EDNA

Quick status of the project

Olof Svensson
Data Analysis Unit
ISDD ESRF
Why do we need EDNA?

• EDNA is the best answer we (developers) have come up with so far for answering these questions:

  - How can we automate data analysis workflows?
    • “pipeline” existing scientific software for (online) data analysis workflows
    • abstract certain calculations to be “generic”, e.g. indexing of a diffraction pattern
    • “flexible” workflows, rapid changes depending on the scientific needs

  - How can we make these workflows robust?
    • easily adapted to new versions of scientific software packages

  - How can we collaborate efficiently?
    • re-use of code without breaking existing functionality
What is EDNA?

- EDNA is about collaboration:
  - Code sharing (SVN)
  - Coding conventions
  - Code reviews
  - Open source (LGPL, GPL)
  - Bug tracker
  - Wiki : http://www.edna-site.org
  - Memorandum of Understanding
  - Executive committee
  - Project manager / coordinator
  - Regular meetings / video conferences

- EDNA is a framework:
  - “Generic” kernel
    - Data modelling framework
    - Support for multi-threaded modules (plugins) development
    - Support for workflow development
    - Testing framework
  - “Specific” applications (MXv1, bioSaxs etc.)
  - Automatic testing and nightly builds
  - Automatic API doc generation
  - No GUI
EDNA Modularity: Plugins and their hierarchy

- Plugin base class:
  - Configuration, working directory, etc.

- Execution plugins:
  - Execution of external programs, e.g. (bash) scripts

- Controller plugins:
  - Control of execution plugins
  - Parallel execution
  - Synchronisation
MXv1 Characterisation v1.1

- MX sample characterisation taking into account radiation damage
- Indexing using MOSFLM or Labelit
- Parallel integration of reference images
- If flux + beamsize:
  - RADDOSE for estimating radiation damage
- BEST strategy calculation
  - taking into account radiation damage
  - multi-subwedge data collection strategies
EDNA / Passerelle Characterisation Workflow

Director

ref-testscale_1_001 → Header001 → MXv1 Indexing → MXv1 Integration → MXv1 Strategy Calculation

Open File → Open BEST log
MXv1 Characterisation v1.2

+ Xtal info
+ beam flux
+ diffraction plan

MOSFLM indexing

Indexing Evaluation
Ok

Failure

LABELIT indexing

Indexing Evaluation
Ok

MOSFLM Predictions

MOSFLM integration

[RADDOSE]

BEST

Data collection plan
Existing scientific EDNA workflows

- Macromolecular crystallography:
  - Characterisation taking into account radiation damage (MOSFLM, Labelit, RADDOSE, BEST)
  - Connection with experiment data base (ISPyB)
  - Parallel execution of characterisation (GRID data processing)
  - Parallel creation of image thumbnails

- Diffraction Computed Tomography
  - SPD: Image correction, fast azimuthal integration
  - Sinograms saved in HDF5 format

- Small Angle Scattering
  - Image correction and fast azimuthal integration

- Full Field XAS
  - Image correction (dark, flat)
  - Image alignment (offset measurements by FFT)
  - HDF5 output
How EDNA will evolve in the future

• Common EDNA developments (Kernel):
  • Improvements of the data model framework (in progress)
  • Improvements of logging (in progress)
  • Full support of Windows and MacOS (in progress)
  • Enhanced support of grid engines / job schedulers
  • Improved documentation (plugin use cases)
  • Graphical workflow editor (Data Analysis Workbench)

• Scientific developments:
  • MX further enhancements of characterisation (kappa, XDS etc)
  • MX auto processing wrappers
  • Biosaxs data analysis (EMBL Hamburg software suite)
  • Tomography
  • More to come...