Final Minutes of the DNA Project Agreement Meeting at ESRF 23rd October 2007

Location ESRF Central Building room 501

Participants

Olof Svensson	ESRF
Peter Keller	Global Phasing
Karl Levik	Diamond
Jose Gabadinho	ESRF
Gerard Bricogne	Global Phasing
Clemens Vonrhein	Global Phasing
Thomas Schneider	EMBL Hamburg
Sandor Brockhauser	EMBL Grenoble
Gleb Bourenkov	EMBL Hamburg
Sasha Popov	ESRF
Darren Spruce	ESRF
Peter Turner	Univ of Sydney
Sean McSweeney	ESRF
Elspeth Gordon	ESRF
Romain Quilici	Univ of Sydney
Katherine McAuley	Diamond
Alun Ashton	Diamond
Andrew Leslie	MRC LMB Cambridge
Romeu Pierritz	ESRF
Marie-Francoise Incardona	ESRF
Bob Sweet	Brookhaven National Light Source
Clemens Schulze-Briese	SLS

Agenda

Session 1: History + current status

 9:00 - 9:15   Introduction (Alun)

 9:15 – 9:35   The first spike (Olof)

 9:35 – 10:15   Recent developments (Marie-Françoise, Romeu)

10:15 – 10:30   Coffee

Session 2: Open discussion around the project agreement

10:30 – 10:45   New name (Andrew)

10:45 - 12:45   Presentation of the project agreement (Alun)

12:45 – 14:00   Lunch in the ESRF restaurant

14:00 – 15:30   Ad-hoc and invited contributions / opinions (chair Alun)

15:30 – 16:00   Coffee

Session 3: Planning

16:00 – 16:15   Planning of activities (Olof) :
                  AALib consodilation / tests
                  Use Cases
                  3 months planning for developing
                  a pre-screening prototype
                  9 months planning of Use Cases and other prototypes
                  Project management tools
                  Next meeting

16:15 - 18:00   Open discussion - resources (chair Alun)

1. Introduction

Alun Ashton welcomed all participants, especially those who were not formally part of the DNA project, and emphasised the importance of the meeting, whose primary objective was to come to an agreed form of working for the new project. A draft Project Agreement had been circulated and it was originally hoped that future participants would be able to sign the document at this meeting. Concerns raised by EMBLEM made this unlikely, but a detailed discussion of the document would still be very valuable.

2. Report on the spike

Olof Svensson gave a report on the spike, explaining that the goal was to produce a prototype for the strategy calculation component of DNA and to use this development to test different frameworks (specifically CCPN and AAlib) and different development tools. He summarised the most important conclusions of the spike as follows:
- A functional prototype that calculated a strategy had been produced.
- A properly defined set of Use Cases (which has not been defined for the spike) would be essential for the development of a data model for DNA 2.0.
- Aalib was mature for future developments.
- CCPN in its current implementation is not appropriate for the development.
- Eclipse and Marratech worked well as tool, but Trak hosted at Diamond was less successful.

At the conclusion of the spike a meeting was held at ESRF (June 5-7th) for which separate minutes are available. A written report on the spike was presented by Olof Svensson. The main outcomes of this meeting were:

- The recognised need for a Project Agreement and Licence
- A decision to start development of initial Use Cases.
- A decision to start examination of methods for developing a data model and producing its implementation
- The need for a meeting in October to mark the official start of the new development and get agreement on methods of working.

There was a discussion on the need for use cases in the spike. In response to comments by Gerard questioning wether this was clearly a requirement, Olof explained that the need was not so obvious to the scientists, reflecting a difference between the developers and the scientists involved. In response to a question from Sean, Olof identified the need for a Project Agreement as the highest priority.

It was agreed that the Spike Report could now be made public.

Action: Olof to make the spike report publicly available initially using the DNA web site till a new project web site has been set up.

Olof identified two Use Cases that were to be tackled first:
- Characterisation taking into account radiation damage
- Data collection taking into account radiation damage.

These were believed to be good targets for a prototype DNA 2.0 as they would be very difficult to achieve within the framework of DNA 1.x.

Several sub Use-cases were identified:
- Prescreening
- Indexing with MOSFLM
- Indexing with MOSFLM, XDS and LABELIT.

Of these, the pre-screening Use Case was the most advanced.

General discussion concluded this section.
Bob Sweet requested clarification of the connection mechanism so it was clear what needs to be done. Sean asked for an estimate of the timescale of the development. Olof felt that with the current resources, it would no longer be possible to achieve this prototype within the timescale of the BIOXHIT funding. Clemens Vonrhein felt that the different web sites for different activities within DNA made it difficult to get an overall picture. Olof explained that this was because the new developments were not public yet and the Use Case development was hidden under a different project name. This would change when the name for the new development was chosen, when this would become publicly accessible (for reading, not for editing).

3. Recent developments

Report on Aalib. Romeu Pierritz gave an introduction to Aalib a library that allows multiple threads (using code written Python) with complete thread safety (http://pyaalib.sourceforge.net). The code is open source and has been used in several applications (ESRF automated beamline alignment, the benchmark and the ranking module in DNA-1).

Since the spike meeting in June, the library has been re-written in Jython, as Java is more portable and its use is more widespread than Python. This should improve the portability of the software.

Thomas asked if this portability was necessary, as it would be simpler to specify that the software had to be run on a Linux box with a particular operating system. However in general discussion it was agreed that this was not a feasible alternative, as beamline scientists did not have the authority to specify beamline computers at some sites.

Some comments were made about the level of documentation available, in particular clear guidelines for usage still need to be provided so that all developers can easily use the package.

There was some concern about the level of risk associated with the dependency on Aalib. As a result, an additional session was scheduled for the afternoon for further discussion.

Marie-Francois reported on the evolution of the Aalib test framework. She listed the requirements of test driven development:
- Quality
- Modularity/flexibility
- Code simplicity and focussing (make unit test pass)

- Code comprehension (should be easy to understand)
- Focus

and described some test extensions have been recently introduced:
ALAssert new class
ALTestCase and ALTestSuite

Gerard questioned the meaning of "Flexibility" in this context. Marie-Françoise clarified that with flexibility she means the modularity of the design which makes changes easy and the test framework which ensures that the software works after the changes.

Discussion continued on the future maintenance implications of using Aalib in case for example if Python or Java APIs change in future versions. It became clear that Aalib would need to be updated to reflect changes both in Python and in Java, but all the software above this (plugins) would be protected from these changes because they would see the same interface in Aalib.

4. The new name for the development

It was felt by several developers that DNA 2.0 was not an appropriate name for the project because the framework will be completely new. Andrew reported that there had been extensive discussion of a new name by Email, and gave a brief summary of the recent Email poll on possible names. However he noted that there had not been universal acceptance of any of the names proposed, and that there was a significant belief that the new name should in some way relate to DNA as this name was now quite well known. The name EDNA (Evolving DNA) had been well received at ESRF, although it had been excluded from the poll as being too close to DNA. It was therefore proposed that the name of the new project should be EDNA. During discussion it was pointed out that there is already a sourceforge project with this name, but in spite of this the name was selected.

It was pointed out that EDNA would refer to the name of the project, and it was still possible for an alternative name to be found for the actual software if there was a strong case for doing so.

Action: The new project is now to be referred to as EDNA. Olof has set up the new edna mailing lists: edna@esrf.fr, edna-dev@esrf.fr, edna-members@esrf.fr,edna-support@esrf.fr , edna-sc@esrf.fr and edna-executive@esrf.fr.

5. The Project Agreement

Alun led the discussion on the Project Agreement. There was an initial discussion on whether there was a need to have the document signed, because some of the changes proposed by EMBLEM (which would be necessary before they would agree to sign) were felt to be contrary to the spirit of the agreeement. The EMBL involvement at this stage comes through the inclusion of BEST. The licensing of BEST is now complicated by the fact that Sasha is now working at ESRF (and Gleb was working for MaxPlanck until recently). It was felt that the situation would be clarified if BEST (and other external software such as MOSFLM) were not considered part of EDNA and in particular were not distributed with EDNA but would have to be obtained independently by sites wanting to use EDNA. However, even if this is done there is a need for a commitment by the developers of these packages to make nay changes required for the development of EDNA.

The need to sign the document appears to depend on whether it is considered as a "legal" document. According to advice from EMBLEM, if the document refers to any licensing terms then it automatically becomes a legal document and requires signatures.
While there was general agreement amongst developers that the EDNA core should have an LGPL license, it was pointed out by Sean that the developers themselves could not decide this themselves, the decision has to be taken by their host institution.

It was therefore decided to go through the document and try to reach agreement on as much as possible. The revised document would then have to be given to host institutions for comments.

It was not clear if the document should be described as a "Memorandum of Understanding" or a "Project Agreement", this may become clearer when institutions give their comments. The change to "Memorandum of Understanding" was proposed by EMBLEM.

The following changes to the Project Agreement were suggested:

1. Section 1.3.2, Give Aalib as an example rather than a definite case.

2. Section 1.4 Delete the paragraph beginning "Acknowledgements" and the first of the two paragraphs beginning "Publications". Retain the second paragraph starting "Publications" and add a sentence stating that partners should inform all other partners about future presentations and publications.

3. Licencing was discussed in relation to section 2. There was general agreement that LGPL was the most appropriate licence for EDNA.
In respect to copyright/ownership, in particular with reference to section 2.3.2, there was a clear concensus that the copyright should be shared between partners (for the common good) rather then belonging to an individual partner. This would simplify how to deal with what happens when a partner leaves the project. It was not clear if this would be acceptable to the various Institutes involved. The terms expressed in the second sentence of the fourth bullet point (For avoidance of doubt, Partners are not allowed ...) was flet to be ambiguous and required clarification.
The remaining issue was how to ensure that required versions of external software (eg BEST, MOSFLM) would be provided.

4. Annex 1. It was agreed that Jose and Darren should be added to the founding membership as defined in the section headed "Project Membership".
It was also agreed that the membership of the executive should be consolidated and those currently classified as "non-voting" members should be removed. The individual institutions, namely DIAMOND, ESRF, EMBL Hamburg and EMBL Grenoble should decide who will represent them on the executive.

5. Annex II. Should be entitled "Software to be used by EDNA" and extended to include MOSFLM.
6. Annex III is now redundant and should be deleted.

7.Annex IV. It was noted that a GUI is not explicitly required or part of the EDNA core.

8. Annex V. Olof explained that this Annex provided a useful definition of working practices and in particular explicitly stated that partners had to agree to code reviews. Peter Keller suggested that an alternative definition of "flexibility" was required.

9. Annex 6. The Title should be altered to remove "Proposal" and to include the date. Peter Keller pointed out that his proposed changes to allow discontinuous ranges had not yet been made to the Scientific Overview.
Action: Andrew to update the Scientific Overview.

Sean suggested that subject to these changes being made, section 1 and the Annexes were acceptable, and the next step would be to get agreement on section 2 from the institutions involved.

Action: Andrew, Alun and Olof to update the Project Agreement. Revised version to be approved by the various institutions.

6. Ad-hoc session.

Every member present introduced themselves and described their involvement in the project or what they hoped to derive from the project. Only substantive comments are minuted.

Olof restated his opinion that the Project Agreement was important to achieving the best way of working and avoiding problems that had arisen with DNA 1.0. The intention was to make EDNA modular and extensible for new science.

Gerard stated Global Phasing's interests as improving the design, control and execution of the experiment. He believed that the project was at a critical phase in terms of getting something tractable into the framework.

Thomas is responsible for 3 beamlines (2 MX) at Petra III. He hoped that EDNA is sufficiently modular to allow easy installation and to make it easy to modify individual modules. He hoped that it would be possible for individual sites to do rapid developments on their own, as collaboraive changes can be slowed down by decision taking. He had some flexibility in recruiting for Petra III.

Peter Turner is connected with the Australian synchrotron in an advisory capacity, and was particularly interested in remote access.

Sean explained that ESRF was currently the prime user and debugger of DNA 1.1, and that the reliability of the existing software was now quite good. He had concerns about the timescale for the new development, and whether EDNA was too ambitious in its concept and needed to be more pragmatic.

Clemens S-B was very interested in automation for SLS and hoped to be able to implement DNA.

Bob Sweet reported that DNA 1.x is used as a server at NSLS and is keen to continue to do this with EDNA, so he hoped that there would be clear definitions of where the "hooks" are in EDNA. He was concerned about the level of funding/resources.

Andy Thompson suggested by Email that European funding should be sought, probably via the STRAP mechanism.

7. Further discussion of Aalib

There was an extensive discussion on Aalib. It was noted that the Aalib development team is small (one person), which raises the question as to whether it is viable as a long term solution for EDNA. Against that, there was thought to be a clear need for a framework that provides the multi-threaded capabilities of Aalib for EDNA and there was no suitable alternative available. The requirement for a new framework (EDNA) to allow extensions to the functionality of DNA 1.1 was explained by Olof and Alun. The size and scope of the scheduler in particular made it very difficult to support. Gerard suggested that an alternative might be to modify the old structure rather than starting from scratch. When asked if Aalib was the only suitable framework available, Romeu commented that Eclipse was an example of the use of plugins, but would not be suitable for EDNA because it does not support multiple threads, which were essential for the complex environment of a synchrotron beamline. In assessing how complete the Aalib library is, Olof felt that internal testing remains an issue as does the lack of complete documentation. It was felt that workshops would help to address this. In response to a question from Sean, Olof felt that Aalib would allow rapid changes to meet new scientific requirements.

It was concluded that although there is a risk associated with the use of Aalib, it is an acceptable one and that for the time being at least Aalib should be used.

8. Planning

Olof presented a planning for the following EDNA developments:
- Project management tools, development practises
- Use Cases
- Development of a pre-screening module

There was a general discussion on the Use Cases. Gerard was concerned that the Use Cases based on simple experiments would be extensible to more complex ones. Olof thought that this should be possible. Romeu asked in the data model would be easily extensible. Olof replied that the data model would need to be carefully designed so that it could be easily updated when the Use Cases were updated.

Planning of Use Cases
In discussion of the proposed Use Cases, there was a general feeling that it would be more appropriate to go directly to the Indexing Use Case and defer work on the PreScreening Use Case. This is partly because the PreScreening module would require significant code development (eg being able to read all the various detector formats).

Thomas agreed to commit help from Hamburg on the development of the Indexing and Integration Use Cases. These would be passed to Elspeth for further comments from ESRF scientists.
This should result in one complete Use case being developed by the end of November. It was felt important to strike the correct balance between having a relatively simple Use Case and a completely general one that would take too long to develop.

Sean felt that it was important to be able to demonstrate a "Proof of principle" with EDNA in order to justify future resources, and to show that it will be able to do things that cannot be done with DNA 1.1 (eg complex experiments).

Olof thought that if the PreScreening were module was postponed, it should be possible to achieve the data strategy and data collection uses cases prototype of EDNA by June 2008, although this did depend on available resources. The developers identified for this work are Marie-Francois, Romeu, Peter Keller and Karl. Alun commented that Karl was heavily committed to installation of DNA 1.1 at Diamond at present, but would be able to contribute more in Jan 2008.

The following actions were agreed:

Action: Fully described Use Cases by end of November for Indexing and Integration from scientists at Hamburg (coordinated by Thomas) and ESRF (coordinated by Elspeth).

Action: Romeu will provide guidelines on usage of Aalib by December

Action: A workshop will be organised in Mid November/early December on how to go from a Use Case to an implementation and to get an introduction to Aalib. This could be hosted by EMBL Grenoble. In order to keep the meeting focussed on EDNA only members of the EDNA project will be invited.

9. Date of next meeting

The next full meeting should be held at end of Feb/beginning of March 2008.