E-Science: Foundations for the European Citizen

Grid2006
Barcelona

Malcolm Atkinson
Director e-Science Institute
& e-Science Envoy

www.nesc.ac.uk
28th September 2006
Overview

- Celebrate Five Years of Success in UK
  - Three Great Strengths Established
- EGEE et al. thriving in Europe
  - e-Science & Cyberinfrastructure everywhere
- Welcome New Projects UK & EU
  - Doing new research
  - A few expert enthusiasts benefit
- How do we establish e-Science
  - Reaching the tail
  - Shape e-Science & e-Infrastructure
    - For e-Scientists in all disciplines
    - For Europe’s Citizens
Defining e-Science

- **e-Science: Systematic Support for Collaborative Research**
  - Multi-disciplinary, Multi-Site & Multi-National
  - All disciplines contribute & benefit
  - Enabling wider engagement
  - Building with and demanding advances in Computing Science
The UK as an Example
UK e-Science Success

- Thriving Community
  - All disciplines & all Research Councils
  - Industry & Academia
  - Many universities & research institutes
  - UK e-Science All Hands Meetings
  - Productive collaboration
Essential Collaboration

- Collaboration
  - Requires Commitment and Strategy
  - A challenge to build and maintain
  - We have done it repeatedly
  - Can we capture and clone the recipes?
  - Can we support it well for all research?
  - For all citizens?

- New Patterns of Communication

But not always or easily
Archaeology & e-Science

- Professor Michael Fulford
  - Archaeology, University of Reading
- Silchester Roman Town: the challenges, aspirations and experience of developing a VRE for Archaeology
Enthusiastic researchers, Many interacting communities & Breadth
UK e-Science Success

- Significant outputs from projects
  - Research results
  - Commercial impact
  - Outreach and international influence
The NERC Success

- Professor Robert Gurney
  - Director, Environmental Systems Science Centre, Reading

- The NERC e-Science experience
  - 11 papers in Nature
  - Enthusiastic uptake of ensemble methods
Predicting Climate Change
Through Volunteer Computing

University of Oxford
Department of Atmospheric Physics

6th September 2006

Slide: Robert Gurney
climateprediction.net Users Worldwide

>300,000 users total (90% MS Windows): >60,000 active
~17 million model-years simulated (as of September '06)
~180,000 completed simulations

The world's largest climate modelling supercomputer!
(NB: a black dot is one or more computers running climateprediction.net)

Impact:
New Science
Understanding of science
Engaging schools
BBC follow on

Slide: Robert Gurney
Construct *in silico* experiments, find and adapt others, manage the experiment lifecycle

- Taverna Workflow workbench
- OGSA-DQP
- Semantic Technologies
- Williams-Beuren Syndrome, Grave’s Disease, Trypanosomiasis in cattle.

- OMII-UK Node, GRIMOIRE Registry, Taverna Workflow workbench
- 12000+ Downloads of Taverna
- Wide transfer to BBSRC (e-Fungi, ISPIDER, ComparaGrid) & MRC projects (PsyGrid, CLEF, CLEFS)
- Semantic Grid pioneer
- WBS gene identification
- Outstanding international links
- Great deal of open source s/w
- Links into BOSC & HGMP
- KT to BT, ComparaGrid, OntoGrid, BBSRC Systems Biology Centre, MIASGrid, Rice Institute etc

Middleware for data intensive *in silico* biology by bioinformaticians

- Carole Goble (Comp Sci, Manchester)
- 7 Universities and institutes (incl. EBI)
- 8 Companies

Slide: Carole Goble & Jim Fleming
Design, develop and implement an advanced infrastructure to support real-time processing, interpretation, integration, visualization and mining of vast amounts of time critical data generated by high throughput devices.

Data mining, text mining

Environmental monitoring, bioinformatics

InforSense, GSK, Oracle

2003: Discovery Net in Action: Fighting SARS in China

2002: Supercomputing 2002 Most Innovative Data Intensive Application Award

2002: KDD CUP 2002 Scientific Text Mining Awards

Yike Guo (Comp Sci, Imperial)

1 university
Workflow Warehousing and Semantic Authoring

Impact
8 out of 10 top Pharmas Co’s are trying it

Slide: Yike Guo
Compute resources on-demand
http://www.realitygrid.org/

Biomolecular chemistry simulation, Computational steerage, Visualization and computation use via supercomputers accessed via Grid.
> Three dimensional Lattice-Boltzmann simulations.
> Closely coupling computation and experiment to speed up scientific discovery.
> Data produced during a single large scale simulation can exceed hundreds of gigabytes to terabytes

- Peter Coveney (Chemistry, UCL)
- 3 Universities and institutes
- 6 Companies

Slide: Carole Goble & Jim Fleming
Aims to manage >1Tb per year of Aero Engine vibration and maintenance data.
Interlinks with search and reasoning services.
Defined and evaluated a distributed search system.
GSI enabled secure engine performance simulation
CBR advisor for diagnostic engineer
A data architecture defined based on Globus and SRB.

BROADEN DTI Project (£3.9M)
Spun out technology exploited through Cybula Ltd., Oxford Biosignals and DS&S.
Successful mid-term demonstrator well received by Rolls Royce
White Rose Grid: experience of building & using production Grids
In Grid Blue Print 2 edition 2

Impact
RR success
B787?

- Jim Austin (Comp Sci, York)
- 4 Universities and institutes
- 3 Companies
Link the X-ray and Opt/IR to understand energetic galaxies at the edge of the Universe

Un**locking new science**

Federating multi-\(\lambda\) data: Deep Field Surveys

NEP J1716.6+6708: an X-ray cluster at \(z=0.81\):
Chandra X-ray image (C. Mullis) overlaid on a deep BRI image (D. Clowe & G. Luppino).
Biomedical Research Informatics Delivered by Grid Enabled Services

Portal

Synteny Grid Service

Publicly Curated Data
- Ensembl
- OMIM
- SWISS-PROT
- MGI
- HUGO
- RGD

Data Hub

Private data
- Glasgow
- Edinburgh
- Oxford
- London
- Netherlands
- Leicester

Private data

http://www.brc.dcs.gla.ac.uk/projects/bridges/

Slide: Richard Sinnott
Healthcare @ Home

GP
Home-mobile-clinic
via PDA-laptop-PC-Paper

Diabetician
Home-mobile-clinic
via PDA-laptop-PC-Paper

Various Clinical Specialists (Distributed)
e.g. Ophthalmologist, Podiatrist, Vascular Surgeons, Renal Specialists, Wound clinic, Foot care clinic, Neurologists, Cardiologists

Patient
Home-mobile-clinic
via TV-PDA-laptop-PC-Paper

Dietitian

Biochemist

Community Nurses / Health Visitors

REFERRAL

ILLNESS

VARIABLES ACCESS MATRIX
IXI normal aging cohort

Available from Spring 2006
www.ixi.org.uk

- Researchers often have difficulty getting access to large cohorts of high quality data
- The IXI project is making high quality MRI scans of 100s of normal subjects (age 18 – 80) available to researchers
- We would be delighted to make these available to researchers in China
Grid-Based Information Models to Support the Rapid Innovation of New High Value-Added Chemicals

- Start 1st Feb 2004
- Fundamental technology supporting highly-dynamic virtual organisations throughout their entire lifecycle
- Strong business focus
- Speciality, agrochemical and pharmaceuticals sector
- Security emphasis
- OGSA-DAI
- Building on myGrid Information Model

Allen Wright (Chem Eng, Newcastle)
- 3 Universities and institutes
- 6 Companies

Slide: Carole Goble & Jim Fleming
Integrative Biology

Tackling two Grand Challenge research questions:

• What causes heart disease?
• How does a cancer form and grow?

Together these diseases cause 61% of all UK deaths

Will build a powerful, fault-tolerant Grid infrastructure for biomedical science

Enabling biomedical researchers to use distributed resources such as high-performance computers, databases and visualisation tools to develop complex models of how these killer diseases develop.
Three New EPSRC Projects

- **CARMEN**
  - Understanding the brain – £4.5m – led by Professor Colin Ingram, University of Newcastle upon Tyne

- **NanoCMOSGrid**
  - Designing nano-circuits – £5.2m – led by Professor Asen Asenov at Glasgow University

- **PMESG (Pervasive Mobile Environmental Sensor Grids) project**
  - Environmental impact of traffic
  - Jointly funded with the Department for Transport
  - £3.5m – led by Professor John Polak at Imperial College
Understanding the brain may be the greatest informatics challenge of the 21st century

- determining ion channel contribution to the timing of action potentials
- resolving the ‘neural code’ from the timing of action potential activity
- examining integration within networks of differing dimensions
CARMEN Consortium

Leadership & Infrastructure

Colin Ingram

Paul Watson

Leslie Smith

Jim Austin
CARMEN Consortium

International Partners

Ad Aertsen
(Freiburg)
Neural network modelling
and large-scale simulations

George Gerstein
(Pennsylvania)
Analysis of spike pattern trains

Sten Grillner
(Karolinska Institute)
Chairman of the OECD,
International Neuroinformatics
Coordinating Facility

Shiro Usui
(RIKEN Brain Science Institute)
Lead for the Japan Node of the
International Neuroinformatics
Coordinating Facility

Daniel Gardner
(Cornell)
Lead for the US NIH,
Neuroscience Information
Framework and Brain ML
CARMEN Consortium

Commercial Partners

- AstraZeneca
  - applications in the pharmaceutical sector

- NeuroServe
  - interfacing of data acquisition software

- Neuralynx
  - application of infrastructure

- Microsoft
  - commercialisation of tools

- CYBULA
  - high performance pattern recognition systems
The Challenge

International Technology Roadmap for Semiconductors

<table>
<thead>
<tr>
<th>Year</th>
<th>2005</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPU Half Pitch (nm)</td>
<td>90</td>
<td>45</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>MPU Gate Length (nm)</td>
<td>32</td>
<td>18</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

2005 edition Toshiba 04

Device diversification

90nm: HP, LOP, LSTP
45nm: UTB SOI
32nm: Double gate

Slide: Asen Asenov & Rich Sinnott
University Partners

Advanced Processor Technologies Group (APTGUM)
Device Modelling Group (DMGUG)
Electronic Systems Design Group (ESDGUS)
Intelligent Systems Group (ISGUY)
National e-Science Centre (NeSC)
Microsystems Technology Group (MSTGUG)
Mixed-Mode Design Group in IMNS (MMDGUE)
e-Science NorthWest Centre (eSNW)

Slide: Asen Asenov & Rich Sinnott
Industrial Partners

Global EDS vendor and world TCAD leader
600 licences of grid implementation, model implementation

UK fabless design company and world microprocessor leader
Core IP, simulation tools, staff time

UK fabless design company and world mixed mode leader
Additional PhD studentship for mixed mode design

Global semiconductor player with strong UK presence
Access to technology, device data, processing

Global semiconductor player with strong UK presence
Access to technology, device data, processing

Global semiconductor player with UK presence
CASE studentship, interconnects

Trade association of the microelectronics industry in the UK
Recruiting new industrial partners and dissemination
Projects: Pioneering methods
Demonstrating results
Infected academia & Industry

Enthusiastic researchers, Many interacting communities & Breadth
UK e-Science Success

- **Reliable e-Infrastructure 24*7**
  - Foundations well established
  - Extending in Function, Scale & Ubiquity
  - NGS
  - e-Science Centres
  - Specialised support centres
    - AHRC Support @ Kings, Text Mining, 2*NERC centres, NCeSS
- **Data Services**
- **OMII-UK**
- **E-Science Institute**
- **DCC**
- **JISC Virtual Research Environments**
- **JISC e-Framework**
Users get common access, tools, information, Nationally supported services, through NGS

UK e-Infrastructure

- GOSC
- NGS
- Integrated internationally

Regional and Campus grids
- HPCx + HECtoR

Community Grids
- BRIDGES
- eMinerals
- GridPP

Nationally supported services
- JISC
- VRE, VLE, IE
- LHC
- ISIS TS2
- DCC
- MIMAS
- EDINA®
- EMBL-EBI
- British Library

Slide: Neil Geddes
NGS Use

Files stored

~400 users

CPU time by user

Usage Statistics (Total Hours for all 4 Core Nodes)

Number of Registered NGS Users

NGS User Registrations
Linear (NGS User Registrations)

Date

Files stored

Number of Registered NGS Users

Date

Files stored

Usage Statistics (Total Hours for all 4 Core Nodes)

Users by discipline

Users by institution

~400 users

Slide: Neil Geddes
Applications: 2

Systems Biology

Econometric analysis

Example: $\text{La}_{2-x}\text{Sr}_x\text{NiO}_4$

Neutron Scattering

Climate modelling

e-Science Centres in the UK

Coordinated by: Directors’ Forum & NeSC

Leicester

Birmingham

White Rose Grid

Lancaster

Digital Curation Centre

Bristol

Reading

Oxford

Cardiff

Cambridge

CCLRC Daresbury

Manchester

Leeds

National Grid Service

York

CCLRC RAL

Sheffield

Southampton

LeSC

Newcastle

Leicester

Glasgow

Access Grid Support Centre

Belfast

National Centre for Text Mining

Edinburgh

National Centre for e-Social Science

York

National Institute for Environmental e-Science

Manchester

Leeds

Open Middleware Infrastructure Institute

Leicester

Digital Curation Centre

Bristol

National Grid Service

Lancaster

National Centre for e-Social Science

Bristol

Open Middleware Infrastructure Institute

+2 years
OMII-UK nodes

School of Computer Science
University of Manchester

School of Electronics and Computer Science
University of Southampton

EPCC & National e-Science Centre

Edinburgh

Manchester

Southampton

+3.5 years
'software and support to enable a sustained future for the UK e-Science community and its international collaborators'

Software
- Provide guidance to the broad UK e-Science community
- Disseminate your e-Science software to a global community

Support
- Software support and training in using e-Science software
- Provide collaborative mechanisms to support the community
- Define, contribute and disseminate best practice and standards

Sustainability
- Provide a *best of breed* software solution
- Partner to provide a sustainable future.
OMII-UK aims to provide software and support to enable a sustained future for the UK e-Science community and its international collaborators.
Digital Curation Centre

- Mission: “… support and promote continuing improvement in the quality of data curation…”
- Vision
  - Centre of excellence in digital curation
  - Authoritative source of advocacy & advice
  - Key facilitator of informed research community
  - Provider of range of resources, tools & services
a centre of expertise in data curation and preservation

Curation?

- Representation information
  - Data integration
  - Changed states
- Preserving and Archiving
- Sustainability and exit strategy
  - Creating/developing the resource
  - Acquiring and ingesting resource
- Linkage and context; metadata of many kinds capture as much as possible from workflow

IPR limits
Ethics issues
AAA

- Access and re-use
  - Publication
  - Discussion
  - Annotation
  - Review, Collaboration Tools
- Authenticity
  - Provenance
  - Lineage
Enthusiastic researchers, Many interacting communities & Breadth

Projects: Pioneering methods
Demonstrating results
Infecting academia & Industry

Mutual Dependence & Growth

e-Infrastructure: Delivering Resources & Services
Connecting & Supporting Communities

Enthusiastic researchers, Many interacting communities & Breadth
The RoW as an Example
Collaboration Pioneer

- Professor Dan Atkins
  - Director of the Office of Shared Cyberinfrastructure, NSF
- The NSF vision of Cyberinfrastructure supporting e-Research, e-learning and engagement
Achieving the CI Vision requires synergy between 3 types of Foundation wide activities:

**Transformative Application** - to enhance discovery & learning

**Provisioning** - Creation, deployment and operation of advanced CI

**R&D** to enhance technical and social dimensions of future CI systems
Size of the infrastructure today:
- 192 sites in 40 countries
- ~25 000 CPU
- ~3 PB disk, + tape MSS
Usage of the infrastructure

EGEE workload

>50k jobs/day

Normalized CPU time

~7000 CPU-months/month
Use for massive data transfer

Large LHC experiments now transferring ~ 1PB/month each
The Important Questions
Q1: Are we succeeding?

- Goals
  - Transforming effect on the way we do research
  - High impact on National / EU performance
  - Widening access to e-Science

- Honest answers
  - Probably not
  - At best we don’t know
Reasons to worry

- Few new players
  - Familiar faces from HPC, etc.
- Only the enthusiastic pioneers
  - Too risky
  - Too hard to use
  - Takes heroes to write applications
  - Takes heroes to install & operate
  - Doesn’t offer what is wanted
  - Doesn’t support incremental change
- Fragmentation of communities & providers
  - Loss of potential cost savings
  - Loss of democratisation
Alternatives to giving up?
Take a Business approach

- Understand users as customers
- Understand our role as suppliers

See Carole Goble’s Keynote at EGEE Conf. 06
27 September 2006
Early Adopters

Early Majority

Late Majority

Laggards

Innovators

Strangers

Acquaintances

Friends

Family

[David De Roure, 2006]
If we form an alliance with e-Minerals then we can build developer tools which will be useful.

NERC community are going to stop using our products if we don’t fix bottlenecks in our workflow.

“If we form an alliance with e-Minerals then we can build developer tools which will be useful.”
Take a Scientific Approach

- Understand e-Science
  - As a phenomenon
- Characterise properties & varieties
- Postulate hypotheses about mechanisms
- Test, validate & refine hypotheses

National Centre for e-Social Science

University of Manchester

University of Nottingham

Lancaster University

Aberdeen

Leeds

Nottingham

Oxford

Bristol

London

Colchester

UK Data Archive

The University of Essex

UCL

University of Bristol
Take an Engineering Approach

• Identify **new** critical issues
  • Effectiveness
  • Usability
  • Adaptability / Flexibility / Driveability
    ▶ The creative research user *must* be in driving seat

• Establish ways of **measuring** them
  • Draw on empirical software engineering

• *Then* demonstrably improve them
Take an Educational Approach

- Embed e-Science in your courses
- Drive it up the agenda in your institution
- Organise summer schools
- Contribute to summer schools
- Go to summer schools
- Organise training
  - ...

4 u 2 do!
Education, Education, Education!

- **Training**
  - Targeted
  - Immediate goals
  - Specific skills
  - Building a workforce

- **Education**
  - Pervasive
  - Long term and sustained
  - Generic conceptual models
  - Developing a culture

- **Both are needed**
• E-Informatics Research Group
• Have established a chapter on E&T in e-IRG white paper
• Workshop in Linz, 10-11 April
• Prepared chapter
• Accepted by formal meeting in Vienna, June
• http://www.e-irg.org/

9. EDUCATION AND TRAINING

9.1. POLICY AREA

Education and training in the use of e-infrastructure are necessary to realise its potential. They will benefit from shared best practice, coordination and compatible policies, and access to shared education and training material and support systems.
ICEAGE launches ET-WG @ OGF

- E&T in the charter of OGF
- ET WG established @ OGF
- Charter accepted (ET-WG)
  - Officers
    - David Fergusson, NeSC
    - Ann Redelfs, EPIC
    - Ruediger Berlich, KMZ
    - Kilian Schwarz, Darmstadt
    - Beth Plale, Indiana U.
    - Malcolm Atkinson, UK
  - Secretariat
    - Kathryn Cassidy, TCD
    - Lilly Hunter, ICEAGE
  - Themes
    - E&T Frameworks
    - E&T Synergy
    - E&T Content

Must combine strength & effect of ICEAGE, e-IRG & OGF
Repository of shared learning resources

The ICEAGE Library facilitates the use of various training resources related to the International Collaboration to Extend and Advance Grid Education (ICEAGE) project.

**Podcast - Audio**
- Podcast Interview with Neil P. Chue Hong
- Research at Forschungszentrum Karlsruhe
- The EGEE Project (Mission / Status / Roadmap)
- The gLite middleware, its architecture and components

**Summer School**
- gLite exercises
- Summary, Certificates & Farewell
- Vision of Future Grids & Research challenges
- GridCC: Real-time instrumentations grids

**IBM Articles and Tutorials**
- Manage multiple resources with a single instance service
- Resource sharing among multiple Web services, Part 2: Incorporating a reusable FactoryService
- Resource sharing among multiple Web services, Part 1: Implement resource sharing using WSRF

**Globus**
- Interoperability Testing for DAIS Working Group Specifications
- Open Grid Services Architecture Glossary of Terms Version 1.5
Start answering Q1

- Learn from business
  - Drive your mission hard
- Recognise the Scientific Question
  - Drive your mission very hard
- Take up the engineering challenge
  - Drive your mission very very hard
- Engage in e-Science Education
  - Every day do what you can
- Start now – Pick your mission
  - Don’t wait for someone else to lead
The Important Questions
Q2: Is e-Science cost effective

- What is the RoI
  - Papers in Nature & Science
  - New knowledge in application domains
  - New Computing knowledge
  - Breakthroughs
  - Nobel prizes

- How does it compare with business as usual
  - The desk top, project & cluster computing
  - The HPC computing
  - ...
Honest answers

- Don’t know
  - Haven’t measured it
- Excuses
  - Too hard to measure
  - Too soon to measure
- Not good enough!
- Time to Grow Up
As Q1 + Standards

- Reduce costs
  - Smarter engineering
- Share cost of each engineering effort
  - Recognise core & common requirements
- Establish standard APIs, Protocols & Policies
  - To capture the common requirement
  - To enable interchangeable solutions
  - To encourage re-use
  - To enable composition & interoperation
  - To foster stability
Start answering Q1

- Learn from business
  - Drive your mission hard
- Recognise the Scientific Question
  - Drive your mission very hard
- Take up the engineering challenge
  - Drive your mission very very hard
- Engage in e-Science Education
  - Every day do what you can
- Help to form / encourage standards
  - If you help they may help you!
- Start now – Pick your mission
  - Don’t wait for someone else to lead

4 u 2 do!
Take Home

- Collaboration
- Three Mutual Strengths
- Sparks of Success
- Danger of Fire going out
- Find your way to fan the flames
- There is Much for you to do
- Enjoy the opportunity

4 u 2 do!