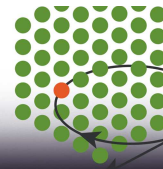




The BIOXHIT Project

www.bioxhit.org

Coordinated by EMBL Hamburg



BioXHit Data Management for PX Structure Determination

Aims of Workpackage 5.2: the need for data management

A key part of the integrated technology platform being delivered by the BioXHit project is the development of automated structure determination software pipelines, which join together computational units (programs or other applications which perform a single part of the process) to cover some or all of the stages from the data processing and reduction through to model building, refinement and model validation.

Within these pipelines it is essential to accurately record, organise and track the input and output data: the individual components need to access the required data on demand, and be able to store their outputs for use by other components downstream in the process. An accurate record of the process is also required when depositing the resulting structures in public databases such as that provided by the eMSD at the EBI.

Different applications may have very different needs and the situation is further complicated by the possibility that data will be stored in a number of different systems at geographically diverse locations (for example a LIMS, facility database or local data store). Workpackage 5.2 will address these requirements, by providing the tools outlined below and integrating them with both the computational units and the automated pipelines being developed within BioXHit.

BioXHit Partners in WP 5.2

Workpackage 5.2 is co-ordinated by **Partner 10**, the **Collaborative Computational Project No4 (CCP4)** based at the CCLRC Daresbury Laboratory in the UK. CCP4 provides a software suite for macromolecular structure determination by X-ray crystallography, which includes basic data management via its graphical user interface system CCP4i, and through technologies such as Data Harvesting. The Partner 10 contribution is led by Peter Briggs.



Partner 12 is the group of **George Sheldrick** at the **University of Goettingen**, Germany. Partner 12 has developed the SHELX suite of programs which are widely-used for crystal structure determination and which form a key component used in the automated pipelines being developed within BioXHit.



The Partner 12 contribution is led by George Sheldrick.

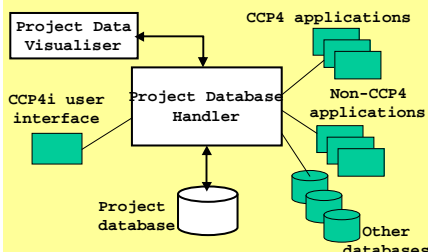
Partner 1C is the **Macromolecular Structure Database Group (eMSD)** which is part of the European Bioinformatics Institute (EBI) based at Hinxton in the UK. It is the flagship European public science institute in the field of bioinformatics. The eMSD was established to ensure that all aspects of 3D structure data are placed in the public domain and served to the scientific community. The Partner 1C contribution is led by Kim Henrick and Avi Naim.



Workpackage 5.2 will fill the need for project tracking within the BioXHit structure solution software pipeline, and consists of three components:

Project Database Handler

The Project Database Handler is a brokering application, which will mediate interactions between the project database and other applications and external databases (local or remote). It will act as a single point of access to the data for the applications that talk to it.



Above: schematic representation showing the interactions with the Project Database Handler

Database for Project & Data Tracking

A database will be designed and implemented which will be capable of storing project history information (the links between the steps performed travelling through a pipeline) and data history (the provenance and evolution of information as the project progresses).

Visualisation Tools

These tools will be interfaces to the database that provide display the project data in selective views, to focus on particular aspects of data-flow or logical flow – for example as work-flow diagram

Current Status of Workpackage 5.2

Partner 10 has so far been working on the development of a prototype of the Project Database Handler, based on the existing data management facility within CCP4i. Partner 12 has meanwhile worked on improving the SHELX information flow for tracking purposes, which includes ongoing work with Partner 10 to provide closer integration of the SHELX programs into CCP4 in order to streamline the data management.

The recruitment of a full-time programmer by Partner 10 will allow work on the Project Database Handler and the tracking database to move more quickly in the New Year. The prototype handler will be released to the BioXHit Partners, and the Partners will be consulted on the specific requirements for the design of the project tracking database. A workshop on data standards organised by Partners 1C and 10 to be held in February 2005 will also help to establish the data exchange protocols to be used in the database schema.

Structural Genomics and Proteomics Research:
first joint meeting of European projects
Barcelona, December 1-4, 2004

