DNA Full meeting 1st July 2004

DRAFT VERSION OF MINUTES, 8th JULY 2004

Location: MRC LMB, Cambridge

Present:

Gerard Bricogne (Global Phasing)
Sandor Brockhauser (EMBL Grenoble)
Liz Duke (diamond)
Phil Evans* (am only)
Joel Fillon (EBI) (am only)
Steve Kinder (Daresbury)
Kim Henrick (EBI) (pm only)
Ludovic Launer (MRC France)
Pierre leGrand (Soleil)
Gordon Leonard (ESRF)
Andrew Leslie (MRC LMB)
Katherine McAuley (diamond)
Lorenzo Milazzo (Global Phasing)
Colin Nave (Daresbury)
Harry Powell (MRC LMB)
Darren Spruce (ESRF)
Olof Svensson (ESRF)
Thorstein Thorsteinsson (Global Phasing)
Graeme Winter (Daresbury)
Martyn Winn (Daresbury)

*As onlooker.

Comments on Previous Minutes for developers meeting 24th Feb 2004

It was noted that Ludovic & Pierre were missing from the list of attendees for the DNA-DEV meeting on 24th February.

Project Overview
(Olof Svensson)

Summary:
     - status of project
     - recent developments
     - report from last developers meeting

Olof reported that in spite of remaining technical issues, DNA is used and appreciated, especially by industrial users at ESRF. It will run on any Linux that is less than about 2 years old.

Over the last six months, there have been a number of developments. In particular, the expertise of the system has been transferred to the Scheduler (but each component provides its own expertise on which the Scheduler relies). The system now gives a more complete summary of the indexing steps, and new expertise has been added to interpret the indexing results. The BEST program has also been incorporated so that suitable exposure times are now suggested … a major advance. The GUI has been modified so that the data collection strategy suggested by DNA is presented in one pane, and these parameters cannot be edited by the user. However, a separate pane, headed "Data Collection" will initially contain the same information and values here can be changed. A "resolution" parameter has been added. If the resolution is changed by the user, data collection cannot be started until a new data collection strategy has been calculated. This feature is not yet fully debugged.

Presently the image numbers are hardwired as 1 and 91 (for the two images used for indexing) - it was thought that this might be too stringent.

In response to Andrew's question regarding the strategy (now that BEST was available), Graeme replied that the start and end phi were used from Mosflm (as it gives a more conservative estimate), but the oscillation width and exposure time were determined by BEST.

It is not yet possible to specify the maximum time allowable for an experiment, but BEST does give information on the total elapsed time for the experiment (including detector readout, although this will require some site-specific information).

Andrew also asked if radiation damage was taken into account by BEST. In particular, did the exposure times suggested by BEST often result in radiation damage so that the final data quality was not as good as predicted. No real data was available to judge this. It was hoped to incorporate Raymond Ravelli's RADDOSE program into DNA at some stage. A BIOXHIT deliverable is to include radiation damage estimation within BEST.

Release of Version 1.0
(Graeme Winter)

Summary:
- why we haven't released 1.0 before the end of June deadline
- what do we mean by a release
- is the part 2 CCP4 license OK
- where DNA is going between now and 1.0, that is what needs to
  be done
- potential problems in release

Graeme reported that the release date had been missed, and suggested that the SHARP/CCP4 model could be followed, i.e. a release that contains components that do work reliably accompanied by some that may still be in development.

Harry pointed out that, during development, individual components were often ready at different times. However, in the delay while tidying up was being performed on those parts not ready, the "completed" parts were often "further improved", and so became "not ready" again. It was decided that it would be sensible to resurrect the notion of a "development" and "release" branch of CVS (which had been attempted earlier and abandoned) to try to avoid this problem in future. The "release" branch would have no new features added, only bugs fixed (it would be a dead branch, not to be remerged with the main trunk).

One of the main problems preventing the release was the issue of system stability. Use of DNA at ESRF had shown a significant failure rate (20-30%) using the latest version of DNA. In many cases DNA had simply stopped for no obvious reason. Although testing the offline system should help to reduce the number of failures, it was not clear what percentage of the failures were due to communication/timing issues which will only show up in online tests. **Improving system reliability of both offline and online versions is seen as crucial to a successful release and must be the major priority**.

Another issue that arose was one of communication between developers and users. Graeme was not aware until very recently that the failure rate was so high. It is important that problems that show up during online use of DNA are made known to all parties (eg through the dnadev bulletin board ?).

Following on from this, there was a discussion about exactly what is meant by a release.

This led to a long discussion about who should make the decision to freeze development and make the release; it was decided that a single person should be appointed to make this decision, but there was also a long discussion about who this person should be and whether or not they should be a developer. The decision was deferred until the afternoon (see Overall Management section).

It was emphasized that the creation of a release branch of CVS was primarily to allow the release date target to be met. Development work should still continue in parallel with testing of the release branch.

It was also decided to limit the release to European synchrotron sites (and specifically not for use with in-house sources). Synchrotron beamlines are seen as by far the most important target for use of DNA. This will greatly simplify the problem of user support, which could easily be unmanageable given the current level of resources if support was to be given to in-house users. Synchrotrons will be expected to supply their own Beamline Control Module. Specifications for this module can be obtained from Olof.

Martyn pointed out that the example of the CCP4 Part II licence in Harry's handout was not, in fact, the part II licence but the source code banner for a piece of code distributed under Part 0 of the CCP4 licence. He believes that Graeme wishes the DNA software to be distributed under the terms of the CCP4 Part II licence which is much more restrictive.

A number of features that were originally specified as being part of the 1.0 release were removed from the specification in order to be able to meet the new deadline. The features removed were:

- Be able to do screening and ranking automatically (hence it should work with sample changers)
- Be able to do space group determination

Broadly speaking these are items on which work is still to be finished. Those items in the original list which are now operational will be included in the 1.0 release. These features are now defined as:

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

For the current purposes, "any site" implicitly means any synchrotron site which is a partner in DNA.

The question was raised as to whether any features not in the current list could be included in the 1.0 release (eg the resolution parameter currently being worked on). After some discussion it was decided that a single person make the final decision on whether new features should be incorporated. (See Overall Management section).

Incorporation of Lims & Sample Changers

   - Sample Changers in DNA - Ludovic
   - DNA & LIMS - Darren/Solange
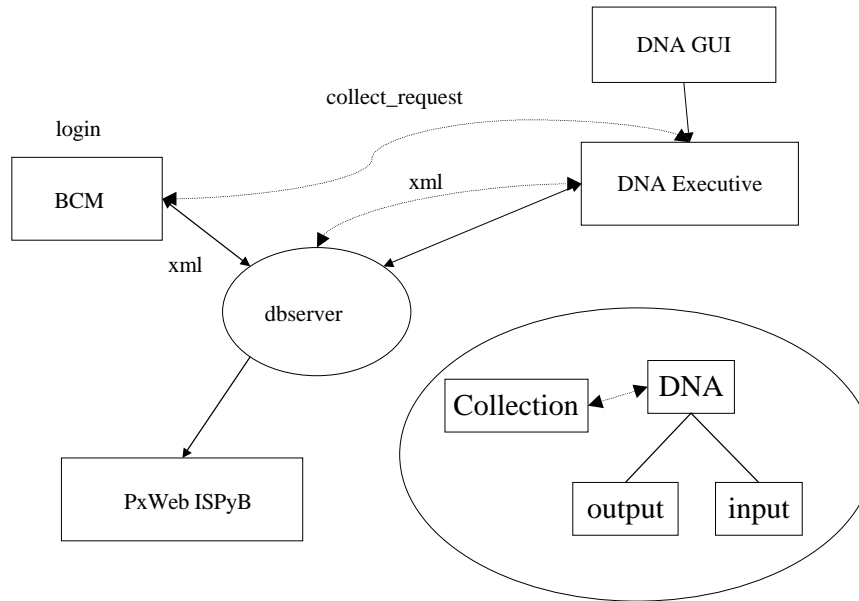   - Databases at the SRS -
     (Graeme speaking on behalf of Karen)

Ludovic reported that the user interface for the sample changer was okay. It can read sample codes on caps inside the dewar.

On EH3, 120 samples have been screened to date, and 46 data collections completed (although this was not done using the LIMS).

Eventually there will be 8 sample changers on PX beamlines at ESRF - the first will be delivered in the Autumn, 2004.

The LIMS is called ISpyB (pronounced I - Spy - B); details are available from www.bm14.ac.uk and www.e-htpx.ac.uk.

* Darren showed a diagram of the basic interconnections between components

DNA GUI

collect_request

login

BCM

xml

DNA Executive

xml

dbserver

Collection

DNA

output

input

PxWeb ISPyB

He will also link DNA to the standard ESRF login.

* Graeme reported that he hadn't spoken to Karen. However, PXWeb and the database are running at the SRS. The DNA gui talks to the database. The BCM connection doesn't yet go to the DB Server.

Database of Images

Harry Powell reported that Olof had accumulated a database of about 1000 pairs of images from the use of DNA on the ESRF beamlines (although this should strictly be referred to as a depository rather than a database). These images should provide a means of testing the robustness of DNA, and in view of the reliability issues raised by Graeme and Olof this was seen as a high priority. It was decided that the images should be split up and distributed to a variety of "testers" who would test the offline version of DNA when it becomes available (mid August). The following agreed to act as testers:

Harry Powell, Katherine McAuley, Gerard Bricogne, Pierre leGrand

The images need to be annotated, to describe features such as the strength of diffraction, presence of ice rings/spots, more than one lattice etc. Harry Powell agreed to draw up a "style sheet" which would be distributed to the other testers.

Olof agreed to rename the images according to a standard format and send them to testers in blocks of 25 pairs.

Testers agreed to use the current DNA documentation (with the GUI) to guide them in processing the images, as a test on the adequacy of the documentation. They will also attempt to index images with MOSFLM if they fail to index with DNA. Any instances where testers feel the indexing should have worked, but does not, should be reported to Harry for further analysis.

The EMBL Hamburg Outstation will also be approached to see if they wish to act as a "tester" in addition to those named above.

FedEx Crystallography

Gordon Leonard reported that Elspeth Gordon and Stephanie Monaco have made extensive use of DNA in the FedEx service that is heavily used by industrial groups. They use DNA to characterise crystals and collect data, but do not process it. They use DNA on all beamlines and, typically, 2 days beamtime a week is allocated (across all beamlines) for this work. While they find DNA very useful in this work, there is a significant reliability issue, with DNA regularly stopping for no apparent reason.

They are very helpful in providing feedback on DNA and suggesting improvements (Olof has a long list).

Olof is aware of the failures, but the reason for them is not clear at present. While an email is sent automatically if DNA crashes, there is no way of knowing when DNA has produced an answer that is incorrect.

There was discussion of providing a tool on the GUI so that users could report failures. It was felt that the most useful option would be to have a list of possible failures so the user could simply identify one (or more), plus a box for comments. Most users probably would not respond if they had to type in the error themselves.

Gordon also commented that many users had complained that DNA was too slow (eg 2-3 minutes for a characterise crystal) and that they could do the same operation faster by themselves. This is a serious criticism that needs to be investigated. At present it is not clear how much time is being taken by each step (ie collecting the images, passing them to DNA, indexing, strategy calculation etc). This information is crucial if the response time is to be improved and Gordon agreed to provide timings for each step in the process. Darren suggested that one partial explanation might be that each of the two orthogonal images was collected as a separate "run" by ProDC, and there is a significant overhead in setting up each run.

Overall Management of the Project

Andrew Leslie introduced this topic. Because the scope of the DNA project has expanded significantly in recent years, and because collaborators are now funded from a variety of different sources (BBSRC eHTPX; EC SPINE and BIOXHIT), it is quite difficult to provide an overall Management structure for the project. One way to tackle this problem would be to establish and executive committee to assist in making strategic and managerial decisions. Proposed membership of this committee would be Sean McSweeny (ESRF), Colin Nave (SRS Daresbury), Gerard Bricogne (Global phasing and representing BIOXHIT), Sasha Popov (EMBL, Hamburg) and Andrew Leslie (MRC LMB). The committee would then address the following issues:

1. Clarification of the input to DNA from the different initiatives in terms of both objectives (goals) **and manpower** or other resources devoted to achieving these goals, including names of Personnel (when known).

2. Resolving any duplication of effort or conflict that becomes apparent from (1).

3. Identify areas that require additional manpower/effort.

4. Ensuring that the project proceeds at an optimal rate and remains focused on the main DNA goals.

5. Ensure that all parties involved receive due credit for their contribution.

6. Decide whether a particular activity (development) should be followed, and at what priority level.

7. Where possible split DNA into a number of smaller tasks with, perhaps, one member of the executive committee being specifically associated with each task. Each task would have a "manager" (normally a developer), and the manager would be responsible for preparing a brief report for the executive committee at regular intervals (e.g. every three months). The report would serve to flag up any issues that might result in a significant delay or that have implications for progress in other tasks.

The main reason for setting up the committee is to improve communication between different aspects of the project, and to provide an "early warning mechanism" for potentially serious problems.

In subsequent discussion, Kim Henrick expressed concern that the exact relationship between DNA and EC funded projects would have to be very carefully worded in the report to the EC. In particular, he felt that describing work done under BIOXHIT (say), as being "collaborative" with DNA would cause difficulties. He was also unhappy with DNA being described as a "project", as he felt that this implies that it already has (independent) funding and, perhaps, describing it as a "grouping" rather than a "project" would be more appropriate. Colin Nave and Andrew Leslie felt that providing a piece of work could be clearly identified as having been carried out using (say) BIOXHIT funding, and that this was a stated objective in the original proposal, then the fact that their work was contributing to the DNA project should not cause any problems. Both BIOXHIT and e-HTPX included the collaboration with DNA in their applications for funding.

From the developers' side, Olof Svensson expressed concern that they felt they had been working "in a vacuum" in recent months, because there was little or no response to emails sent to the DNA bulletin board. Andrew Leslie suggested that one way to address this issue would be for one member of the executive committee to be responsible for ensuring that a reply was forthcoming. Different members of the committee could be responsible for different types of query, perhaps matching the smaller tasks envisaged in (7) above.

In relation to the release of DNA version 1.0, Andrew Leslie agreed to take on responsibility for making the final decision about which features would be included in the release and which should be held back in the development version. However, the more technical decision on whether the code itself is sufficiently stable for a release would be made jointly by Olof and Graeme, as it was felt that they have the best overall view of the system.

Although details of exactly how the committee will function remain unclear, it was agreed that it was worthwhile setting up such a committee and seeing if it would help to improve communication between developers and PIs and more generally across the whole project. Olof Svensson's suggestion that at least one member of the executive committee should also be present at the developers meetings, was also seen as an excellent way of improving communication. Typically, the committee member who is geographically closest to the site of the developers meeting would attend.

The existing assignment of individuals responsible for different modules (as detailed on DNA web site) was discussed and it was felt that no changes were required at present.

Tasks and deadlines

The following tasks and deadlines were assigned:

1) Preparation of off-line version of DNA to be made available for testers by Graeme (Mid August). This will be a beta release of version 1.0. The developers themselves should decide on the mechanics of whether and how to create a branch of CVS for the beta release.
2) Test images from ESRF to be renamed and made available to testers. Olof/Graeme, Mid July.
3) Off-line version of DNA to be tested with the test images. Images to be annotated using a style sheet drawn up by Harry. DNA User documentation should be used by testers. 25 pairs of images for each tester to be done by mid September. Harry, Katherine, Pierre, Gerard, plus someone from EMBL Hamburg ?
4) Timings for each step of crystal characterisation by DNA using online version at ESRF. Timings to be broken down as far as possible (it is possible to see when each image has been indexed/integrated etc ?). Gordon.
5) Version 1.0 release of DNA for European synchrotron beamlines by 1st December. (Note that each synchrotron is expected to provide their own beamline control module). Everyone !

6) Documentation. Should be available for release on 1st December, with separate documentation for Developers/Installers/Users. Developers should provide "Developers" documentation for their own code. Graeme to provide installation documentation. Users documentation will be tested in (3) above, necessary changes to be implemented by Olof.
7) Executive committee to be set up. Andrew

Any other business

Date of next meeting: First two weeks in December suggested for next full meeting. Liz Duke will look into possibility of holding it at the Rutherford site.

Andrew Leslie thanked all those who attended the meeting.