Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

JISC: Models of Sustainability Workshop

December 3rd 2007

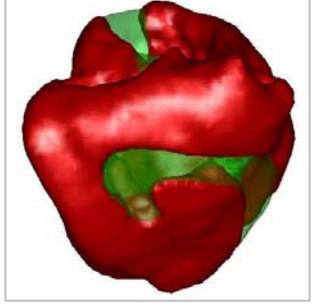
Integrative Biology: Experience and Sustainability

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU
IBM
UCL
CCLRC
EPSRC
The University of Nottingham
THE UNIVERSITY OF BIRMINGHAM
UNIVERSITY OF LEEDS
University of Sheffield

www.integrativebiology.ac.uk

Agenda

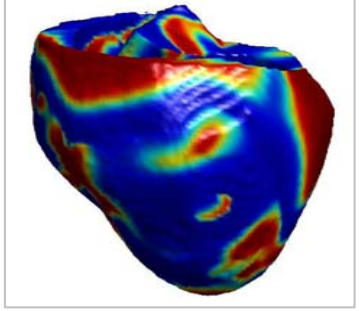
- What did we expect to build?
- [Who we built it for]
- [What were their needs]
- What did we end up building?
- Level of Infrastructure achieved (maturity etc).
- Is the infrastructure being used beyond the project?
- How is the infrastructure being maintained?
- Lessons learnt about sustainability



Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk



What we expected to build

•To build an Integrative Biology Grid to support applications scientists addressing the key post-genomic aim of determining biological function

•To use this Grid to begin to tackle the two chosen Grand Challenge problems: the *in-silico* modelling of heart failure and of cancer

•To leverage the global Grid infrastructure to build an international "collaboratory" which places the applications scientist "within" the Grid allowing fully integrated and collaborative use of:

- HPC resources (capacity and capability)
- Computational steering, performance control and visualisation
- Storage and data-mining of very large data sets
- Easy incorporation of experimental data
- User- and science-friendly access

=> Predictive *in-silico* models to guide experiment and, ultimately, design of novel drugs and treatment regimes

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU

IBM

UCL

CCLRC

EPSRC

The University of Nottingham

THE UNIVERSITY OF BIRMINGHAM

UNIVERSITY OF LEEDS

University of Sheffield

Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

Who we built it for (Our user groups)

Degree and/or post grad qualification in
Industrial engineering, maths, biology, physiology

Typically…

Computing skills developed over time to allow them to develop models (C, Fortran, Delphi…).

Not computer scientists. Not grid savvy. ***Competitive.***

Licensing/IPR issues

Keen to use and adapt other scientists work

'Married' to their favourite home grown technologies e.g. Meshalyser, Memfem, Coolgraphics…

Based in Oxford, Nottingham, Birmingham, Auckland, Tulane, Washington Lee, Calgary, Johns Hopkins, Sheffield, Utrecht, Graz…

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU

EPSRC

The University of Nottingham

THE UNIVERSITY OF BIRMINGHAM

University of Sheffield

IBM

UCL

CCLRC

UNIVERSITY OF LEEDS

What were their needs?

- Data management problematic – large datasets generated and tying information together an art. Need an easy to use solution for Data Management.

- Current simulations tie up desktops for many hours. Need better compute facilities without having to become a ‘grid’ expert. Requirement for parameter sweeps common.

- Visualising results on desktop limited by local facilities and *ad hoc* development of suitable tools

- Laptop to HPC migration for most users a huge leap not a small step. Need to consider how to educate/support.

- Cannot exclude scientific community who have not progressed to computational models. Need to consider innovative ways of enabling collaboration.

What have we ended up building?

a) IB Interface (IBI) and underlying core services:

- Job submission
- Job monitoring
- Workflow
- Data management
- Advanced visualization for data analysis
- Computational steering

b) VRE prototype for Memfem heart modelling to support graduate life scientists

c) Portal Interface for job submission and management

d) Advanced Visualisation Techniques

e) CHASTE – Cancer, heart and soft tissue environment

f) Image analysis techniques e.g cancer cells on pathology slides, high resolution MRI and histology data for the heart

Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

Interfaces to services and utilities (IBI)

IBI is a portable, cross platform dedicated IB Interface (IBI) through which services are realised.
The Interface is based on FLTK, it is modular, has a WYSWYG builder, so you can add more panels to build your applications on IBI.
It supports OpenGL-based graphics for visualization of geometry and images.
It is C++ based.

Command Line .. Still used by some

API enables more proficient users to write own interfaces/applications

Portal for users who want lightweight flexible access based on uPortal e.g. check status of a simulation at a conference

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU
IBM
UCL
CCLRC
EPSRC
The University of Nottingham
THE UNIVERSITY OF BIRMINGHAM
UNIVERSITY OF LEEDS
University of Sheffield

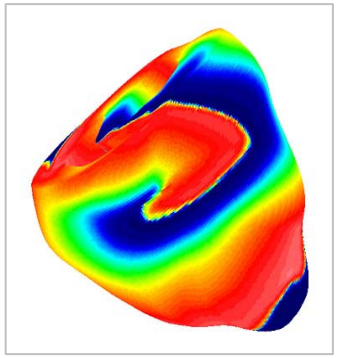
Integrative Biology

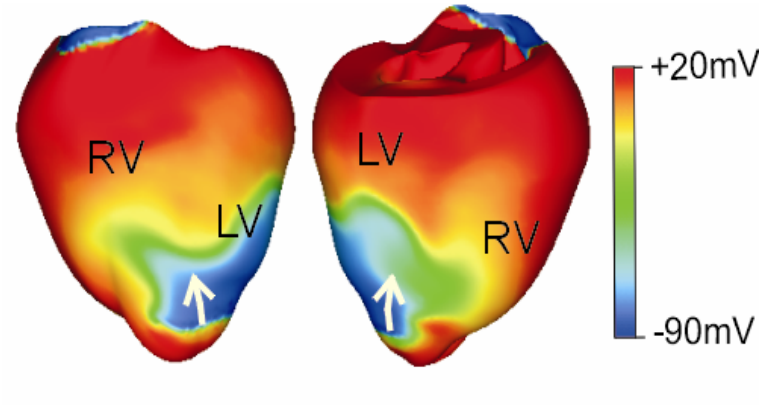
- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

IBI: Managing Simulations on the Grid

- Project utilises NGS and HPCx as well as local compute facilities.
- Utilises Web Service interface
- Uses Grid Security with myProxy for certificate management





-The job monitoring uses Hibernate to provide persistent objects which hold information about the jobs of the user.

- This allows the user to track the status even if the compute cluster or the web services server are crashed and rebooted.

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU
IBM
UCL
CCLRC
EPSRC
The University of Nottingham
THE UNIVERSITY OF BIRMINGHAM
UNIVERSITY OF LEEDS
University of Sheffield

IBI: Workflow

Utilise IBRun script that takes as input a string that contains all the necessary information such as the

- input, data file and any other input e.g. mesh file
- output directories on SRB
- executable name
- the number of processors
- the clock time
- the compute cluster that the user wants the job to be run.

It creates and manages a simple workflow to get the input data and executable, execute the job, return the job handle to the job monitoring service and at the end of the job marshal the output to SRB and/or the visualization toolkit

Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

IBI: Data Management

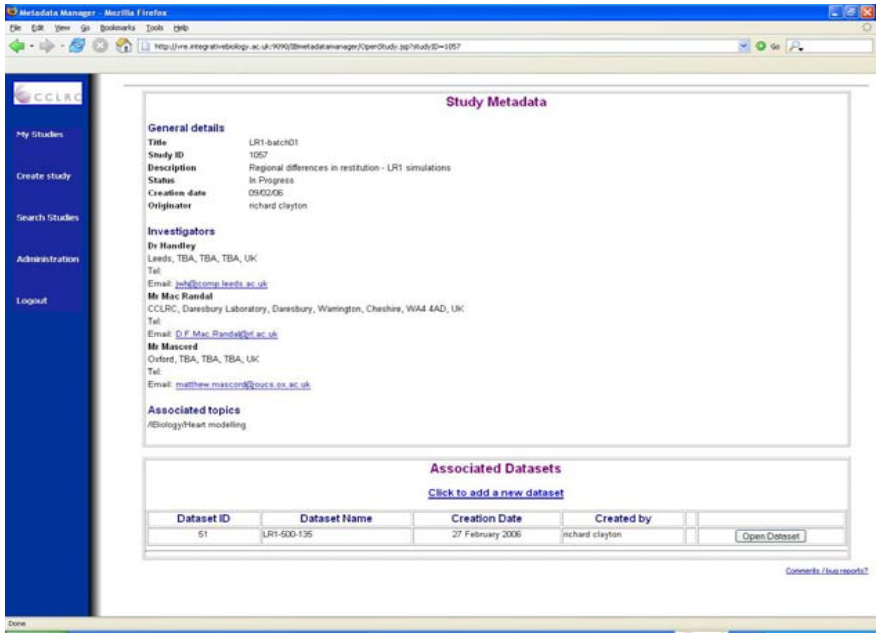
Utilises SRB on NGS to allow users to manage their data

The IBI allows the user to browse and manipulate his/her files and data as an extension of their local file store.

The job submission service supports the creation and ingestion of metadata based on the CCLRC metadata schema which is adopted as de facto standard across various scientific disciplines.

- 150 jobs/project
- 250 output files/job
- 4Mb/output file

⇒ **150Gb/project**



Metadata editor allows ‘experiments’ to be setup and metadata stored against simulation runs

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU
EPSRC
The University of Nottingham
THE UNIVERSITY OF BIRMINGHAM
IBM
UCL
CCLRC
UNIVERSITY OF LEEDS
University of Sheffield

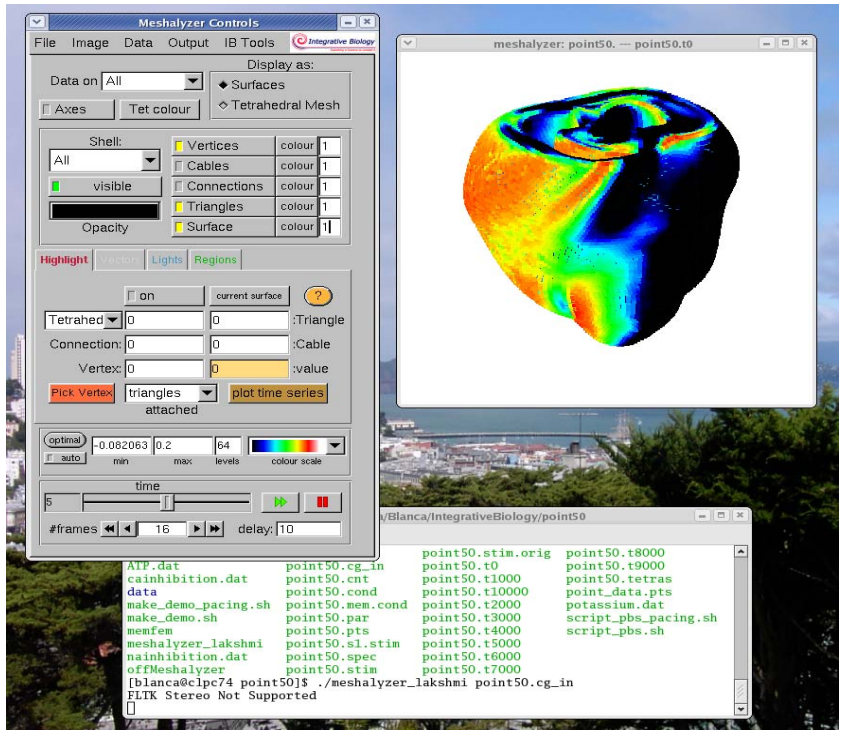
Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

IBI: Visualising Results

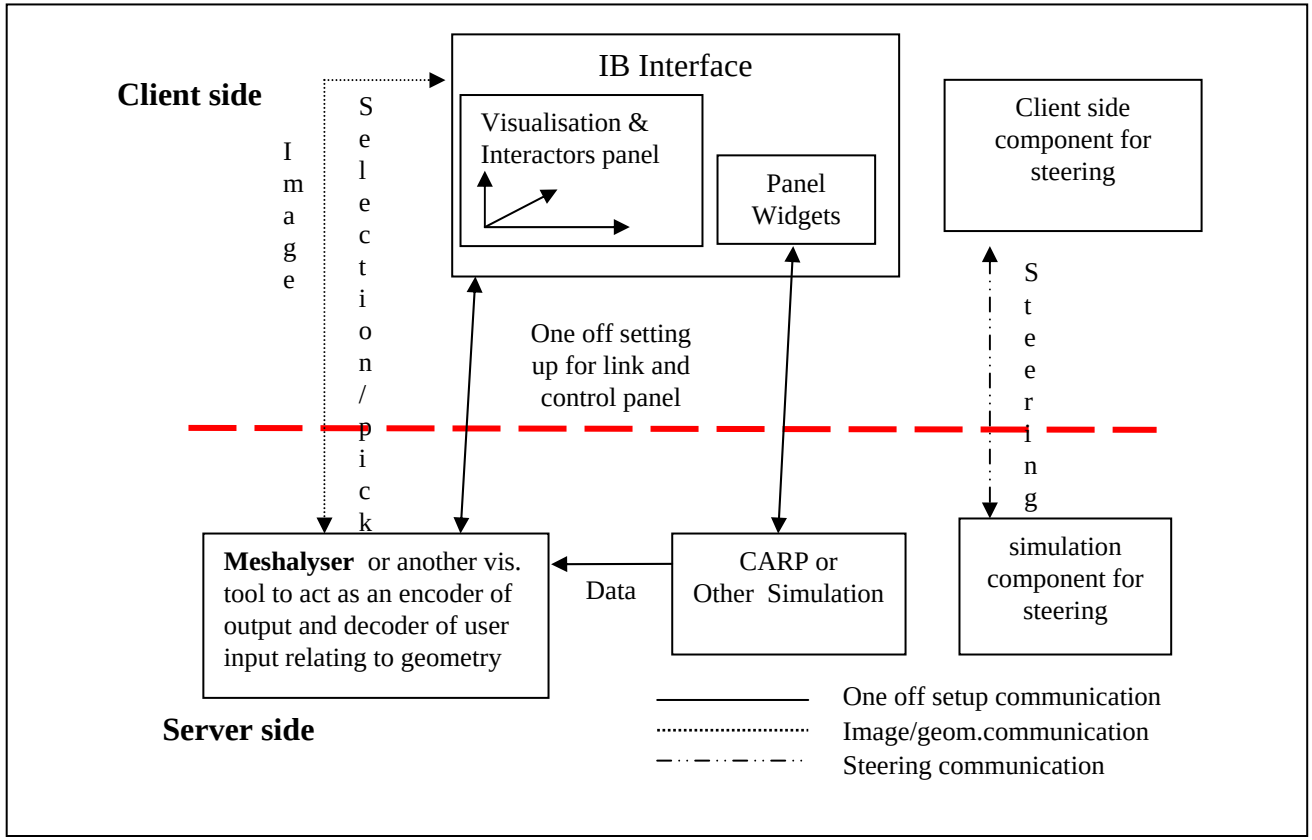
- IB visualisation services process the data closer to where it is generated or stored and create images on the fly on the server and render it to the user's desktop.
- Use both the dedicated CCLRC visualization cluster as well as standard compute clusters to generate high resolution images and visualization on the fly.
- Implemented as an example with 'Meshalyser'



Additional services for server side camera orientation-dependent animation, MPEG movie generation, as well as extension of existing functionality of tools such as CoolGraphics to include cutting planes and C++ classes for image processing

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU
IBM
UCL
CCLRC
EPSRC
The University of Nottingham
THE UNIVERSITY OF BIRMINGHAM
UNIVERSITY OF LEEDS
University of Sheffield

Visualisation Architecture

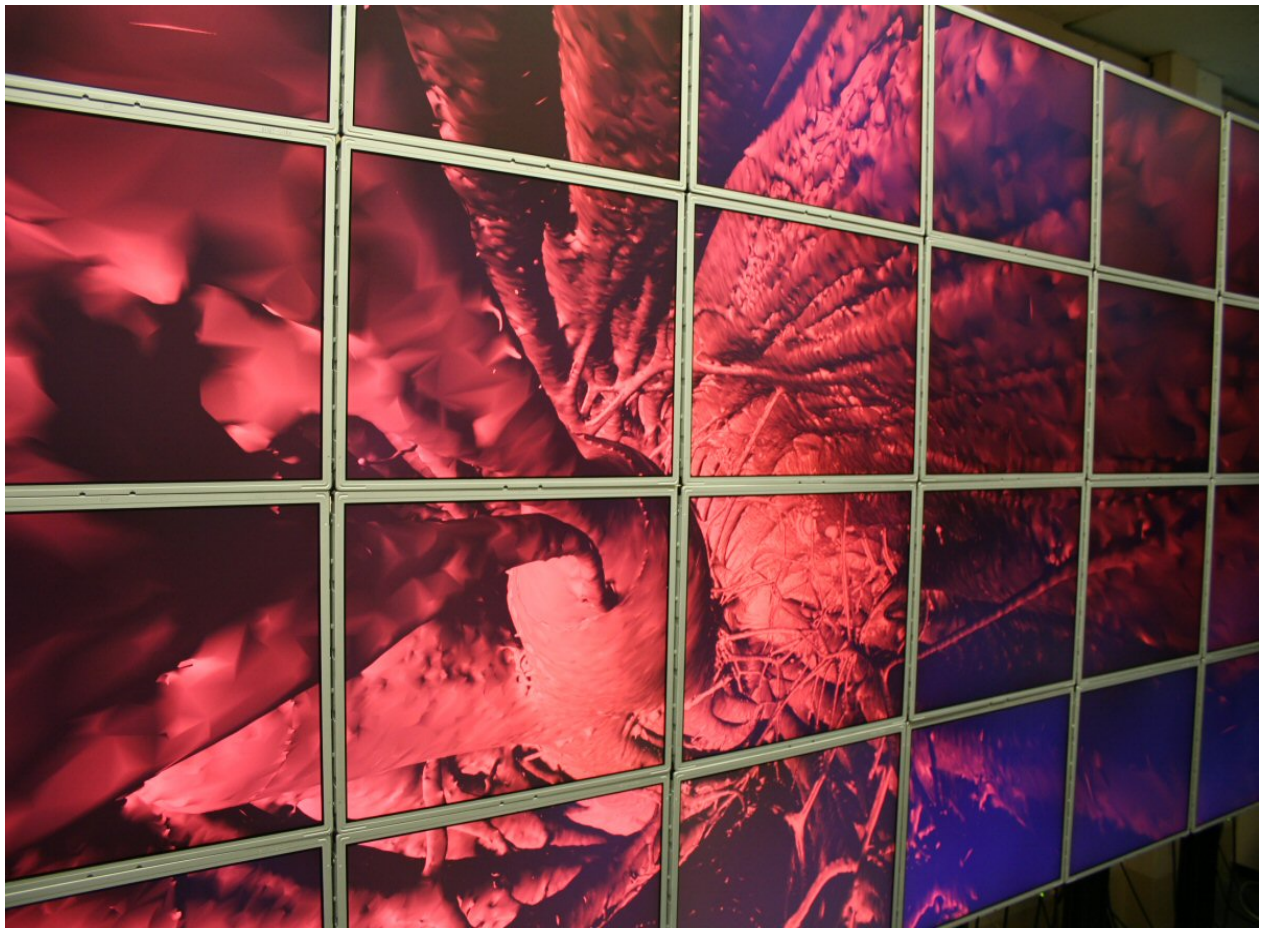


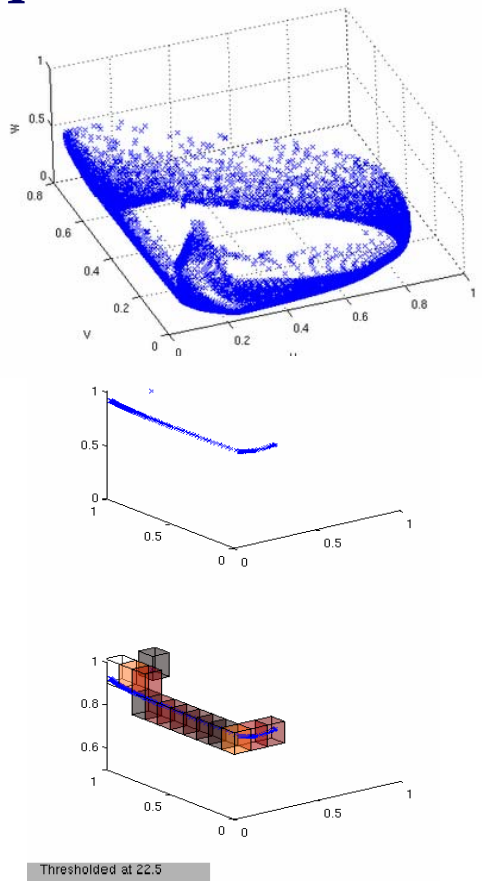
Integrative Biology

- exploiting e-science to combat fatal diseases

www.integrativebiology.ac.uk

Advanced Visualisation Techniques





The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU

EPSRC

The University of Nottingham

THE UNIVERSITY OF BIRMINGHAM

University of Sheffield

IBM

UCL

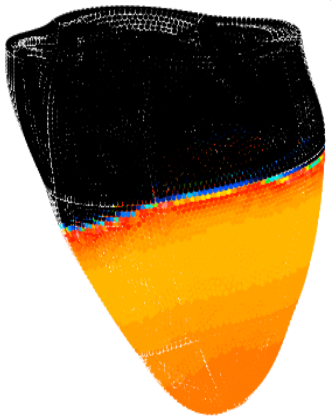
CCLRC

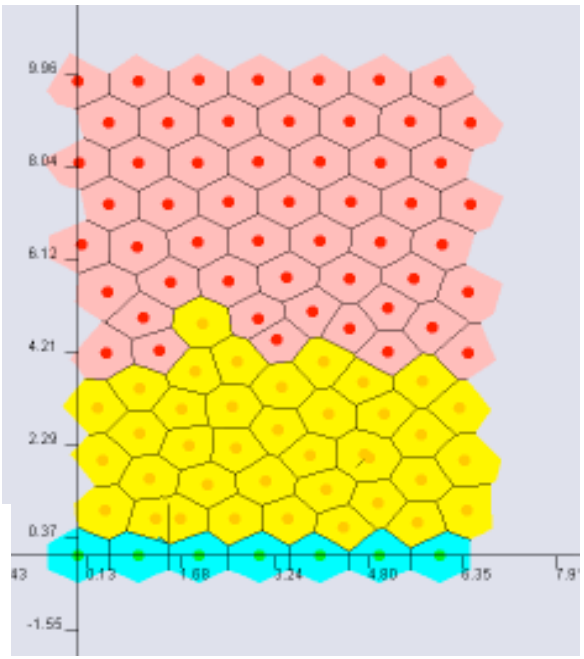
UNIVERSITY OF LEEDS

Integrative Biology
- exploiting e-science to combat fatal diseases
www.integrativebiology.ac.uk

CHASTE – Cancer, Heart and Soft Tissue Environment

Current features:
- Bidomain simulation of electrical activity in the heart
- Meineke crypt simulation of colorectal cancer
- Object orientation for extensibility
- Exploitation of parallel hardware (MPI)





Architecture:
C++, MPI, PETSc, Boost, deal.ii

Teaching environment:
We teach people how to develop software using our tools, processes and software libraries, by pair-programming with them

The University of Auckland
TE WHARE WANANGA O TAMAKI MAKAURAU
IBM
UCL
CCLRC
EPSRC
The University of Nottingham
THE UNIVERSITY OF BIRMINGHAM
UNIVERSITY OF LEEDS
University of Sheffield

Use of infrastructure beyond the project?

- Initial project consortium in heart modelling involved only Oxford and Auckland, whilst that in cancer involved Nottingham, Oxford and Sheffield.
- Consortium now extended to include over 20 HEIs, research institutes and industrial partners in 10 countries (UK, Holland, Austria, Ireland, Germany, US, Australia, Canada, New Zealand and Japan).
- Includes both modelling and experimental groups.
- Have provided access to UK NGS and HPCx to these groups.
- Well over 100 publications in journals and conference proceedings.
- IB (we think) is the largest biological user of HPCx facilities (received large top-up in October to allow us to complete projects).

How is it being sustained?

- Project ends at the end of Feb 08, so we have been addressing this issue over last 6 months
- In negotiation with STFC over details of Open Source release of IBI
- In final stages of negotiation with project partners over Open Source release of Chaste
- IBVRE middleware already released as Open Source
- No direct follow-on project (as no obvious avenue for support) but several applications building on IB have already been submitted including:
  - FP7 VPH NoE – Oxford leading WP3 which is developing the IT infrastructure and Sharon will be Technical Project Manager if funded
  - FP7 PREDICT (VPH Strep in collaboration with leading Pharma)
  - STFC (Lakshmi Sastry) applying to OMII to develop robust version of IBI
  - EPSRC e-Science Network application just submitted to maintain the consortium
  - EPSRC HPC Software Development funding received for Chaste with string input from Fujitsu
  - EPSRC Hector funding application submitted and second to follow
  - EPSRC Platform Grant application in preparation to extend colorectal cancer research

Lessons learnt about sustainability?

• It is difficult given the current funding model!

• Academia is not geared to developing robust software and middleware (so doing something about this will be very expensive).

• Funding agencies have not generally (until very recently) funded software development itself, and even more so have not funded the additional development required to turn research software into robust production level software

• Career progression/reward for those developing software is difficult (RAE etc), and software engineering training in academia is pretty-much non-existent

• Staff recruitment/retention (developers, project managers etc) difficult given academic pay structures

• Take home: sustainability is not even on the radar screen of most (all) HEIs, and most academics (e.g. response to infrastructure at ESF Workshop)