AGENDA and MINUTES
DNA-DEV meeting
3rd-6th of May 2005
Hamburg

Tuesday May 3rd

13.30	Beamtime 0 and arrival. Breakout on Background processing and feedback if Olof present?		
Wednesday May 4th			
Session 1: Current Status			
9.30	Welcome and opening of meeting	Victor Lamzin	
9.35	Review of Full DNA meeting actions	Alun	
9.50	Feedback from recent meetings e.g. Brookhaven, up and coming summer calendar	All	
10.10	Release 1.0.1	Graeme/Alun	
10.30	Review of pipeline experiments	Karen, Olof and Darren	
10.45	Coffee		
Session 2: Near Future 1			
11.00	The Offline version and related issues (e.g. support and how wide)	Graeme to lead	
11.15	Release 1.1 including Branches	Alun/All	
12.00	Data Access Layer 1		
12.30	Lunch		
Session 3 Near Future 2			
13.30	Data Access Layer 2 - how it will be done e.g. UML and tools		
14.00	Kappa		
14.30	XREC autocentering and DNA connections	Victor Lamzin	
15.00	Coffee		
Session 4 Near Future 3 and Local			
15.30	"Scaling and Merging" and "Spacegroup determination"		
16.00	Discussion on local Hamburg DNA issues (BCM, dose mode, user feedback, strategy improvements). (incl MAR integration?)		
17.00	Beamtime 1: breakout on Radiation damage		

Thursday May 5th			
Session 5: software updates			
9.30	Updates from data reduction, scaling and merging etc packages (i.e. Mosflm, BEST, XDS etc) and discussion on their usage or inclusion or more in DNA.		
10.15	Scheduler/XIA.		
10.45	Coffee		
Session 6			
11.00	Ranking		
11.30	General DNA software development. Are we using CVS etc happily? ESRF would like to make a presentation on agile programming(?) to include unit testing, validation etc		
12.30	Lunch		
Session 7			
13.30	autoRICKSHAW and DNA connections	Santosh Panjikar	
14.00	AOB and mopping up.		
15.30	Beamtime 2		
20.00	Late Dinner		

Friday May 6th

9:00 - 14:30 Kappa workgroup meeting

MINUTES,
Kindly taken by Harry Powell, Steve Kinder and Sandor Brockhauser

Attending:
Alun Ashton
Graeme Winter (GW)
Darren Spruce (DS)
Sandor Brockhauser (SB)
Pierre le Grand (PlG)
Lorenzo
Romeu Andre' Pieritz
Olof Svensson (OS)
Harry Powell (HRP)
Steve Kinder (SK)
Karen Ackroyd (KA)
Victor Lamzin
Gleb
Sasha Popov (SP)
Parthasarathy Venkataraman

Observers
Manfred Weiss
Clemens Schulze-Briese
Santosh Panjikar
Tilo strutz
Binoy Mathew
Uwe Mueller
Andrea Schmidt
Uwe Ristau
Brice Kauffmann

4th May 2004

Review of Full DNA meeting actions

GW liaised with PRE re: pointless. Didn't make much progress. HRP - pointless is closer to being ready, is available for testing on some platforms - needs cctbx.

Action: GW to make further enquiries on POINTLESS.

GUI was blocking - OS et al to look at this. GW & OS will discuss here (as suggested at the video meeting).

PlG: XDS - discuss later, but will be in 2.0

Andy Thompson - to send input parameters to Avi re: raddose etc.

Sasha & Gleb to report back to full meeting on maximum doses at different beamlines. SP - have info on ID29 & ID14-4. In principle it can be done, but data is currently lacking.

OS - passing info back to ES apropos downstream processing. OS – must be discussed with GW.

Sample ranking - Romeu will present 050505.

Karen Ackroyd - testing on 7.2 & 14.2 - will be discussed later.

DS & OS - Darren and Olof have been coordinating developments of database links as necessary for pipelining tests and the ranking module.

MAD - OS to develop plans for a separate module - 2.0 feature, no timescale, refer to full meeting.

Standard set of images - CVS & make them available, action GW. HRP - what are the images for exactly? Answer - to check that successive iterations of DNA still work- we should detail the requirements. PlG - two types of tests - performance testing and something else.

Olof – will be integrating CHOOCH at some stage.

Publication - HRP to remind AGWL.

Andrew will communicate with the following on these topics, who will provide their contribution by the end of June:

GW - scheduler
DS, SK & KA - BCM
OS & DS - Pipeline & ES
HRP - Mosflm mosds
SK & KA - GUI & XML
all - features (or was it "futures"?)
SP & GB - BEST modifications

ACTION Above to e-mail Andrew and CC: to HRP

Feedback from recent meetings e.g. Brookhaven, up and coming summer calendar

GW: week at Brookhaven - set up DNA & tutorials. Getting it working was relatively straightforward when installing from the release candidate but the installation and setup was more problematic if taken from CVS or copied

from another beamline - due to abstraction. Worked on 2 beamlines, collect & characterise and collect & integrate. BNL want to install on all 6 beamlines. No feedback since visit. Support may or may not be an issue as there are individuals at BNL keen on DNA.

GW has talk at ACA as well. GW pleas for repository for talks – e.g slides etc.

OS - IUCr - not giving a talk but is giving poster.

HRP - EMBOcourse - talk @ course at ESRF in June, maybe practicals.

Release 1.0.1

GW has produced a release candidate which has already been tested offline in Soleil and Diamond.
GW - but it has been found to be slightly broken - broken PATH, didn't include scripts directory. Now fixed though (since yesterday) - not CVS committed yet. Just about ready - confirm through testing this week.

ACTION: Video conference - list of bugs fixed - action GW.

Review of pipeline experiments

OS – Presented slides as below derived from Johan's report:

Pipeline tests 2-3 March 2005

Notes from report written by Johan Turkenburg

- In York 30 samples of 4 proteins were frozen (by hand) using ESRF/SPINE loops and Fedexed to Grenoble in baskets in a dry shipping dewar.
- Details of samples were put into PXWEB using a web browser in York (by hand). No diffraction plans were submitted at this stage.
- All samples were loaded into the sample changer at once. Barcodes were scanned by the sample changer, and apart from one all were recognised.
- Barcodes were linked to sample information in PXWEB automatically. A list of samples to screen was drawn up and displayed in DNA.
- Crystals were put on the diffractometer, centred and 2 diffraction images were taken at 90 degrees apart (all automatically).
- Images were indexed by DNA using mosflm, and a strategy determined. Results were retrievable using PXWEB.

Results and problems

- In the longest uninterrupted run 27 samples were screened without any human intervention
- Turnaround per sample was in the order of 4 minutes. In a parallelised process the sample is now unloaded and a new sample is loaded while DNA analyses the test images. Most time spent doing the sample transfer and the centring.
- One pin was lost by the sample changer. Given the number of sample changes over two days (at least 200?), this is remarkable, and better than most human users would achieve.
- Most of the interruptions to the process were caused by a failure to unload the sample.
- Centring of the crystal was extremely successful! In a few cases the software bailed out and asked for manual intervention.
- DNA never fell over but failed in a few cases
- All results were easily retrieved using PXWEB.

Suggestions for improvements

- It would be useful if the list of samples to be screened could be manipulated more extensively in DNA before a screening command is issued. Ordering according to prefix, alphabet, symmetry etc were discussed, or simply the order in which samples are clicked.
- Remote input of diffraction plan needs to be tested, and DNA should be able to read this.
- Multiple programs to index/assess images (mosflm/XDS/labelit).
- Presentation of results to enable user to rank samples. Graphs from MOSFLM/BEST.
- 'Collect' button in DNA which allows collection of selected samples after screening (requires a means to select samples in an interface).

. On ID14-EH3

Follow up tests for reproducibility

KA - 18 May - SRS test of DNA. Generic data acquisition software PXGen will be controlling a sample changer. No barcode reader, so all information must be in DB, won't be read at BL. ISpyB as DB for test.

DS - other ISpyB developers - enhanced version for DL, but not installed yet.

SK - screening results should go to DB

Plans for Hamburg - once sample changer installed - after 1st July. So maybe by the autumn.

GW - BNL people interested in pipelining.

Offline version of DNA and related issues

The important discussion point was: can we meet support requirements that would follow from a general release of an offline DNA version? Previous releases have always been to collaborating synchrotron sites. Graeme is however asked regularly for an offline version. The subject of what people want from an offline version was discussed. DNA is designed for online use at synchrotrons. People could however use an offline version at synchrotrons that do not have an online version installed, or with lab sources, using images collected online. Use as a learning tool was also mentioned. For 1.0.1 a list of non synchrotron sites for distribution has previously been agreed

- Structural Genomics Oxford
- LMB Cambridge
- York
- Diamond Light Source

It was agreed that no release further than this would be possible. Support is not available currently. A release without support would also not be advisable. From experience, problems can be foreseen and this would only create bad feedback. Following on from release to the above sites the issues involved in a more general release should be evaluated.

Action: All to evaluate offline success following release.

Graeme introduced the subject of a CCP4i interface to XIA. He demonstrated what he has already created. Graeme felt that this would be easier to support than a full version of DNA and allow wider use of the functions provided through XIA. He needs a mechanism to release elements of his work not used in DNA. Graeme stated 98% of the code would currently be unchanged. He stated that he had seen or been informed of problems using the GUI with some versions of Java (GNU Java explicitly). However now these are known of they can be investigated and addressed. A number of developers expressed concerns about the general approach.

- Possible divergence of code.
- Extra overhead maintaining a second GUI. Although simple now there could be a temptation for this to grow.
- Elements of XIA that are integral, and developed as part of DNA, being released separately and without recognition.
- A different interface removes the possible benefit of people using the standard GUI prior to arrival at a collaborating synchrotron.
- Facilities available outside XIA would not be available e.g. ranking.

To progress it was felt we need to get a list of known problems that exist using the standard release as an offline version.

Action: Graeme to enter known problems in Bugzilla and circulate.

It would be several months before such a release could, or would, be made. This subject should be continued at subsequent video conferences. The executive committee should be consulted.

Action: Alun/Graeme to email Executive committee.

Release 1.1 including branches

The basic list of major new features for 1.1 was reviewed.

1) Scaling of data as collected
2) Merging of data as collected
3) Inclusion of pointless
4) Hard limits for radiation damage
5) Modelling and some beamline description
6) Simple sample ranking based on I/sigI
7) Some data saved to database

The possibility of adding the ability to start XDS as a completely separate process was also requested later in the meeting. It was felt this was fairly simple and valuable enough to be added to this list.

1 and 2 have been in the head and Graeme's branch for some time now. Graeme uses them in his own data processing but they have not been available as a standard part of DNA, or used much outside Daresbury. Significant discussion followed on what sort of scaling was in DNA and/or actually required by DNA. The scaling implemented currently can take significant time and blocks further action in DNA. This had been felt unacceptable and was the cause of it not being released. The possibility of some 'quick and dirty' scaling was raised. Merit was seen in this but no decision made. Another possibility was to allow the option to run scaling separately via a button, with the proviso that user would be warned that this could take a long time and would block data collection. Concerns were raised that such a warning would simply put people off ever using it. It was felt that this needed further discussion and should be put to the executive committee.

Action: Alun to request guidance from Executive.

Pointless is still under development and not released yet. No support has yet been built into DNA. It was however felt to be achievable for 1.1.

Adding a hard dose limit should be fairly simple. A standard base limit could be stored in the system defaults file. Corrections would need to be applied based on factors such as fill mode of the synchrotron, attenuation being applied, wavelength in use etc. The BCM could provide data on these factors via a parameters request. Andy Thompson was to have looked into important parameters for radiation damage and forward to Avi Naim (EBI) following the Feb 14th DNA meeting.

Action: Pierre to follow this up with Andy and submit to DNA mailing list.

No ranking has been implemented yet but work on the software framework is underway. Olof and Romeu presented this later in the meeting. It should be feasible for 1.1 and would be implemented through the executive system.

Item 7 refers to results from screening, as required for ranking, and would also be feasible for 1.1.

Olof presented a further list of enhancements requested by an ESRF BAG meeting. The priority of these and their feasibility were discussed individually, below:

Indexing

- Would like to have the option of using more than 2 images for indexing **low priority**
- Would like to read in orientations found from different indexing procedures to get strategy (DNA is good as this is then saved to database) **low priority**
- Not happy with DNA indexing procedure – not as robust as other software for some projects? **done.**
- DNA criteria for a good indexing solution are rather strict and we would like to be able loosen these up (perhaps through expert mode a la CCP4i "not often used options"). **We agree, can be changed easily, should be reviewed! (Harry, Graeme to send email) Documentation!**

Strategy

- Strategy is not calculated since the estimation of the mosaicity fails. **Not a simple one answer problem. Failure of code must be addressed. Allowed to collect data - exec decision?**
- When will anomalous strategy be working ? **1.1? Only strategy!**
- The strategy is not very good – Ravelli's strategy is better. Can we implement this stratgey program in DNA? **Not a problem – definition…**
- Can we run DNA including a previous mtz file(s) to calculate the strategy ? **Low priority.**

Resolution limit of the crystal

- Can DNA automatically take several test images at different resolutions ? **No, not 1.1.**
- The I/σI at the edge of the detector always seems to be 3, is this a bug ? **A feature, to be improved in 1.1.**

Integration

- Can DNA assess number of saturated spots on image and predict a high resolution/low resolution pass strategy? **Yes, but not 1.1.**

Integration and Data Processing

- Is there a time lag between data collection finishing and the data integration finishing? **No, but scaling time is arbitrary (2 min -> hours). Quick scaling by limit the resolution and/or the number of cycles if time allows. Warn user that if scaling is used the DNA system will be blocked till the scaling and merging is done. Run the scaling in the background? Parallel scala? User to choose? Info to user?**
- When can we use the clusters for parallel MOSFLM? **Concurrent MOSFLM jobs – not necessary for 1.1.**
- The information from the integration step – to the DNA window is not very helpful. Can we have an executive-style summary with tracking of rms deviation/spot shape, number of saturated spots? **RMS deviation, saturated spots easy – spot shape difficult. Needing more investigation…**
- How can we tell if the crystal is suffering from radiation damage? **How to distinguish between crystal and beamline? Scaling needed for doing it properly. It's possible to do it by running BEST on each batch of images. Not 1.1!**
- Should we go back and collect a reference image – we should have on-line scaling and warning messages when the Rmerge (or other metric) passes a pre-set threshold. **Not possible – at least not for 1.1.**
- Can we have some recommended pre-set exposure times which are beamline specific? **Possible for 1.1!**

Alun presented a list of 1.1 features that had previously been compiled and these were also commented on:

1. BCM development documentation – DNA developers @ Hamburg (after release)
   Full documentation still outstanding. BCM simulator was however, very helpful,
2. Strong diffraction at the end of the detector and spots well separated move detector closer -> DNA 1.X
   Previously mentioned
3. Some images seem to be marked icy by mistake
   Yes. Graeme? Test set please and yes for 1.1. low priority
4. Indexing failed image 1. No orientation matrix. Not integrating images – mosaic estimation failed. (Should have failed – multiple crystals)
   Covered previously

5. Result images even if indexing failed (needs more thought)
   Done? This may have already been done.
6. Indexing failed. Spots getting "smudgy". Not getting max. likely cell @ edge.
   Unclear what this is???? check last dev meeting.
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!
   Cured by indexing improvements. Needs confirming?
8. MOSFLM XML to contain spot profiles -> DNA 1.1
   Agreed
9. Dynamically allocate ports
   Darren? ;OK; can be in 1.1
10. Mosflm log stored in system defaults file if used by multiple users permissions cause failure.
    Change needed in config file?
11. Move mosflm logfile each time an auto index performed.
    Graeme will check
12. Add ability to set mosaic spread value to GUI
    See earlier DEBATE.
13. executive should not go on to strategy if too many bad spots (BEST)
    probably doesn't now. Graeme will check later today
14. Enhance BCM simulator to simulate data collection by copying images.
    DONE!!!
15. Improve error reporting and split into program failure and operational failure.
    TO be done. Follow up as not likely for 1.1
16. Add MaxOsc keyword to strategy testgen.
    Into config file?
17. Configurable max and min exposure time
    Part of strategy redesign stuff.

Action: Alun/Steve/Olof to enter 1.1 features into bugzilla.
Action: Alun to compose a single coherent list for submission to executive committee.

Data Access Layer (DAL)

DS - lots of links to databases from labs at ESRF. DAL generates a link to an identity and a handle.

See appendix for Darren's slides.

Data handling

(see DS's slides, but it's something like this -

http request <------> ISpyB request <---------> MySQL server <----> ISpyB DB
handler python sql
dictionary

gave e.g. of DB code - Python coding of SQL calls

DbHTTPServer.py - XSchema autogenerates the database XSD

Graeme requested if these links could be done without http servers for simplicity on a stand alone machine. This was thought possible and both mechanisms would be required.

Lorenzo - want to keep a really general description, so if we can define "detector" "sample" etc we can just share the description of the structure.
DS? pointed out that Ludovic Launer has added some completely general descriptors, e.g. of "machine".
Lorenzo - develop a good data model, if possible using a pre-existing model (Poirot?).
GW - should we examine the DNA data model?
OS - agree it should be cleaned up – top down rebuild using proper OO techniques.
LM - description should be independent of software
PlG - do we need DB in DNA?
DS - use instead of a configuration file. This has advantages.
GW e.g remote access, eg via PXWeb.
Manfred Weiss - databases allow examination of the procedures in retrospect.
Romeu - Something about magic numbers, design parameters, typecasting, compatibilities.
LM - Use of clear data models is useful because it allows easy sharing.
PlG - define the internal structure of DNA so that it reflects what DNA does (or is intended to do). Need to define the data structure but not the methods.
standards - e.g. imgCIF (CBF) - hierarchical format, IUCr standard. contains sufficient information for ready translation to XML.
Lorenzo - try to keep a consistency in style
DS Don't see how this maps onto actions
Lorenzo - something about CCPn's needs for XML and databases
GW – Felt that some commercial UML tools were inflexible
SB Next action should be to decide how to proceed (in small group) - defining a data model for data storage in DNA
DS - his model is defined by eHTPX (eh? does this make sense?); immediately data is entered it gains inertia & doesn't move much.
GW - will this affect DNA directly?
SB - shouldn't. A later developed model should be able to deal with the current model.

DS - this is to provide an interface to the DB, not to replace the current code.
SK - But that's an important goal, especially for the scheduler.
HRP - who's going to be working on the data model?
OS - everyone involved...
AA/PlG - translate imgCIF to UML
SK - Do we want Object Domain? Why?

It was generally decided that the best way to proceed initially would be with pen and paper. KA - need to deconstruct the current dna and organize in the hierarchical fashion. Decided on an additional videoconference to discuss this, need to deconstruct dna prior to the next dna-dev and dna meetings.

ACTION Alun to organise specialised DNA Video conference for DAL

Date suggested/decided on dna-dev 7th July @ESRF
Date suggested/decided on dna(full) 8th July @ESRF

ACTION Alun to find if these dates are ok.

GW - people should talk about the other bits of data, e.g. diffractionimage, bucket of stuff in the scheduler.

All were in agreement on this last point.

SB - Kappa

Have some beta users - very enthusiastic, e.g. Venki Ramakrishnan.

Overview - aims - align xtal on arbitrary axis
- smart data collection
  - equiv images from different crystals
  - etc

talked about software structure - has an internal experimental model, suggests using a DNA_kappa_strategy_request.

Uses the GONSET convention for aligning the crystal.

Integration into DNA

1st milestone - December 2005

Will need to add extra panes to the DNA GUI if Kappa is enabled.

ES - check if kappa is available - get kappa_orientation,

alignment_request, strategy_request etc.

GUI requires 4 new panes
- Strategy
  - Desired orientation
  - crystal vector selection
  - kappa strategy selection

Scheduler/XIA
---
new activities - kappa alignment calculation
- kappa strategy calculation

communication possibilities - prepare i/p files and calling external unit & cahnges to BCM to accommodate extra motors if kappa is available. -

Also integrate into pipelining.

Requirements - need to calibrate mini kappa
- directions of rotation axes
- locations of rotation axes
- motor limits
- collision prevention - not implemented yet, currently uses circle slippage to identify collisions - this is unsafe.

SK noted that work will need to be done on the XML as this did not exist for Kappa and SB will now open a new CVS branch for developing and interface to DNA. SB stated he will make his developments to work with the scheduler as the core of the scheduler will not change on implementing XIA

5th May 2004

Discussion on local Hamburg DNA issues

Report from Sasha and Gleb:
- DNA has been successfully installed on two Hamburg beamlines (BW6, and X13). After a disk crash, DNA could be easily be reinstalled.
- A new local BCM had been developed on top of the BCM simulator by implementing its interfaced functionalities. Gleb wants to replace this current BCM solution by a more elegant one.
- Until now there has not much feedback, but lots of interests from the user side.

problems occurred:
- Correct permission check in DNA (e.g.: after the images has been created by BCM) is missing.
Alun suggested reproducing this problem and registering it to bugzilla

ACTION: DNA needs to check image file permissions. Bugzilla.

- Starting with modified resolution does not work properly
It must also be put to bugzilla

ACTION: problem with modified resolution needs to go into Bugzilla.

- Dose mode (instead of declaring the exposure time, one should be able to define the number of photons /between 10^1, and 10^4/)
Gleb suggested using the "dose" value instead of the "exposure time" with an optional checkbox "Dose Mode" that would make the meaning of this value clear. Graeme suggested storing this value separate in the XDS scheme.
Gleb has also asked for visualizing the estimated time left in case of dose mode. Although it can be done somehow, Alun believed that it is not DNA's duty.
- Failure with indexing
Harry asked to be immediately informed about such cases.
- Collect and Integrate cannot be stopped when mosflm is waiting for a new file (Scheduler bug)
Graeme acknowledged this bug, and promised a fix

ACTION Graeme to fix bug with Scheduler not being able to stopped when Mosflm is waiting for new file.

- DNA process communication error handling scheme is statically defined by a timeout value.
BCM should automatically handle beam-loss situations, but currently DNA could not follow what is happening. DNA communication error handling scheme ought to be changed (eg: by asking BCM first about the reason of the delay...)
- Gleb complained about the DNA Manual
- User requests arrived: I/sigma, and Redundancy factors should be made available for users.
- Image handling issues (automatic symmetry determination must be refined)
Graeme has some ideas. Gleb also do so. Harry thinks that pointless may help. Finally, Graeme suggested to have a new button "Collect low resolution" next to "Collect and Integrate" especially for point group determination

ACTION Graeme and Gleb to work on issues with automatic symmetry determination

Updates from data reduction, scaling and merging etc packages (i.e. Mosflm, BEST, XDS etc) and discussion on their usage or inclusion or more in DNA.

Mosflm developments
==========
+ Harry has found a bug in his latest version. Fix is promised by the end of the next week.
+ Graeme reported a mosflm bug in the XML output. Harry has kindly asked again to be immediately informed about such cases.
+ Pointless is available in more platforms
+ No more SCALA developments
+ Mosflm win32 binary is available for use with cygwin. It is still unstable on windows.

* BEST developments
==========
+ Resolution can be calculated by BEST
+ Resolution limits can be also calculated from images taken with the detector far from the sample
+ indexing, and strategy are also better with images taken far from the sample
+ redundancy request is available for BEST strategy
+ taking account the radiation damage will also be available (with a basic approach) in half a year

* XDS integration (Pierre's report)
==========
+ XSCALE is available. Its integration by scheduler can be done in 2 minutes as Graeme believes.
+ XDS, and XSCALE are using parallel CPUs. They will be available for running on clusters soon.
+ full XDS integration can be done for DNA v2.0
+ Pierre suggests that DNA could start XDS integration automatically (if configured so) parallel without reading the XDS results back. Graeme believed that it can easily be done for v1.1.

Scheduler and XIA developments (Graeme's report)

+ Labelit WILL be compatible with mosflm. Executive must be asked if it should be added to the autoindex process (e.g.: in a case of a mosflm failure). Sasha would rather encourage the beamline scientist to do a proper job during the beam centre determination.
+ Graeme has finished the developments in his branch. Before leaving this branch, he has removed some irrelevant parts to make it cleaner and more easily mergeable to the main branch.

+ Graeme promised that XIA with a new architecture will be available for DNA v2.0
+ Graeme reported that XIA has a CCP4i interface for the functionalities of the current offline scheduler.
Alun highlighted that we made a decision (May 4) that DNA-offline would not contain a CCP4i interface.
+ XIA v0.1 will be available for testing at the end of the summer.
Graeme wants to release v0.1 to get important feedback for further XIA developments. Karen would like him to be more focused on important DNA developments. Alun highlighted that some of Graeme's XIA development is an independent activity.
+ XIA-core is under the control of a Committee (Graeme, Olof, and Pierre)
+ Scheduler is a part of XIA, but not a part of XIA-core Sandor believed that XIA Committee is responsible for DNA-XIA connectivity, XIA integration to DNA by replacing the current DPM-Scheduler, and the Scheduler emulation functionalities of XIA. He became confused if the XIA-core Committee as mentioned above was able to fulfil the responsibility of XIA Committee in case the Scheduler would lie out of XIA-core that it only has control of.

(NOTE from Alun: the Scheduler is the predecessor of XIA DPM and XIA Core. Major developments on the Scheduler have now stopped and is therefore 'stable'. Connectivity from other DNA modules to the Scheduler have been guaranteed to work in the same way with XIA as the interaction is through the XIA core. In addition to XIA-Core and XIA-DPM there are other XIA modules which are not a part of DNA as they are for downstream processing e.g. MR)

+ Graeme has invited Olof, and Romeu to make suggestions, and even developing and supporting XIA-core in the future.

Licensing

========
+ Victor suggested making the licensing issues clear.
He has highlighted the importance of licensing (e.g.: legal issues; keeping track of download/usage; acknowledgement of authors;...). For third party integration, he suggested to use the arp/warp licensing technique, but described two scenarios:
a) Distribute DNA and BEST separately, apply plug-in technique for integration after installation
b) (more paperwork) supply DNA and BEST in one downloadable package
+ DNA developers mentioned that DNA release 1.0 is licensed under the standard CCP4 Licence, but later solution must be discussed at the next DNA-Full Meeting
+ Harry was wondering if DNA and XIA have any agreement.

Graeme reassured us that the current CCP4 licensing scheme used by DNA allows him to perform XIA development on the way of licensing it under the same CCP4 licence scheme.

* Branching and technical issues
==========
+ Graeme's branch has been reduced and prepared for merging
Sandor has suggested merging the Head to Graeme's branch first, and testing it before final migration from the merged Graeme's branch to the Head branch. He can be asked for technical details. Steve wanted to highlight the importance of checking all the changes before actually performing their merge. We have decided to ask the Executive about the procedure to follow.

ACTION Alun to consult with Executive that the remerging of the Graeme branch is acceptable in this way.

+ Sandor's account does not work. Graeme promised to check and make it alive again.

ACTION: Graeme to ask CCP4 support about Sandor's Account.

Ranking and Executive System

Romeu, and Olof performed presentations (see Appendix) of Object Oriented Programming, and its application for Data Ranking, and the reengineering of the Executive System. They suggested using this technique for a better understanding/portability/... of the DNA. Darren expressed his being resolved in following them in restructuring the DB module. Graeme has also agreed on its main benefits, and asked Olof, and Romeu for cooperation in XIA.

Auto/RickSHAW

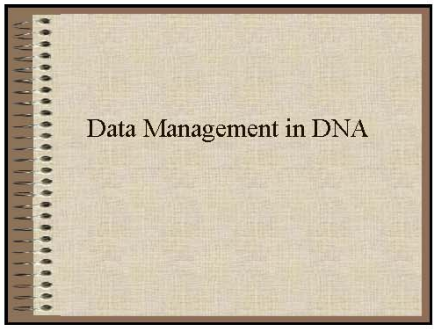
Santosh gave a presentation. DNA would need a binary/scaled feedback from rickSHAW to decide if the data collection should be stopped, or continued. autoRickSHAW is glued together in a csh script, so it does not contain any abstract pipelining design that could be adoptable for Scheduler/XIA design.

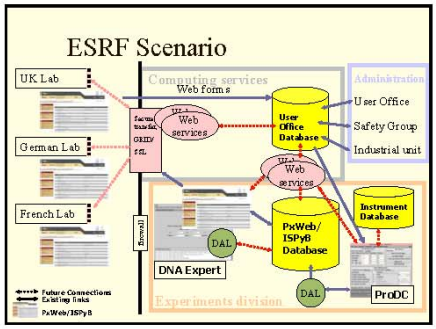
Testing

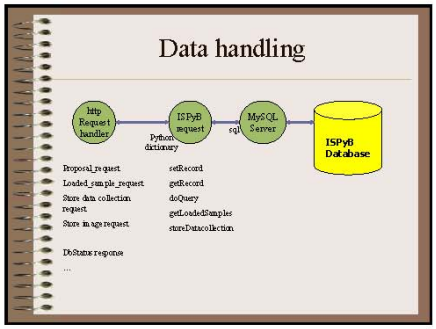
+ Pierre has volunteered for manually testing images if it is necessary for building standard tests
+ Pierre suggested to define different test classes and build separated sets of test images for them Sandor described the test as the result of comparison of a predefined correct answer and the current output generated by the software under the test.

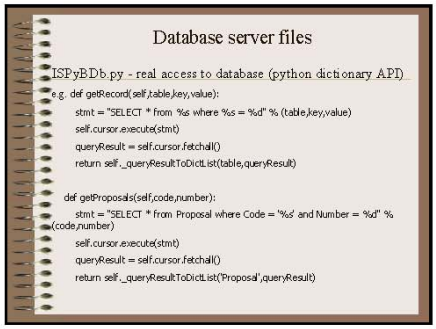
+ Darren has suggested building an automatic standard test framework that handles all the test classes and all the test images...
+ Sandor volunteered to design such a testing work frame. Since he has been asked to indicate the deadline, he does so. Sandor will give a short proposal of the DNA-testing framework on the DNA-full meeting on July 8, 2005 in Grenoble.

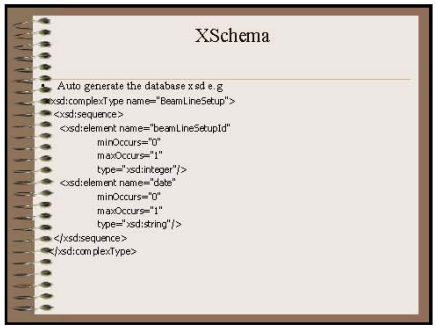
APPENDIX
Presentation from Darren.











Xschema

•Add compound objects for common requests to reduce work in the client

E.g.

```
<xsd:complexType name="Proposal_response">
  <xsd:sequence>
    <xsd:element name="status"
        minOccurs="1"
        maxOccurs="1"
        type="status"/>
    <xsd:element name="Proposal"
        minOccurs="0"
        maxOccurs="1"
        type="Proposal"/>
    <xsd:element name="Person"
        minOccurs="0"
        maxOccurs="1"
        type="Person"/>
    <xsd:element name="Laboratory"
        minOccurs="0"
        maxOccurs="1"
        type="Laboratory"/>
    <xsd:element name="Session"
        minOccurs="0"
        maxOccurs="unbounded"
        type="Session"/>
  </xsd:sequence>
</xsd:complexType>
```

Steps to add to model

1. Define model
   - Linking with existing model
2. Generate database tables
   (From model)
3. Add methods to handle request/response
   If new methods are needed.
4. Implement in client, keep track of results.

What do we need

- Centralised data model
  - E.g web site with definitions
- Utilities to build from model

Presentation from Olof.

Disclaimer

- This is **not** a try to dictate how to develop and write code, it's just suggestions of how we think the **collaborative** software development could be improved.

Problems with Python:

- Python is a scripting language
- Python is the best scripting language
- However, for big projects involving many developers (like DNA)

Problems with Python 2:

- No built-in polymorphism like in Java / C++:

```
void foo(int x)
{
    printf("I'm in void foo(int x): x is %d\n", x);
}

void foo(char *x)
{
    printf("I'm in void foo(char *x): x is "%s"\n", x);
}
```


Problems with Python 3:

- No compilation – use of unittests absolutely necessary!

Suggestions 1

- Object oriented design:
  - Everything is an object. Try to encapsulate everything.
  - Derive all classes from a base object which provides the magic number and the id.
- If you follow strict object oriented design more people can work on the same module without the risk of exploding the program.

Suggestions 2

- Follow consistent naming scheme:
  - Use name space for naming of classes and add subclasses to the name of base classes, examples:
    - ESObject, ESAction, ESActionCom, ESActionComBCM, ESActionComBCMCollect
  - Use "o" for objects, "str" for strings:
    - _oesAction, _strOutput
  - Use "m_" for desgning class / instance variables:
    - self.m_oesAction, self.m_strOuput

Suggestions 3

- Don't use language specific tools like lists and dictionaries directly – encapsulate them:
  - Python list object:
    - a = [1, 2, 3]
    - a = ESList(1, 2, 3)
- If you do this your code will:
  - Be very easy to port to another language
  - Be very easy to understand by someone who doesn't know the programming language.

Conclusion

- The DNA project is a collaboration. Even if entire modules of the system are developed by individual developers / partners it is important that the structure of the programs is easy to maintain and further develop by any developer.
- This presentation is meant to be a starting point for discussions on how we can improve the code we write so that it is easier to maintain, expand and in general to understand for other developers.

Presentation from Romeu

Data Ranking - Code Project

BioXHIT :: DNA-Dev Meeting

Dr. Romeu André PIERITZ
ESRF - Exp.Div./So Soft Group
pieritz@esrf.fr

Objective / Content

- How can we Rank?
- What is the main feature that we need?
- What kind of software architecture can we implement?
- How it works?
- How can it be used with DNA?
- What is inside?
- Conclusions...

Scope and Limits

- No discussion about Crystallographic physics inside: to be defined by the scientific committee;
- Only Introduction of the Code Architecture Philosophy.

How can we Rank?

...good question!!!
We can't define it yet....

"It depends on the data set and the experiment! (of course: it depends on the scientist...)"

What is the main feature that we need?

"FLEXIBILITY" !!!

- Flexibility... to change the rank strategy on the fly;
- Flexibility... to rank different data sets...by different methodologies;
- Flexibility... to compare different ranks;
- Flexibility... to change....

What kind of software architecture can we implement?

Pure Virtual Mechanism
(true OOP-Object Oriented programming):

- based on a "Rank Project" container
- analysed by a "Virtual Rank Method";

How it works?

The rank container: store all crystals and their parameters (evaluated by DNA);

Virtual Rank Method: rank the data set by its engine;

How can it be used with DNA?

(#1 version) - Simple and Easy:

"Process control" =
single process execution by command line arguments...

>DNA stores the crystal file data on the disk and requests the processing with a project.
When the rank software ends, DNA reads the project and shows the results

Scenarios:

#1- Rank a Single Crystal:
> dnrnk.exe -ip projectA.drp -ic crystal098.cry -resu mad.est

#2- Rank two Crystal Sets:
> dnrnk.exe -ip projectA.drp -ip2 projectB.drp -op projectC.drp

#3- Rank Different Methodology Test:
> dnrnk.exe -ip projectA.drp -resu strategyB.est -op projectD.drp

#....

What is inside? (top-down)

What is inside? (top-down)

What is inside? (top-down)

```
# dr_engine_mad.py

from dr_engine          import DREngine
from dr_factor_mad      import DRFactorMAD


class DREngineMAD( DREngine ):
    """Class for DR engine MAD rank objects."""

    def __init__( self ):
        DREngine.__init__( self )
        self.m_odrnormalize.setReference( 100 )

    def rank( self, _odrProject ):
        objfactor  = DRFactorMAD( self.m_odrnormalize.getReference() )
        objproject = objfactor.rank( _odrProject )
        return objproject
```


What is inside?

class DRObject ("pure" basic virtual class)

class DRManager
    class DRManagerRank, ...

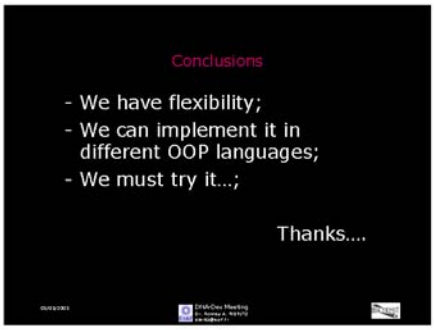
class DRCommand

class DRProject
    class DRProjectCrystal, class DRProjectCrystalSet

class DRFactor
    class DRFactorIce, class DRFactorResolution, class DRFactorStrength, ...

class DRRank
    class DRRankEngine01, class DRRankEngine02, ...

class ...,


Conclusions
- We have flexibility;
- We can implement it in different OOP languages;
- We must try it...;
Thanks....

Acknowledgements
Dr. Olof SVENSSON
ESRF - SciSoft Group
Dr. Claudio FERRERO
Head of: ESRF - SciSoft Group
and the SciSoft and BioXHIT Project