and have it tighly coupled to the data collection.
- Ability to run the data analysis pipeline automatically within the data collection scheme.
- Hardware integrated to collection algorithm off site, allowing testing and setup before the visit
- Increase user throughput as experiment is run in simulation before visit.
- Supports SPEC, Tango, EPICS, GDA etc. As experiment configuration better abstracted from hardware.
- Multi-threading issues reduced.


ESRF Workbench
File Edit Diagram Run Search Window Help
Data
Workflow
100%
Beamline_Control_BM29-StandardScan.moml
Palette
Select
Marquee
Connection
Hardware
Motor
Detector
Measurement Group
Counter Timer
Mirror
Edna Plugins
Edna Plugin
EDPluginControlInterf...
IO and Plotting
Data Import
Data Export
Plot Data
Scripting
Communication
General
Workflow
XML
Director
darkIt
shutter
Scan
mirror
mono-energy
darkI0
Counter Timer
I0
It
Detector
Nexus File
shutter


ESRF Workbench
File Edit Diagram Run Search Window Help
Data
Workflow
100%
Beamline_Control_BM29-StandardScan.moml
Palette
Select
Marquee
Connection
Hardware
Motor
Detector
Measurement Group
Counter Timer
Mirror
Edna Plugins
Edna Plugin
EDPluginControlInterfacev1_2
IO and Plotting
Data Import
Data Export
Plot Data
Scripting
Communication
General
Workflow
XML
I0
It
Detector
Nexus File
shutter
Data Import
Edna Plugin
Plot Data

Conclusions

- Collaborations are good and should continue.
- Testing is required of code written in collaboration.
- Developers in collaboration need to develop resuable decoupled code.
- Collaborations need to consider all users and code should be created as reusable APIs.
- Eclipse/RCP is ideal platform for this.