Minutes from the DNA developers meeting
held at the ESRF 27th – 29th of September 2004
(draft version 2004-10-22)

Present:

Karen Ackroyd[1], Alun Ashton[2], Gleb Bourenkov[3], Gerard Bricogne[4], Sàndor Brockhauser[5], Steve Kinder[1], Pierre Legrand[6], Sean McSweeney[7], Lorenzo Milazzo[4], Alexander Popov[3], Bill Pulford[2], Harry Powell[8], Darren Spruce[7], Olof Svensson[7], Thorstein Thorsteinsson[4] and Graeme Winter[9]
Solange Delageniere[7] and Ludovic Launer[10] (database session only)

Meeting agenda:

Date	Time	Activity
Monday 27th	09:00 - 10:15	DNA-DEV session 1
	10:15 - 10:30	Coffee
	10:30 - 11:30	DNA-DEV session 1 continuation
	11:30 - 12:30	Seminar
	14:00 - late	ID14 EH4 beamtime
Tuesday 28th	09:00 - 10:30	DNA-DEV session 2
	10:30 - 11:00	Coffee
	11:00 - 12:30	DNA-DEV session 3
	12:30 - 14:00	Lunch
	14:00 - 15:30	Scientific DNA-DEV meeting
	15:30 - 16:00	Coffee
	16:00 - 17:00	Meeting summary
Wednesday 29th	08:00 - 16:00	Beamtime on ID29.

---

[1] CCLRC SRS Daresbury Laboratory, UK
[2] Diamond Light Source, Oxford, UK
[3] EMBL Hamburg, Germany
[4] Global Phasing, Cambridge, UK
[5] EMBL Grenoble, France
[6] Soleil, St Aubin, France
[7] ESRF, Grenoble, France
[8] LMB, MRC Cambridge, UK
[9] e-HTPX, Daresbury Laboratory, UK
[10] BM14, MRC France, Grenoble, France

1. DNA session 1

1.1 Overview of actions taken since the last DNA full meeting on July 1st:

1) Preparation of the off-line version of DNA to be made available to testers by Graeme: **Done**
2) Test images from ESRF to be renamed and made available to testers. Olof/Graeme, Mid July: **Done**
3) Off-line version of DNA to be tested with the test images. Harry, Katherine, Pierre, Gerard, plus someone from EMBL Hamburg: **Partly done, Harry and Pierre have started, however Gerard and Katherine will start soon**.
4) Timings for each step of crystal characterisation by DNA using the on-line version at ESRF: **Done**
5) Version 1.0 release of DNA for European synchrotron beamlines by December 1st: **In progress**
6) Documentation. Should be available for release on December 1st. Developers should provide "Developers" documentation for their own code: **Not done**
   Graeme to provide installation documentation: **Done**
   Users' documentation will be tested in (3): **Not done**
7) Executive committee to be set up. Andrew: **Done**

1.2 Feedback from test image processing

Pierre gave a brief description of the results currently in the data base. Harry also pointed out that it is important to use the DNA version of MOSFLM which was at the time of this meeting version 6.2.5 18th of August 2004. There was though a problem with this version because images indexed with standalone MOSFLM didn't give the same results as when indexed with DNA. This problem has been solved (MOSFLM version 6.2.5 version 27th of September became available during the meeting but only for DNA developers and at the time of writing these minutes the latest version is 6.2.5 1st of October).

One problem that was addressed was that the DNA system stopped when the screening resulted in "very icy" images. It was agreed that the DNA system should continue to try to index the images even if they were flagged as "very icy", since the ice region would be excluded. In these cases MOSFLM should handle images without any spots gracefully.

It was also felt that the bug reporting should be improved.

Many suggestions were made for improving the database: The MOSFLM output should be separated from the visual inspection, and it should be possible to search for certain features/results in the database.

Harry said that the data base of results from the test images will be available when there's useful information made available within four weeks.

1.3 Feedback from the beta release

Many suggestions where made in order to improve the DNA system, e.g. to rename the "Start Auto Indexing" button, clearer output etc. All these points are summarized in the Section 4.1 (List of actions).

We agreed that the installation script of release 1.0 should leave a blank space for the error reporting email address with instructions how to fill it in. No complaints where recorded concerning the installation scripts/procedure which was interpreted as no major changes have to be done before the 1.0 release. Graeme said that he would like to have more people involved in the actual release.

1.4 Features for DNA 1.0

This discussion rapidly became a discussion of whether any new features should be added to DNA version 1.0. We agreed that the best would be not to add any new features to the list presented in the July 1st meeting:

- Integrate and scale collected data
- Work with single wavelength data (MAD will be considered for DNA 2)
- Automatically determine the optimum exposure time (BEST)
- Run offline
- Run on any site (the distribution version will run however only on Linux computers)

1.5 Error reporting

It was agreed that the best way to report and track bugs is to use bugzilla. Alun urged to put all problems in bugzilla, even local problems like for example communication problems between DNA and ProDC which happened to be troublesome during summer 2004.

This discussion was cut short because of the seminar about automatic data collection at the SSRL which was organized in the same seminar room.
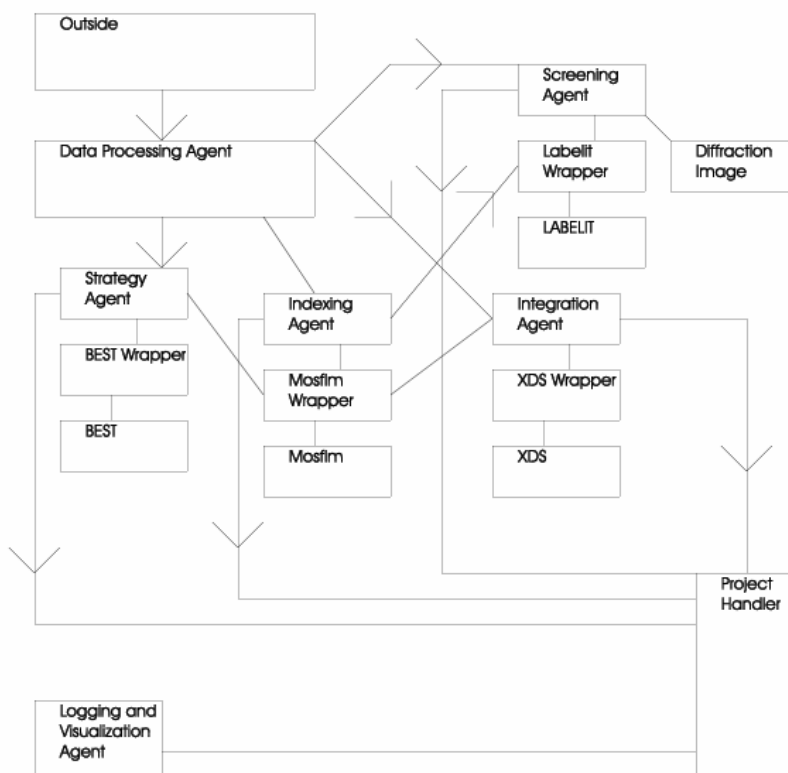
2. DNA session 2

2.1 XIA – Crystallographic Infrastructure for Automation

On the agenda this item was originally named "DPM rethinking". Graeme presented a new framework called XIA that would replace the current scheduler in the DNA project. His idea is that XIA is the scheduler for DNA 2.0. He presented several reasons arguing for a new framework:

- The current scheduler has grown a lot and is starting to be difficult to manage due to its (lack of) design
- With a new design it will be easier to add both new programs and programmers to the framework.
- The idea with XIA is also to be shared amongts many collaborations, notably the CCP4 automation project that can provide man power.

Graeme presented this draft of a possible XIA design:

The XIA was welcomed in the meeting as a very interesting project, however some concern was raised:

- Timescale: Graeme thinks that it will take him 9 months to get a first working version
- Sharing of goals: If more projects adopt XIA, how will we be sure that the DNA project goals are met?

- Impact on current development: Olof pointed out that the design and implementation of XIA shouldn't take resources away from the maintenance and development of DNA versions 1.0 and 1.x.

Graeme did say that the project is in the requirements phase - i.e. it is not a final idea yet.

The meeting did ask Graeme to prepare extensive documentation on the idea which would impact so heavily on DNA. The documents should then be circulated and developed further with DNA collaborators. The final document/idea will be presented to the DNA exec and will incorporate comments, feedback and any outstanding concern from other developers whose work are dependant on the scheduler.

2.2 Project manager / coordinator

Everybody agreed that the DNA project needs some kind of project management. A recurring problem is that till now in almost all dna-dev meetings ambitious task lists have been done but between the meetings very little has been done to follow up the work. Olof mentioned as an example the task list made at the dna-dev meeting in Daresbury, in February 2004. Everyone in the meeting agreed on a task list for implementing automatic screening by the DNA 1.0 release planned for June 2004. However, it turned out that in June almost none of the points on the task list was accomplished. One reason was the fact that the new version available in May 2004 did not work as well as anticipated, however a project manager could have had spotted this anomaly well ahead of June 2004 and tried to take action to remedy the situation.

We also agreed that the term project manager is not a good one since the person in question will not have the authority to take decisions. We therefore agreed that project coordinator is a better term. Graeme suggested himself to be the first rotating project, however Sean told him that he already had too many things to do for the DNA release 1.0. Sean suggested instead that Alun Ashton would be proposed to the Executive Committee for being the first DNA project coordinator. Olof pointed out that the task list of the project coordinator must be well defined and it was agreed that a proposition for a task list would be circulated on the dna-dev mailing list for being later submitted to the Executive Committee.

2.3 Database

In the BioXHIT meeting at the EBI on July 2nd it was agreed that two pipelining experiments would be performed, one in December using test crystals and one test in February using real crystals. However, it was not clear to which extent the DNA project committed itself to these tests. Gerard pointed out that for the February test there will be many eyes closely following it and we have to be very careful not to be pointed out as a source of failure for the test.

For the pipelining tests many developments are needed in the DNA system: login to a database, access the diffraction plan, automatic screening and writing back the screening results to the database. Some developments have already been performed, there are prototype database login panes in the GUI and the executive has a prototype database access function. Darren estimated the time needed for reaching the February test goals to be about 6 months which would be just enough. It was agreed that for the December experiment simply copying

and pasting information forth and back from the database interface and the DNA GUI would suffice. If the database access integration into DNA fails for the February experiment we agreed to simply import and export information to and from the DNA system using text files. Darren was nominated for the coordinator for the database integration.

2.4 XDS integration

Pierre presented the current status of his efforts to integrate XDS into DNA. He said that he had many difficulties since the scheduler is very MOSFLM oriented. Pierre has managed to write an interface to XDS that can be run in parallel to the MOSFLM interface. He said that it is for the moment not possible to replace MOSFLM with XDS.

It was agreed that the XDS integration will not be a DNA 1.0 feature but should be included in versions 1.x. Many questions need to be answered, for example how to make XDS strategy work with BEST and how to make it possible to choose between MOSFLM and XDS from the GUI / diffraction plan.

List of Tasks:
1. Header information needs to be made available to XDS (action on Graeme, Harry and Pierre)
2. XDS will be run in parallel to MOSFLM.
3. To run XDS a user will select a check button in the GUI.
4. A second results page will be needed.

ESRF staff offered to help testing the XDS module.

3. Scientific discussions

It was proposed in the last DNA full meeting to devote a part of the DNA developers' meeting to scientific discussions. Gordon Leonard, Elspeth Gordon and Gerard Bricogne were invited to this first dna-dev scientific discussion to give their opinions from a scientific point of view:

3.1 DNA Version 1.0 – what must be improved (Gordon Leonard)

- Reliability
  - Improved (perfect?) communication between 'expert system' and BCM. Lack of this is reason for most 'failures' of DNA.
  - Better STRATEGY option (does it search for 100% anom completeness?)
- Improved Speed
  - 18.35:59   Start
  - 18.36:37   Start of second Image
  - 18.37:18   Auto indexing starts
  - 18:37:52   Integration starts
  - 18:38:01   Strategy calculation starts
  - 18:38:10   BEST strategy starts
  - 18:38:21   Ready for data collection

- Better output
  - Some sort of summary that is easily accessible
  - Why did DNA 'fail' (beam-centre; too high r.m.s deviation, ice rings (Q: why should it not at least try to index icy images)
  - results of Integration, STRATEGY, TESTGEN (is default % overlaps too high?)
- More Flexibility
  - What happens if dmin less than than input? How should this be defined? In any case, should change XtoD accordingly, re-run strategy etc
- Crystal Ranking
  - Urgently needed. Maybe a proper summary would suffice for time being
- Proper Communication with LIMS
  - Entering results by hand will soon become impossible if SCs start to work as advertised!

DNA v1.x (x > 0) what I want!

- Properly refined cell dimensions
  - Should we take four images ($\Phi = 0,1; 90, 91$) and not two? Would take longer but time made up elsewhere could allow this. Ready for Integration as soon as real data collection is started.
- Automatic Integration & Scaling
  - See above. Need to decide how best to scale data (i.e at ESRF often best to use integrated intensities at low resolution).
  - Produce (if asked) two files: One merged mtz file & one 'unmerged polish' file
- Integration of Absorption edge scanning

- Collection of XANES scan and transformation of this to automatically determine wavelength choice (& collect data)

3.2 DNA – What the Industrial ladies want ! (Elspeth Gordon)

Since Elspeth's presentation contains several big images I've copied only the text here in order not to make the size of this document too big. The full presentation of Elspeth can be found here: http://www.dna.ac.uk/... (to be specified!)

Elspeth started by pointing out that DNA has become a very important tool in the "MXpress" service. The MXpress service collects industrial data for companies who only send their crystals together with instructions how to collect data from these crystals. The DNA system not only speeds up this data collection and removes tedious repetitive tasks, it also makes it possible to use less skilled persons to collect the data. Elspeth then described briefly the dewar form and the data collection/report in the PXWEB system.

This is how a typical MXpress data collection scheme looks like:
- Input crystal ID
- Collect 2 images
- Autoindex
- If images can be indexed and diffraction is to required resolution…
- Strategy
- Collect Data

There are only two steps not covered by the DNA system today, these steps are "Input of crystal ID" and "If images can be indexed and diffraction is to required resolution…".

This is what the MXpress staff want from DNA
- Reliability – communication with other processes
- Continuity – validation of each step and move rapidly to the next step
- Flexibility (– degree of desperation differs)
- Communication with our (PXWEB) database

Elspeth ended her presentation with the following wishlist:
- DNA to decide to collect data and do it.
- Database driven
- Results go directly into the database

3.3 Protocol descriptors (Gerard Bricogne)

Waiting for brief summary from Gerard…

4. Meeting summary

4.1 List of actions

Edited version by AWA 30/09/04

1. Fixed:

   1. Allow resizing GUI vertically – check if really needed (repeated below)
   2. Output displays 2 best data collection strategies -> Fixed
   3. MOSFLM mosaic spread estimation failed – Fixed!
   4. Different behavior between command line and GUI driven -> Fixed
   5. Output displays 2 best data collection strategies -> Fixed
   6. If mosaic spread estimation fails – index response should fail.
   7. Truncate to estimate number of residues based on 50% solvents
   8. "veryicy" status should not stop processing (bug 808)
   9. "Sticky" icy flag in the scheduler (bug 808)
   10. Error in BEST strategy, number formatting. To fixed …

2. DNA Version 1.0 – list of possible actions to be approved by the executive committee

   1. Number format exception with blank fields in GUI (Steve, Karen)
   2. BEST and BEST configuration file must be distributed with DNA (Graeme, Olof)
   3. BEST error was table not found data collection strategy. No stf file? Crashed. Attempted fix - inclusion of MOSFLM keyword "separation close". BEST problem?
   4. Unhappy return value from MOSFLM 256 Harry puts in extra bit of DNA output and supplies new MOSFLM, Graeme implement messaging. (Harry, Graeme)
   5. Ensure startup is clean – no old processes left behind. (Steve has a script that might help)
   6. Abort in DNA -> abort in BCM -> DNA 1.0 (High priority Olof, Darren)
   7. Documentation for the release – "minimal" developers documentation. User documentation configuration in install script. (Graeme)
   8. Remove Error report button, remove Sample identifier and database comment. (Steve, Karen).
   9. Change Auto Index button to Characterize and annotate (Steve, Karen)
   10. Check for $CCP4 in install script (Olof) (Note from AWA: if this is to build programs i.e. the need for CCP4 libs then $CCP4_LIB is also needed)

   11. Collect and integrate – integrate while collecting (try for 1.0 with easy solution. If fails move to 1.n) (Olof low priority)
   12. Reduce precision in outputs -> DNA 1.0 (Graeme low priority)

13. Allow resizing GUI vertically (low priority – already implemented? Steve, Karen)
14. Reduce number of error messages sending emails email goes out of the local configuration file (low priority Olof)
15. Create two MTZ files normal & unmerged Polish (low priority Graeme)
16. Integration, scaling & merging, distribution of "DNA" specific versions of SCALA, TRUNCATE etc into "third_party" (not top priority, it shouldn't stop 1.0 going out, Graeme)

3. DNA Version 1.n (a.k.a. 1.1 or 1.x :)) This list is just a reminder for post-DNA 1.0 features to be discussed after the release of 1.0.
   1. BCM development documentation – DNA developers @ Hamburg (after release)
   2. Strong diffraction at the end of the detector and spots well separated move detector closer -> DNA 1.X
   3. Some images seem to be marked icy by mistake
   4. Indexing failed image 1. No orientation matrix. Not integrating images – mosaic estimation failed. (Should have failed – multiple crystals)
   5. Result images even if indexing failed (needs more thought)
   6. Indexing failed. Spots getting "smudgy". Not getting max. likely cell @ edge.
   7. Index failed? Parameter 2x over limit? Cell edge to 11 Å. Use higher resolution to reduce fractional coordinate errors. User needs assistance? More information!
   8. MOSFLM XML to contain spot profiles -> DNA 1.1
   9. Dynamically allocate ports
   10. Mosflm log stored in system defaults file if used by multiple users permissions cause failure.
   11. Move mosflm logfile each time an auto index performed.
   12. Add ability to set mosaic spread value to GUI
   13. executive should not go on to strategy if too many bad spots (BEST)
   14. Enhance BCM simulator to simulate data collection by copying images.
   15. Improve error reporting and split into program failure and operational failure.
   16. Add MaxOsc keyword to strategy testgen.
   17. Enhance BCM simulator to simulate data collection by copying images.
   18. Configurable max and min exposure time

4. DNA Version 2.0 – as well only a reminder for future discussions.
   1. Automatic use of attenuators for reference images -> DNA 2.0?
   2. Calculated strategy -> too many overloads. How to automate detection & change resolution.
   3. MOSFLM XML to contain spot profiles -> DNA 1.1?

4.2 Next DNA developers' meeting

It was suggested to organize the next dna-dev meeting in Hamburg in February 2005.