UK e-Science People & Data WWW Conference Edinburgh

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

www.nesc.ac.uk
25th May 2006

Overview

- ## Brief History
  - ### Investment, Engagement & Community
- ## Present
  - ### Projects, Activities & Technologies
  - ### Now & New
- ## Future
  - ### Usability & Abstraction
  - ### Interoperation & Federations

What is e-Science?

- Goal: to enable better research in *all* disciplines
- Method: Develop *collaboration* supported by advanced distributed computation
  - to generate, curate and analyse rich data resources
    - From experiments, observations and simulations
    - Quality management, preservation and reliable evidence
  - to develop and explore models and simulations
    - Computation and data at all scales
    - Trustworthy, economic, timely and relevant results
  - to enable *dynamic* distributed collaboration
    - Facilitating collaboration with information and resource sharing
    - Security, trust, reliability, accountability, manageability and *agility*

Commitment to e-Infrastructure

- **A shared resource**
  - That enables science, research, engineering, medicine, industry, ...
  - It will improve UK / European / ... productivity
    - ▸ Lisbon Accord 2000
    - ▸ e-Science Vision SR2000 – John Taylor
  - Commitment by UK government
    - ▸ Sections 2.23-2.25
  - Always there
    - ▸ c.f. telephones, transport, power

Science & innovation investment framework 2004 - 2014

July 2004

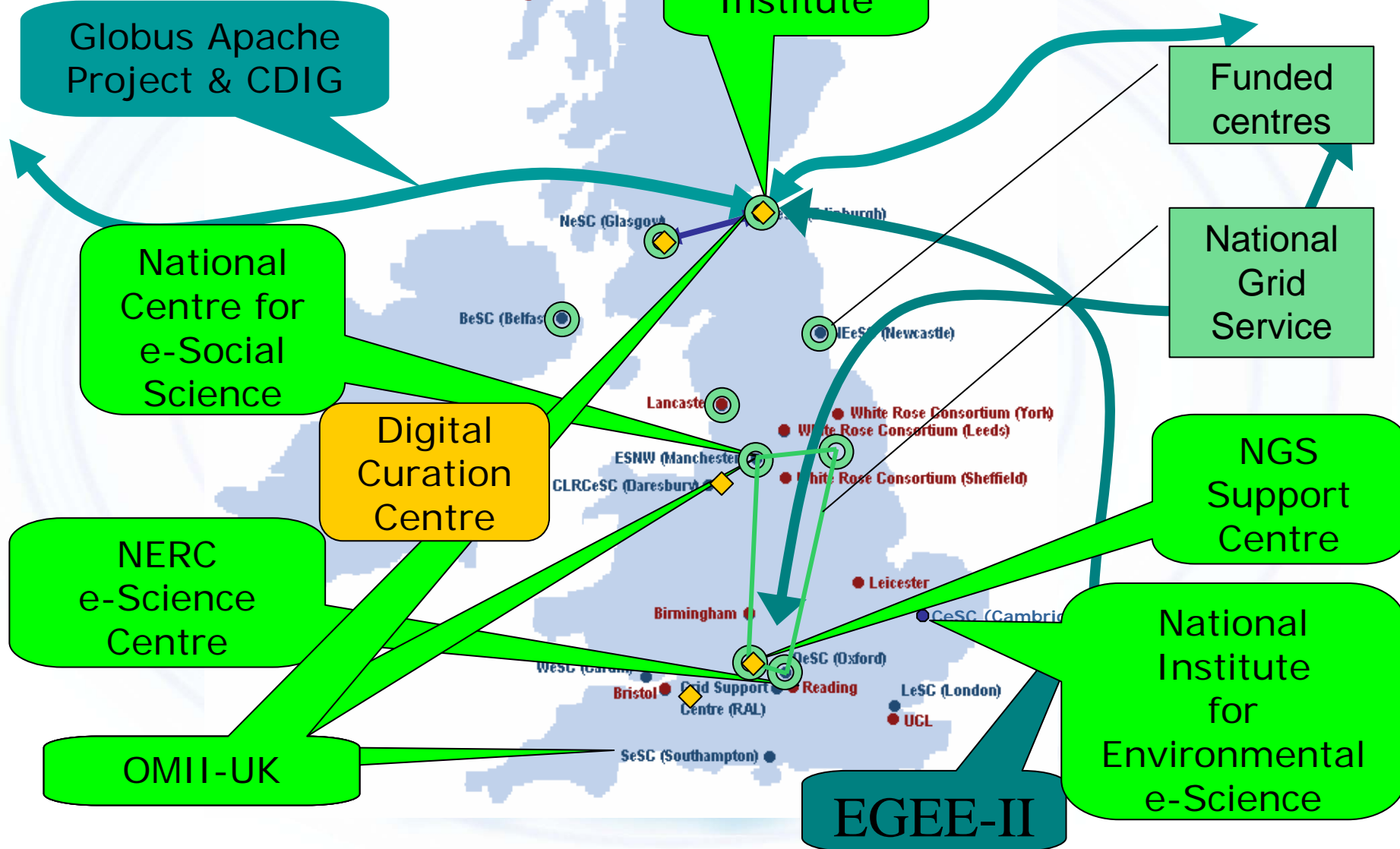HM TREASURY   dti   department for education and skills

Gordon Brown
Chancellor of the Exchequer

Charles Clarke
Secretary of State for Education and Skills

Patricia Hewitt
Secretary of State for Trade and Industry
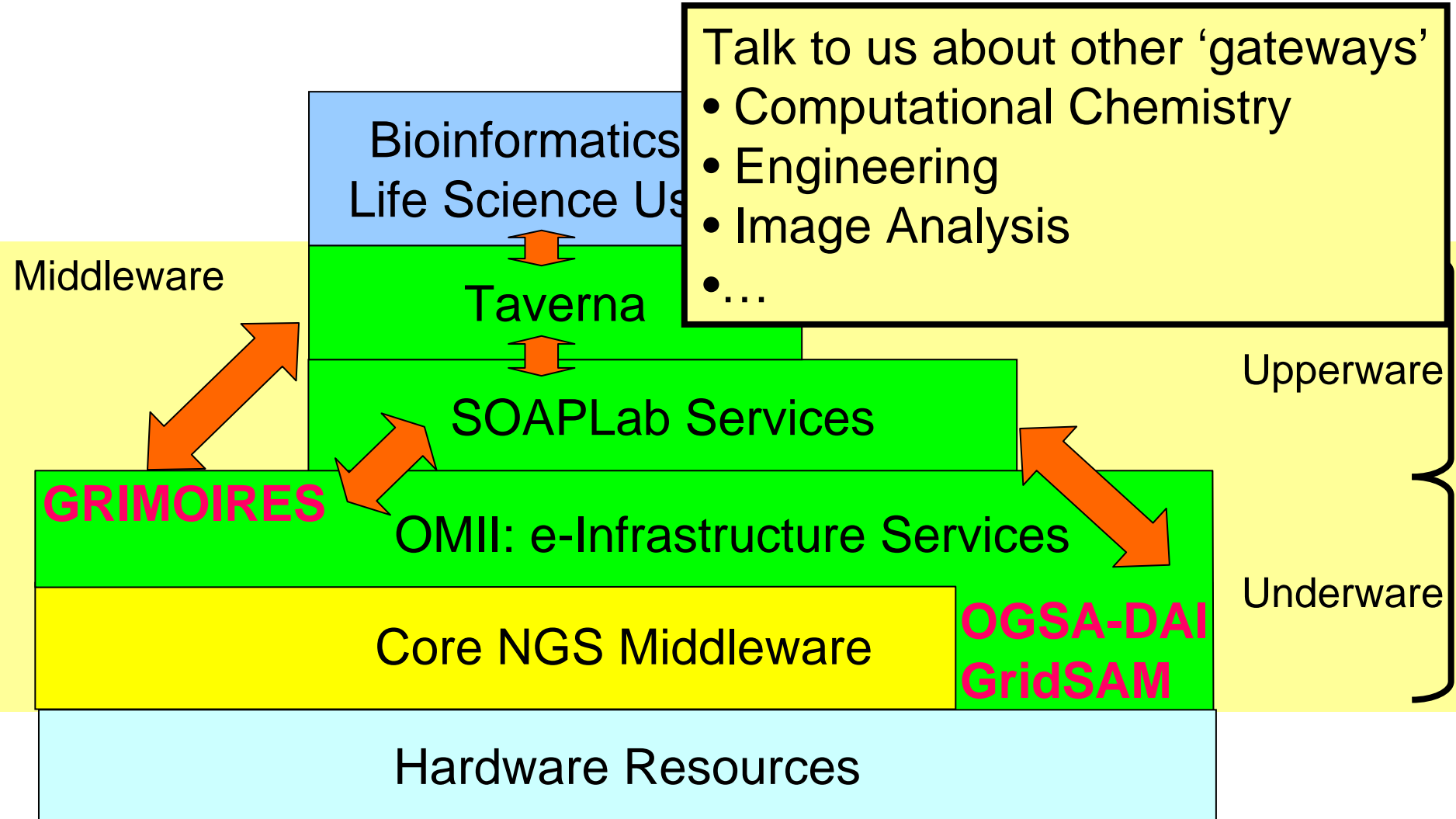
The e-Science On The Map Today

e-Science Institute

Globus Apache Project & CDIG

Funded centres

National Grid Service

National Centre for e-Social Science

Digital Curation Centre

NGS Support Centre

NERC e-Science Centre

National Institute for Environmental e-Science

OMII-UK

EGEE-II

NeSC (Glasgow)

BeSC (Belfast)

NEeSC (Newcastle)

Lancaster

White Rose Consortium (York)
White Rose Consortium (Leeds)

ESNW (Manchester)
CLRCeSC (Daresbury)

White Rose Consortium (Sheffield)

Leicester

Birmingham

CeSC (Cambridge)

WeSC (Cardiff)

Bristol

OeSC (Oxford)

Grid Support Centre (RAL)

Reading

LeSC (London)

UCL

SeSC (Southampton)

OMII-UK & National Grid Service: Life Sciences Gateway

Bioinformatics
Life Science Us...

Taverna

SOAPLab Services

GRIMOIRES

OMII: e-Infrastructure Services

Core NGS Middleware

OGSA-DAI GridSAM

Hardware Resources

Middleware

Upperware

Underware

Talk to us about other 'gateways'
- Computational Chemistry
- Engineering
- Image Analysis
- …

The National Grid Service

Core sites
White Rose (Leeds)
Manchester
Oxford
CCLRC

Partner sites
Bristol
Cardiff
Lancaster

Access to HPC facilities
HPCx
CSAR

Capacity
300 + CPUs
30+ Tera Bytes
Specialist facilities
Cardiff 4x16 proc SGI
Bristol: Intel
Lancaster SUN Cluster
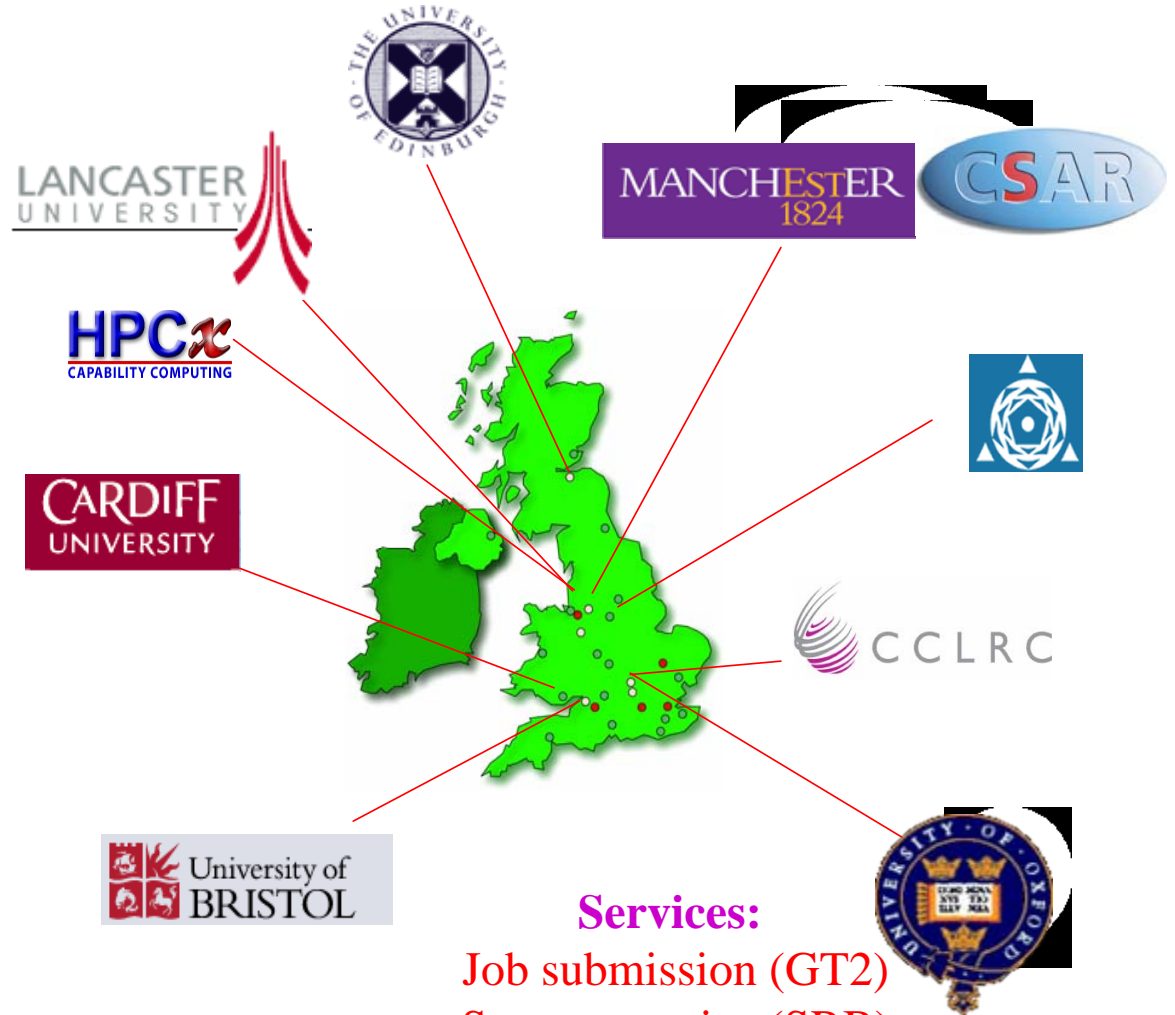
Services:
Job submission (GT2)
Storage service (SRB)
Oracle Service
OGSA-DAI

About Grid Computing Now!

A Knowledge Transfer Network project funded by the DTI Technology Programme aimed at transferring knowledge about Grid Computing Technologies to Public and Private Sectors in the UK.

Partnership between Intellect, the UK Hi-Tech Trade Association; National e-Science Centre, a world leader in Grid Computing research; and CNR Ltd, a consultancy focused on SME organisations and business intermediaries.

Substantial number of industrial, business and academic partners

Website
- Background Information
- Industry News/Events
- User Case Studies

Events programme
- Technical Overviews
- Multiple vendor perspectives
- User Case Studies

Sector Agenda
- Healthcare; Government; Telecoms; Services; etc..

User Community
- Network with peers
- Find useful contacts
- Contribute experience

www.gridcomputingnow.org

Collaboration is the key to e-Science

Biomedical Research Informatics Delivered by Grid Enabled Services

http://www.brc.dcs.gla.ac.uk/projects/bridges/
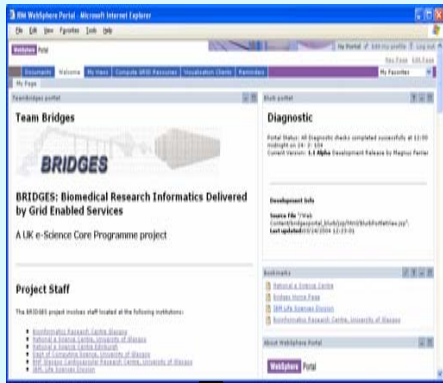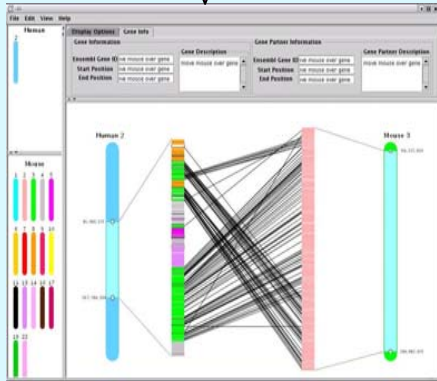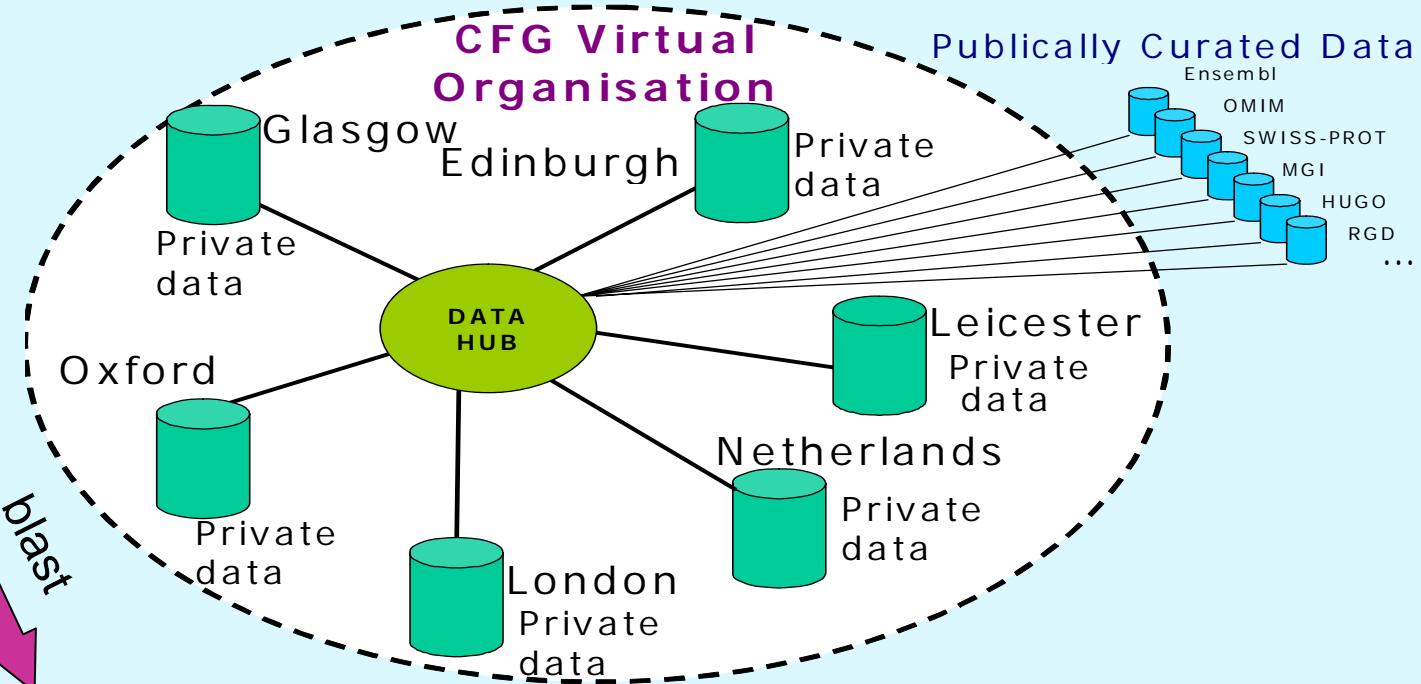
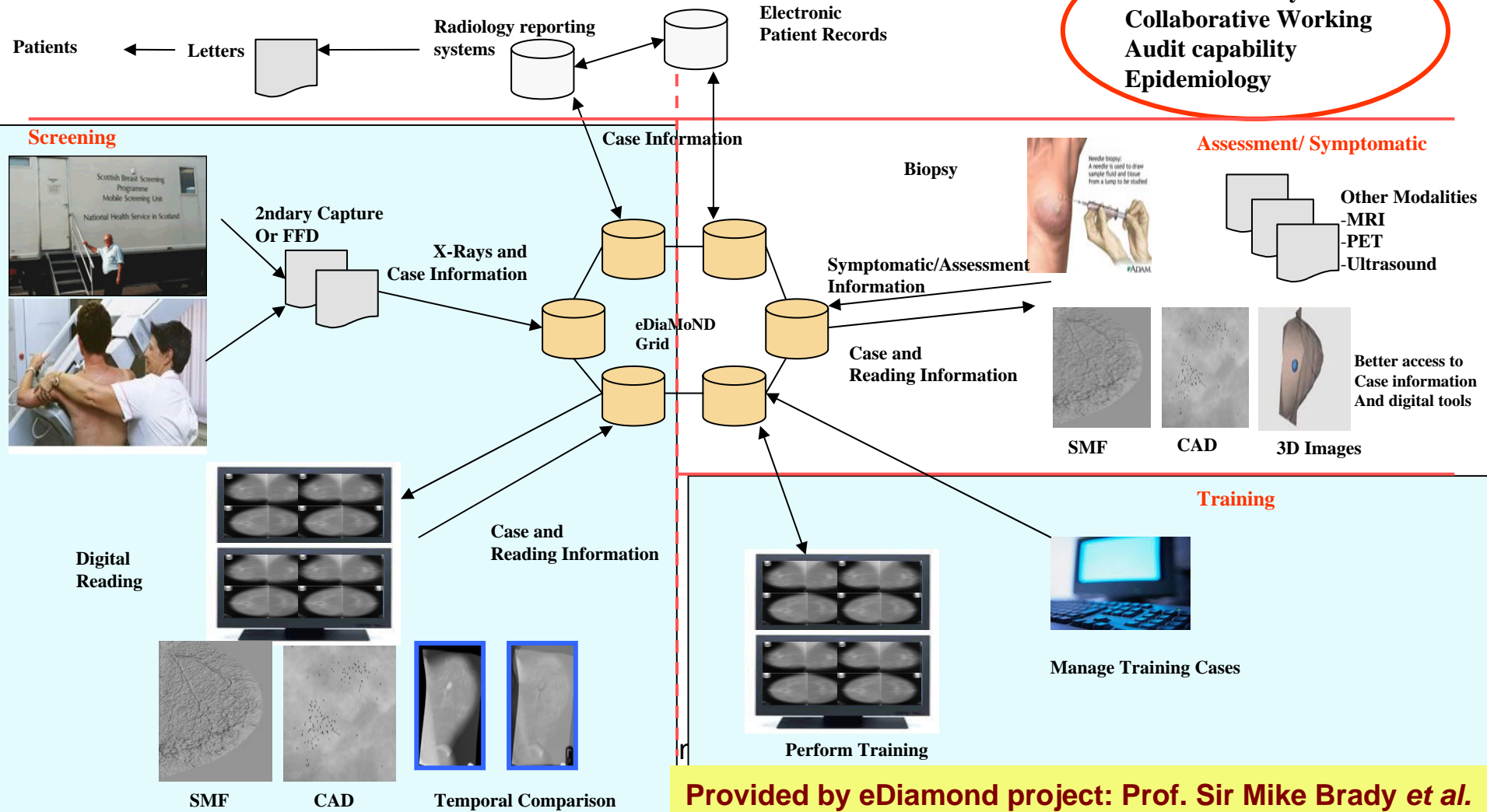# eDiaMoND: Screening for Breast Cancer

1 Trust → Many Trusts
Collaborative Working
Audit capability
Epidemiology

Patients ← Letters ← Radiology reporting systems

Electronic Patient Records

Screening

Case Information

Assessment/ Symptomatic

Biopsy

2ndary Capture Or FFD

X-Rays and Case Information

eDiaMoND Grid

Symptomatic/Assessment Information

Other Modalities
-MRI
-PET
-Ultrasound

Case and Reading Information

Better access to Case information And digital tools

SMF   CAD   3D Images

Digital Reading

Case and Reading Information

Training

Manage Training Cases

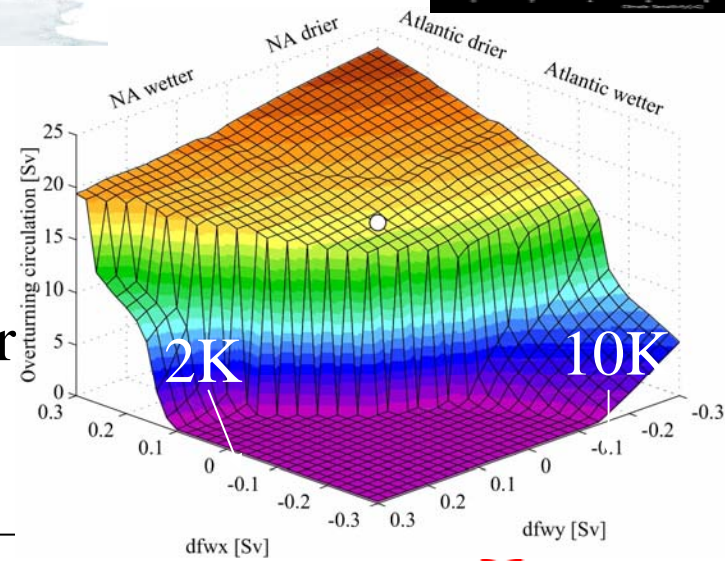SMF   CAD   Temporal Comparison

Perform Training

climate*prediction*.net and GENIE
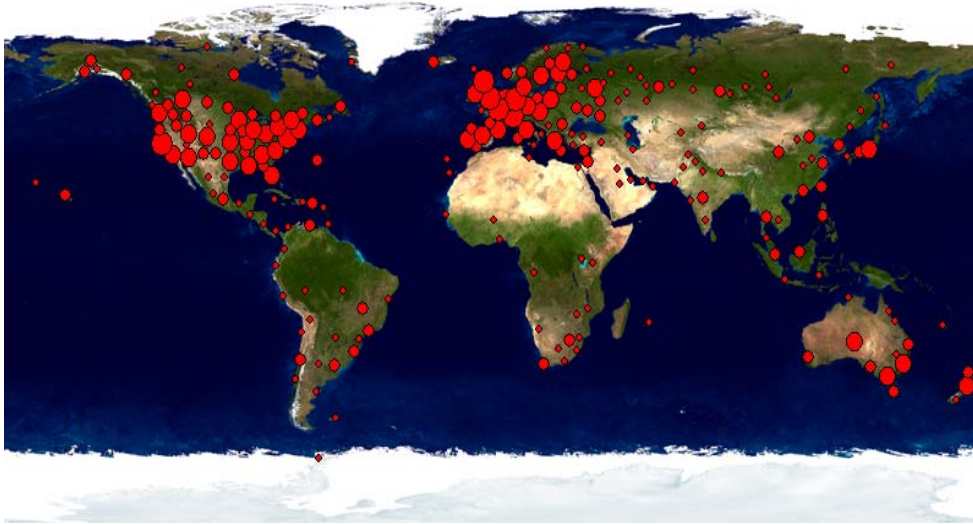
- Largest climate model ensemble
- >45,000 users, >1,000,000 model years
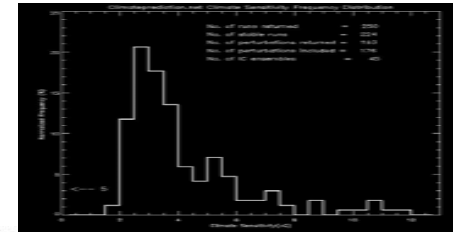
Response of Atlantic circulation to freshwater forcing

NATURAL
ENVIRONMENT
RESEARCH COUNCIL

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.

©Bioengineering Research Group   The University of Auckland

Courtesy of David Gavaghan & IB Team
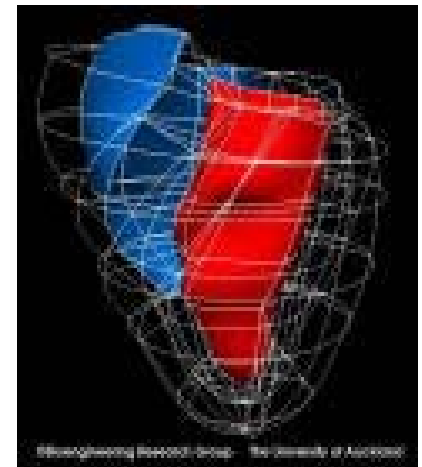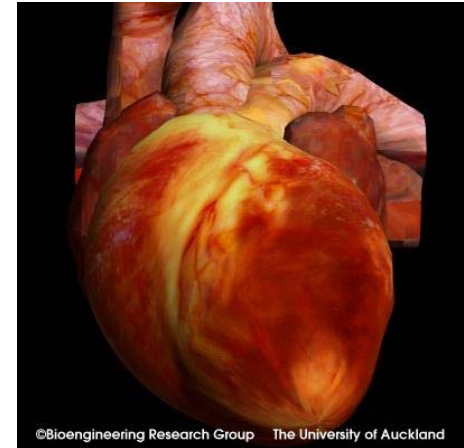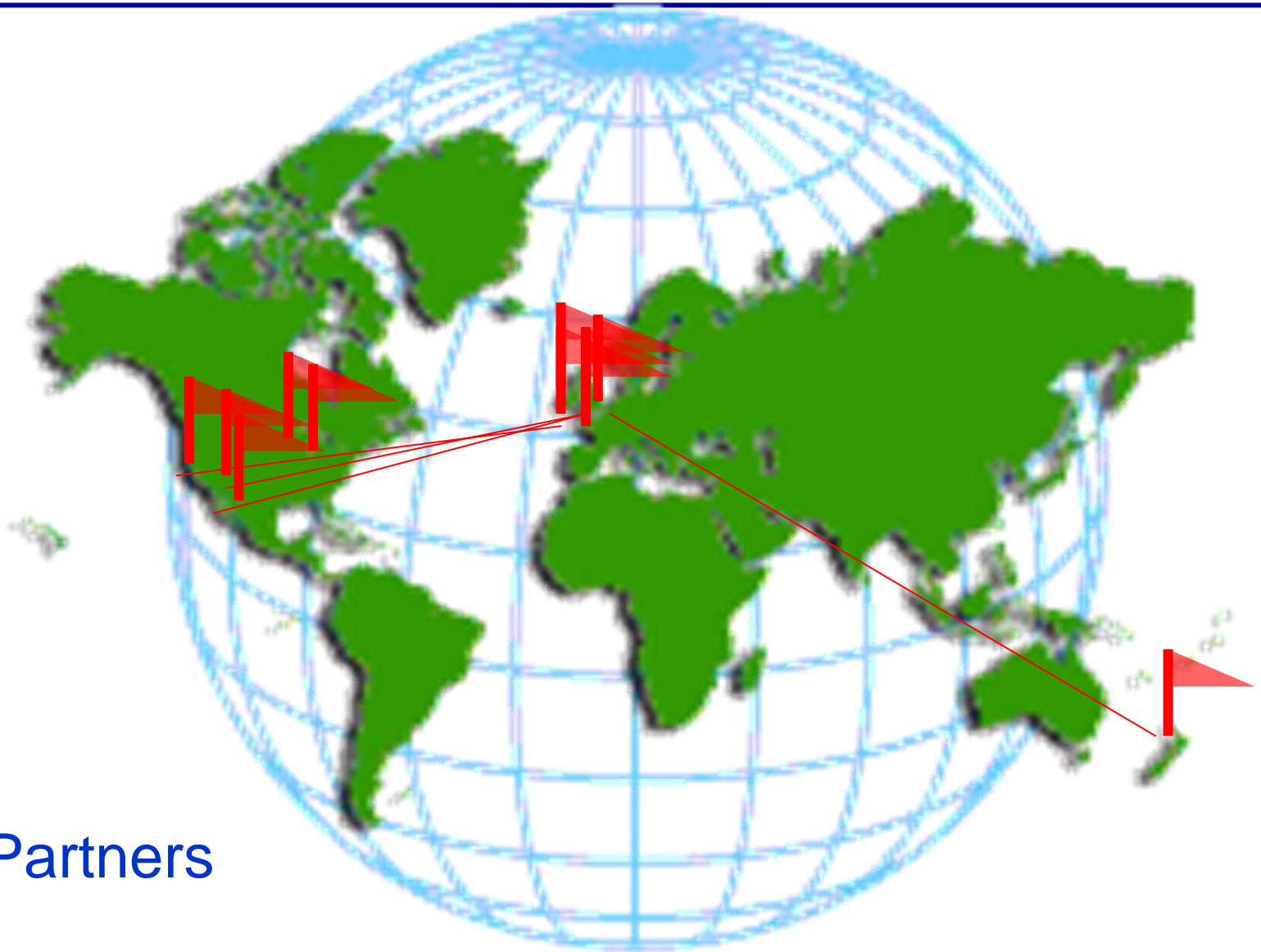
IB Partners

Foundations of Collaboration

- **Strong commitment by individuals**
  - To work together
  - To take on communication challenges
  - Mutual respect & mutual trust
- **Distributed technology**
  - To support information interchange
  - To support resource sharing
  - To support data integration
  - To support trust building
- **Sufficient time**
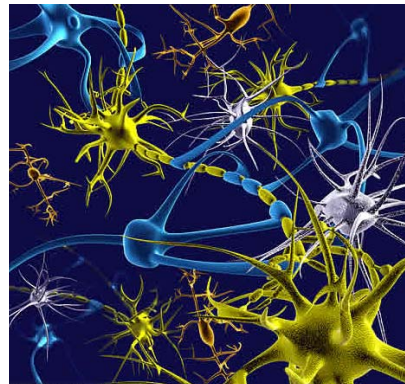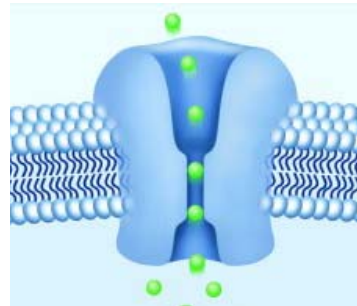- **Common goals**
- **Complementary knowledge, skills & data**

CARMEN - Scales of Integration

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**
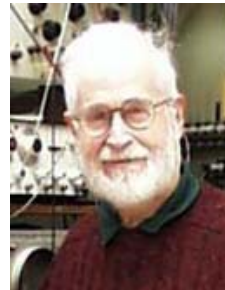
CARMEN Consortium

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

Neuralynx
High Density Electrophysiology / Recording Systems

- interfacing of data acquisition software

Microsoft®

- application of infrastructure
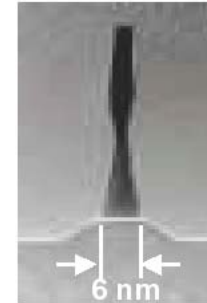
CYBULA
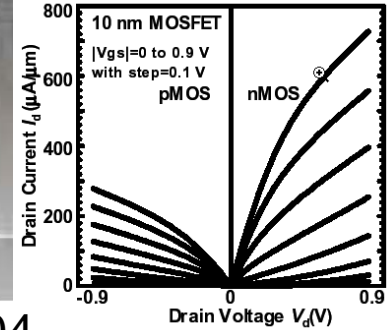high performance pattern recognition systems

- commercialisation of tools

The Challenge

International Technology Roadmap for Semiconductors

Year	2005	2010	2015	2020
MPU Half Pitch (nm)	90	45	25	14
MPU Gate Length (nm)	32	18	10	6

2005 edition

Toshiba 04

Device diversification

90nm: HP, LOP, LSTP

45nm: UTB SOI

32nm: Double gate

230 nm

Bulk MOSFET

Standard

Single Set

25 nm

FinFET
UTB SOI
FD SOI
Bulk MOSFET
LSTP
LOP
HP(MPU)

Stat. Sets
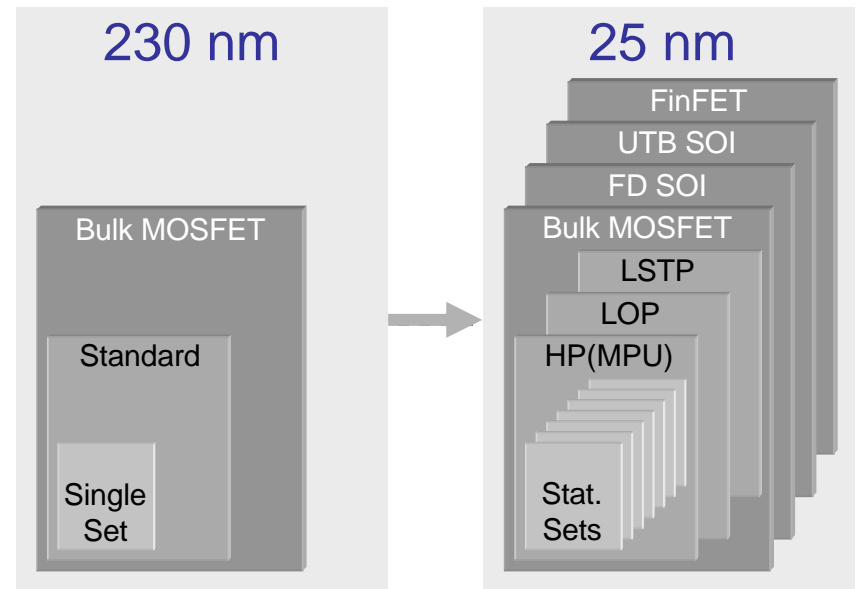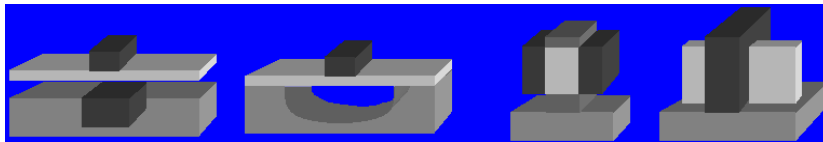
University Partners

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)

Industrial Partners

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

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

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

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

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

FUJITSU
Global semiconductor player with UK presence
CASE studentship, interconnects

Trade association of the microelectronics industry in the UK
Recruiting new industrial partners and dissemination

Next steps

in

e-Science

Collaboration

- **Essential to assemble experts**
  - Multi-discipline, Multi-organisation, Multi-national
- **Hard to achieve**
  - Instinctive competition
  - Trust slow to build
  - Communication is difficult

  Address these issues

- **Requirements**
  - Leadership
  - Investment
  - New culture
  - Technology
  - Cross commercial – academic boundaries

  Focus here

Towards Accessible e-Science

- **High-level tools**
  - Abstraction
  - Metadata-driven interpretation & guidance
  - Well-defined semantics
  - Automated data management is key

- **Convenient user-controlled composition**
  - Lego-brick convenience + Precision control
  - Understood by scientists, engineers, modellers, diagnosticians, …

- **Responsibility & Credit**
  - Provenance tracking automated & understood
  - Culture developed for consortia and collaboration
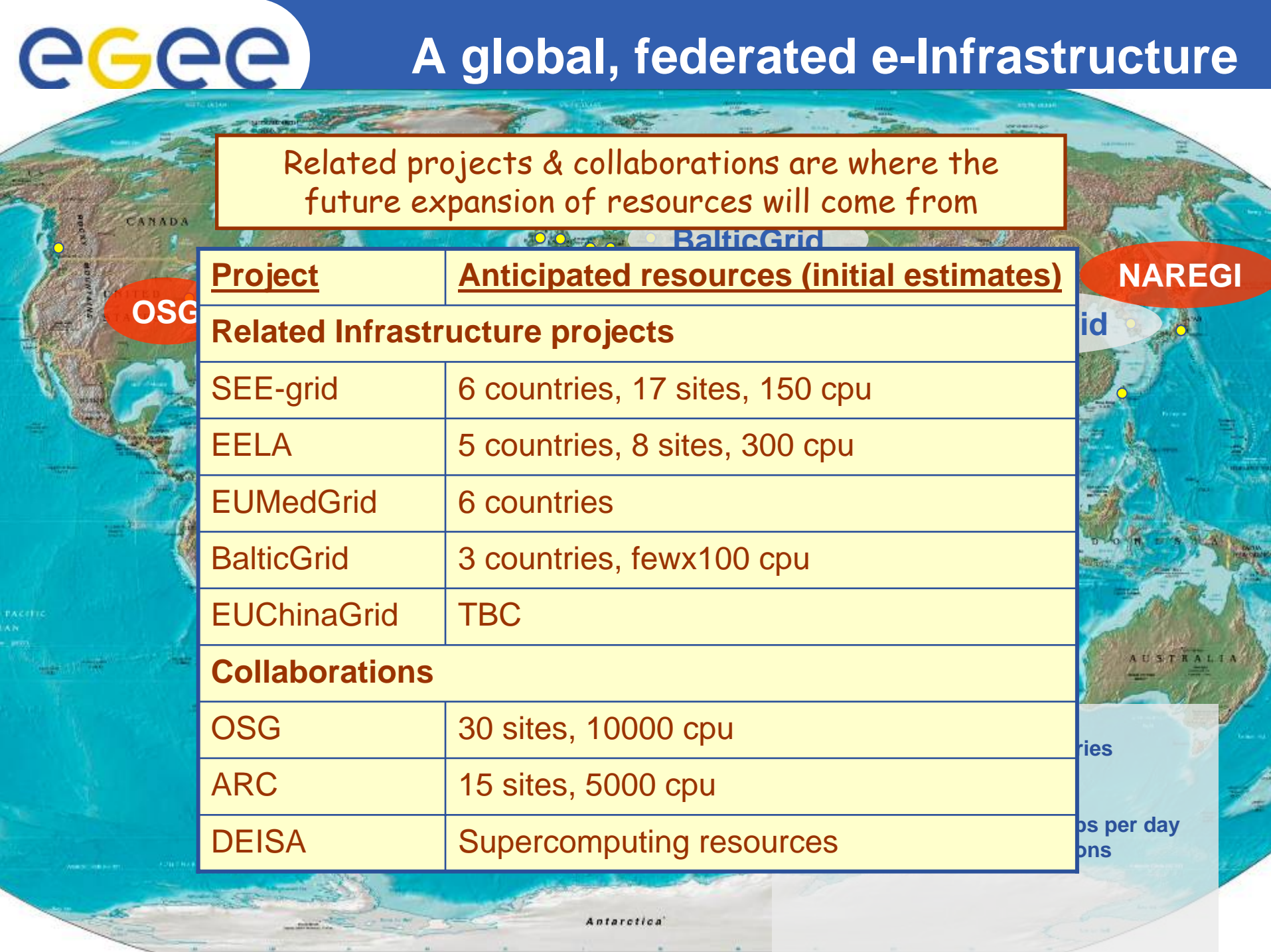
Built on Dependable Infrastructure

- **Global Federations**
  - Dependable and persistent facilities
  - Always there & always on
  - Consistent for mobile users
  - Consistent for mobile code & mobile information
- **Affordable e-Infrastructure**
  - Based on well-established standards
  - Based on well-honed operating procedures
  - Investment preserved through stability
  - Utility improved through agile development
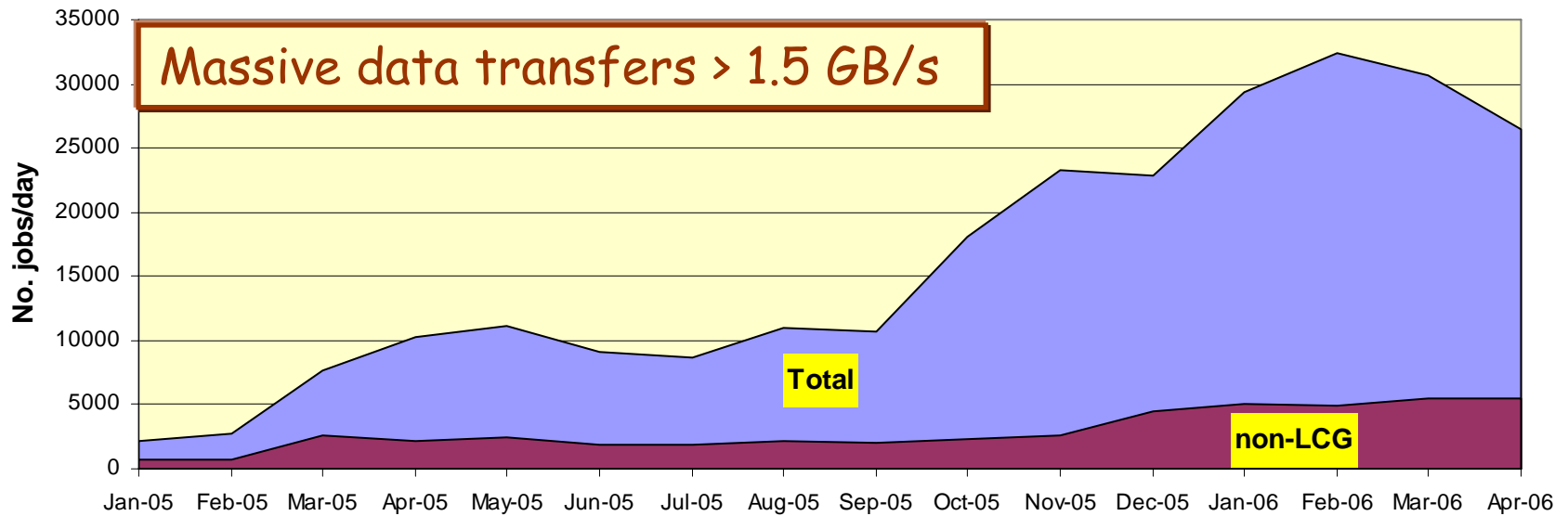- **Trustworthy management of information**

A global, federated e-Infrastructure

egee

Related projects & collaborations are where the future expansion of resources will come from

BalticGrid

NAREGI

OSG

Project	Anticipated resources (initial estimates)
Related Infrastructure projects	
SEE-grid	6 countries, 17 sites, 150 cpu
EELA	5 countries, 8 sites, 300 cpu
EUMedGrid	6 countries
BalticGrid	3 countries, fewx100 cpu
EUChinaGrid	TBC
Collaborations	
OSG	30 sites, 10000 cpu
ARC	15 sites, 5000 cpu
DEISA	Supercomputing resources

Massive data transfers > 1.5 GB/s

Total

non-LCG

No. jobs/day

35000
30000
25000
20000
15000
10000
5000
0

Jan-05 Feb-05 Mar-05 Apr-05 May-05 Jun-05 Jul-05 Aug-05 Sep-05 Oct-05 Nov-05 Dec-05 Jan-06 Feb-06 Mar-06 Apr-06

Sustained & regular workloads of >30K jobs/day
- spread across full infrastructure
- doubling/tripling in last 6 months – no effect on operations

Statistics:
Submitted: 617
Waiting: 85
Ready: 846
Scheduled: 9566
Running: 12166
Done: 6427
Aborted: 3184
Cancelled: 89
Active Sites: 156 : 33963

Developed by e-Science, HEP
Imperial College
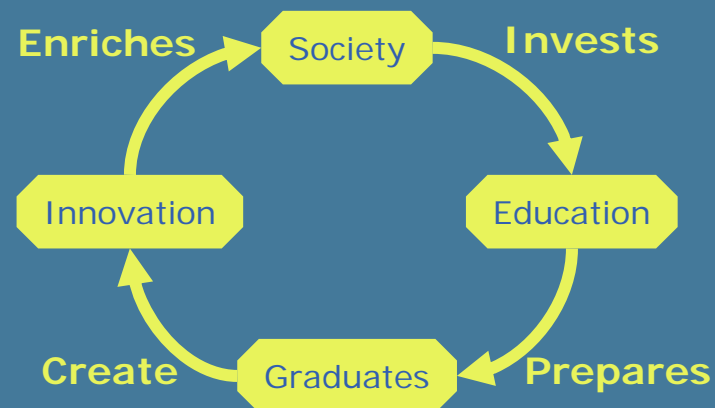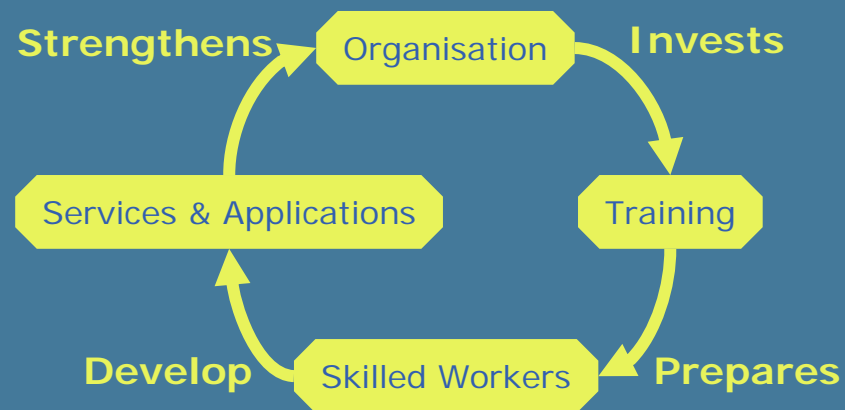London

Invest in People

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

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

- **Both are needed**

Strengthens → Organisation → **Invests**

Services & Applications · Training

Develop · Skilled Workers · **Prepares**

Enriches → Society → **Invests**

Innovation · Education

Create · Graduates · **Prepares**

Take Home Message

- **UK e-Science**
  - Has delivered innovation and scientific results
  - Has e-Infrastructure & support in place
  - Works with industry & generates new businesses
  - Has a thriving community
- **UK e-Science is ready for more**
  - New, intellectually challenging projects
  - Collaboration will make this possible
- **Collaboration is a Key Issue**
- **Trustworthy data sharing key for collaboration**
  - Plenty of opportunities for research and innovation