CCP4 and BIOXHIT

Data Management and Project Tracking in Structure Solution

Peter Briggs, CCP4

Outline

- Background
- Staff involved in CCP4 effort
- Aims and Deliverables
- Current Status
- Links to other projects
- Next steps

Background: the BIOXHIT Project

- **BIOXHIT**
  - Bioxtallography on a Highly Integrated Technology Platform for European Structural Genomics
  - EU Framework Programme 6 “Integrated Project”
  - 20+ partner institutions

- **Aim**
  - *“To provide platform for high-throughput structure determination from crystallisation to structure solution”*

- **Timeframe**
  - started 1st January 2004 for 4 years
  - problems with recruitment meant late start for CCP4

- **Website**
  - **`http://www.bioxhit.org`**

CCP4 Contribution

WP 5.2: Data Management & Project Tracking in Structure Solution:

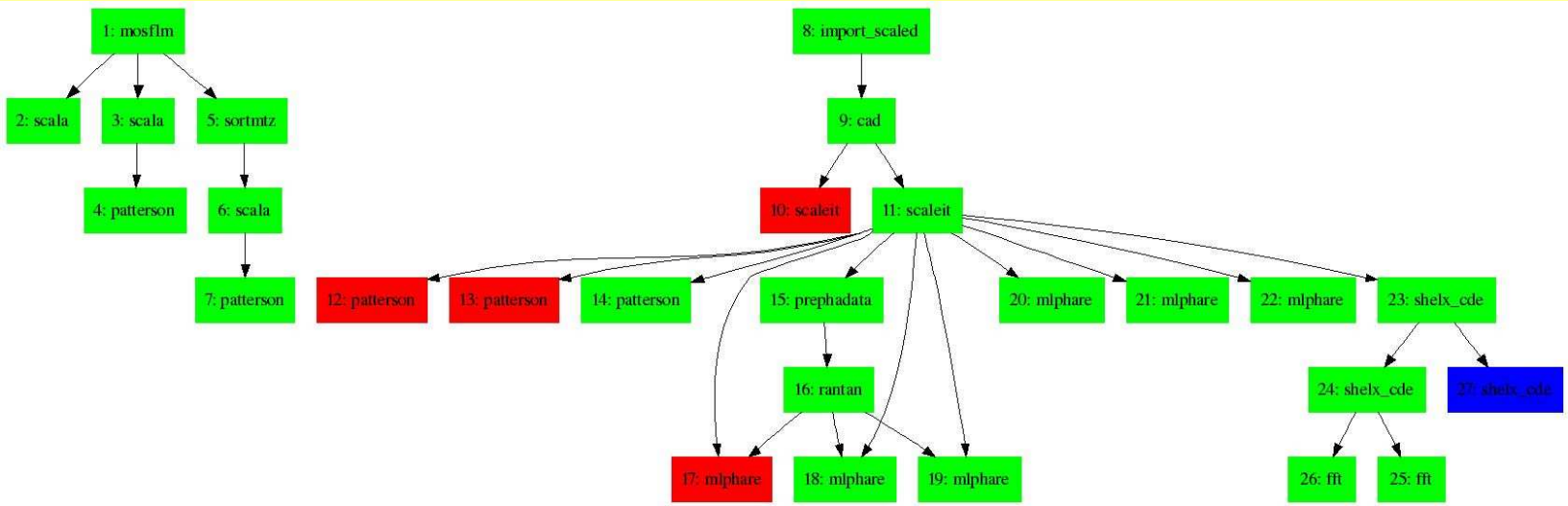
- "*To fill the need for project tracking within the BIOXHIT structure solution software pipeline"*
- Pipeline covers software components post-data processing (scaling and merging, phasing, model building, refinement)
- Complementary to PIMS and DNA

Staff for CCP4 effort at Daresbury:

- Peter Briggs
  - *project coordinator for CCP4*
- Wanjuan (Wendy) Yang
  - *full time programmer*

The need for data tracking

Manual structure determination can generate large quantities of data rapidly:



For automated systems this is likely to be even more of a problem

Aims

- Implement system for both manual and automated structure determination
  - *Use CCP4i as a starting point*
  - *Accommodate non-CCP4(i) applications*

- Implement multiple database backends
  - *Don't force user to have mySQL*

- Gather as much information as possible automatically
  - *using this system gives you tracking "for free"*

- Recognise that structure determination will most likely not be performed exclusively within a single software package and that data will most likely not be stored in a single database
  - *exchange of data between systems requires standards for transfer e.g. standards developed in BIOXHIT WP 5.1*

Project Deliverables

Project Database Handler

- broker application to mediate interactions between database and client applications
- hides implementation of backend
- aim to provide client APIs to handler from different languages
- deal with multiple users/clients within/outside CCP4 system

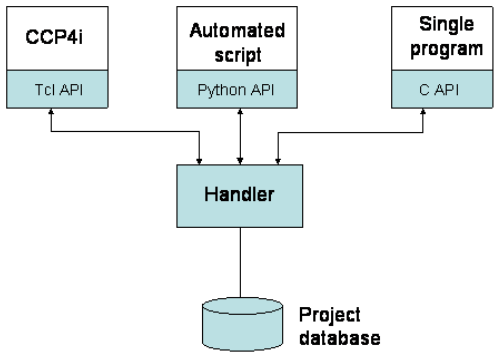
Database for Project Tracking

- project history (steps in the determination process)
- project data (“knowledge base”) & data history
- “operational” data
- aim to provide database schema and multiple implementations

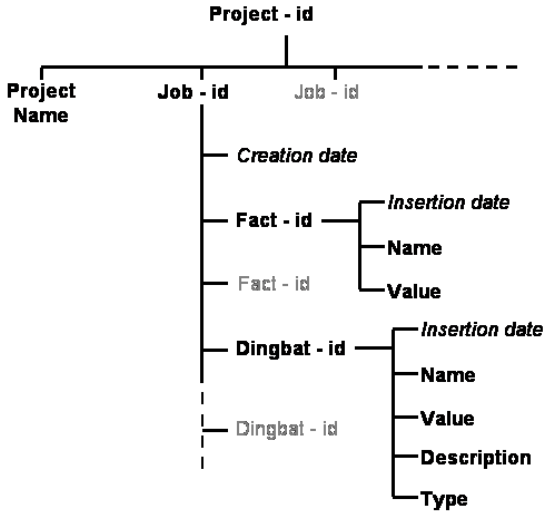
Visualisation Tools

- provide views of data to facilitate review and analysis

Prototype architecture and database



Prototype client-server architecture



Prototype basic database schema

Current Status/Work so far

- Currently prototyping to gather requirements
  - working with Graeme Winter
  - key criterion: *“needs”* rather than *“would like”*
- Wendy has produced prototypes for:
  - basic database schema implemented in mySQL
  - database handler and client API for Python
  - visualiser (“project data explorer”)
- Principal outcomes:
  - define requirements & build a list of issues to be resolved in full implementation

Project Data Explorer

Links to other projects

Current collaborators:

- **e-HTPX**: Graeme Winter & **XIA** (ongoing)
- **CCP4i**: obviously

Other relevant projects:

- **CCP4 Automation:** requirements still largely to be determined
- **PIMS**: requirement to able to exchange data
- Other projects interfacing with CCP4(i):
  - **CCP4MG** & **Coot**
  - **CRANK** (Steven Ness/Leiden)
  - **MOSFLM** (transferring data from processing)

Next steps

Immediate plans:

- Work with Graeme to use prototype in XIA code
- Prototype implementation of CCP4i-compatible backend
  - *integrate original handler prototype into CCP4i?*
- Develop prototype visualiser for project tracking in CCP4i
- Talk to other projects for possible links/requirements