

WorkPackage 5.2: Implementation of Data management and Project Tracking in Structure Solution

Peter Briggs, CCP4

Data Management and Project Tracking in Structure Solution

Aim of task 5.2.1:

- To fill the need for project tracking within the BIOXHIT structure solution software pipeline
- Pipeline covers software components post-data processing

Lead Partner CCP4-CCLRC (partner 10)

Contributions also from

- eMSD/EBI (partner 1C)
- ELETTRA (partner 5)
- University of Goettingen (partner 12)

Poster 50: BioxHIT Data Management for PX Structure Solution

Requirements for WP 5.2

- Diverse software components used within automated pipelines need to be able to store and retrieve data on demand
- Decision making, reviewing, deposition requires tracking of data
- Understanding of project status and history requires visualisation tools
- ***Structure determination will most likely not be performed exclusively within a single software package or at a single site***
 - exchange of data between systems requires standards for transfer
 - compliance with standards developed in WP 5.1



Starting point – CCP4i

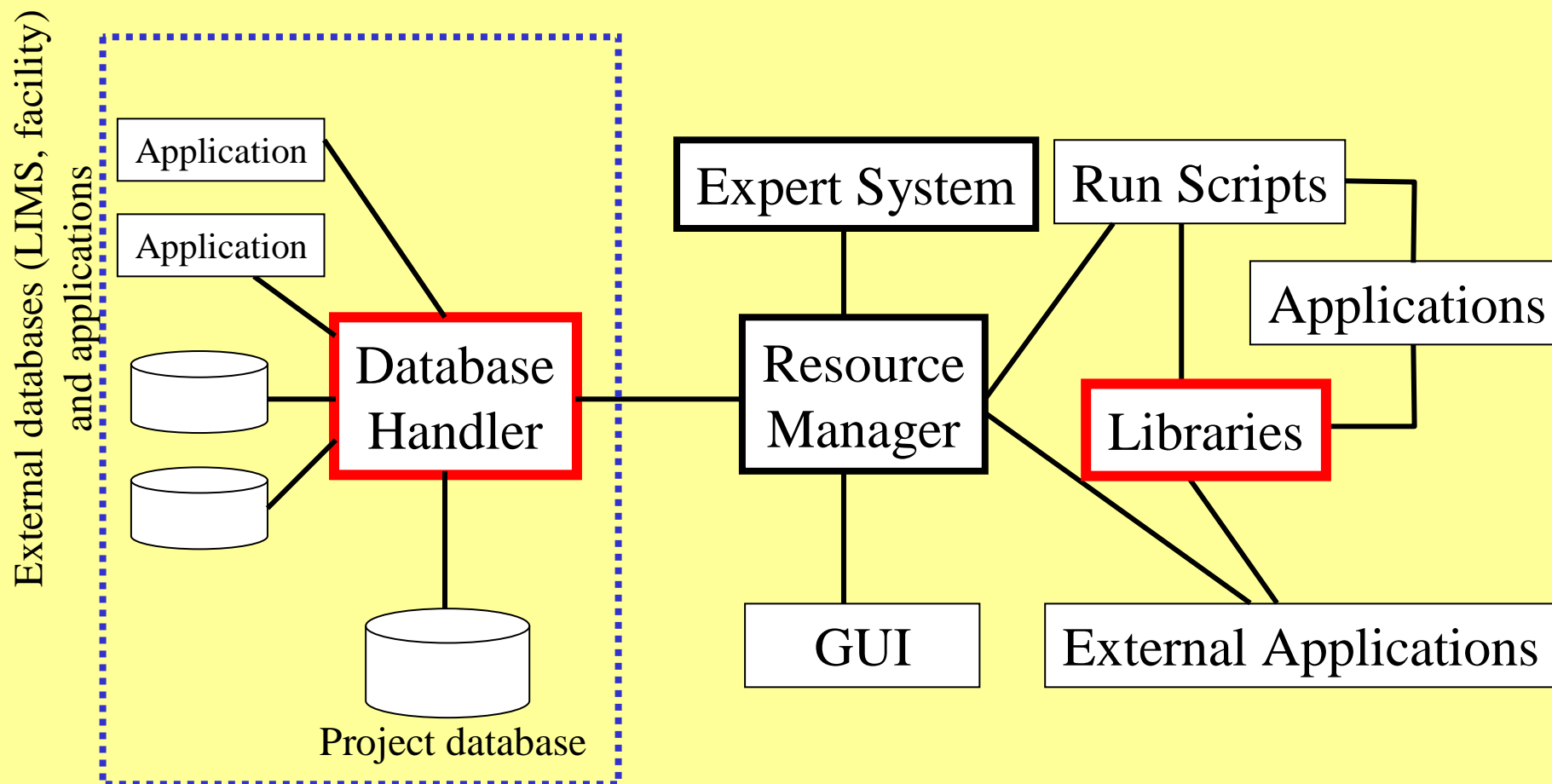
CCP4 currently offers a basic project database facility within the graphical user interface system CCP4i

- automatically records metadata about tasks run in a project
- offers file organisation and a rudimentary project history

This has limitations for automated pipelines e.g.:

- Only accessible from within the CCP4i system
- Tracking capability and scope of data is limited
- Flat-file database implementation scales poorly
- Cannot be accessed remotely or by multiple “users”

Towards an open architecture



Original plan for WP5.2

Project Database Handler

- mediate access to the database by arbitrary applications
- separate the existing CCP4i handler into independent server process
- prototype handler is a milestone
- handler is 18-month deliverable D 5.2.1

Database for Project Tracking

- project history (tracking of “logical steps”)
- project data and data history (“data provenance”)
- expand the scope of the existing CCP4i project database
- report on requirements for database system is 18-month deliverable D 5.2.2

Visualisation tools

- allow different views of data history, query functionality ...
- production of prototype visualisation tools is a milestone

Progress with project database handler

Progress was severely limited by delays in recruiting full-time programmer

- will be starting 6th December
- impacts on immediate future plans:
 - 18-month milestones & deliverables pushed back
 - task description & aims remain as written in first 18-month plan

Meanwhile ...

- Worked on prototype handler within the context of CCP4i
- useful for exploring implementation issues (such as communication protocols)
- October 2004: started looking at querying of the project history within the context of CCP4i
- planned standards workshop for February 05

Project handler: next 18 months

- Most immediate milestone: develop prototype handler with minimal database (developed independently of CCP4i)

Deliverables:

- Project handler application with well-defined communication protocols
- Version 1 of design for project tracking database
- Implementation of version 1 of project tracking database within handler
- *Incorporation of data exchange standards from WP 5.1*

Milestones:

- Develop prototype visualiser
- Integrate project handler into CCP4i

Other progress in WP5.2

Partner 12 (Goettingen):

- made improvements to SHELX information flow for tracking purposes
- on-going effort to integrate SHELX programs with CCP4i with Partner 10 to facilitate tracking

Joint workshop to be held 9th-11th February 05 will focus on standards within software/automated pipelines:

- **“Automation standards and frameworks: *from data reduction to structure*”**
- BioXHIT (CCP4 + EBI) + e-HTPX + CCP4 Automation Project
- links to Section 4
- will feed into data exchange standards in WP 5.1 to be used by handler to exchange data with other systems (*amongst other things*)



Workshop on “Automation Standards and Frameworks”

Joint BioXHIT/e-HTPX/CCP4 Automation Project workshop
Cambridge 9-11th February 2005

Organisers: Kim Henrick, Peter Briggs, Graeme Winter, Charles Ballard

Aims:

- Set standards to ensure compatibility between the various automation initiatives within BioXHIT, e-HTPX, CCP4 and others
- ***Standards for frameworks for automation***
 - BioXHIT task 5.1.1
- ***Standards for data exchange between computational units in the structure determination software pipeline***
 - BioXHIT task 5.1.2
- ***Toolboxes for Automation***
 - BioXHIT task 4.7.1