Grid Overview and Context

John Gordon

j.c.gordon@rl.ac.uk
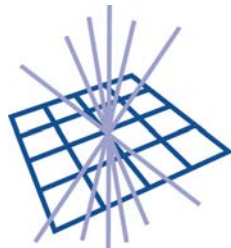
GridPP
UK Computing for Particle Physics

eGee
Enabling Grids for
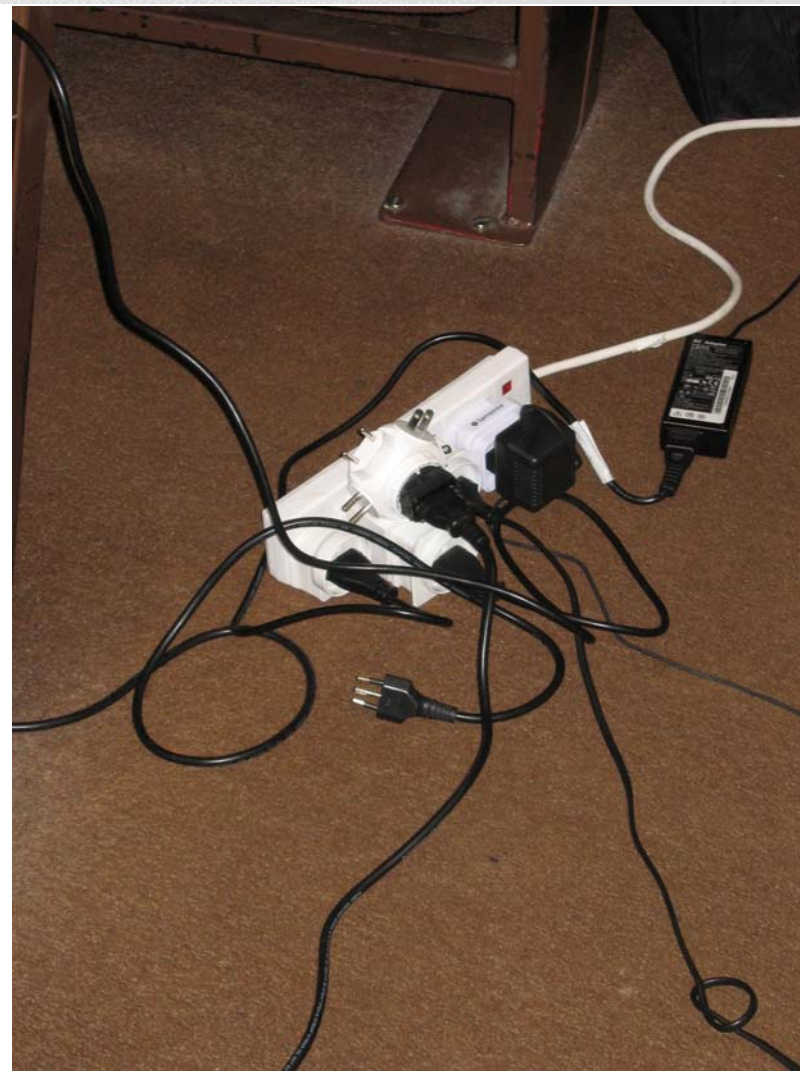E-science in Europe

http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf

1. Coordinates resources that are not subject to centralized control

2. … using standard, open, general-purpose protocols and interfaces

3. … to deliver nontrivial qualities of service

1. YES. This is why development and maintenance of a UK-EU-US collaboration is important

2. YES... Globus/CondorG/EDG meet this requirement. Common experiment application layers are also important here

3. NO(T YET)… Experiments define whether this is true (currently only ~100,000 jobs submitted via the testbed c.f. system internal tests of up 10,000 jobs per day. Await LCG-2 deployment outcome…)

John Gordon
e-Science Centre

What is a Grid?

- Distributed resources
- no common management
- standard protocols
- flexible organisations

- …and which ones are we interested in?
- EGEE, LCG, GridPP, NGS
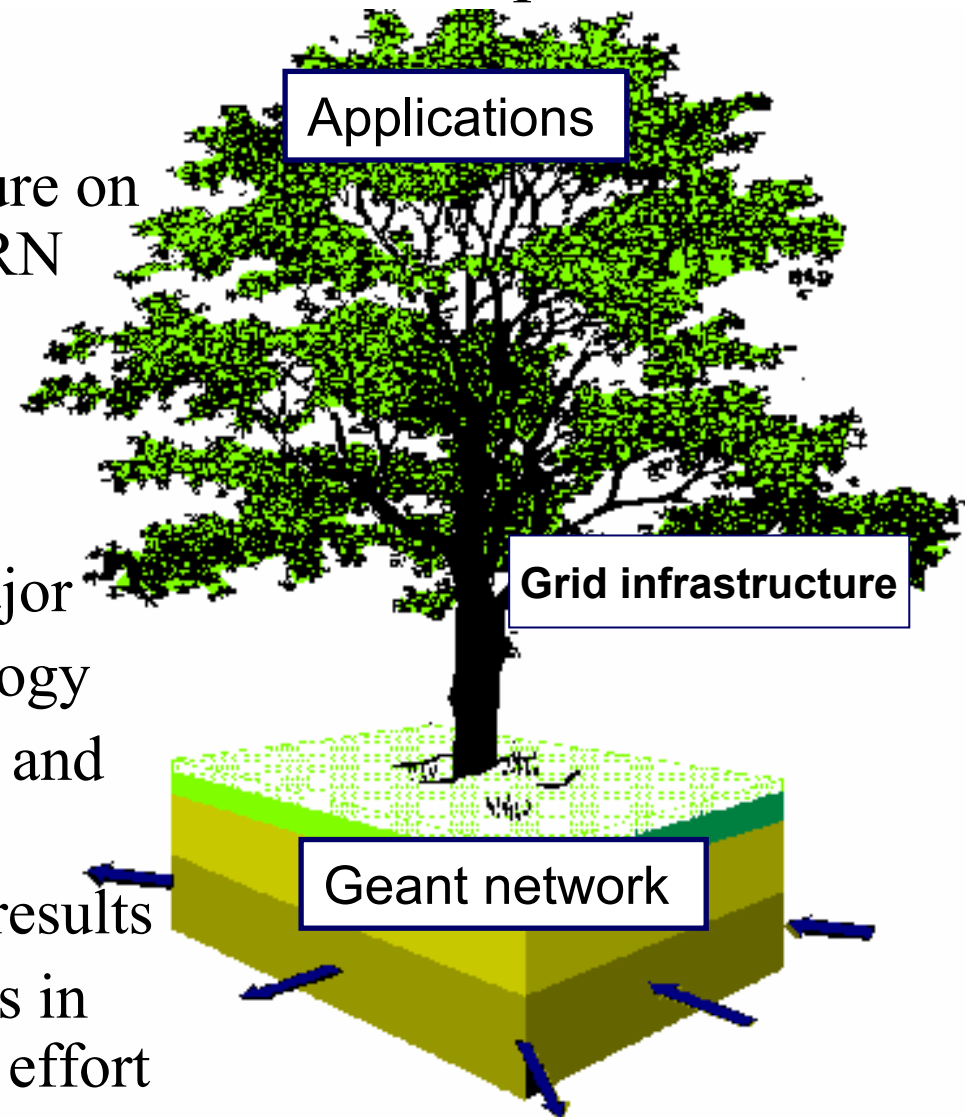
John Gordon
e-Science Centre

CCLRC

Enabling Grids for E-science in Europe

- **Goal**
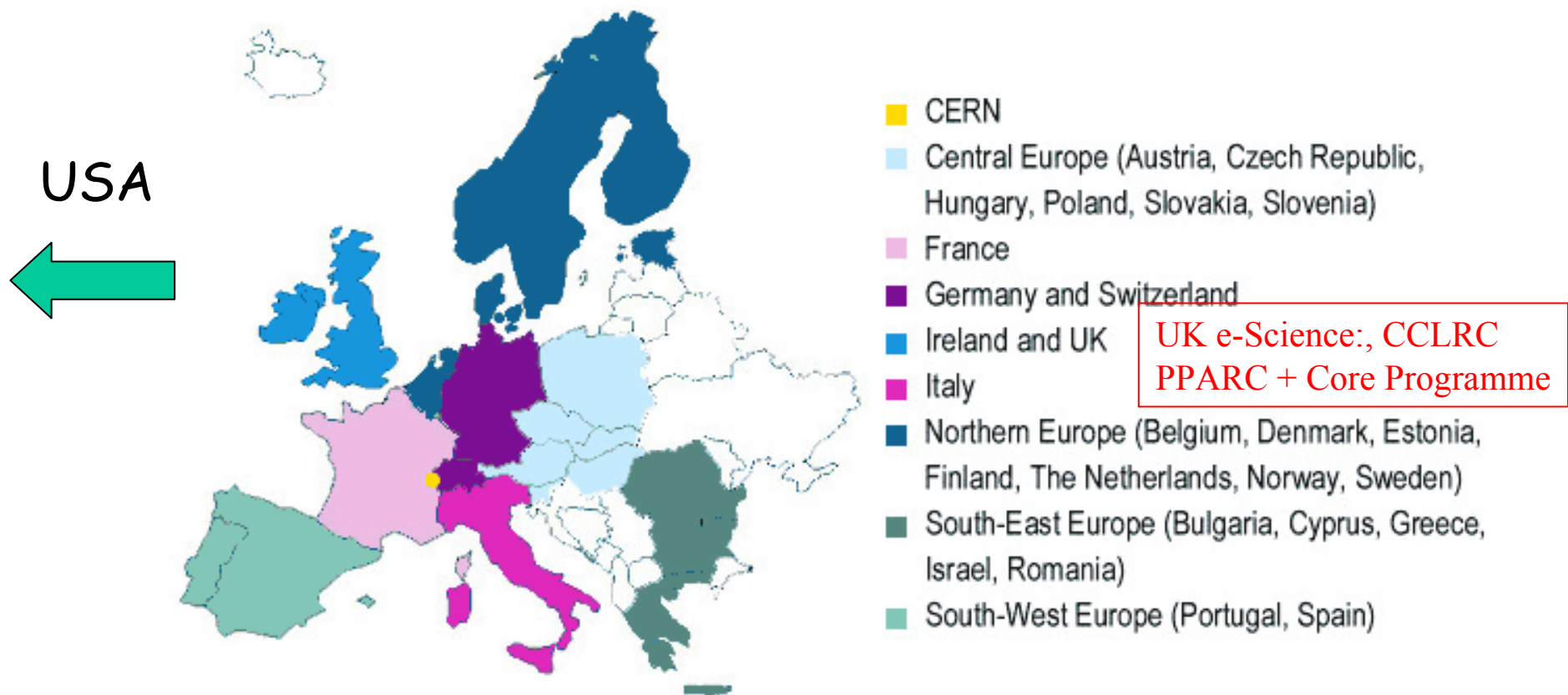 - Create a wide European Grid production quality infrastructure on top of present and future EU RN infrastructure

- **Build On:**
  - EU and EU member states major investments in Grid Technology
  - International connections (US and AP)
  - Several pioneering prototype results
  - Large Grid development teams in EU require major EU funding effort

Applications

Grid infrastructure

Geant network

CCLRC

Enabling Grids for E-science for ~~Europe~~ Everyone

USA

- CERN
- Central Europe (Austria, Czech Republic, Hungary, Poland, Slovakia, Slovenia)
- France
- Germany and Switzerland
- Ireland and UK
- Italy
- Northern Europe (Belgium, Denmark, Estonia, Finland, The Netherlands, Norway, Sweden)
- South-East Europe (Bulgaria, Cyprus, Greece, Israel, Romania)
- South-West Europe (Portugal, Spain)

UK e-Science:, CCLRC
PPARC + Core Programme

10 Consortia (incl. GEANT/TERENA/DANTE) → 70 Partners

John Gordon
e-Science Centre

EGEE Challenge

- A large investment in a short time(32M€ in 24 months)
  - rationale is to mobilise the wider Grid communityy in Europe and elsewhere and be includive
  - demonstrate production quality sustained Griod services for a few relevant scientific communities (at least HEP and Bio-Medical)
  - demonstrate a viable process to bring other scientific communities on board
  - propose a second phase in mid 2005 to start early 2006
- Move from R&D middleware and testbeds to industrial quality software and sustained production Grid infrastructre performance
- Implement a highly distributed software engineering process while maintaining efficiency and a fast release cycle (development clusters)
- Harmonise EGEE activities with national and international activities
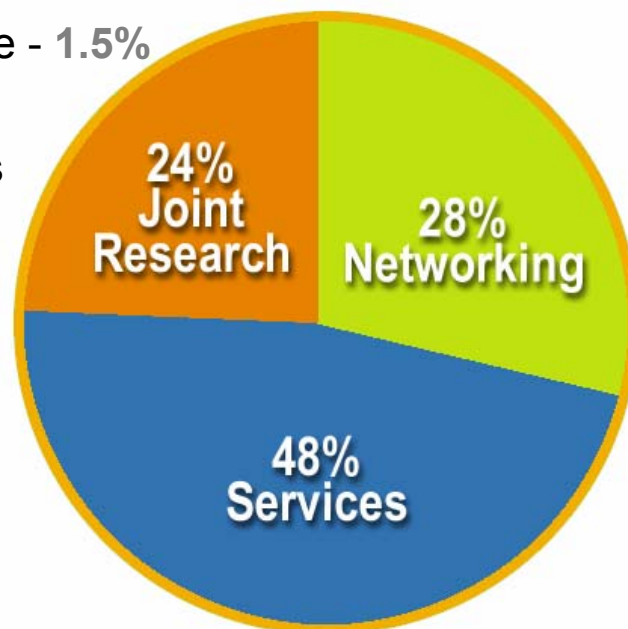- Cope with new FP6 rules and different and often conflicting EU Grid Plans and actvities

John Gordon
e-Science Centre

EGEE Project Structure

24% Joint Research

JRA1: Middleware Engineering and Integration - 17%

JRA2: Quality Assurance - **1.5%**

JRA3: Security - **3%**

JRA4: Network Services Development - **2.5%**

28% Networking

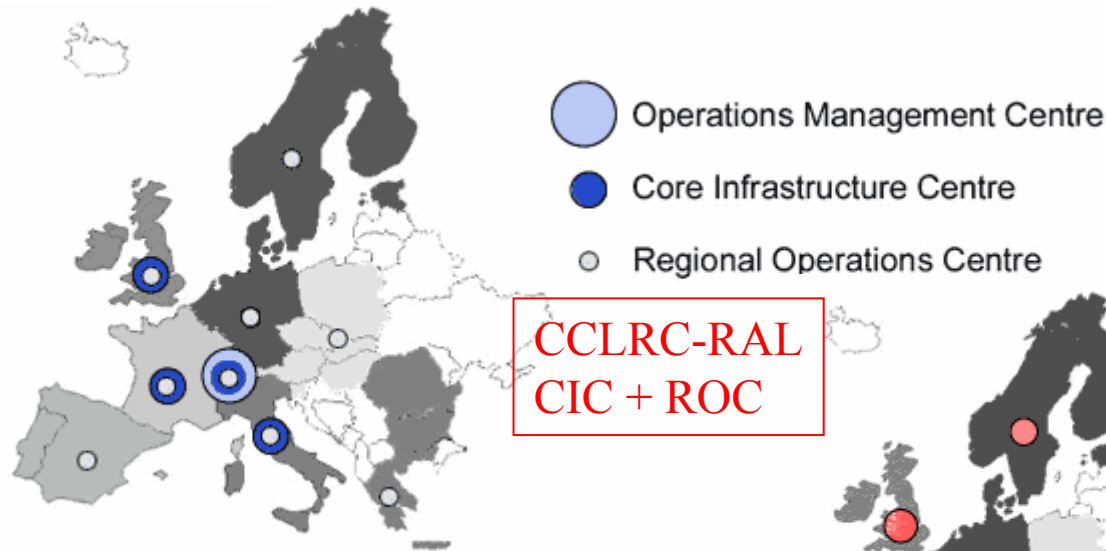**NA1**: Management

NA2: Dissemination and Outreach

NA3: User Training and Education

NA4: Application Identification and Support

NA5: Policy and International Cooperation

24% Joint Research

28% Networking

48% Services

48% Services

SA1: Grid Operations, Support and Management

SA2: Network Resource Provision

Emphasis in EGEE is on operating a production grid and supporting the end-users

John Gordon

e-Science Centre

EGEE – Key UK Activities

LRC

Operations Management Centre

Core Infrastructure Centre

Regional Operations Centre

Service activities

CCLRC-RAL
CIC + ROC

Middleware Integration and Testing Centre

Middleware Re-engineering Centre

Quality and Security Centres
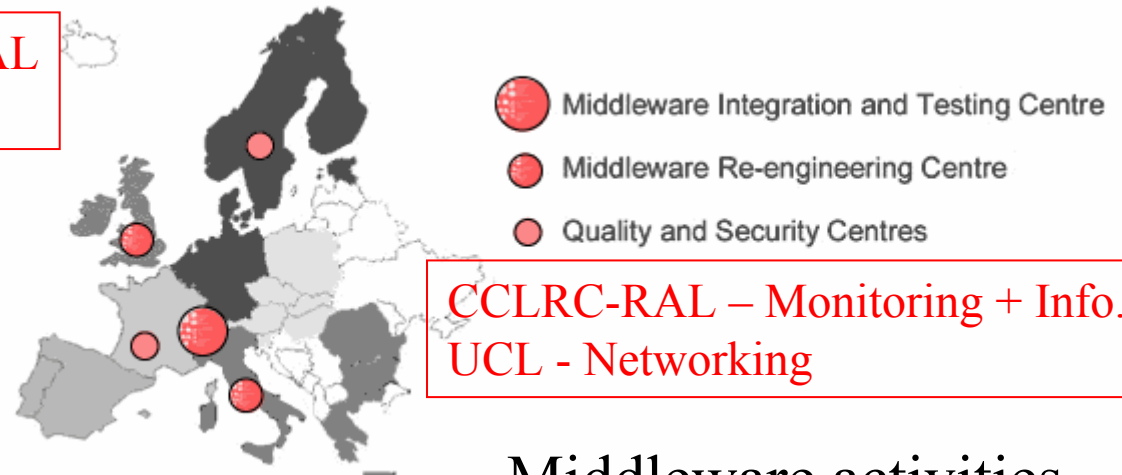
CCLRC-RAL – Monitoring + Info.
UCL - Networking

Middleware activities

Lead Networking Centre

Pilot Application Centre

Regional Networking Centre
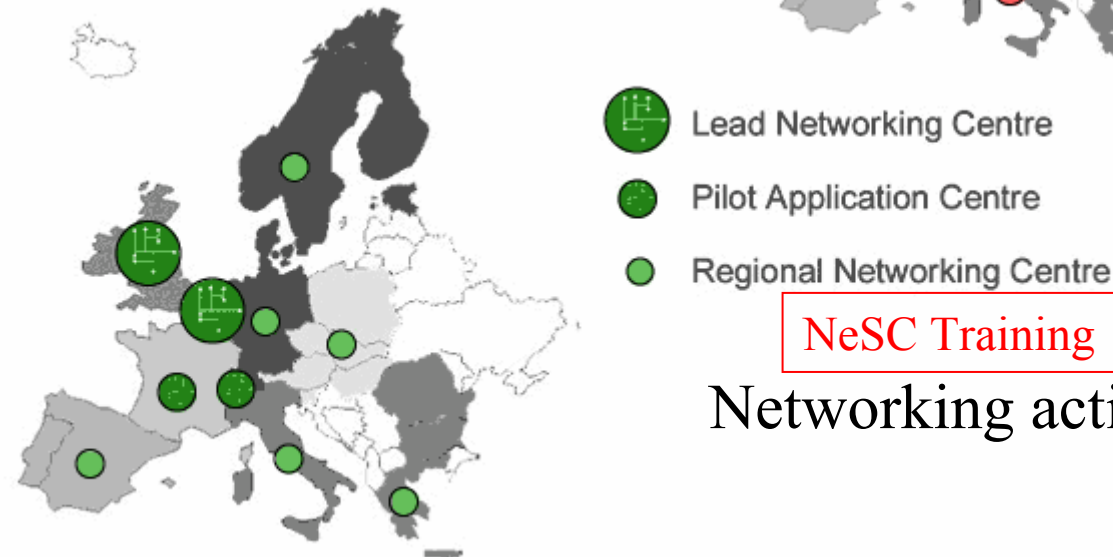
NeSC Training

Networking activities

John Gordon
e-Science Centre

What is LCG?

- Building a Production Grid to do the computing for the Large Hadron Collider
- Wider than Europe but only Particle Physics
- Four areas of work
    - Applications
        - Libraries, data management, interfaces
    - Fabrics
        - infrastructure for Tier0 Centre at CERN
    - Grid Technology
        - the computing model,
    - Grid Deployment
        - deploying a Production Grid

John Gordon
e-Science Centre

What is LCG?

Status for Resource Broker CERN_lxn1188: Fri Jun 4 07:45:54 BST 2004

- RAL is LCG Operations Centre
- Also an EGEE ROC
- Monitor GridPP (and NGS and GridIreland)
- Developed tools for LCG, reuse for GridPP
- Continue developing for EGEE
- EGEE CIC running grid-wide services
- Accounting

John Gordon
e-Science Centre

LCG Summary

- **LCG is geographically wider than EGEE, but limited to Particle Physics Applications**
  - EGEE has a wider set of VOs
- **LCG2 middleware forms the base release for EGEE**
  - but EGEE middleware will provide the next version for LCG
- **LCG Infrastructure in Europe forms the initial EGEE infrastructure**
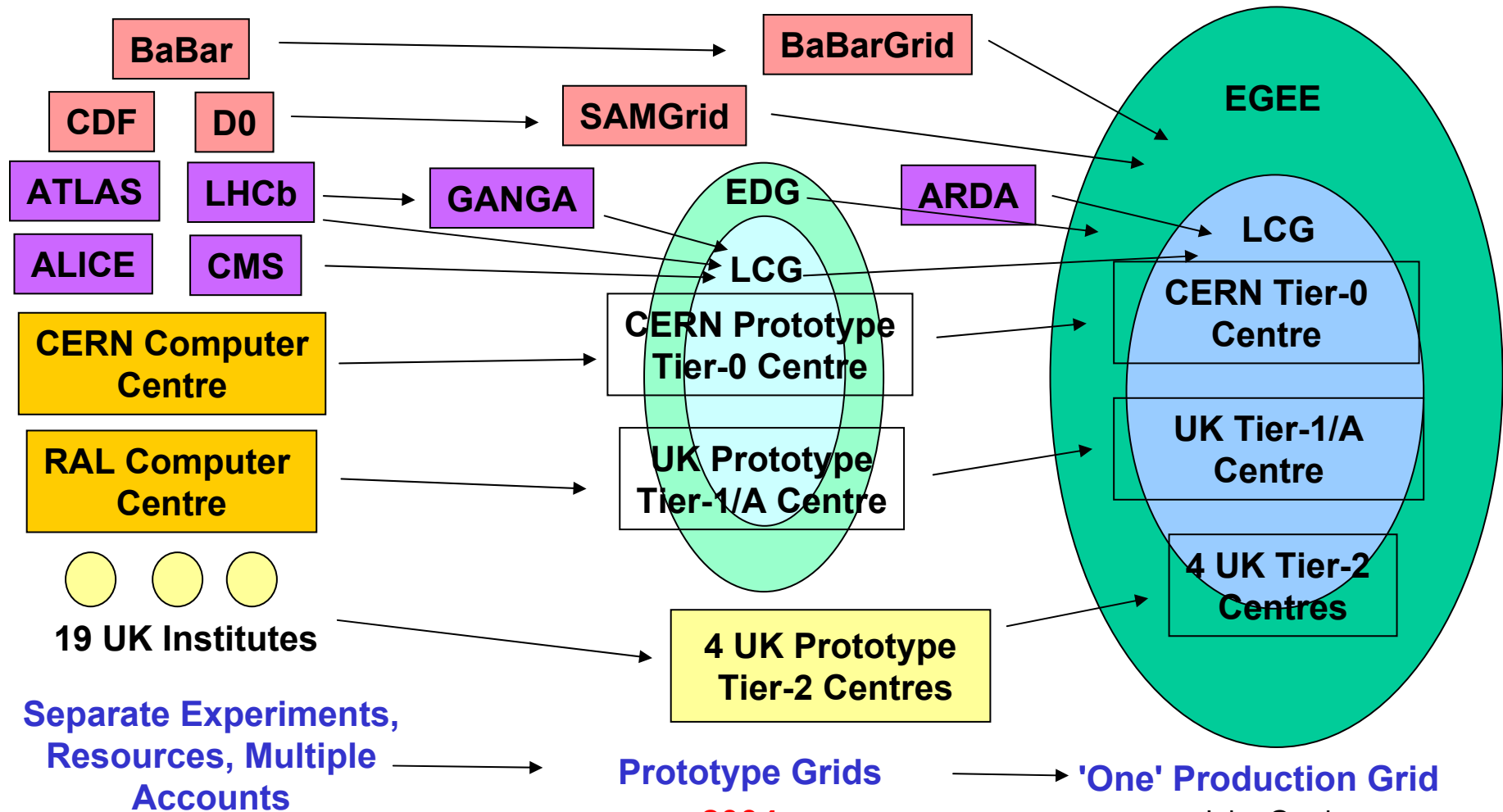 - but it will soon extend beyond this

John Gordon
e-Science Centre

CCLRC

GridPP

- 19 UK Universities, CCLRC (RAL & Daresbury) and CERN
- Funded by the Particle Physics and Astronomy Research Council (PPARC)
- GridPP1 - Sept. 2001-2004 £17m "From Web to Grid"
- GridPP2 – Sept. 2004-2007 £16(+1)m "From Prototype to Production"

GridPP Summary:
From Prototype to Production

CCLRC

BaBar → BaBarGrid

CDF   D0 → SAMGrid

ATLAS LHCb → GANGA

ALICE CMS

EGEE

EDG   ARDA

LCG

CERN Computer Centre → CERN Prototype Tier-0 Centre

RAL Computer Centre → UK Prototype Tier-1/A Centre

LCG

CERN Tier-0 Centre

UK Tier-1/A Centre

4 UK Tier-2 Centres

19 UK Institutes

4 UK Prototype Tier-2 Centres

Separate Experiments, Resources, Multiple Accounts → **Prototype Grids** → **'One' Production Grid**

2001

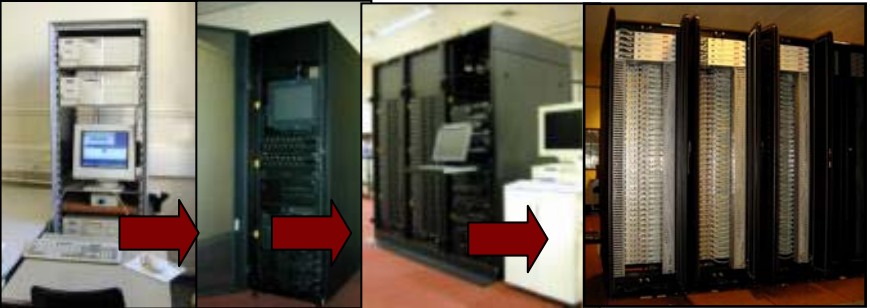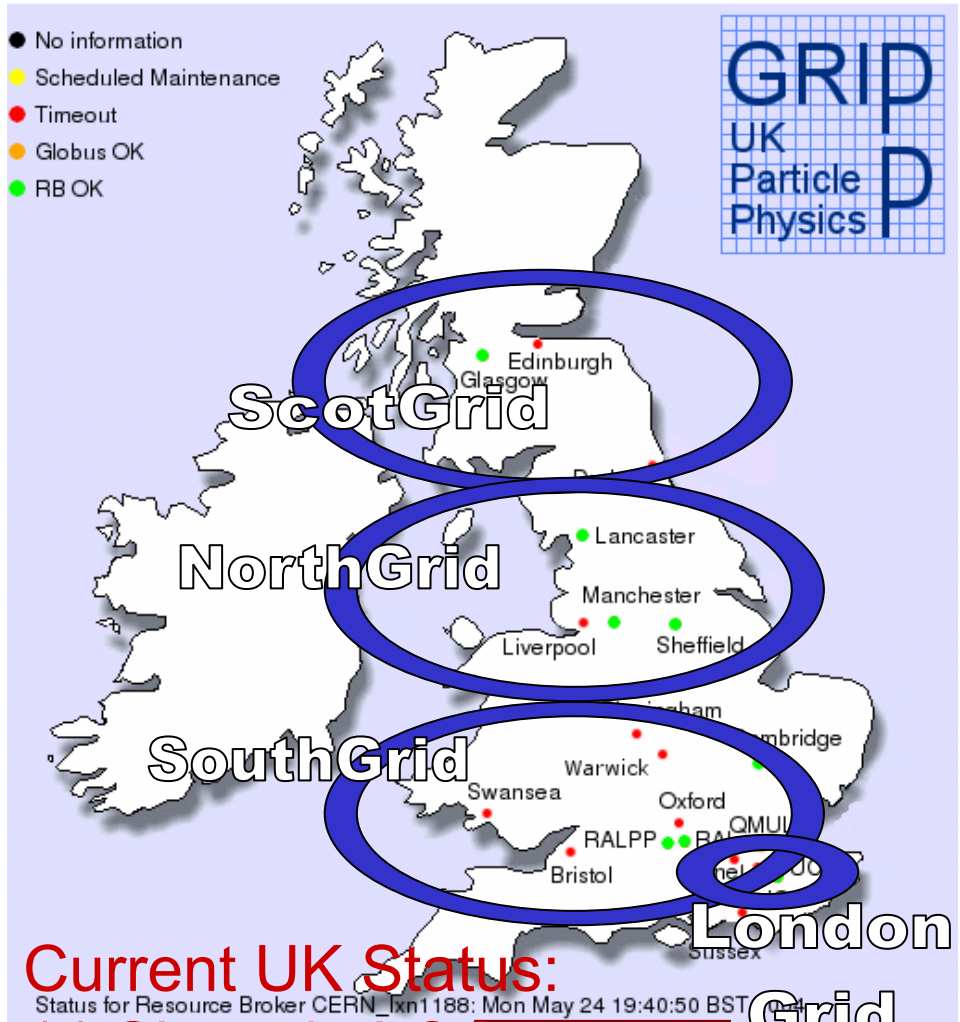**2004**

John Gordon
e-Science Centre

2007

CCLRC

NorthGrid ****
Daresbury, Lancaster, Liverpool, Manchester, Sheffield

SouthGrid *
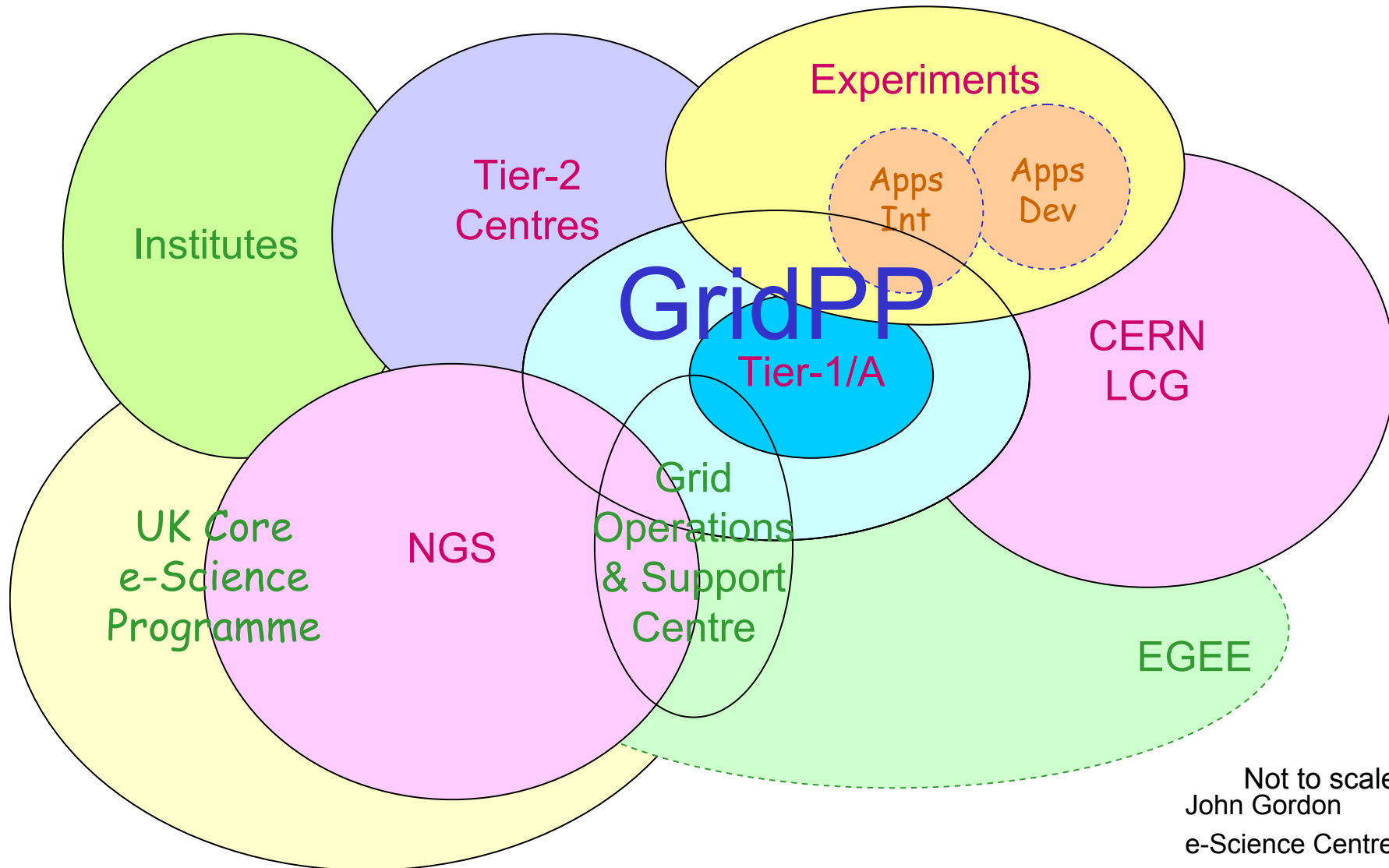Birmingham, Bristol, Cambridge, Oxford, RAL PPD, Warwick

ScotGrid *
Durham, Edinburgh, Glasgow

LondonGrid ***
Brunel, Imperial, QMUL, RHUL, UCL

- No information
- Scheduled Maintenance
- Timeout
- Globus OK
- RB OK

GRID
UK Particle Physics
P

ScotGrid

NorthGrid

SouthGrid

London Grid

Edinburgh
Glasgow
Lancaster
Manchester
Liverpool
Sheffield
Warwick
Swansea
Oxford
RALPP
QMUL
Bristol
Sussex

Current UK Status:
11 Sites via LCG

Status for Resource Broker CERN_lxn1188: Mon May 24 19:40:50 BST

e-Science Centre

- GridPP closely follows LCG
- Supports more than just LHC experiments
- Hopes through EGEE to participate in a single UK grid

John Gordon
e-Science Centre

CCLRC

Experiments

Tier-2 Centres

Institutes

Apps Int

Apps Dev

GridPP

Tier-1/A

CERN LCG

UK Core e-Science Programme

NGS

Grid Operations & Support Centre

EGEE

Not to scale!
John Gordon
e-Science Centre

John Gordon
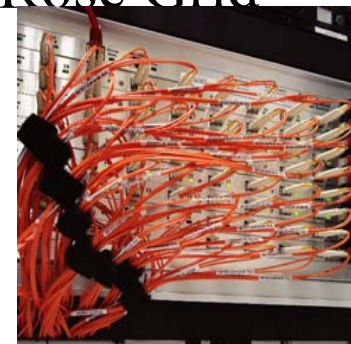e-Science Centre

CCLRC

- **UK – National Grid Service (NGS)**
  - JISC funded £1.06M Hardware (2003-2006)
    - Oxford – Compute Cluster, 128 CPUs, Myrinet
    - Manchester – Data Cluster, 18TB SAN, 40 CPUs Oracle, Myrinet
    - White Rose Grid (Leeds) - Compute Cluster, 128 CPUs, Myrinet
    - E-Science Centre funded 420k Data Cluster, 18TB SAN, 40 CPUs Oracle, Myrinet
    - CSAR and HPCx are also core members
  - JISC funded £900k staff effort including 2.5SY at CCLRC e-Science Centre
    - The e-Science Centre leads and coordinates the project for the JISC funded clusters
  - Production Grid Resources
    - Stable Grid Platform – gt2 with experimental gt3/4 interfaces to Data Clusters
    - Interoperability with other grids such as EGEE
    - Allocation and resource control
      - unified access conditions on JISC funded kit
    - Applications
      - As users require and licenses allow

e-Science Centre
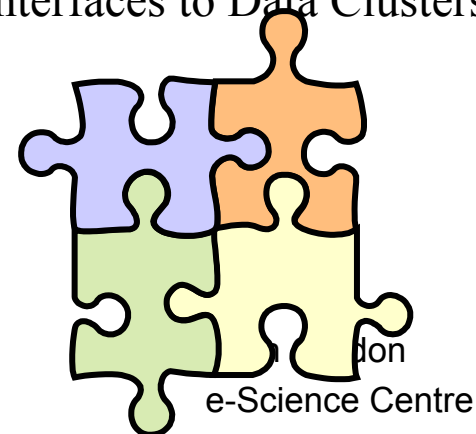
CCLRC

- **NGS today**
  - 2 x Compute Clusters (Oxford and White Rose Grid – Leeds)
    - 128 CPUs (64 Dual Xeon Processor Nodes)
    - 4TB NFS shared filespace
    - fast programming interconnect – Myrinet
  - Data Clusters (Manchester and CCLRC)
    - 40 CPUs (20 Dual Xeon Processor Nodes)
    - Oracle 9iRAC on 16 CPUs
    - 18TB fibre SAN
    - fast programming interconnect – Myrinet
  - In total 6 FTE's effort across all 4 sites.
    - just enough to get started
  - Plus HPCx and CSAR – National HPC services

John Gordon
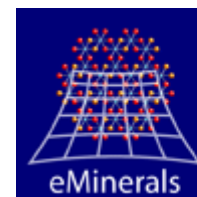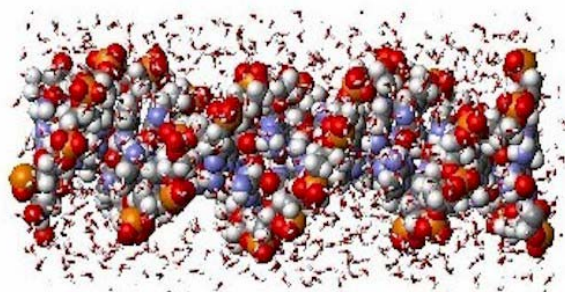e-Science Centre

Examples of NGS applications

Gaussian.com
THE OFFICIAL GAUSSIAN 03 WEB SITE

GEODISE
Grid Enabled Optimisation and Design Search for Engineering

Ab Initio Molecular Orbital Theory

eMinerals

RealityGrid
eMaterials

emboss

Fasta
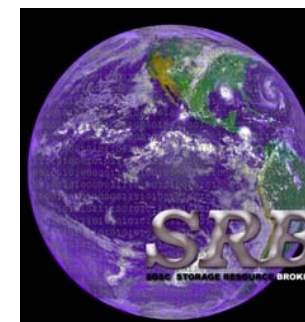
The DL_POLY Molecular Simulation Package

NC BLAST

NAG
Numerical Algorithms Group

ORACLE

SRB
e-Science Centre

Grids in Context

CCLRC

- Try to standardise on a common set of middleware and shared infrastructure
    - GT2 based now
- All Grids have plans to develop middleware, most based firstly on W3C Web Services and then WSRF
    - these will happen roughly in step

John Gordon
e-Science Centre

Summary

- Aim for a common Grid in the UK

- Participating in EGEE and wider projects

- A common set of middleware

- A shared infrastructure

- but not freely available to all

- Supporting a variety of Virtual Organisations with access control as determined by the local sites

- … and your job is to instrument the resources and connect them to the infrastructure

John Gordon
e-Science Centre