Final Version, 14TH FEB 2005

Minutes - DNA meeting 25th January 2005, Daresbury Labs

Present: Andrew Leslie (MRC-LMB Cambridge), Sandor Brockhauser (EMBL Grenoble), Pierre Legrand (Soleil), Liz Duke (Diamond), Alun Ashton (diamond), Steve Kinder (DL), Mylrajan Muthusamy (DL), Karen Ackroyd (DL), Sue Bailey (DL), Gordon Leonard (ESRF, Grenoble), Sean McSweeney (ESRF, Grenoble), Harry Powell (MRC-LMB, Cambridge), Keith Wilson (CCP4, York), John Cowan (DL), Lorenzo Milazzo (Global Phasing), Gleb Bourenkov (Max-Plank, Hamburg), Sasha Popov (EMBL, Hamburg), Graeme Winter (DL), Colin Nave (DL), Gerard Bricogne (Global Phasing), Martyn Winn (CCP4, Daresbury), Charles Ballard (CCP4, Daresbury), Avi Naim (EBI), Oleg Dolmanov (EBI), Darren Spruce (ESRF)
Via Web Link: Olof Svenson

1. Current Status

General
Alun Ashton: the poster has been produced; version 1.0 has been released and is working. Feedback has been generally positive.

(i) A fixed release (1.0.1) is needed by February or March.
(ii) New features will go in version 1.1 - release date to be decided but not before Summer 2005.

Problems will be logged via Bugzilla (for logging and assignment of bugs)

CVS will be used to record changes.

The last developers' meeting was held in Grenoble (26 - 29th Sept 2004). The next will be held in Hamburg, with suggested dates of March 10-11th.

There was some discussion about offline testing of images. Harry reported that it hasn't really been pursued - perhaps try to look at more images. There was an extended discussion about whether the offline testing of this particular image set was worthwhile. However, it was decided that a limited set of test images was required to be able to test further releases of DNA, and this would need to include images that have, in the past, given problems (eg blank images, very icy images etc (vide infra).

ESRF
Sean: There have been 850-900 real tests at ESRF of which about 500 succesfully indexed the data, leading to 240 datasets. This equates to an average of 5 runs of DNA a day, suggesting there is some reluctance to use it. It is available on BM14, BM16, ID23, ID29 and the ID14 beamlines. There will be beamline test on 2 & 3rd March of a pipelining experiment (although originally scheduled for ID23, this will probably now be done on ID14-3 because it requires a sample changer.

Gordon: Most failures are due to beamline problems rather than DNA, e.g. incorrect beam position in the image header.

Each crystal takes ~2 minutes for initial characterisation and it then takes ~5 minutes to the start of the data collection. Although the total time (5mins) was longer than desirable, the consensus was that 2 minutes to characterise a crystal was satisfactory, at least for unattended operation

SRS
Steve Kinder: The installation seems well worth it. Version 1.0 is installed on beamlines 14.2, 7.2 and 9.6.

14.2 - good comments from users, ~50% have used it in December 04
7.2 - Mar345 IP used for test purposes only (not a scheduled beamline)
9.6 - too slow on the old Alpha workstation currently on the beamline.

10.1 has the Mar DTB and CCD detector. Steve is currently evaluating the effort required to install PXGen++ without losing significant features of the hardware, for example by writing a http server to sit between DNA and the Mar hardware.

Dual CPU systems have been bought to upgrade 9.6, 14.1 and 14.2, and will be installed in February. Steve felt that there was a need to increase the level of use. Reliability and faster machines will encourage this.

Hamburg (EMBL and MPG)
Sasha: DNA is installed for offline use on BW6 (MPG, variable wavelength) and X13 (EMBL fixed wavelength) - each has a Mar CCD. Computing hardware needs to be purchased for online operation. On BW6, a socket is available to allow communication between DNA and the Mar software controlling the spindle/detector. Work to achieve this communication is in progress and will take approximately another month.

Documentation is lacking but this does not seem to be a major problem - examples seem to suffice mostly.

2) Prioritising future developments & assigning tasks

Point group determination – POINTLESS (written by Phil Evans) could be thought of as ready for inclusion in testing. POINTLESS works well with small amounts of data, but will probably require more than the two segments of data collected for cell refinement (this needs to be tested). If POINTLESS suggests that the true Laue symmetry is lower than that initially assumed, this would be dealt with by completing the data collection that was underway and then calculating the strategy required to complete the dataset, taking account of the data already collected. With multi-axis goniometers, could design a strategy to specifically test the assumed symmetry. When collecting in batches, POINTLESS could be run after each batch.

Action: Graeme to liaise with Phil Evans on incorporation of POINTLESS.

Scaling/merging. This has been tested for a couple of months at DL, including on the development station 7.2. A twinning test based on the 2nd moment is available though it would be better if twin detection was recorded in truncate.

Action: Graeme/Alun. Requires GW's branch to be merged back in to the main branch. See also section headed “Merging CVS branches” in section 5 of these minutes.

In connection with integration/scaling/merging, it was noted that the GUI is currently “blocking”, so that if an integration/scaling job has been submitted it is not possible to ask DNA to collect further data until this job has finished. It would be preferable to have a mechanism by which such jobs could be run in the background leaving the GUI free to run other tasks. On completion of the background job, a flag would be raised in the GUI alerting the user so that the output could be inspected. In fully automated operation, the ES itself would have to check for such flags and take appropriate action (ie screen the output of scaling/merging and decide if further data collection from that sample is required). It is not clear at how present how best to do this.

Action: Olof (and others ?) to look into possible mechanisms.

Pierre Legrand raised the issue that it was currently difficult to integrate the use of XDS into DNA, even to perform the integration stage, because the current structure was “MOSFLM centric”. As the use of alternative processing programs was always an intended feature of DNA, this is not desirable, even if this is not currently a high priority.

Action: Graeme to discuss the issue with Pierre and determine how much work would be involved in using XDS for integration. In particular, to clarify if the successful integration of XDS relies on making changes to the XDS code to provide additional/alternative output for DNA.

Incorporation of radiation damage modelling. Sasha presented ideas on how to proceed - on a specific beamline it should be possible to find some empirical characterisation of the damage using an initial crystal of the same type as to be used for data collection. Preliminary work on this approach carried out at the ESRF looked very promising. In addition, a more theoretical estimate could be provided using RADDOSE, which needs information on both the sample & source. Essentially this is research and not yet ready for inclusion in DNA, though it will be implemented and tested in BEST. Further experiments are required before inclusion in a DNA release could be contemplated. It was suggested that a report on progress could be presented at the next DNA full meeting.

To provide the necessary information for RADDOSE, there needs to be a standard description of the beamline.

Action: Andy Thompson to send list of important parameters to Avi Naim (EBI) for inclusion in the EBI data model.

In the meantime, an interim "maximum exposure for each beamline" would be useful, which would be imposed as an upper limit to exposure times suggested by BEST. This would have to take account of attenuators in the beamline and would require experiments at the ESRF.

Action: Sean, Gordon, Sasha and Gleb to conduct further experiments and report back to next DNA full meeting. Darren to investigate passing information on attenuator settings to DNA.

Feedback from downstream processing (eg substructure solution)
e.g. updating Laue group in light of processing, radiation damage problems - action Gerard as part of BioXhit. may also interface with ccp4 automation project.
A mechanism needs to be found to provide this feedback to DNA, possibly via the Executive System.

Action: Olof to investigate ways of passing information back to the ES that would probably result in collection of additional data.

Sample ranking Probably initially based on I/sig(I), but including ice ring analysis. Olof has some plans on how to fit this module into DNA (see below). Ranking could be based on resolution at some defined I/sigI or on I/sigI at some defined resolution (e.g. if defined by resolution at edge of detector).

Action: Olof to develop plans for a ranking module, to be circulated to developers prior to implementation.

Interaction with IspyB database (Update from Darren).
Testing of a pipeline on ID14EH3 is due to take place on 2-3 March 05 with intermediate tests in between (e.g. 10 Feb). People involved are Ludovic Launer, Jens Meyer, Olof Svensson, Solange Delageniere, Vincente Ray Bakaikoa, Darren Spruce, and Bernard Lavault.

Gordon Leonard and Martin Walsh are driving the ISPyB development

Earlier tests in December used the pipeline
PDA -> PXWeb & ISpyB -> DNA -> ProDC -> SPEC -> sample changer -> microdiff

ISpyB will be delivered for the next delivery cycle – 14Feb but is not yet sufficiently complete to be used in the pipelining test.

All these changes are in the main branch.

The tests will be repeated at the SRS, first on station 7.2 (with dummy sample changer) and then on 14.2. This needs planning.
Action: Karen Ackroyd.

The database development is a crucial part of the planned ability of DNA to perform sample screening and ranking. It is therefore important that the links with the database are developed in a timely manner for use in the screening module.

Action: Darren and Olof to coordinate development of database links as necessary for pipelining tests and the ranking module.

MAD experiments. Olof to develop plans for a separate module to enable DNA to deal with MAD experiments. An outline plan to be distributed to developers for feedback prior to implementation.
Action: Olof

Kappa Goniostats. Sandor has written an improved strategy module that allows for multiple scans at different Kappa angles. This would need to be incorporated into DNA. He would like to have a partial DNA-dev meeting at the same time as the next Kappa developers meeting at ESRF in April or May.
Action: Sandor to discuss with Graeme

Kappa goniostats will be installed on ID23 and ID14-4 soon - Feb/March 05.
Keith Wilson pointed out that strategy for SAD/MAD (aligned) might be different to the strategy for normal data collection (no blind zone). It will be assumed initially that data will be collected using an omega scan.

Standard set of images Gerard pointed out that that test data and result of tests could be archived (to allow them to be repeated when software updated) in CVS i.e. CVS gives a record of validation tests.
Action: ???

3) Timescale & content of releases

Version 1.01: Bug-fix release, end of March 2005.
Action: Alun to coordinate

Version 1.1: Release not before Summer 2005. Include Scaling, merging and POINTLESS. Include hard time limit for radiation damage modelling and some beamline description file. Simple sample ranking (probably based only on I/sigI).with some data in database.

Version 2.0: End of 2005. Review realistic radiation damage model for possible inclusion. Include MAD module (which can depend on radiation damage) and improved ranking. Include initial kappa functionality.

Radiation damage modelling - report on conclusions before deciding when it should be included.

4) Standards for feeding back data from downstream calculations (Gerard)

Gave us some thoughts about "an expert system for smart data collection". These involved a "protocol descriptor" which must be amenable to -

(1) generation (i.e. the "imagination" of the expert)

(2) ranking by simulation ("judgement" " " ")
(3) execution ("executive power" which acts on the advice of the expert)
(4) revision (which keeps the expert on call)
(5) interpretation and data annotation (important for complex protocols).

DNA related issues -

* new capabilities - multi axis, multi wavelength, multi pass, etc. Will require extension of DNA in terms of (a) contents and (b) accessibility & communicability.

Topics for discussion -

* management of this evolution
* choice of management tools
* embedding in a data model (eg UML, CCPn) which would also provide tools for defining beamline configuration files and their querying.

Lorenzo Millazo then discussed:
* the CCPN data model
* UML from ObjectDomain
* CCPN meta model
* rules governing names, organization of data structures, inheritance, operations & methods.

see www.ccpn.ac.uk for more details.

This was followed by a discussion started by Keith re: PIMS and data models used by CCPN. Although there were some concerns, it was hoped that the various parties (DNA, CPP4, e-HTPX, PIMS, BIOXHIT, CCPN etc) would be able to work together using a common data model and, where appropriate, common tools.

5) Architecture of DNA (Graeme and Sean)

The scheduler (Graeme)

Talked about the scheduler and the extra functionality included in his branch which is being tested on station 7.2 at the SRS. This included scaling and merging of data as well as some diagnostics based on these stages. For example, the diagnostics for radiation damage (from the scaling of the data) were complementary to the prediction proposed within BEST and RADDOSE.

There was concern that the branch had diverged too far from the main trunk. Gerard suggested that in future the separate development branch could usefully be tagged to identify divergences from main branch. Alun agreed to work with Graeme on incorporating the desirable features of his branch in to the main trunk. Graeme agreed to provide information about the functionality of the added modules.

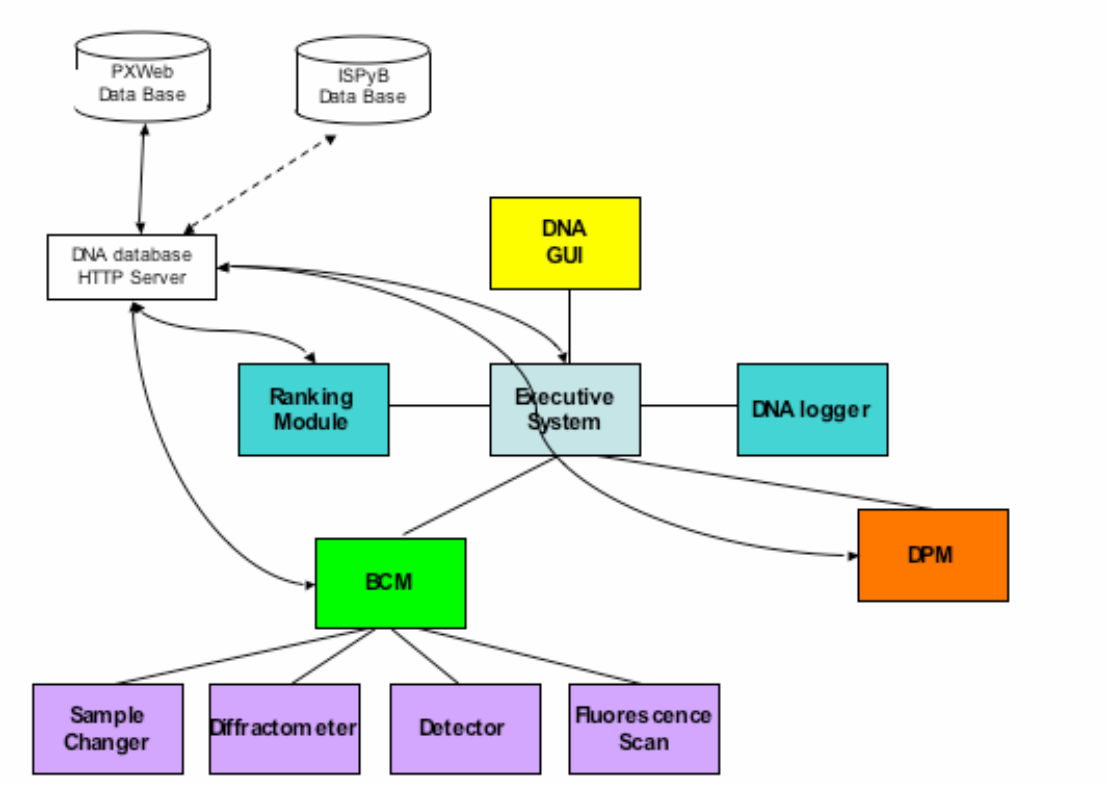
Graeme also mentioned XIA which was a possible basis for downstream automation associated with the CCP4 automation project. It was agreed that the present scheduler in DNA should not be rewritten at this stage to incorporate the ideas from XIA. If experience with CCP4 indicted that XIA gave advantages it could be incorporated in the future.

The Ranking and MAD modules (Sean)

The DNA Ranking Module - this will have several development steps;
(i) simple ranking based on I/sig(I)
(ii) more advanced involving several parameters
(iii) perhaps change the scheme for MAD, SAD, etc - could be set either by the GUI or the diffraction plan.

The DNA MAD Module - will implement;
(i) an edge scan
(ii) running CHOOCH
(iii) Selection of wavelengths
(iv) communication with beamline hardware.

The proposals for the position of the ranking module and MAD module were new to some developers. Sean said that Olof intended to provide more details for discussion among the developers.



Merging CVS branches

This caused considerable discussion following on from Graeme's contribution above.

It was decided that only those items in Graeme's branch that will be part of the version 1.1 release should be merged at present. While merging the whole branch would be "possible", Graeme suggested that only merging in the required parts "might be difficult".

Alun & Graeme would review GW's CVS commits to determine the current state of his branch.

The criteria for inclusion in the main branch were that the developments should be

a) Documented
b) Tested
c) Useful

The question of peer review of the code was raised and it was concluded that this would not be practical. However a peer review of the three criteria above should be carried out.

Darren suggested that the required features from GW could be recoded and implemented in the main branch, but it was felt that this was probably not the most efficient route.

6) DNA publication (Andrew)

Another publication is required to describe and advertise the recent developments for the DNA package as a whole. Those responsible for individual modules should write up a section describing them. Timescale 2-3 months. Individual contributions to Andrew by end of March. Aim for article in Acta D. The article will describe the collaborative bits but people could publish details of specific parts separately.

Action: Andrew to coordinate.

An abstract to the IUCr was suggested with the aim of having someone give a talk (post meeting note Olof Svensson to submit on behalf of DNA)

7) Beta-site testing of off-line version of DNA

It was decided to distribute DNA to selected non-SR sites and non European SR sites based on the likelihood of useful feedback. Suggested sites were Bob Sweet at NSLS, Structural Genomics Consortium (Oxford), LMB, York.

8) Any Other Business

Date of next full meeting: Three dates were suggested for the next full meeting: 7/8th July and 12th July. This meeting will take place at ESRF. Meeting closed at about 4:30pm.