UK e-Science
OGC Technical Committee
Edinburgh

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

www.nesc.ac.uk
28th June 2006

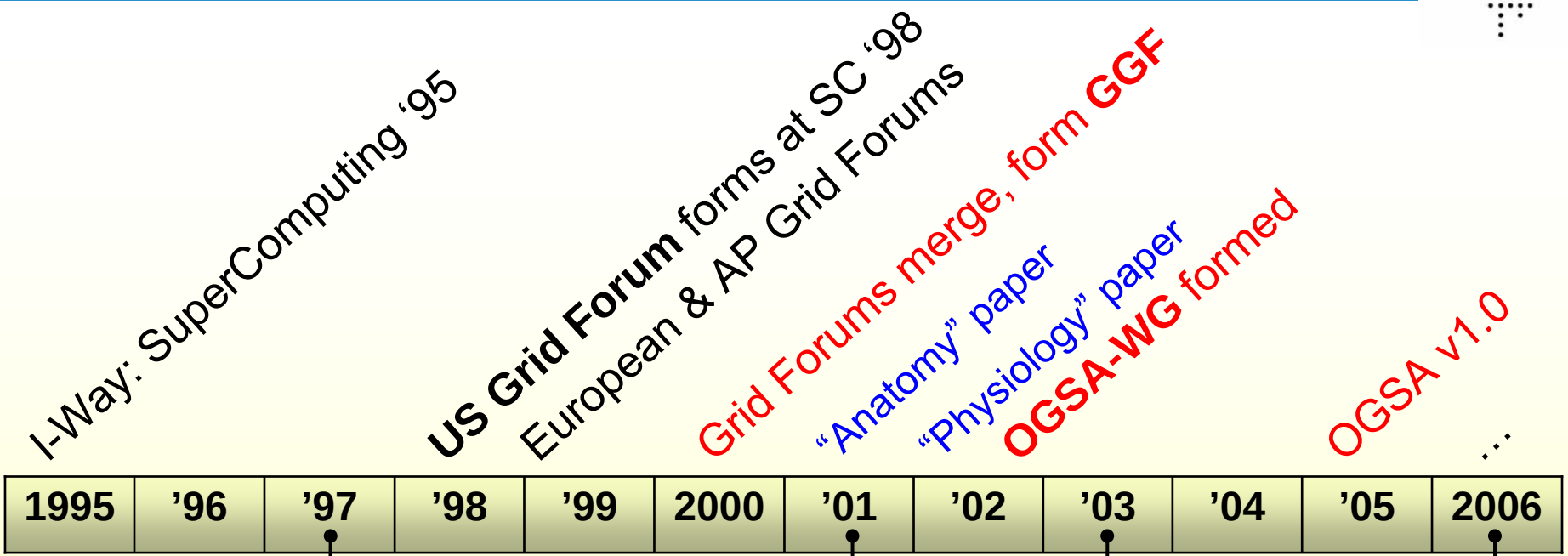
Overview

- Brief History
  - E-Science, Grids & Service-oriented Architectures
- (Geo)Data Deluge
  - Causes of Growth
  - Interpretational challenges
- Crucial Issues
  - 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*

A Grid Computing Timeline



- **DARPA** funds **Globus Toolkit & Legion**
- **EU** funds **UNICORE** project
- **US DoE** pioneers grids for scientific research
- **NSF** funds **National Technology Grid**
- **NASA** starts **Information Power Grid**

- **UK e-Science** program starts

Japan government funds:
- **Business Grid** project
- **NAREGI** project

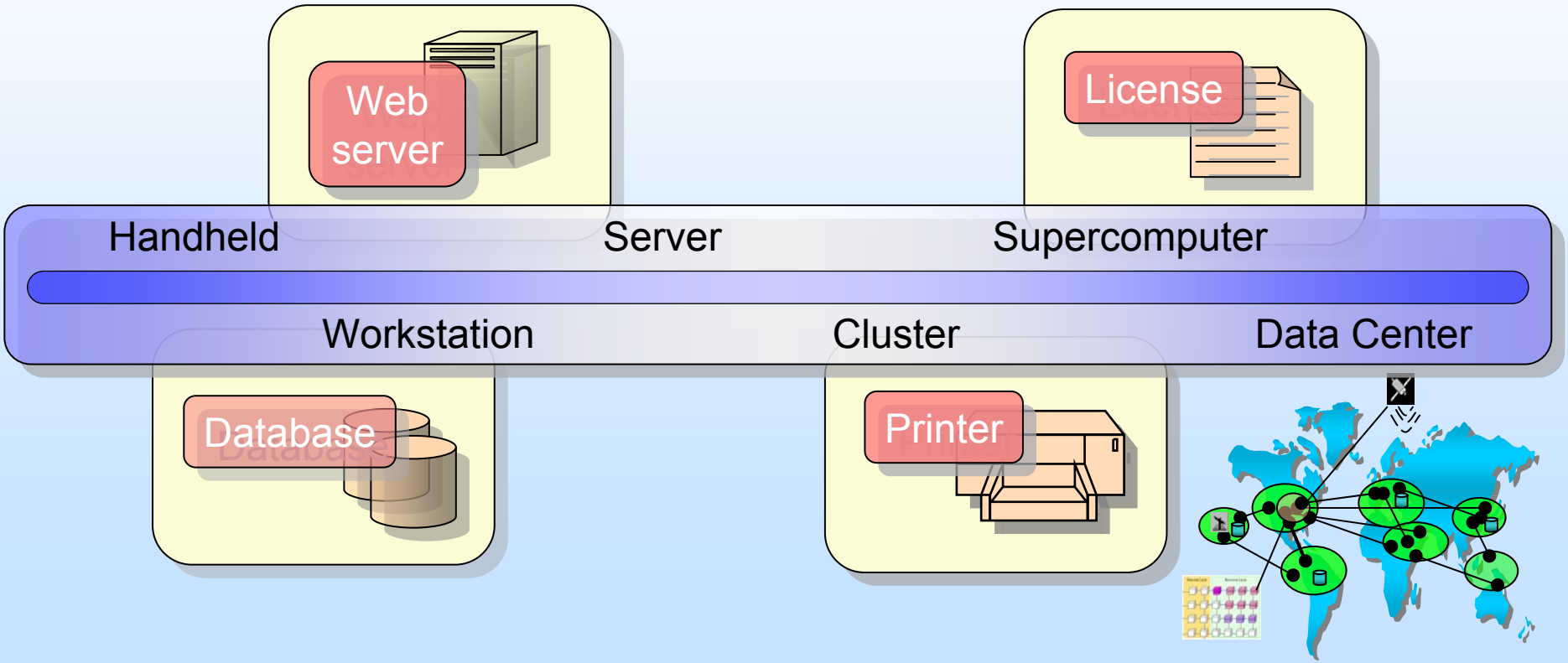
Today:
- Grid solutions are common for **HPC**
- Grid-based **business** solutions are becoming common
- Required **technologies & standards** are evolving

Source: Hiro Kishimoto GGF17 Keynote May 2006

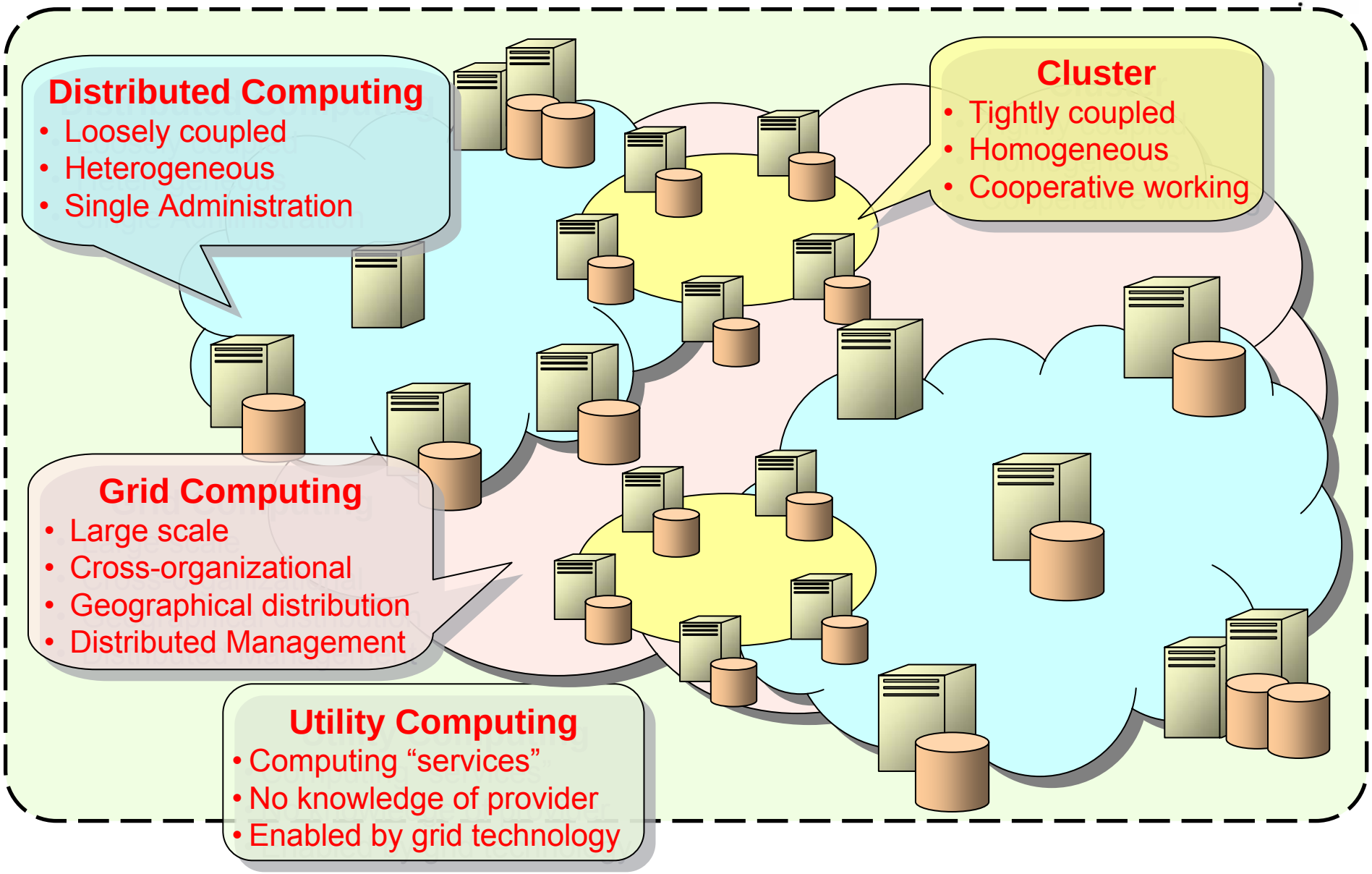
What is a Grid?

A grid is a system consisting of

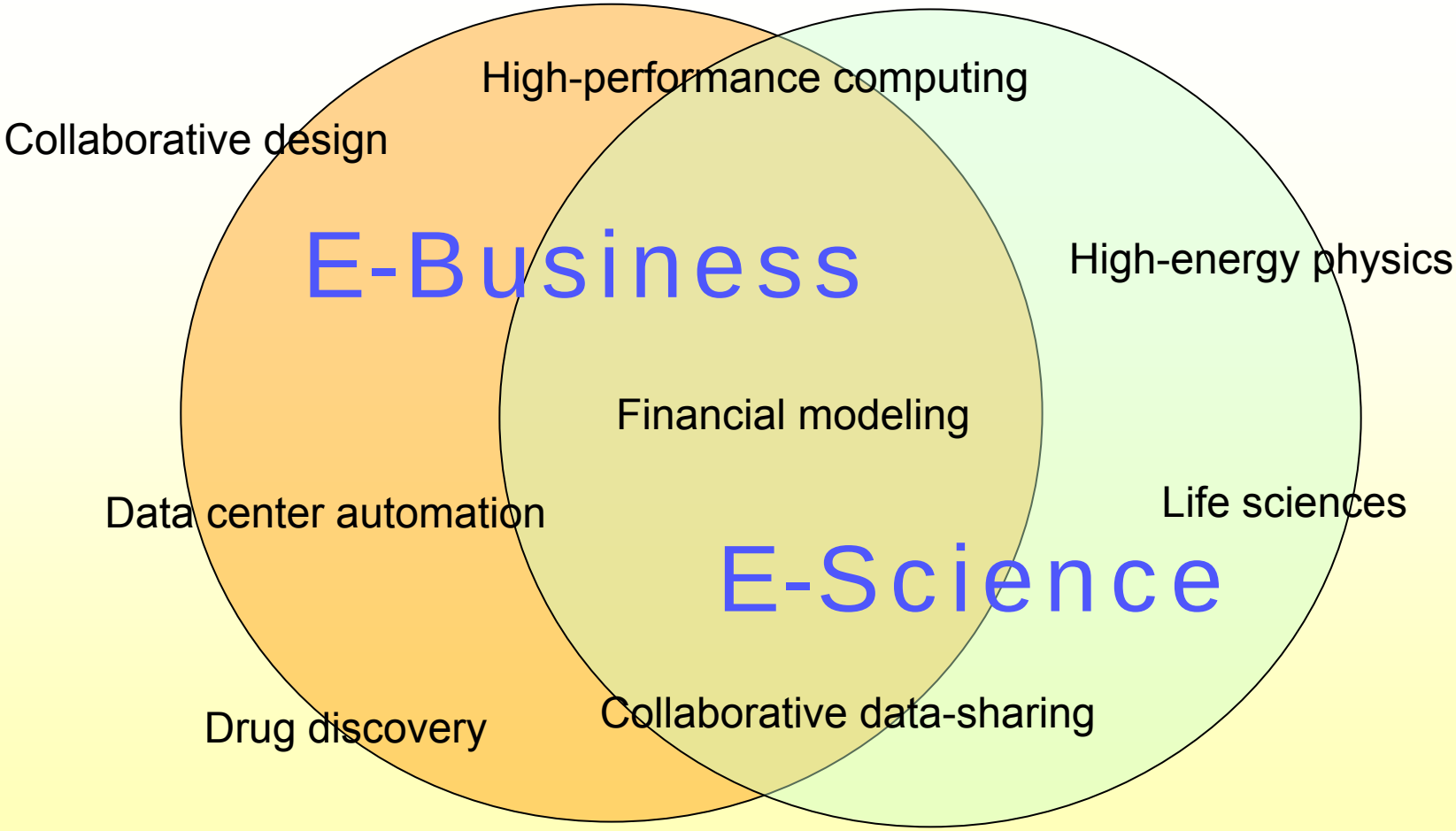
- Distributed but connected resources and
- Software and/or hardware that provides and manages logically seamless access to those resources to meet desired objectives



Grid & Related Paradigms

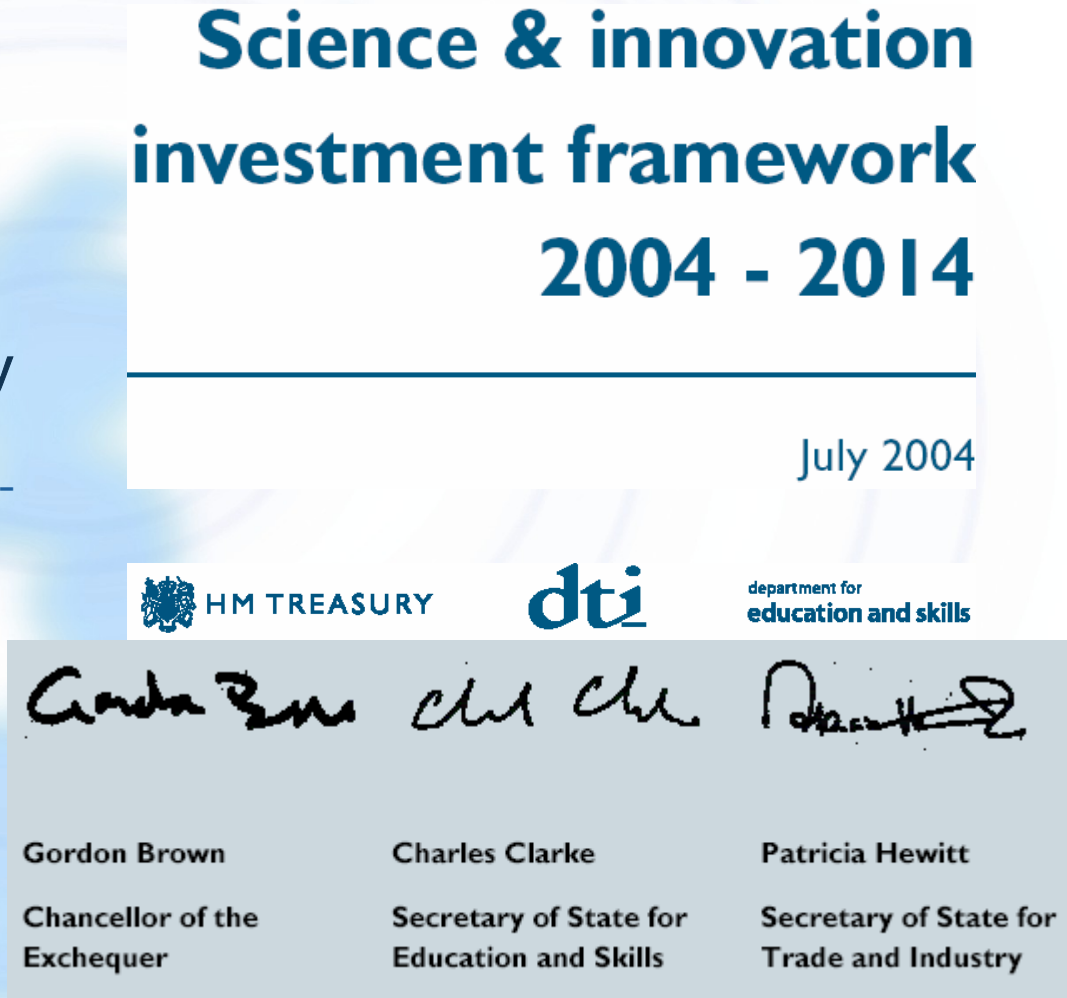


How Are Grids Used?



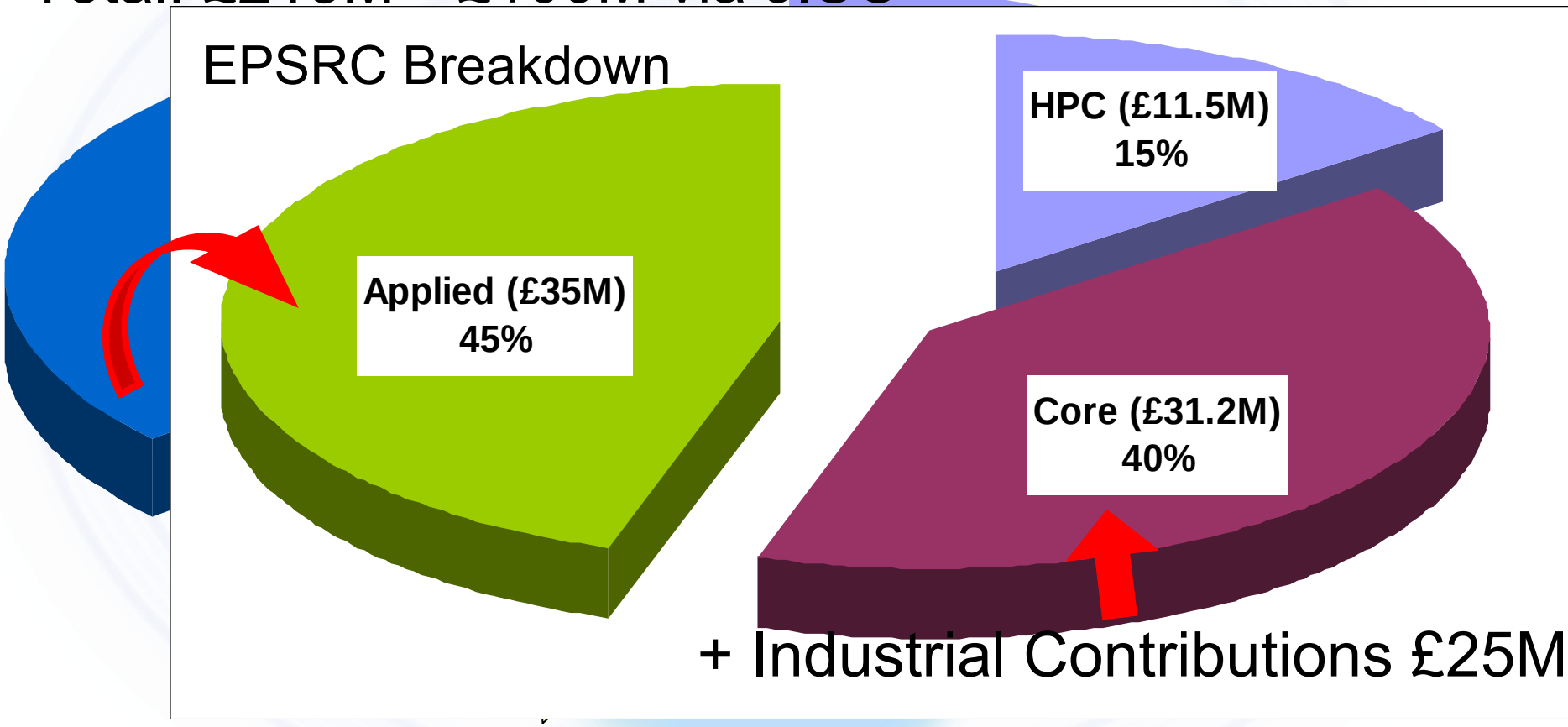
Commitment to e-Infrastructure

- 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



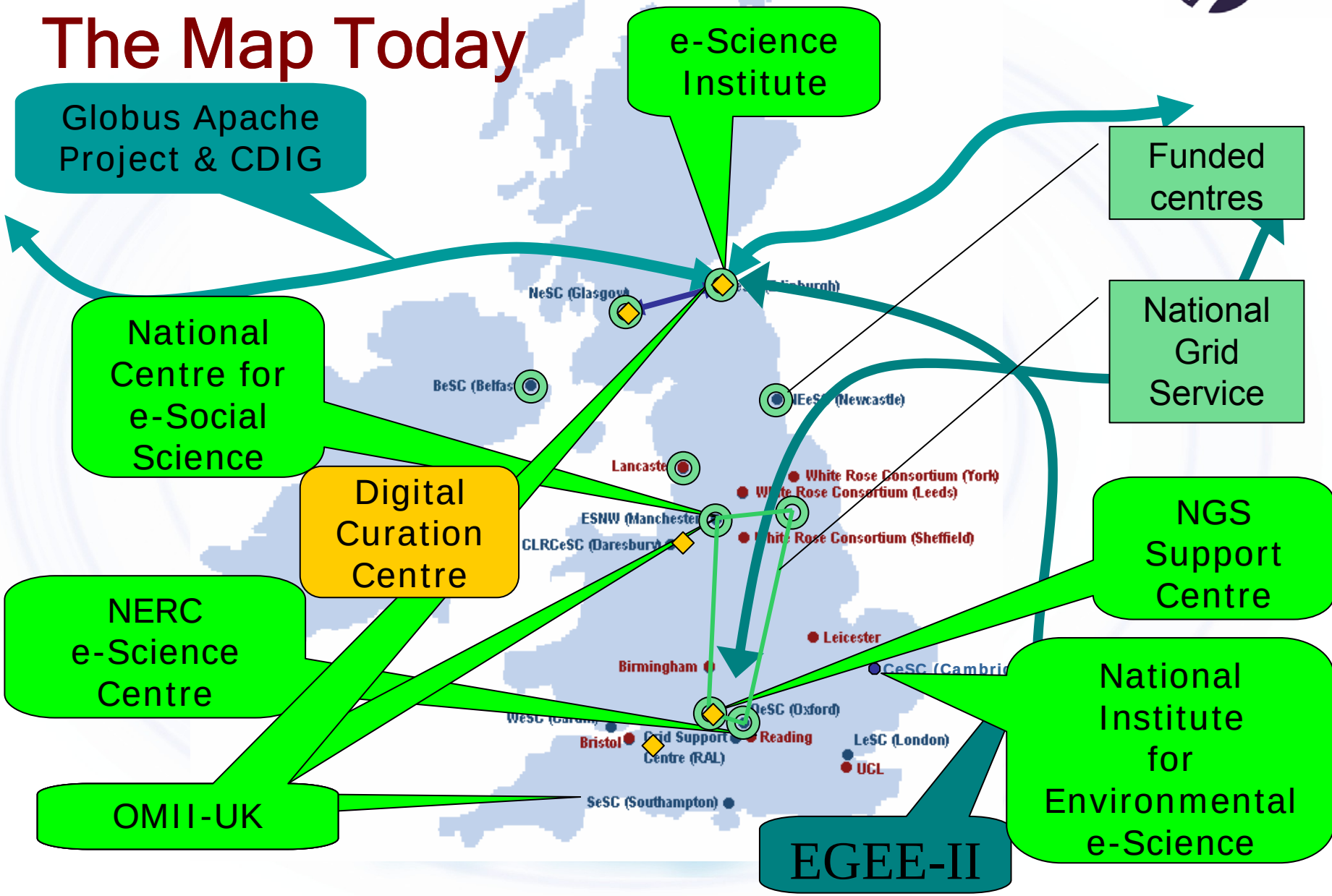
UK e-Science Budget (2001-2006)

Total: £213M + £100M via JISC



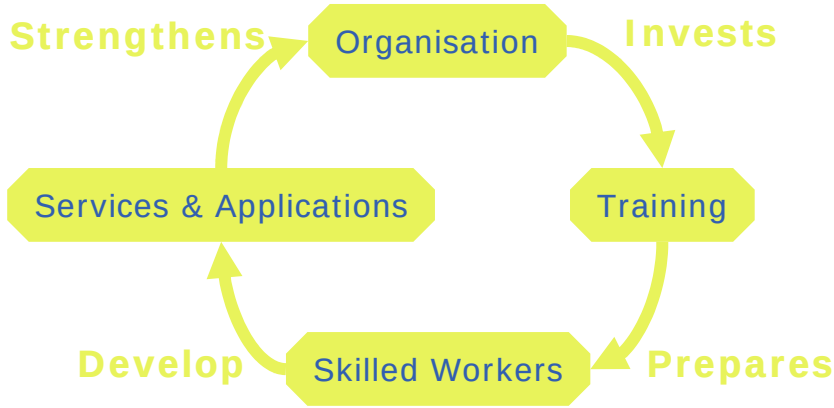
Source: Science Budget 2003/4 – 2005/6, DTI(OST)

The e-Science On The Map Today

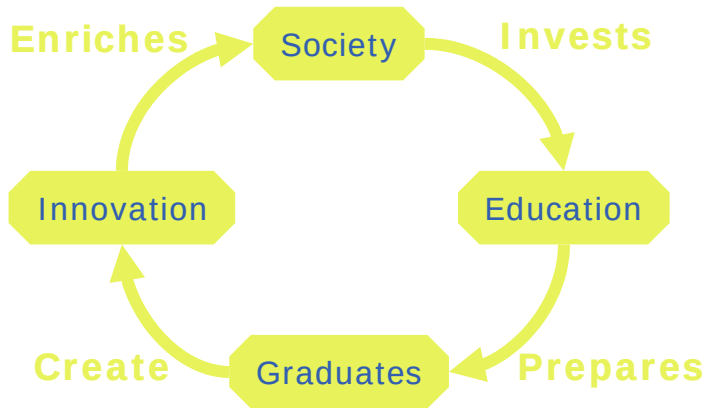


Invest in People

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



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



- **Both are needed**

(Geo)Data Deluge

Compound Causes of (Geo)Data Growth

- Faster devices
- Cheaper devices
- Higher-resolution
  - all ~ Moore's law
- Increased processor throughput
  - ⇒ more derived data
- Cheaper & higher-volume storage
- Remote data more accessible
  - Public policy to make research data available
  - Bandwidth increases
  - Latency doesn't get less though

Interpretational Challenges

- Finding & Accessing data
  - Variety of mechanisms & policies
- Interpreting data
  - Variety of forms, value systems & ontologies
- Independent provision & ownership
  - Autonomous changes in availability, form, policy, …
- Processing data
  - Understanding how it may be related
  - Devising models that expose the relationships
- Presenting results
  - Humans need either
    - Derived small volumes of statistics
    - Visualisations

Interpretational Challenges

- Finding & Accessing data
  - Variety of mechanisms & policies
- Interpreting data
  - Variety of forms, value systems & ontologies
- Independent provision & ownership
  - Autonomous changes in availability, form, policy, …
- Processing data
  - Understanding how they may be related
  - Devising models that expose the relationships
- Presenting results
  - Humans need either
    - Derived small volumes of statistics
    - Visualisations

Variety & Autonomy Essential

Interpretational Challenges

- Finding & Accessing data
  - Variety of mechanisms & policies
- Interpreting data
  - Variety of forms, value systems & ontologies
- Independent provision & ownership
  - Autonomous changes in availability, form, policy, …
- Processing data
  - Understanding how it may be related
  - Devising models that expose the relationships
- Presenting results
  - Humans need either
    - small volumes of statistics
    - visualisations

Standards & Collaboration Essential

Crucial Issues

Collaboration

- Collaboration is a Key Issue
  - Multi-disciplinary
  - Multi-national
  - Academia & industry
- Trustworthy data sharing key for collaboration
  - Plenty of opportunities for research and innovation
  - Establish common frameworks where possible
    - Islands of stability – reference points for data integration
  - Establish international standards and cooperative behaviour
    - Extend incrementally
- Trustworthy code & service sharing also key

Federation

- Federation is a Key Issue
  - Multi-organisation
  - Multi-purpose
  - Multi-national
  - Academia & industry
- Build shared standards and ontologies
  - Require immense effort
  - Require critical mass of adoption
- Trustworthy code & e-Infrastructure sharing
  - Economic & social necessity

Major Intellectual Challenges

- Require
  - Many approaches to be integrated
  - Many minds engaged
  - Many years of effort
- Using the Systems
  - Requires well-tuned models
  - Well-tuned relationships between systems & people
  - Flexibility, adaptability & agility
- Enabling this
  - Is itself a major intellectual challenge

Questions & Comments Please