EDNA Meeting June 02-03 2009
EMBL-Hamburg
DESY, Bulding 25b, Room 109

DRAFT MINUTES version 3, 16th July 2009

Pre-meeting event: Tuesday, 02.06.09, 11:00 - 12:00 Room 109/25b
Seminar by Olof: "EDNA: A tool for online data analysis and a collaborative effort"

Agenda

Tuesday, 02.06.09

12:00 -14:00 EMBL Coffee Room, Building 25a
Lunch and warm up

14:00 -15:00 Demo of MXv1 + description
Session leaders: Olof + Marie Francoise + Gleb
15:00 - 15:30 Coffee
15:30 - 18:00 MXv1 future priorities and scientific goals
Session leaders: Andrew + Sean

- Characterisation with different workflow (including symmetry determination)
- Processing of multi-wedge strategies: post refinement, integration and scaling
- Tool for radiation sensitivity measurement
- SAD data collection strategy and processing

18:00 - 18:30 Thomas
Guided tour to PetraIII

Wed. 03.06.09

09:00 - 11:00 MXv2 future priorities and scientific goals
Session leaders: Thomas + Gerard

- Re-implementing MXv1 with XDS and (more) generic data model
- Kappa/STAC integration
- Processing to be extended to use kappa strategies
- Inverse beam processing
- Calibration options

11:00 - 11:30 Coffee
11:30 - 13:00 ISPyB, DNA, AutoProcessing and non-MX applications
Session leader: Alun

- "Documentation" of results (graphs etc) - replacement of DNA
- Ranking. Demo of current ISPyB ranking - do we want more?
- Status of ISPyB and persistence layer for EDNA
- Summary of MX AutoProcessing and non-MX EDNA applications

13:00 - 14:00 EMBL Coffee room, Building 25a.
Lunch and wrap up.

Participants:

Alun Ashton	DLS	Alun.Ashton@diamond.ac.uk
Mark Basham,	DLS	Mark.Basham@diamond.ac.uk
Gleb Bourenkov	EMBL Hamburg	gleb@embl-hamburg.de
Gerard Bricogne	Global Phasing	gb10@globalphasing.com
Sandor Brockhauser	EMBL Grenoble	brockhauser@embl.fr
Michele Cianci	EMBL Hamburg	cianci@embl-hamburg.de
Gwyndaf Evans	DLS	gwyndaf@gwyndafevans.co.uk
Michael Hellmig	BESSY	michael.hellmig@bessy.de
Peter Keller	Global Phasing	pkeller@globalphasing.com
Gerrit Langer	EMBL Hamburg	gerrit@embl-hamburg.de
Andrew Leslie	MRC-LMB	andrew@mrc-lmb.cam.ac.uk
Karl Levik	DLS	karl.levik@diamond.ac.uk
Sean McSweeney	ESRF	seanmcs@esrf.fr
Uwe Müller	BESSY	umue@bessy.de
Ezequiel Panepucci	SLS	ezequiel.panepucci@psi.ch
Santosh.Panjikar	EMBL-Hamburg	Santosh.Panjikar@embl-hamburg.de
Venkataraman Parthasarathy	EMBL-Hamburg	pvenka@embl-hamburg.de
Romeu Andre' Pieritz	ESRF	pieritz@esrf.fr
Alexander Popov	ESRF	apopov@esrf.fr
Thomas R. Schneider	EMBL-Hamburg	thomas.schneider@embl-hamburg.de
Olof Svensson	ESRF	svensson@esrf.fr
Takashi Tomizaki	SLS	takashi.tomizaki@psi.ch
Johan Unge	MAXLAB	Johan.Unge@maxlab.lu.se

Apologies were received from Marie-Francoise Incardona

Summary of the Meeting

This summary was the last session of the meeting, but has been included here as a form of executive summary, as the full minutes are quite detailed.

Chair: Andrew

1. BEST functionality

- Multi-wedge strategy. Done and implemented in EDNA. Finished task
- SAD strategy. Already done in BEST, should be implemented in EDNA as a very high priority.
- Radiation sensitivity estimation tool. Will require development in BEST, timescale ~6 months. Requires modified work flow in EDNA. Work on EDNA component could start now, but low priority because of timescale of BEST work.
- Modifying strategy for a very small beam with a very large crystal (ie a fresh crystal volume for each image). Needs a proper use case. Simple case (eg 1 micron beam, >20micron xtal) quite rare. More usual case (10micron beam with 20-50 micron crystal) much more useful but much more complex. Needs to be done, but not a high priority for EDNA now.

2. EDNA tasks

Olof to coordinate and contribute as appropriate.

In order of priority:

1. html markup of results so that they can be stored in ISPyB and viewed by the user. The first target should be the overlay of the prediction on the diffraction image for the reference images. This should also be viewable from CCP4i interface. This requires coordination with ISPyB team.
   Identified manpower: Karl

   Extension to other detector types. This requires new modules for ImageReader (can be based on existing modules for ADSC and MarCCD images). This provides an excellent introduction to EDNA coding styles. The Pilatus detector is required with a high priority, and the Mar555 (Flat Panel) with lower priority. Because this is independent manpower, this should be carried out in parallel with html markup task.
   Identified manpower: Ezequiel Panepucci for Pilatus, Gleb for Mar555.

2. Changes to EDNA necessary to implement the BEST SAD function.

3. Pointless execution plugin. Based on one or several input MTZ files (eg two small

wedges).
Identified manpower: Sandor
Thomas offered to test feasability of calling this from MxCube (?)

4. Kappa/STAC plugin. This requires a modified workflow. Required reorientation is calculated by STAC after indexing of initial images (is BEST needed to calculate an initial strategy as well, so that the best choice is made of which crystallographic axis to align on rotation axis ?). Crystal is re-oriented and new reference images collected (at a different resolution if predicted by BEST ?), and these are used to calculate a new data collection strategy.
Identified manpower: Sandor

5. Testing.
   (i) Check robustness of indexing and error message handling using an annotated suite of reference images.
   Identified manpower: Andrew
   (ii) Scientific testing with appropriate samples, data collection using a strategy provided by EDNA. These results should be part of a publication on EDNA mxv1, planned for Acta Cryst D.
   Identified manpower: Alun for experiments at Diamond, Olof for experiments at ESRF. Coordination required to carry out the same style of experiment.

6. Scala and MOSFLM cell refinement plugins.

7. Persistence layer. Should be discussed further at future developers meetings. The radiation sensitivity tool, which requires handling multiple small datasets, could be a good test case for development of the persistence layer. Olof to set up a working group (Including Peter K., MarK ...) to consider this issue in advance of the next developers meeting.

3. Future Meetings

The next developers meeting should be scheduled for September. At this meeting the lack of documentation should be addressed. Ideally, Olof will prepare new documentation, including a skeleton plugin, for discussion at the meeting.

The next EDNA Full Meeting will take place early in 2010, site to be determined.

4. Congratulations

Sean congratulated Olof and the EDNA team on achieving the first EDNA release, and this was endorsed by all present. Olof highlighted the importance of the collaborative nature of the project.

Brief summary of sessions 2 (MXv2) and 3 (ISPyB, DNA, AutoProcessing and non-

MX applications)

This summary was prepared after the meeting.

There was an extensive discussion of the general data model that forms the backbone of MXv2, and is required both for more complex data collection strategies and to allow integration of alternative data processing packages such as XDS.

Thomas and Sean expressed some misgivings about whether such a general description was really required and what the likely timescale for the development would be. Peter, Gleb and Gerard explained the need for the general data model and the importance of getting feedback on the model from (beamline) scientists. It was agreed that some text description of the model would be necessary to ensure that it was readily understandable. Peter said that he hoped that it would be possible to have a suitable description of the general data model in the EDNA wiki before too long for comments. It emerged form the discussion that there was an urgent need for a capability to collect "single axis" data from kappa goniostats, and that this could be done as an extension to MXv1 rather than waiting for MXv2.

Action: Peter agreed to summarise the current data model as soon as possible so that others could provide comments.

A discussion of necessary calibration protocols for MXv2 followed, and the importance of selecting a suitable sample and defining an appropriate experimental protocol was highlighted.

Finally, the issue of divergent versions of MxCube at different synchrotron sites was raised. Sean agreed to encourage developers at ESRF to organise a workshop for MxCube developers in order to help to coordinate these developments.

Action: All Sites. Provide Sean with contact details for a nominated MxCube developer at each site.

Action: Sean to discuss organisation of a workshop with ESRF staff.

In the final session, Alun discussed alternative ways of displaying the results of EDNA, storing information in ISPyB, and the feasibility of replacing DNA with EDNA on MX beamlines, concluding that there were no serious challenges to doing so. He outlined the issues involved in developing a persistence layer for EDNA. He also described other applications for the EDNA framework being developed at Diamond, including DArc (a data archival system), tomography, automated diffraction data processing (could be done as a spike) and mentioned that there was some interest from CCP4 staff looking for a replacement to CCP4i.

Introduction

Andrew welcomed participants, and explained that a major objective of the meeting was to prioritise the EDNA tasks for the next six months, emphasising the importance of doing this because of the limited manpower available. A secondary objective was to encourage others to actively contribute to the EDNA project.

The meeting will initially discuss EDNA MXv1, which is restricted in scope to simple "single axis" experiments and the use of MOSFLM for integration and MOSFLM/LABELIT for indexing.

The following session will discuss MXv2, which has a more detailed generic data model which is needed to allow more complex data collection strategies (eg about several axes) and the incorporation of other processing packages (XDS, DENZO).

Demo of MXv1

Olof and Gleb gave a demonstration of the EDNA software, using a variety of diffraction images.

Session 1. EDNA MXv1 future priorities and scientific goals

Joint chairs: Sean and Andrew

MXv1 will be available from Friday 5th June, all participants were encouraged to install and use the release.

Manpower is a serious issue for EDNA, and threatens the future of the project unless new developers can be persuaded to contribute.

The resources issues were considered separately for developments required for BEST and for EDNA MXv1.

For BEST, three functionalities were considered:

- SAD strategy. Differs from conventional strategies in that a different target is used, to optimise the signal to noise for anomalous differences. This work is essentially complete, but requires changes to EDNA because in the standalone implementation it requires BEST to be run interactively, with a first run giving tables from which an optimal resolution is selected and then a second run works out the strategy for this resolution. An alternative would be to provide a default value for Ranom (eg 5%) and then do the rest automatically. It may require changes to the RADDOSE input to specify f" values as a function of energy.

  Gwyndaf suggested that one could also use images at phi and phi+180 to get an experimental estimate of the anomalous signal.

- Multiwedge strategies. This option is now fully functional and used in MXv1. Integration of the images is not yet possible, but this can be done in other ways (automated processing pipelines such as Xia2 and Autoproc).

- Radiation sensitivity tool. This is based on the idea of a "sacrificial crystal" which would be used to quantify the radiation damage limit for a type of crystal and used to devise the optimal strategy for another crystal of the same type. This would require significant development, with a time frame of approximately six months.

A general discussion about the desirability (and hence priority) of performing integration in MXv1 followed. Gerard made the point that a proper experimental record was required which would allow a seamless integration of the data. This was agreed, but thought to be part of MXv2. Andrew pointed out that the advantage of doing integration was that it would allow Pointless to be run to check the Laue/Space group symmetry. It is not clear yet how many images would be required to do this.

The issue of limited resources and their impact on the progress of EDNA arose again. However, Gerard pointed out that EDNA should not be judged by its limited content, but on the merits of the underlying framework. Sandor agreed, commenting that development with EDNA was much simpler than when working with DNA, as one could be confident that new features would not break existing ones. Olof observed that it was no longer necessary to discuss "how" to do something in EDNA, only what should be done next. Alun asked if the plugins for Cell Refinement, Pointless and Scala could be done by programmers with no real familiarity with MX. Olof believed this should be possible. Alun thought that if this was true, then some manpower could be found, but there was still a question about the level of documentation available to help new developers code plugins. Marc agreed, saying that after an unavoidable interuption to his initial efforts to write a plugin, he found it difficult to make progress because of a number of uncertainties.

Olof responded that he was aware that the documentation was incomplete. He would be motivated to spend time on additional documentation now that MXv1 has been released, however he pointed out that this would mean that he would spend less time developing new plugins. The goal would be to provide sufficiently good documentation in order to help new developers who commit themselves to providing additional plugins.

<u>Progress of Plugins</u>

Scala: minimal

Pointless: A use case has been prepared, but no code written

Integration: This has been written, as it is required for characterisation. There should be no difference between integrating a single image (as in the

characterisation) and a series of images (as in a (partial) dataset). Parallel processing of multiple blocks of images can be done. There is still a question of what output (other than the MTZ file) is required for presentation to the user. One feature not yet implemented is the "waiting for images" option, for the case where data collection takes longer than data processing.

Cell refinement: not done.

Integration into the beamline GUI

Sean believed that EDNA would have improved acceptance by incorporation into the beamline GUI, and as several sites are using, or plan to use, MxCube, there is a strong argument for supporting this. He also wanted to preserve the same (or a very similar) way of interacting with EDNA for users at different sites. There should be standardised templates for the workflow (eg the number of images used for characterisation).

Error handling and reporting

A concerted effort was run EDNA in a way that would make it fail in order to test the error handling. This proved quite difficult to do, but a failure was observed when using 3 blank images supplied by Alun, and the error message in this case was not appropriate (a failure in indexing rather than trapping the images as blank). Olof explained that full error propagation is a feature of EDNA, so that appropriate error messages should appear in the EDNA logfile and in another test this was indeed the case. However, there is room for improvement here.
Where possible, appropriate error messages should be produced by third party software (rather than by EDNA) as these are likely to be more informative.
Thomas was in favour of providing "hints" to help trouble shoot cases of failure.

Operation of Pointless within EDNA

The best way of handling cases where Pointless suggests (with a certain level of confidence) that the chosen Laue group is incorrect was discussed, in particular whether it was best to abort the current data collection and restart it, or complete the data collection and then work out the best strategy for collecting the remaining data. If the strategy has been devised based on the expected lifetime of the crystal and the true symmetry is lower than that selected, it would be necessary to abort data collection as soon as possible, otherwise it would not be possible to collect the additional data within the "lifetime" of the crystal. In cases where the chosen symmetry is lower than the true symmetry, the best course of action is not so clear.

This was not thought to be a serious issue, as it was believed that in 95% (or 99% according to one estimate) of cases the spacegroup is known before the crystal is characterised.

Collection of additional images for cell refinement

With DNA, if the data is to be integrated, additional images are collected at phistart+90^{o} prior to data collection proper, to allow cell refinement to be carried out with MOSFLM. Although it is not clear that the images will be integrated in the same way with EDNA, it was felt that the possibility of collecting these additional images should be retained in the workflow, but only carried out if required.

Additional Features required

It is important to extend the range of detectors that can be handled by EDNA in MXv1, in particular to deal with images from Pilatus detectors (minicbf) and the Mar flat panel detector (Mar555).

There was also a call for EDNA to be available on a broader range of platforms, in particular for Macs. This would allow potential users to test the software in their own labs prior to a synchrotron trip.
Note, however, that in subsequent discussion of the release policy for MXv1, it was decided only to inform synchrotron sites because there is insufficient manpower to support users installing the software in home labs.

Importance of inclusion of integration/scaling in MXv1

There was further discussion on the need to include integration (using refined cell parameters)/scaling in MXv1. The radiation sensitivity aspect of BEST would work better with them included but does not depend on it.
Sean felt that other issues should have a higher priority.
Olof estimated that approximately one months effort was required per EDNA plugin, although more time might be needed to optimise things.

This brought the first session to a close

??? Is anything important missing ... I have some gaps in my notes but can't remember what happened then !!

Session 2. EDNA MXv2 future priorities and scientific goals
Joint Chairs: Gerard and Thomas

Thomas intends to use a mini-kappa device on the PetraIII beamlines, but realises that this is unlikely to be taken up by users unless easy-to-use software is available. Thus the integration of a kappa capability in EDNA is a high priority for him.

Sandor described the use of the kappa software (STAC) within DNA, and very recent changes he had made to be able to use it with EDNA. The essential difference between DNA and EDNA is that there is no interaction possible with EDNA. However, it is possible to run STAC to give possible new orientations for data collection and a strategy

based on Raimond Ravelli's STRATEGY software. There is a need to know the current goniometer angles, these are not in the EDNA MXv1 data model. Also, at present, STAC relies on information from its calibration file. It would also be possible to work out the goniometer angles for data collection and then pass this to BEST to derive a full strategy.

Olof commented that the information required from the STAC calibration file could be stored in the EDNA configuration file.

Sean questioned whether a fully general data model was really necessary, particularly given the manpower restrictions.

Both Sandor and Peter felt that a general data model was indeed important. The current status of the data model was queried, and it was suggested that a proposal for the data model should be circulated as soon as possible, with a fixed time limit for feedback. Olof warned against imposing arbitrary time limits as this would encourage taking short cuts.

Sean requested that the data model be accompanied by text documentation so that it could be understood by non-specialists. Peter agreed that this would be valuable, as he felt it was important to get feedback from scientists on the data model.

Thomas emphasised that he had an immediate need for a simple single axis kappa experiment, which would not need the complexities of the full data model. Peter explained that while this would be OK for now, it would not meet the long term goals of EDNA, and the data model needed to be future proof. Also, that general descriptions are not necessarily more complex, and that work with the collision maps in BioXDM had demonstrated this.

Action: Peter agreed to summarise the current data model as soon as possible so that others could provide comments.

Gerard suggested that a restricted feature general model for simple kappa experiments (single rotation axis) would be useful, but that this would need, for example, a very careful treatment of the calibration.

Thomas suggested that the EDNA testing framework could be used to test the data model prior to its full implementation, to help speed up development.

Gleb responded to Sean's concern that the data model was too abstract by explaining that it should really just correspond to an engineer's view of the experimental setup.

There was general agreement that the workflow for kappa should be changed so that the strategy was calculated by BEST rather than a combination of STRATEGY for the geometric component and BEST for the exposure times.

Olof suggested that it would be worth including the kappa strategy (as a single axis

experiment) into MXv1, rather than having to wait for MXv2. Gerard agreed, saying that this would be a good step forward which would help the community and could help to justify additional resources in the future. As a single axis experiment, the subsequent processing of the data would also be straightforward.

When asked about the current status of the data model, Peter explained that so far there had only been two video conferences dedicated to the data model, and the description of the experimental setup (Goniostat and Detector) was nearing completion. When complete, this would be put in the EDNA Wiki for comments. Text comments on the model would also be given. As yet there is no model for the Strategy or results of existing data.

Representatives from Max Lab, BESSY and Gerard for SOLEIL, all of whom have kappa goniostats, expressed a strong interest in kappa software within EDNA. Thomas asked for a simple version in the first instance, with incremental improvements to follow. Users would need to be educated to make use of the kappa strategies, and it must be easy to use without worries about possible collisions etc.

The issue of where the interface to the user would appear was raised. MxCube or the Diamond GDA were possibilities.

For the single axis kappa, the issues of extending the processing and inverse beam processing (as itemised on the agenda) do not arise, and so these were not discussed.

Gerard then introduced the topic of calibration of the beamline, which is an important issue for the generalised data model. A simple experimental protocol needs to be devised to determine the appropriate parameters and an appropriate test crystal would be needed. Such a test would also exercise the data model and description of the geometry.

Sean suggested that the CORRECT step of XDS could provide the necessary information (when supplied with appropriate images).

The required frequency of the calibration was discussed, and the question of whether the experimental protocol had to be defined by EDNA (as some beamlines scientists might to define their own). Peter argued that EDNA should define the protocol because this must properly define the relevant parts of the data model. Discussion with beamline scientists will be needed to define this, to ensure that the experiments are practical and feasible. Sean also commented that the required accuracy of the parameters needed to be defined, to ensure that this was achievable. In addition, EDNA should be able to extract the parameters from the processing results (eg with an "edna --calibrate" command). For the mini-kappa, the calibration in STAC needs to be re-implemented in EDNA and stored in EDNA language (ie data model). It will also be necessary to define how this information can be supplied from an external calibration procedure).

Discussion of the MxCube user interface followed. The interface is in use at ESRF, PetraIII, Max Lab, SOLEIL and BESSY. There was some concern that local versions

were diverging, making future common developments more difficult. There was a need for coordination in development. Sean commented that there had been some discussion of organising an MxCube workshop, although nothing had been arranged yet. Sean requested that each site provide him with a name of a contact person who is responsible for MxCube activities, and he would provide these names to staff at ESRF in the hope that this would spur the organisation of a meeting. If sufficient outside interest is shown, this could help to persuade ESRF management to increase the priority for support of MxCube.

Development has started by Darren Spruce (ESRF) and Krister Larsson (MAX-LAB) for incorporating the EDNA characterisation with chemical composition input and multi-subwedge data collection into the mxCuBE GUI. A prototype version should be ready for tests in July 2009. Thomas Schneider suggested that some programming effort would be available from Hamburg to help with the incorporation of EDNA within MxCube.

Session 3: ISPyB, DNA, AutoProcessing and non-MX applications
Alun

Replacing DNA on the beamlines with EDNA
The only development in DNA since the start of the EDNA project has been to allow for remote operation. In terms of replacing DNA, as currently run on the beamlines, with EDNA the major issue is the absence of html summary that can be loaded into ISPyB, or any easy way of displaying results of individaul steps to the user.

In terms of displaying results of EDNA, two possibilities are:

- using ISPyB
- using the CCP4 Baubles utility (requires appropriate markup)

It is not necessary to duplicate the "Collect and integrate" option of DNA before it could be replaced by EDNA as this option is very rarely used.

The ranking can now be done within ISPyB (and in the latest implementation the selection criteria can be weighted). There are still the following limitations to ranking with ISPyB:

- Cannot create groups
- Cannot export results to the BCM
- Limited flexibility (only five criteria available for ranking)

In spite of these limitations, the ranking facility within ISPyB was seen as a very promising start.

Overall, there seems little reason why EDNA (v1.0) should not replace DNA on the beamlines.

Persistence layer

Because ISPyB is an archive rather than a running record, there is still a question over developing a suitable persistance layer for EDNA. "Pickle" (as used by PHENIX developers) is an option but this does not address concurrency issues.
This topic should be addressed at the next developers meeting.

Other Applications that use the EDNA framework

- Alun described the DArc (Data Archive and registration) server that has been developed at Diamond using the EDNA framework. This is a straightforward application that could serve as a good example for new developers learning how to use EDNA.
- EDNA was also being used to develop an application for tomography at DLS.
- There is interest in using EDNA to develop automated processing of MX data, running third party packages such as Xia2 and AUTOPROC. This could be done as a standalone spike.
- CCP4 have also expressed an interest in using EDNA in the context of a new CCP4i application. However, following an introduction to EDNA provided by DLS, they had a number of issues, most notably:
  a. Platform dependence (eg Windows)
  b. Aspects of the setup that could be automated
  c. Lack of documentation and examples
  d. Windows specific "costly" UML editor.

  None of these were thought to be blockers, and the cost of the UML editor (approx £100) was not seen as a difficulty.

Release notes for EDNA MXv1

A preliminary version of the release notes has already been circulated. Comments and suggestions have to be provided to Alun by 12:00 on 4th June.

Distribution of EDNA MXv1

Following discussion, it was decided to limit the distribution to MX groups at synchrotrons. While it was felt that there may well be interest from lab-based MX groups, the manpower was not available to support such a broad community.
Following a suggestion from Thomas, Alun agreed to try to find a suitable list of groups, possibly making use of the LightSources.Org website.
Olof reported that the installation documentation would be ready by 4th June.