The National e-Science Centre: Role and Activities

Dave Berry, Research Manager

Indo-UK Workshop on e-Science
Delhi, February 2004
Outline: The National e-Science Centre

- The National e-Science Centre
  - Role and mission
  - The e-Science Institute
- NeSC projects
NeSC Roles

- Help coordinate and lead the UK e-Science Programme
  - Community building activities, regional support & outreach
  - Skill building through training events & support centre

- Help establish the UK’s international role
  - International meetings, standardisation work & presentations

- Run the e-Science Institute
  - Knowledge building through workshops and conferences
  - Research visitors and events

- Undertake R&D projects
  - To deliver reliable middleware
  - To engage industry
  - To stimulate the uptake of e-Science technology and methods
  - Emphasis on scientific data management
e-Science Institute

- A meeting place
  - The focus for presenting UK e-Science
- Visiting researchers
  - Collaborate in our research and development
  - Engage in and develop our event programme
  - Build bridges with their community
  - Visits last between one week and six months
- Research-oriented event programme
  - e-Science research topics
  - Training to e-Science research teams
eSI Events held in the 2nd Year
(from 1 Aug 2002 to 31 Jul 2003)

We have had 86 events:

- 11 project meetings
- 11 research meetings
- 25 workshops
- 2 “summer” schools
- 15 training sessions
- 12 outreach events
- 5 international meetings
- 5 e-Science management meetings

(though the definitions are fuzzy!)
eSI Workshops

- Space for real work
  - Crossing communities
  - Creativity: new strategies and solutions
  - Written reports
- Scientific Data Mining, Integration
- Grid Information Systems
- Portals and Portlets
- Virtual Observatory as a Data Grid
- Imaging, Medical Analysis and Grid Environments
- Open Issues in Grid Scheduling
- Data Provenance & Annotation
- e-Science Workflow Services
- GeoSciences & Scottish Bioinformatics Forum

http://www.nesc.ac.uk/events/

Suggestions always welcome!
eSI Industrial Involvement

- 194 registered users from 97 companies
- 133 different delegates have attended events (169 event registrations) from 64 companies including not only
  - IBM, Microsoft, Oracle, Sun, Hewlett-Packard
  but also ...
  - Apple, Astra Zeneca, BAE, Cisco, Honeywell, Motorola, Organon, Pfizer, Siemens
Research visitors

- Establish a collaboration with NeSC
  - Pre-established mutual interests
- We encourage diversity of disciplines and interests
  - Complementary experience, knowledge and skills
  - We can help match interests and develop a plan
- Visitors must already be engaged in relevant R&D
  - This is not a training opportunity
- Our support depends on the length and value of visit
  - Typically covers travel and/or local living costs
  - Application via our web site
eSI in Edinburgh
Outline: NeSC Projects

- The National e-Science Centre
  - Role and mission
  - The e-Science Institute
- NeSC Projects
e-Science Projects

- OGSA-DAI/DAIT, MS.NETGrid, SunDCG, GridWeaver, BRIDGES, PGGrid, FirstDIG, ODD-Genes
- EGEE, NextGrid
- OGSA Test Grid, IBM Early Evaluation
- edikt
- Digital Curation Centre, Publishing Scientific Data
- GridPP, AstroGrid, QCDGrid, RealityGrid Portal
- Biological Spatio-Temporal Databases
- CoAKTinG, Grid-enabled Modelling Tools and Databases for Neuroinformatics, e-Diamond
- TIES, TIES-II
Digital Curation Centre

communities of practice: users

Collaborative Associates Network of Data Organisations

Industry

curation organisations

research collaborators

standards bodies

community support & outreach

management & co-ordination

research

devolution

testbeds & tools

services
The Virtual Observatory

• International Virtual Observatory Alliance

UK, Australia, EU, China, Canada, Italy, Germany, Japan, Korea, US, Russia, France, India

How to integrate many multi-TB collections of heterogeneous data distributed globally?

Sociological and technological challenges to be met
Data Services

- **GGF Data Access and Integration Svcs (DAIS)**
  - OGSI-compliant interfaces to access relational and XML databases
  - Needs to be generalized to encompass other data sources
- **Generalized DAIS becomes the foundation for:**
  - Replication: Data located in multiple locations
  - Federation: Composition of multiple sources
  - Provenance: How was data generated?
1a. Request to Registry for sources of data about “x”

1b. Registry responds with Factory handle

2a. Request to Factory for access to database

2b. Factory creates GridDataService to manage access

2c. Factory returns handle of GDS to client

3a. Client queries GDS with XPath, SQL, etc

3b. GDS interacts with database

3c. Results of query returned to client as XML
edikt

Standards

Requirements analysis

Technology matchmaking

E-Science Apps

Gap filling

Rigorous engineering

Grid Services for e-Science Data Management

Commercial SW components and skills

CS Research

Edikt project

- The team: 8 professional software engineers, support staff, project manager, commercialisation manager, architect, and SAB
- SHEFC funded research and development grant
  - 3 years funding: May 2002 – 2005
  - +3 years funding upon successful project and review
**ELDAS – Data Access Service**

- ELDAS runs anywhere
- Suitable for grid & web

- Implemented using Enterprise Java Beans
- Data Access Components interface to distinct DBMSs
- Accessible as a grid data service or a web data service
BinX – accessing legacy binary data

• The Problem:
  • Many binary data files
  • Applications must “know” the data format
  • Binary data formats are machine-specific

• The Solution:
  • Write a “stand-alone” format description in XML
  • Provide a library to
    ▶ Interpret the description
    ▶ Provide file access across different machines
  • Build higher-level services
Problems specific to Bio-Informatics Community

- DBs growing exponentially!!!
  - Biobliographic (MedLine, ...)
  - Amino Acid Seq (SWISS-PROT, ...)
  - 3D Molecular Structure (PDB, ...)
  - Nucleotide Seq (GenBank, EMBL, ...)
  - Biochemical Pathways (KEGG, WIT...)
  - Molecular Classifications (SCOP, CATH,...)
  - Motif Libraries (PROSITE, Blocks, ...)
More genomes ......

Yersinia pestis
Arabidopsis thaliana
Buchnerasp. APS
Aquifex aeolicus
Archaeoglobus fulgidus
Borreli urgorferi
Mycobacterium tuberculosis

Caenorhabitis elegans
Campylobacter jejuni
Chlamydia pneumoniae
Vibrio cholerae
Drosophila melanogaster
Escherichia coli
Thermoplasma acidophilum

Helicobacter pylori
Mycobacterium leprae
mouse
Neisseria meningitidis Z2491
Plasmodium falciparum
Pseudomonas aeruginosa
Ureaplasma urealyticum

rat
Rickettsia prowazekii
Saccharomyce cerevisiae
Salmonella enterica
Bacillus subtilis
Thermotoga maritima
Xylella fastidiosa
System Usage Scenario

Browser based clients...

Client Site X

Secure access for CFG VO

Shared/Private Data Sets

BRIDGES Portal

Personalised Services

BLAST

Smith W

SV

Push relevant data to Secure Data Repository

Authorisation

Per user, per site

Secure Data Repository

Remote data in Oracle, DB2, Sybase, Excel, flat files, XML...

BRIDGES Portal

Generic services used by other projects

BLAST

Smith W

SV

Generic services used by other projects

Secure Data Repository

Secure Data Repository

DL

Wrappers

Remote data in Oracle, DB2, Sybase, Excel, flat files, XML...

Epcc

e-Science
dri

University of Glasgow
Website

- **National e-Science Centre**  http://www.nesc.ac.uk/
  - Mission, Background, Foundation
  - Locations, Staff, Resources, Projects
  - Register interest, Mailing lists, NeSCForge
  - Regional associations and Collaborations
  - News, Notices
  - Presentations and Lectures  http://www.nesc.ac.uk/presentations/
- **e-Science Institute**  http://www.nesc.ac.uk/esi/
  - Mission, Events (Future and Past)
  - Register for Events, Visitor Programme
- **UK e-Science**
  - Map and Index of Centres  http://www.nesc.ac.uk/centres/
  - Technical Papers  http://www.nesc.ac.uk/technical_papers/
  - Index of >100 Projects  http://www.nesc.ac.uk/projects/
  - Task Forces  http://www.nesc.ac.uk/teams/
- **General Information**
  - Glossary, Bibliography,
  - Who’s who
  - E-Science job vacancies
Questions?

daveb@nesc.ac.uk