Minutes from the DNA developers meeting held at the ESRF 24th – 26th of September 2003

Present: Karen Ackroyd (DL), Steve Kinder (DL), Graeme Winter (DL), Darren Spruce(ESRF), Olof Svensson (ESRF), Ludovic Launer (MRC France, BM14), Pierre Legrand (EMBL) and Harry Powell (MRC).

This DNA developers meeting can be summarized as follows: discussion on Wednesday morning, continued discussion and coding session in the ID14 graphics room on Wednesday afternoon, beamtime on all Thursday on ID14 EH2, and continued beamtime till noon on Friday morning.

In the discussion meeting we agreed on the following agenda[1]:

1. Connection between DNA and a beamline LIMS – sample changer
2. Parallelisation – how to collect data at the same time as indexing and integration
3. New Expert System organization
4. Programs beyond MOSFLM (SCALA etc)
5. GUI requirements
6. DNA Distribution package
7. Integration of BEST and XDS in the DNA system
8. Andrew Leslie's questions
9. WEB pages improvements
10. Process collected data
11. Go through list of points from last DNA developers meeting (June 2003)
12. Next DNA full meeting

1. Connection between DNA and a beamline LIMS

Ludovic started by resuming the LIMS meeting he had organized on Friday the 19th of September. In this meeting Ludovic presented his plans for building a generic PX beamline LIMS system that will be used primarily on BM 14 and later on Diamond beamlines. The goal of the LIMS meeting was to receive feedback on the approach he chose. The participants found the meeting to be very fruitful and it was agreed to make a working plan for facilitating the interaction between PXWEB (existing ESRF beamline LIMS) and the future BM14 LIMS as well as to find areas where a common development could be envisaged (collection of information about the experiment, information passing via the LIMS system outside the ESRF etc.).

---

[1] In reality we treated the points in a slightly different order but I thought it would make sense for these minutes to reshuffle the real order

After this brief resume we discussed at what level the DNA system should come into contact with a beamline LIMS. We agreed on the following three interactions and the need of XML definitions for the communication between the DNA system and the LIMS:

1.1 Sample changer: The DNA system should get information from the LIMS about the contents of the samples in the sample changer. The XML definitions for this information transfer should be based on the definition developed by the e-htpx project. It should also send commands to the sample changer through the BCM. The corresponding XML definitions are already defined by Ludovic (in the CVS repository).

1.2 Data processing: The DNA system needs to write back figures of merit to the LIMS for each sample. Darren, Graeme and Olof should define the XML for this data transfer.

1.3 User identification: In order to prevent users from using by mistake another user's sample and to associate a crystal to a user, the user should be identified before starting to use DNA. Karen, Darren and Ludovic should look into the XML definition of the user's identification.

At the ESRF the LIMS system is PXWEB. Darren will make the necessary changes for incorporating the three interactions mentioned above with PXWEB. At DL no LIMS is yet implemented on the beamlines. Graeme suggested to make a beamline LIMS only for DNA. Deadline for the XML definitions and the DNA DL LIMS: 4 weeks.

2. Parallelisation – how to collect data at the same time as indexing and integration

Since the last DNA developers meeting the DNA system can now do post refinement of the cell and integrate collected data. This was successfully tested on the ID14 EH2 beamtime. Graeme has implemented a parallel processing for the data integration. Thanks to his developments the DNA system can considerably speed up the data integration.

However, the data collection and the integration are done sequentially, i.e. the DNA system waits for the accomplishment of the data collection finishes before starting to integrate the data. Without a sample changer, this was not penalizing – DNA finished always integrating the data before the next sample was mounted[2]. This will not be the case if a sample changer is used. We therefore discussed how to parallelise data collection and data processing. We agreed on the picture below (see Figure 1).

As can be seen from Figure 1, there are two distinct scenarios: crystal characterization and data collection. We felt that the best would be to implement a parallelization scheme in the data collection scenario because it would have a larger impact for a user on a beamline without sample changer and because it is easier to implement. We agreed on a 4 weeks

---

[2] Even if Darren tried hard to mount the crystal faster…

deadline for the data collection parallelization and a 12 weeks deadline for the crystal characterization parallelization.
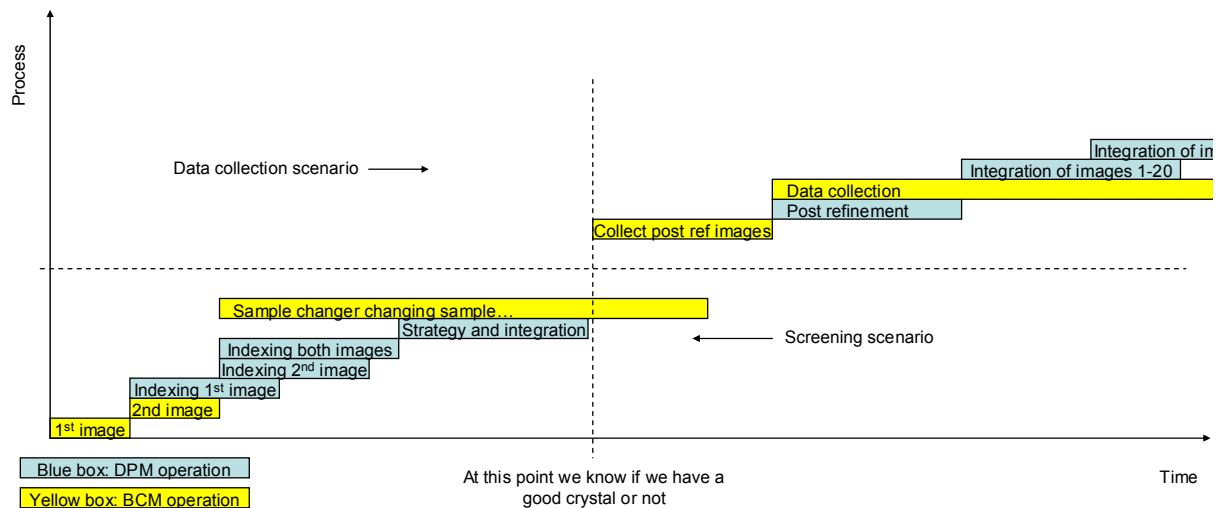

Figure 1

3. New Expert System organisation

Olof has participated in a meeting organised by Andrew Thompson concerning automation of PX beamlines using intelligent systems. Andrew's plan is to have fully automatic PX beamlines when Soleil will be opened to users in three years time. He has therefore started a collaboration with experts in artificial intelligence and automatic control from IRISIA at Rennes in France. The collaboration started 6 months ago and there have been three meetings so far.

Even though the main matter of the meeting was automatic beamline alignment, Andrew let Olof briefly present the DNA system structure. Olof asked the question whether expert system technologies are necessary since we're aiming for building an Expert System. The answer was rather surprising – the IRISIA people pointed out that what we call the "Expert System" is not at all where the expertise should be implemented. They argued that the expertise should be implemented in the different "agents" which have particular tasks to perform, i.e. the expertise for checking if an auto-index solution is good or not should be implemented in the DPM and the expertise for knowing if a resolution is reachable or not should be implemented in the BCM.

We agreed that this is indeed a sensible way of developing the DNA system. It will make development easier because what we previously called the "Expert System" will be

considerably easier to develop and maintain. In order to keep the acronym "ES" we decided to rename the "Expert System" to "Executive System". It will have a minimum of expertise but should execute the decisions taken by the agents – i.e. if the DPM says that the integration of data being collected is not good enough, the Executive System should tell the BCM to stop data collection.

The term "Expert System" could still be used but should then be related to the whole DNA system, see Figure 2 below:
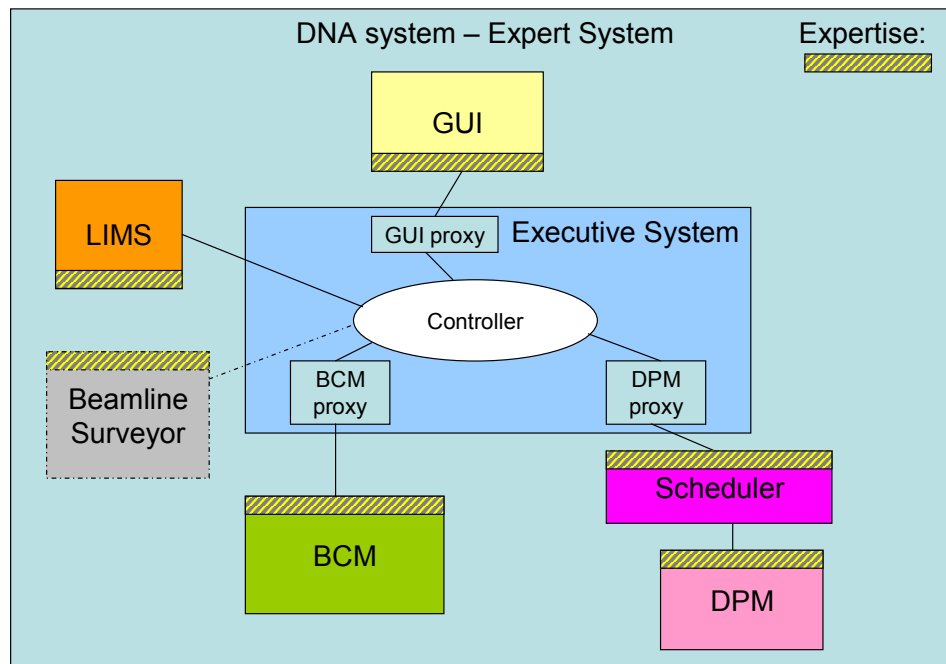

Figure 2

The major work to be performed is to take out the expertise already implemented in the ES and implement it in the scheduler. Olof and Graeme will do this, and this work will go hand in hand with the parallelization work. Time estimated to finish the ES reorganization: 12 weeks.

4. Programs beyond MOSFLM (SCALA etc)

As mentioned in Section 2, the DNA system can already integrate collected data. However, to be truly useful for (online) analysis of collected data, the DNA system must also run the CCP4 programs SORTMTZ, SCALA and TRUNCATE on the integrated data and extract relevant information to be presented to the users. The DNA system should also be able to determine the correct space group for the crystal since only the Laue group is known after the auto indexing. This work will be performed mainly by Graeme who estimates that in 12 weeks we will have a first version of the DNA system that runs SCALA and in about 6 months time all programs (in parallel whenever possible).

5. GUI requirements

We discussed which developments are needed for improving the GUI and we came up with the following list:

- HTML pane – this was tested briefly in the last DNA dev meeting at DL but is not yet implemented. The idea is that the ES provides the GUI with an URL (either pointing to a file or to a web server) whenever there's new information to be displayed. The GUI will then load the corresponding HTML page. Karen, Steve and Olof will work on this; estimated time 12 weeks.
- Pierre Legrand pointed out that ADXV distribute free software that can be used for visualization of diffraction images and can also overlay the images with predicted spots from the DPM. Pierre argued that the advantage of ADXV over an HTML page is that one can view the image interactively, zooming in etc. Olof said that this is a valid point only if the user is in front of the computer when the DNA system is collecting data, and that the interactivity is not a priority for an automatic system. We decided not to take any action yet on this point but to make an inquiry with the DNA community first.
- It was generally agreed that a progress bar on the GUI with an estimate of the remaining collection time would be appreciated.
- As mentioned in Section 1, a proposal number field is needed in the Reference Image Panel.
- We also agreed that we need an "anomalous" checkbox in the Reference Image Panel.
- We discussed the need for an user to choose between several strategy options – MAD etc. However, we agreed that for the time being we will not implement this choice.
- Pierre pointed out that for a MAD experiment the DNA system must also do edge scans and hence an edge scan panel is needed in the GUI. We agreed however that a MAD capable DNA system is too far in the future (> 6 months) so it's too early to start any development.
- Olof said that some users (e.g. Sean) complained that they couldn't copy and paste from the GUI to other X windows on ID14. This problem was solved by upgrading the Java version from 1.3.1 to 1.4.1.

6. DNA Distribution package

Graeme pointed out that the DNA system is now working well and is sufficiently stable for the creation of a distribution package. The content of this package should be:

- A manual for getting the DNA system running
- All the programs needed to run the DNA system (including compiled Java classes).
- Test images for testing the installation.

Graeme thought a first version of the distribution package could be ready within 4 weeks.

7. Integration of BEST and XDS in the DNA system

Harry reported that the latest version of MOSFLM now writes input files for BEST. If the command "best on" is given, MOSFLM produces a file called "best.hkl" that should be working with BEST.

Pierre reported that he has almost finalized the XDS Python wrapper and that he can start to work with Graeme in order to incorporate it into the scheduler.

8. Andrew Leslies questions

Harry came with a list of questions from Andrew. I include the questions and the answers here because they might be of interest to many people:

8.1. Which scoring system do we use in the DNA system today?
We have not yet implemented a scoring system – this will probably be a major discussion point in the next full meeting (see Section 12). Harry pointed out that we need a database of images from similar crystals, he argued that the crystals should be mounted by the same person who should be an experienced crystallographer in order to make differences in diffraction images rely only on the crystal quality, not on the way the crystals were mounted.

8.2. Are we treating multiple projects within the same sample changer puck?
No! We decided that we start with screening of samples in a sample changer belonging to the same project. A sample changer puck with samples from different projects will have to wait before being automatically screened by the DNA system.

8.3. Do we have a data base of test images?
Not yet – Olof mentioned that Sean had already started to create a set of reference images as a screening test. We did collect some screening images during the beamtime however we will need more. We decided that we should put together a web page describing what we have as test images and collected data sets and how to get hold of images and data sets (see Section 10).

8.4. How do we integrate images?
As mentioned in Section 2, the DNA system can now integrate images, however more work is needed (see Sections 2 and 4).

8.5. Criteria for rejecting images?
The criteria for rejecting images is at the moment only based on thresholds for the RMS spot deviation, fraction of rejected spots in the cell refinement and the move of the beam centre. This needs definitively to be improved in order to do a robust

screening. For example, the scoring needs to take into account cell edges, distance, wavelength, spot shape after integration etc.

8.6. Next DNA meeting?
See Section 12.

9. WEB pages improvements

The DNA web site needs to be improved. We agreed on the following distribution of tasks for providing web pages:

- A "run through" manual of how to use the DNA system (Karen and Steve), deadline 4 weeks.
- Installation manual (Graeme, see Section 6)
- Overall DNA system description (Olof)
- DNA connection to LIMS (Darren)

A first version of these web pages should be in place within 4 weeks. A more comprehensive DNA system design document should be finished within 12 weeks (linked to Sections 2 and 3).

10. Processing of collected data

We used the DNA system on ID14 EH2 to collect reference images and data sets from several crystals, mainly lysozyme but also ribonuclease (provided by Raimond Ravelli) and NimA (provided by Hanna Kristin-Leiros). The NimA data set was a challenge for the DNA system because of the small crystal size. After an initial indexing failure we pulled back the detector and the DNA system managed to index, run strategy, collect and integrate the data. However, the images and results cannot be distributed because the NimA crystal is a part of a current project.

The lysozyme crystal data sets will be used as a simulation of the crystal screening. The idea is to try to relate screening parameters with the actual quality of the integrated data. Graeme will process the lysozyme data sets and Olof will process the data sets of the ribonuclease crystals together with Raimond Ravelli. We agreed that the best would be to do this as soon as possible, thus we agreed on a 2 weeks deadline.

11. List of points from last DNA developers meeting

Here's the status of the job list we agreed upon in the last DNA meeting. The text in blue is taken from http://www.dna.ac.uk/minutes/110603/notes.html.

11.1. KA: Avoid data being accidentally overwritten. Auto increment runnumber at start of each data collection. Done but not checked in. Test during the beamtime revealed that this solution was not appropriate as the run number was automatically increased before the data collection had started.

11.2. OS, KA, DS: Where resolution is not achievable no check is made. Requested resolution to be checked by expert system and user warned. Min & max resolution to be calculated from BCM parameters etc. Darren will implement this in ProDC (already implemented in PXGEN++).

11.3. OS, SK, KA: BCM parameters request must include synchronous/asynchronous and this must be dealt with in GUI. Done.

11.4. GW: MOSFLM filenames are incorrect when new filename is shorter. Must reset filename string. Done.

11.5. GW: MOSFLM filenames are incorrect when new filename is shorter. Must reset filename string. Not done – 4 weeks deadline.

11.6. GW: When process is aborted MOSFLM doesn't recover. Reset MOSFLM after aborts. Probably done…

11.7. GW, OS: Index image 1 while collecting image 2. Requires parallelizing of MOSFLM - non-urgent. See Section 2.

11.8. GW, HP: Use and keep spot file to avoid re-indexing in strategy calculation. Not done – 4 weeks deadline

11.9. GW, HP: Fixed format file writing from MOSFLM to enable output to be used by scheduler. Done

11.10. KA: Pxgen++ data collection removes "_" from file templates. Correct filename parsing. Is it done? Tests needed…

11.11. SK: Search facility in message output windows. Not done – 4 weeks deadline.

11.12. GW: Reflections rejected reported by scheduler using wrong parameter. Done.

11.13. KA: Check requested exposure time for DNA collections is achievable. Done.

11.14. KA: Beam centre input in pxgen++ data collection pane not saved from DNA. Done.

11.15. HP: MOSFLM should check for minimum number of reflections and deal with this. Graeme said he's working on a Python module for pre-screening images

before running MOSFLM – it should be ready within 12 weeks.

11.16. GW: Trap any MOSFLM crashes and deal with them appropriately. Work under way.

11.17. OS: Expert to make quality assessments. Done -> DPM, see Section 3.

11.18. GW: Mosflm doesn't get space group information from auto index pane. Pass this information to Mosflm. Done.

11.19. HP: Mosflm is not reliably picking the best solution - may be picked up later when collecting full dataset. Try to improve initial space group selection. Work under way.

11.20. OS: Separate configuration editor for java.properties, xml files, config files and dna.setup for setting environment variables 4 weeks deadline.

11.21. OS: Expert to create new directory of form <directory>/<prefix>_<runnumber>_<dnaFiles> and tells GUI it's available. Where names are re-used keep old data by renaming directories using old<n>. Done.

11.22. SK, KA: Display information from 21. We were not sure why this point was added to the list…

11.23. OS: email for DNA use. Done.

11.24. OS: Implement DNA log server. Not done – LIMS 4 weeks, see Section 1.

11.25. OS, GW: Implement integration of images in batches of 10. Store batch summary files in sub directories of 21 with /batch_<n-n+10>/scaling /integration. Done but needs improvements.

11.26. OS: Implement collection of post refinement images. Index image at 0 degrees while collecting images at 91 & 92 degrees. Index images at 91 & 92 degrees while collecting images at 1 & 2 degrees. Post refinement image collection implemented, see Section 2 for parallelisation.

11.27. GW: Update DNA web pages. Not done – see Section 9.

11.28. GW: Create DNA manual. Not done – see Sections 6 and 9.

12. Next DNA full meeting

The date of the next full DNA meeting has been fixed to the 4th of November at the ESRF.