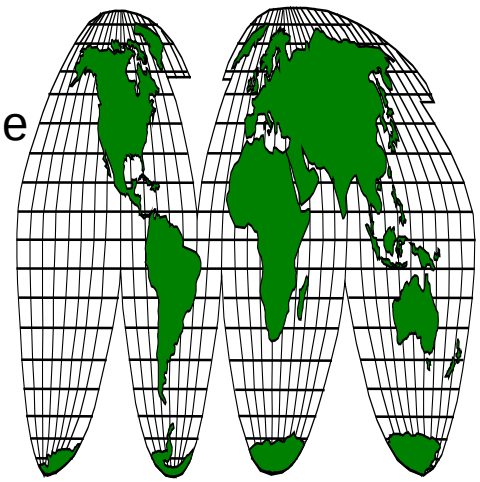
Oxford University
e-Science Centre

Digital Curation

e-Science Centre University perspective

Paul Jeffreys
Director Oxford e-Science Centre
http://e-science.ox.ac.uk/

paul.jeffreys@oucs.ox.ac.uk



Oxford University
e-Science Centre

What's in a name?!

- 'Oxford e-Science Centre' vs 'Oxford Regional Grid Centre'
- However:
  - IBM Press Release quote (Aug. 2001):
    - "The driving force behind the evolutionary Grid Project is the global scientific community'
  - e-Science funding!

- But in Oxford, much more than e-Science
- Aim of talk:
  - Explain e-Science regional centre
  - Oxford's work in data curation → connection

Front page FT, 7th Mar 2000

UK scientists expect £100m to build super-fast internet

By Clive Cookson, Science Editor

British scientists are preparing an ambitious plan to build a super-fast internet for the next generation.

"The grid", as it is provisionally known, will work far more quickly and reliably than today's internet. It should eventually enable computer users to receive exactly the information they want from anywhere in the world within seconds – and without having to go through a tortuous search process.

The government is expected to announce about £100m in public funding for the grid proposal in this year's spending review.

The concept of the grid as a computing infrastructure emerged in the US in the late 1990s. A few experimental grids are being built there with funding of about $100m (£62m) a year from federal agencies.

The grid introduces a new set of "middle-ware" – programs to facilitate collaborative working.

The UK work will be carried out in collaboration with Cern, the

times more processing and networking power than Cern can handle today.

"The challenge is real and must be met on time," said Chris Jones, Cern's head of technology

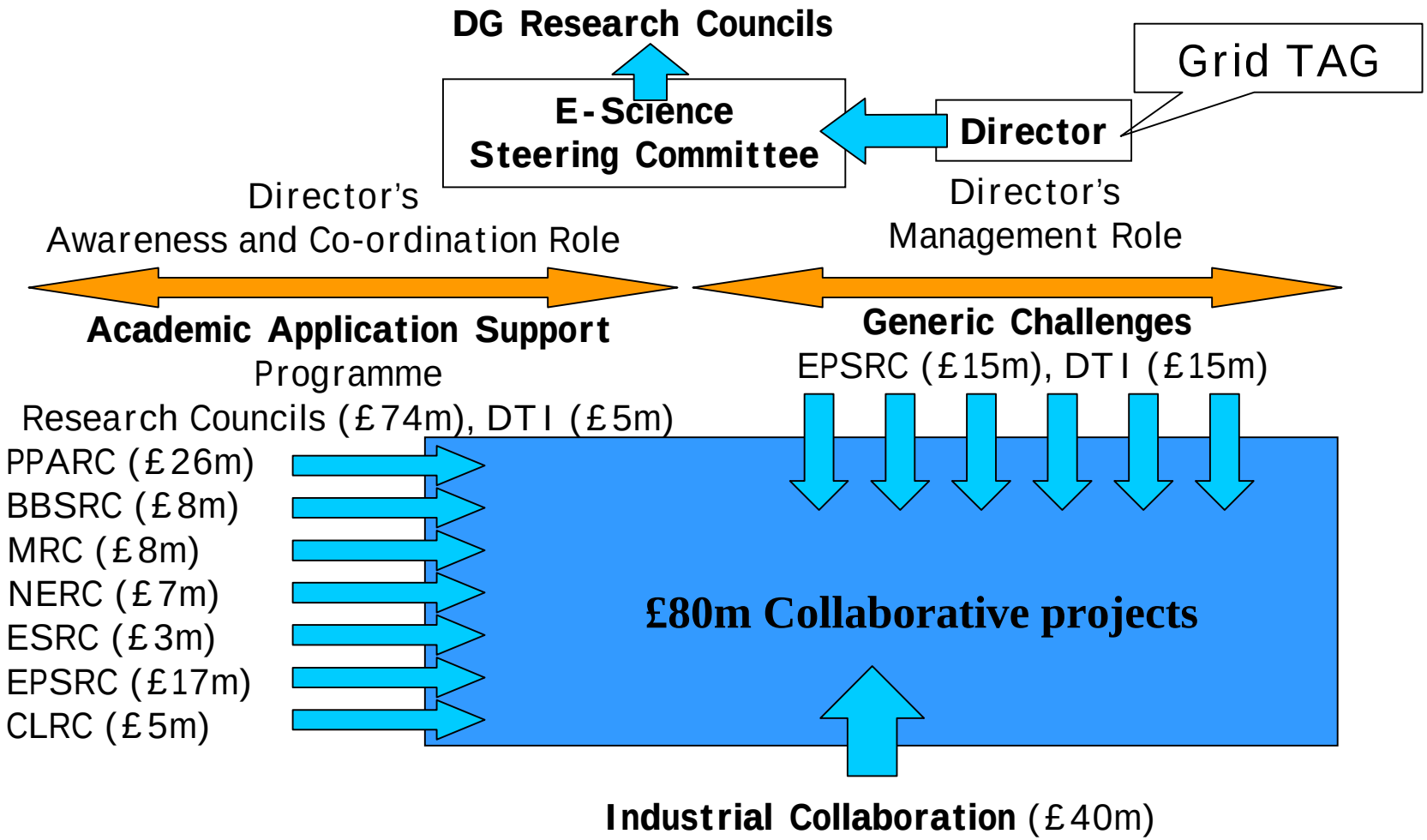
best way forward.

Cooking up "the grid", Page 13

"'The Grid', as it is provisionally known, will work far more quickly and reliably than today's internet. It should eventually enable computer users to receive exactly the information they want from anywhere in the world within seconds – and without having to go through a tortuous search process."

CONTENTS

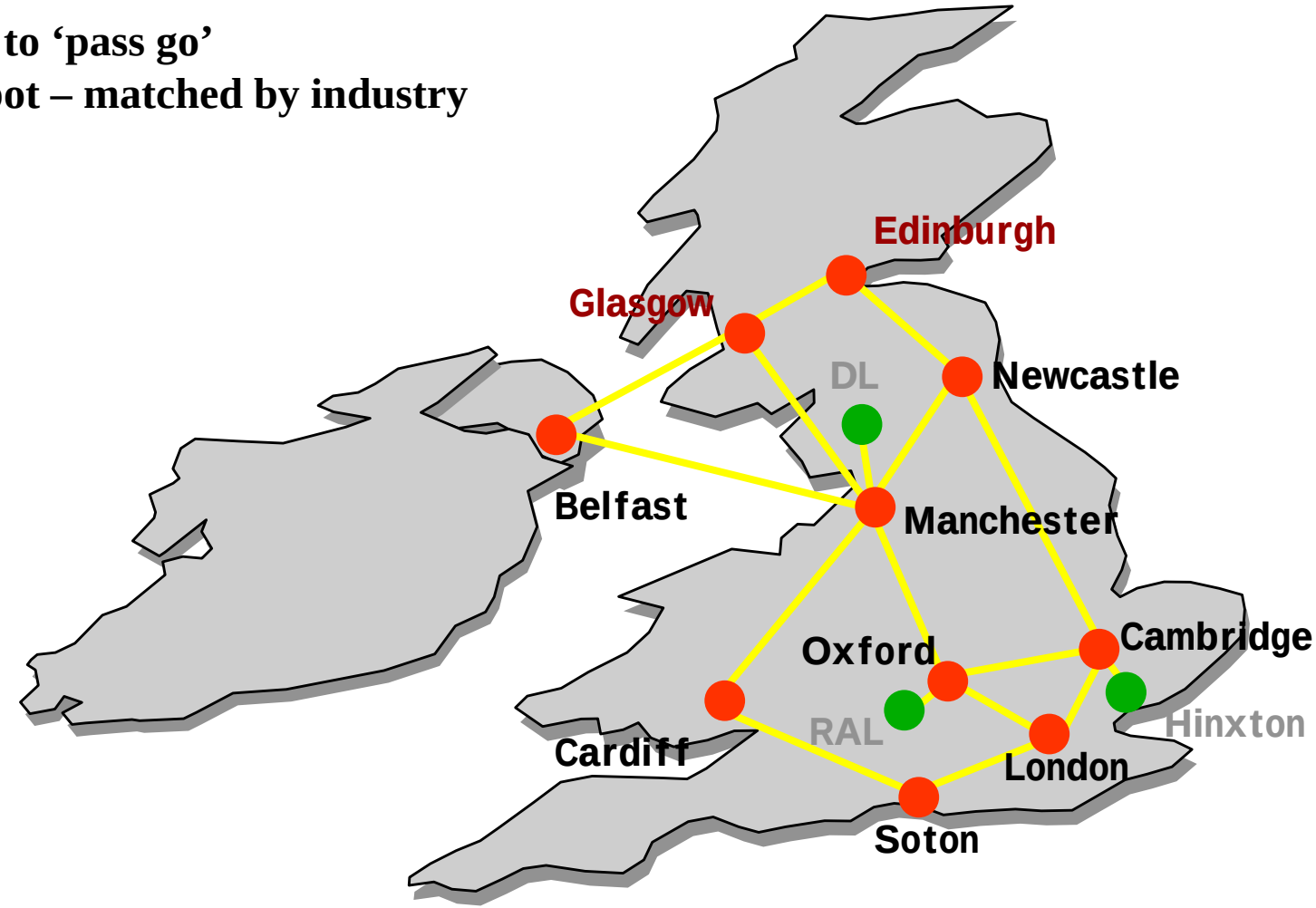
SR2000 e-Science Allocation



UK Grid Network

Oxford University
e-Science Centre

£470k to ‘pass go’
£1M pot – matched by industry



Status of Oxford e-Science Centre

Oxford University e-Science Centre

- Formally launched 17 August (Mike Brady, Mike Giles, Todd Huffman and Mark Sansom, David Gavaghan, David Shotton, and Paul Jeffreys)
  - Web site … http://e-science.ox.ac.uk
  - Administrative Centre and Point-of-Contact (Oxford University Computing Services)
    - Document repository (considering within Grid context!)
  - Grid afternoon organised … 17 September
  - Interview with BBC World News24
  - Commissioning underway

OeSC Team

Oxford University
e-Science Centre

- Director: Paul Jeffreys
- Associate Technical Directors:
  - Technical — Mike Giles
  - Dissemination — David Gavaghan
  - Application — Mark Sansom
- Technical Contacts:
  - GLOBUS — Ray Miller
  - Condor — Charles Curran
  - Web — Malcolm Austen and Sebastian Rahtz
  - 'Web service' — Matthew Dovey
  - Digital Curation — Lou Burnard, Dave Price
- Point-of-Contact:
  - Sue Crowley, sue.crowley@oucs.ox.ac.uk, 01865 273229

Oxford University
e-Science Centre

OeSC Management

- Substantial/important activity; needs appropriate structure
  - Management Board
    - Overall 'responsibility' for delivery and operation of functional e-Science Centre, resources invested, accountability, outside relations and integration, and overseeing all e-Science and Grid activities in OU
  - Technical/Development Board
    - High level management of e-Science/Grid activities, internal resources invested on OeSC, coordination
  - Technical Committee
    - Detailed internal technical management and organisation
  - User Committee
    - Forum for Users working on e-Science and Grid projects, and for dissemination across the University

[Must have representation wider than just e-Science!]

Access Grid

- Collection of resources that support formal and informal group-to-group interaction across the grid
- Supports large-scale distributed meetings and collaborative work sessions

http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm

Oxford University
e-Science Centre

OeSC Access Grid Commissioning

- Commissioned in Computing Laboratory
- Team established from Computer Science and OU Computing Services; advice taken
- Purchasing complete – kit arriving
- Operational 'in November'
- Dedicated connection to OU backbone

- Second installation in hospital area
  - funds requested as part of OeSC bid!
- Oxford Internet Institute interest…

Grid Starter-kit/Commissioning

- OeSC Commissioning
  - Team established in Computing Services (OUCS)
  - Installed on equipment housed in OUCS machine room
  - GLOBUS and Condor installed on PC, and on Linux cluster shortly
  - Aim to specialise on 'full service aspects' of Grid operation
    - Special interest to IBM

Equipment Offered

- Linux cluster (16 node, Pentium IIIs)
- PC running standard software
- OSCAR to be upgraded (Jan-Feb) .. At least 5%
- Share of resources collected through RC e-Science funds
  - Particular effort to 'connect' with GridPP
  - … and to make GridPP resources available

Oxford University
e-Science Centre

Oxford and digital curation

- Oxford partner in CEDARs project
  - OAIS reference model used for exemplars in Digital Archiving
- Humanities Computing Unit
  - Oxford is key player in national DNER via Oxford Text Archive (part of AHDS) and Humbul (part of RDN)
  - Oxford Text Archive –
    - Collecting and curating since 1976 (mainly textual)
    - Preserve digital outputs of UK scholars in areas of literary and linguistic research
    - Working with Humbol Humanities Hub to develop an Arts and Humanities Portal using Dublin Core and Z39.50
    - The joint activity will investigate ways in which Grid activities can be used to manage and deliver digital content
- Participating in the EEBO Text Creation Partnership
  - Aims to produce full-text transcriptions of selected texts from Early English Books
    - Handling the images is the challenge for storage

Oxford University
e-Science Centre

Related Library Activities at Oxford

- Norbert Lossau recruited:
  - Optimise digitisation process for Library
  - Spend Mellon Foundation funds for faculty led initiatives
  - Define intellectual and technical infrastructure for Oxford Digital Library
- Encoded Archival Description being developed for describing at 'collection level' complemented by 'Master' XML-encoding for item level description
- DIENST, METZ under investigation
- Oxford is one of four world wide hosts for the Text Encoding Initiative consortium, maintaining the *de facto* standard for digital libraries
- ***The 'Challenge': develop search and retrieval technologies which can seamlessly and sensibly combine information from texts, images, databases and multimedia sources***

Grid-related Library Activities

- Library Information Management
  - *Distributed* storage for Digital Copyright Deposit
    - Intermemory
    - DIENST
  - Integration of existing library protocols
    - Z39.50 (already part of SRB)
    - OpenURL
    - → to ensure co-existence
  - Web Service Model

    (function-specific services are built into an overall service which results in an architecture which is similar to the Grid)
    - Service Discovery – UDDI (directory)
    - Service Description – WSDL/WSCL (interaction with service)
    - Service Interactions – XML/SOAP

Possible OeSC (Industrial) Projects

- Web Service based Grid approach to Information Access
  - Explore models for locating, accessing, manipulating data from multiple heterogeneous sources
  - Explore new models for searching/indexing very large datasets

- A Grid Service
  - Understand the implications of operating within a University environment, possibly focusing on a specific application, and study the service issues using in-house security and network expertise

Oxford University
e-Science Centre

Conclusions

- e-Science Centres are becoming established across the country
  - OeSC is well advanced, managerially and technically
- Will become 'core' part of University life
- Although global science is driving the initiative, the interest is much wider
- Oxford (Digital) Library and OTA/Humbul potentially important part of the Grid activities
  - Lou Burnard, Michael Popham, Dave Price and Matthew Dovey
  - See real opportunities for collaboration; areas of overlap which may cross-fertilise
  - Aim to construct 'Industrial Project' in the Digital Curation area (strong commercial interest)