

Collaborative Mapping and Analysis

Tools for Biological Spatio-Temporal Databases

www.hgu.mrc.ac.uk

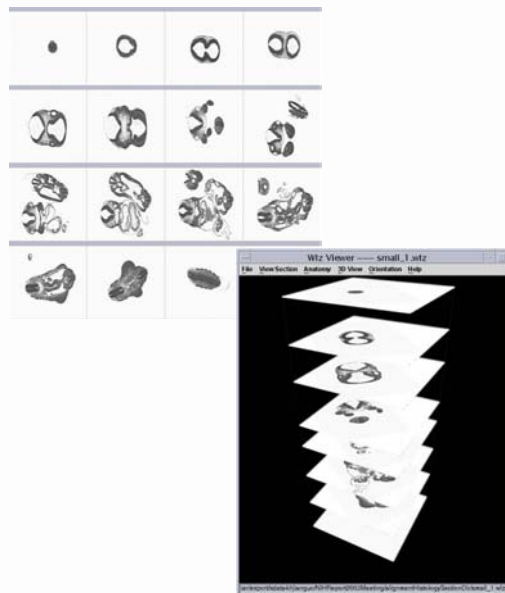
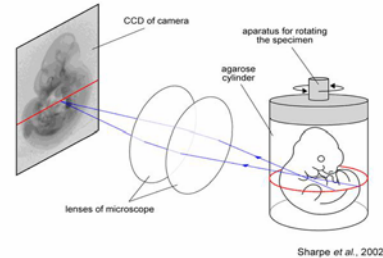
Introduction

e-Science and the GRID are new terms to describe an extension of the use of the Internet not in specific attribute but in scale. The key benefits of the GRID are transparency, security and access to large scale compute and database resources. This is achieved by using high-bandwidth networks to share resources and to access central provision of large scale systems for example the high-performance computing at Daresbury (HPCX). Most e-Science funded projects are aimed at developing the infrastructure or at large scale problems for which the advantages of the GRID are clear. In this project we are focusing on the potential small scale benefits. In particular the benefits for biomedical scientists "at the lab bench" and specifically in terms of the tools developed as part of the Mouse Atlas Programme at the MRC Human Genetics Unit.

The tools range from simple image processing for example 3D reconstruction of biological material through data mapping to collaborative working both on reconstruction and analysis of data. A particular application we will consider is the use of a secure GRID for collaborative discussion of gene-expression data submitted to the Mouse Atlas Gene-Expression database.

This project has been running for only 2 months so there are no results to report. In this poster we present the application area and the likely target problems that we will address.

3D Reconstruction



Biomedical Image Analysis

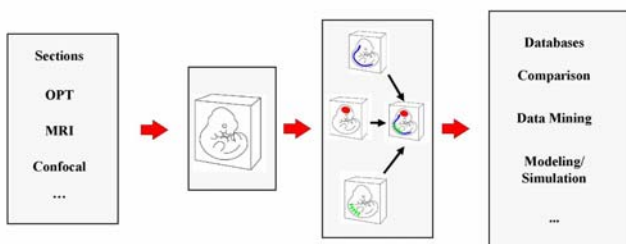
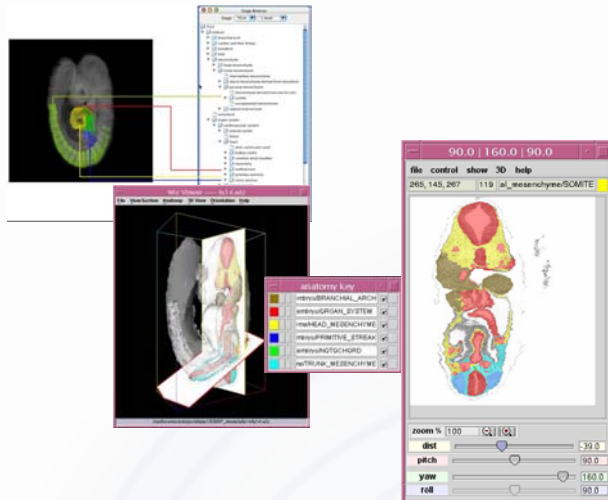


Image analysis in the context of development is typically to describe the observed patterns in terms of the recognizable structural features or to make a number of pattern feature measurements. In the advent of standard atlases there is now more emphasis on spatial mapping of the data with the possibility of much more sophisticated analysis, comparison and modeling. Some aspects of this process have been studied as part of the Mouse Atlas Program and where the GRID could be important is outlined below.

3D reconstruction is relatively routine in some laboratories but full reconstruction from high-resolution serial sections required complex and compute intensive calculation. With the development of Optical Projection Tomography (OPT) at the HGU this requirement will become widespread as experiments demand the geometric fidelity of OPT with the high-resolution and histology of microtome sections. GRID aspects:

- Remote HPC service
- Transparent high-speed data transfer
- Data security

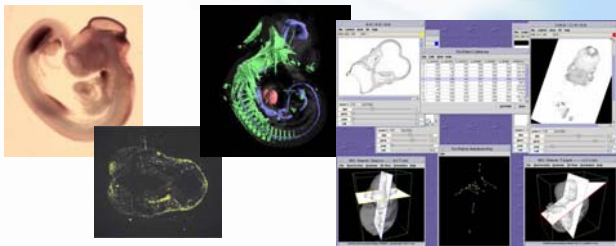
Atlases and Ontologies



The Edinburgh Mouse Atlas Project (**emap**) is an early example of a new generation of spatio-temporal frameworks for model organisms. The key components of the framework are grey-level voxel models of each developmental stage, an ontology of anatomical development and a mapping between the text ontology and the geometric space of the model embryos. These demand a high level of expertise and significant development time. For these community resources to survive we need high-quality tools for the community to use and contribute, particularly to the ontology mappings but also for inter- intra- and temporal transformations. GRID aspects:

- Collaborative image tools
- Complex transformation calculation - HPC & deformation modeling
- Database interoperability

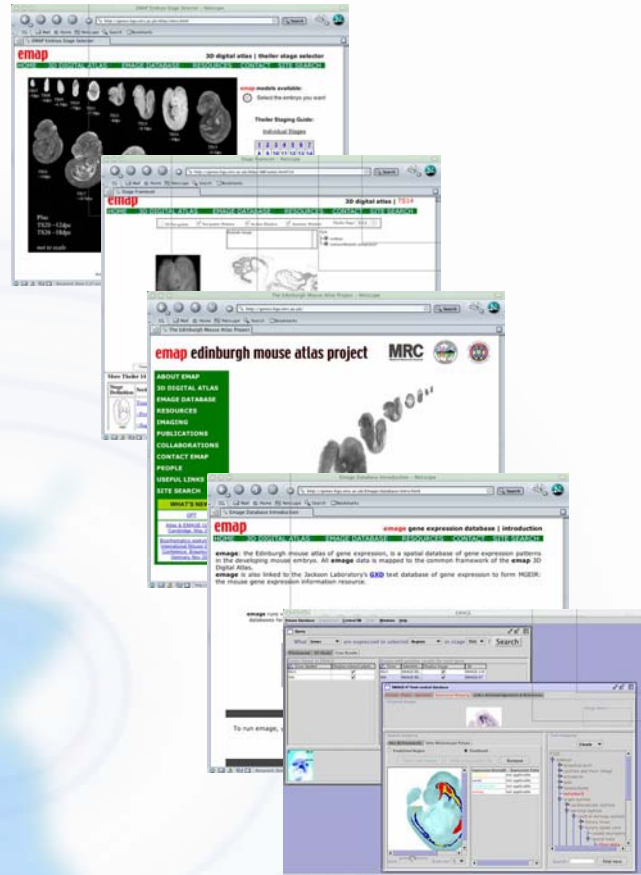
Data Mapping



In situ gene-expression data comes in many forms, 2D wholmount or sections, 3D volumes, single/multi channel etc. A common requirement is to be able to spatially map the data onto the standard framework or onto other images. In general this is a complex non-linear warp transformation established by manual interaction or automated processing. In either case the calculation is compute intensive and if interactive warping is required access to HPC is necessary. GRID aspects:

- HPC
- High-speed and secure data transfer

Spatio-Temporal Databases



Once the data is reconstructed and mapped it can then be submitted to the central database. The Mouse Atlas gene-expression database **emage**, is curated to the extent that an editorial group will check each submission for completeness and quality.

This process is made easier by validity checking within the submission interface but the key checks on the data interpretation and mapping can only be done by "expert-eye" and the editorial procedure may include an interchange between the submitter and editor. This needs to a secure discussive environment including image manipulation and mapping.

Query of the database does not need the discussive aspect by many users may desire data security and privacy. The key requirement for query is interoperability and query translation. For this type of data the translation may involve complex spatial and temporal transformations both of the query domain and the data return. GRID issues:

- Interactive discussive and secure environment
- Interoperability
- Data-mapping services