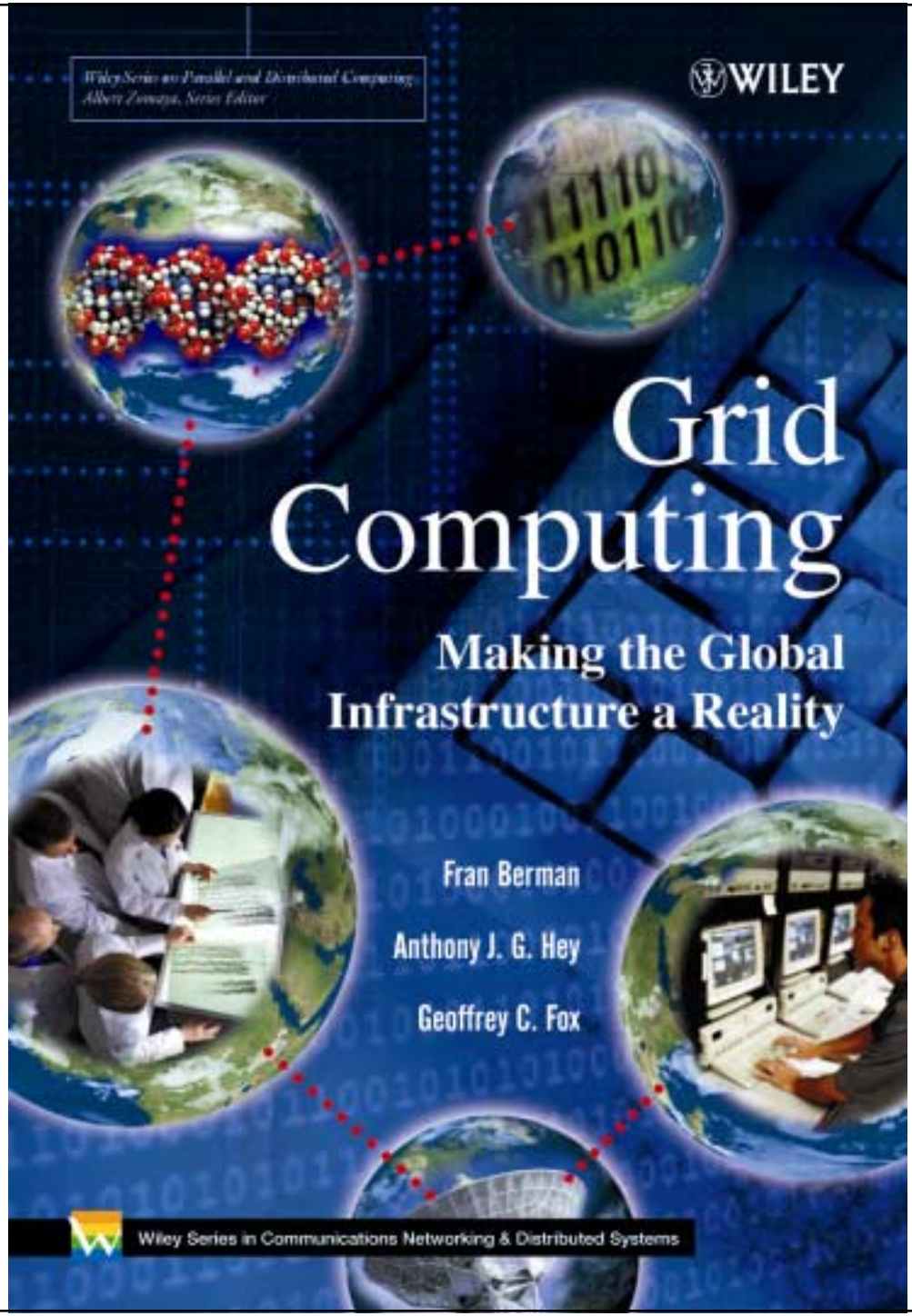
e-Science, Virtual Organisations and the Grid

Tony Hey,
Director of UK e-Science Core Programme
EPSRC, UK

e-Science and the Grid

‘e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.’

John Taylor
Director General of Research Councils
Office of Science and Technology


Wiley Series on Parallel and Distributed Computing
Albert Zomaya, Series Editor
WILEY
Grid Computing
Making the Global Infrastructure a Reality
Fran Berman
Anthony J. G. Hey
Geoffrey C. Fox
Wiley Series in Communications Networking & Distributed Systems

What is the Grid?

- *"A new technology infrastructure, which we call the GRID, will emerge to supersede the Web, and form the bedrock of the new economy" (Deutsche Bank)*
- *"Web and computing services shall go the way of the electric power grid – always available, easily accessible, and predictably expandable" (PricewaterhouseCoopers)*

The Grid as an Enabler for Virtual Organisations

- Ian Foster, Carl Kesselman and Steve Tueke

  'The Grid is a software infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources'

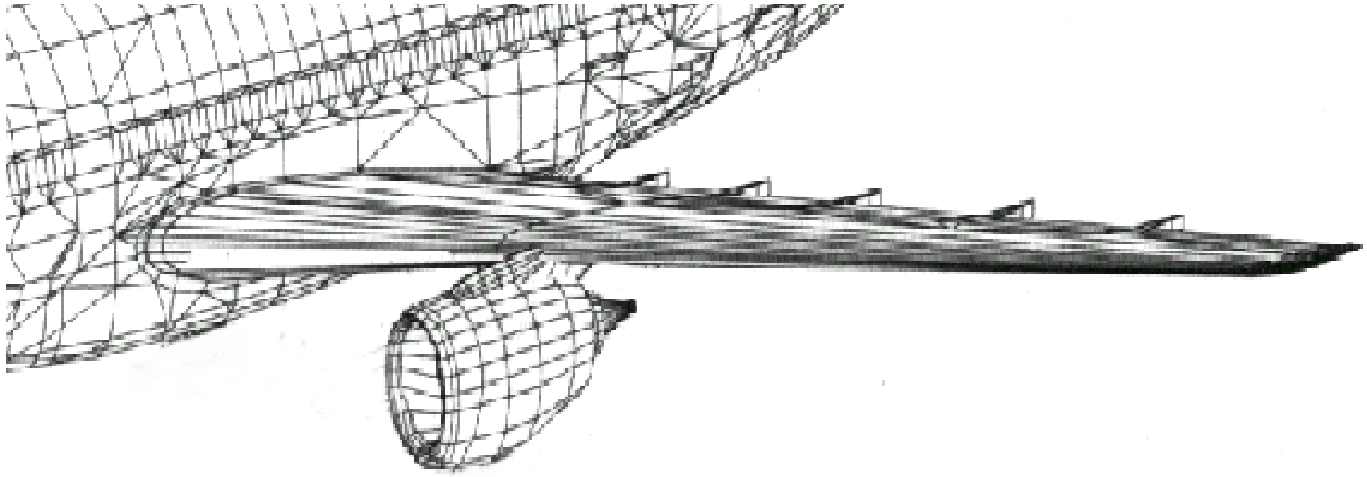
- Enabling infrastructure for transient 'Virtual Organisations' not only for e-Science but also c-Commerce and e-Utilities

NASA's IPG

Vision for the Information Power Grid is to promote a revolution in how NASA addresses large-scale science and engineering problems by providing persistent infrastructure for:

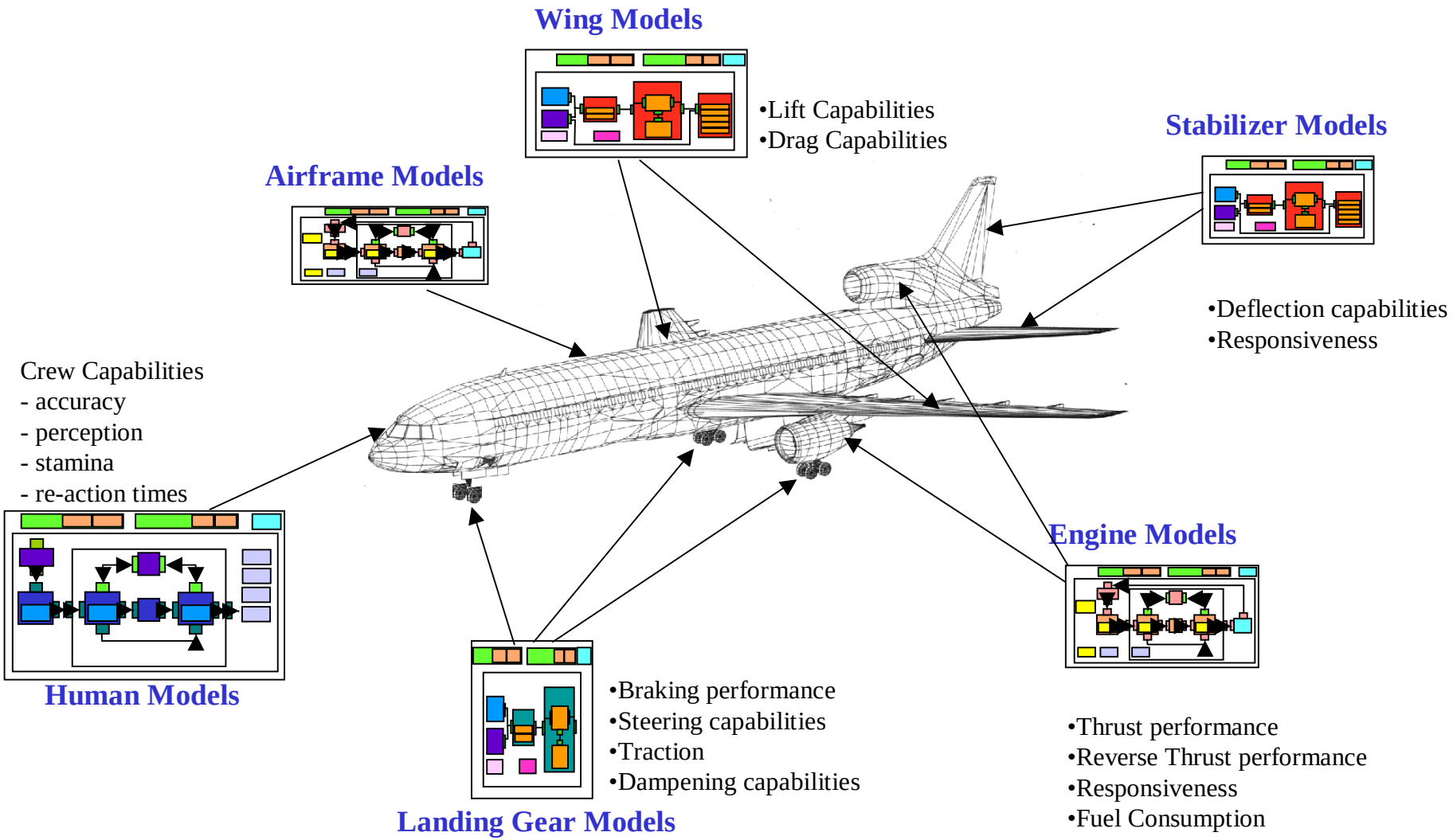
- "highly capable" computing and data management services that, on-demand, will locate and co-schedule the multi-Center resources needed to address large-scale and/or widely distributed problems
- the ancillary services that are needed to support the workflow management frameworks that coordinate the processes of distributed science and engineering problems

Multi-disciplinary Simulations



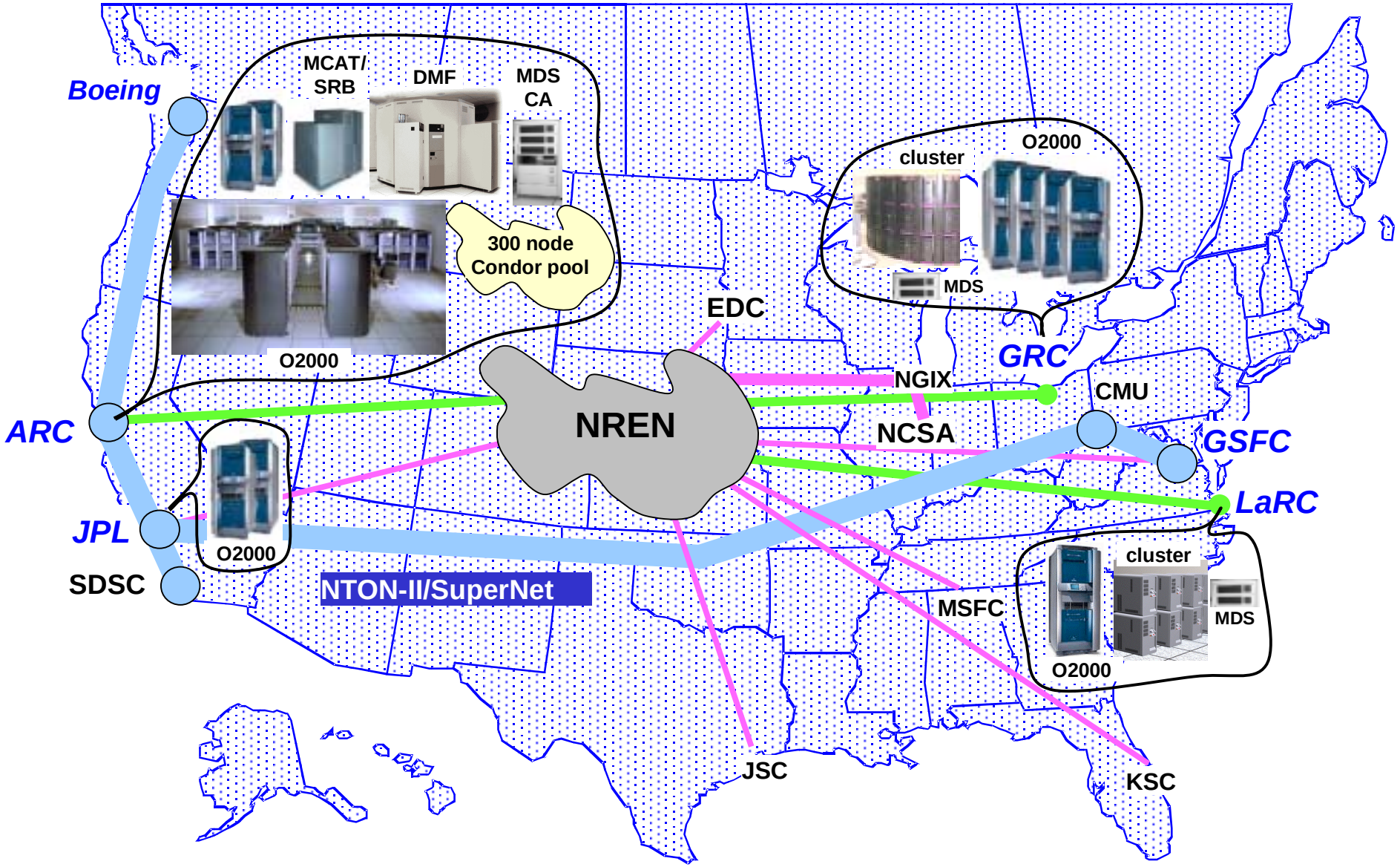
Multiple sub-systems, e.g. a wing lift model operating at NASA Ames and a turbo-machine model operating at NASA Glenn, are combined using an application framework that manages the interactions of multiple models and uses IPG services to coordinate computing and data storage systems across NASA.

Multi-disciplinary Simulations

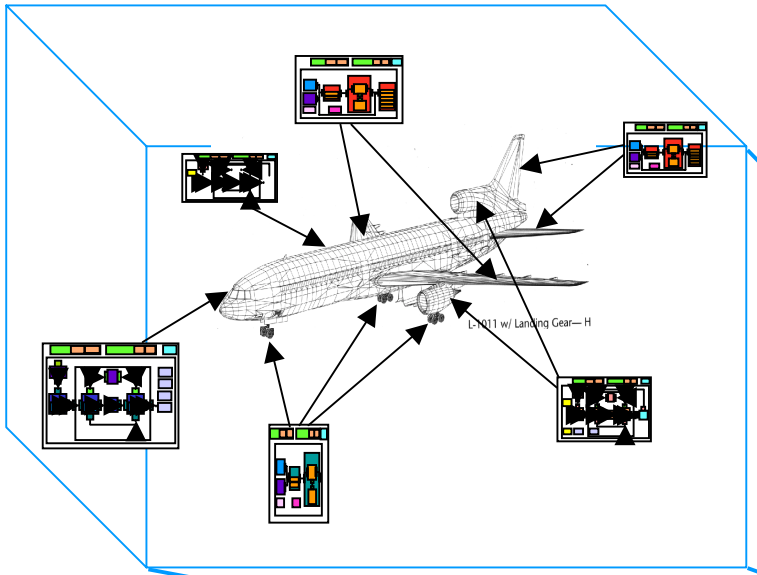


Whole system simulations are produced by coupling all of the sub-system simulations

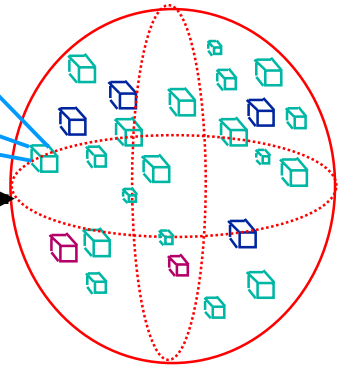
IPG Baseline System



Multi-disciplinary Simulations



Virtual National Air Space (VNAS)

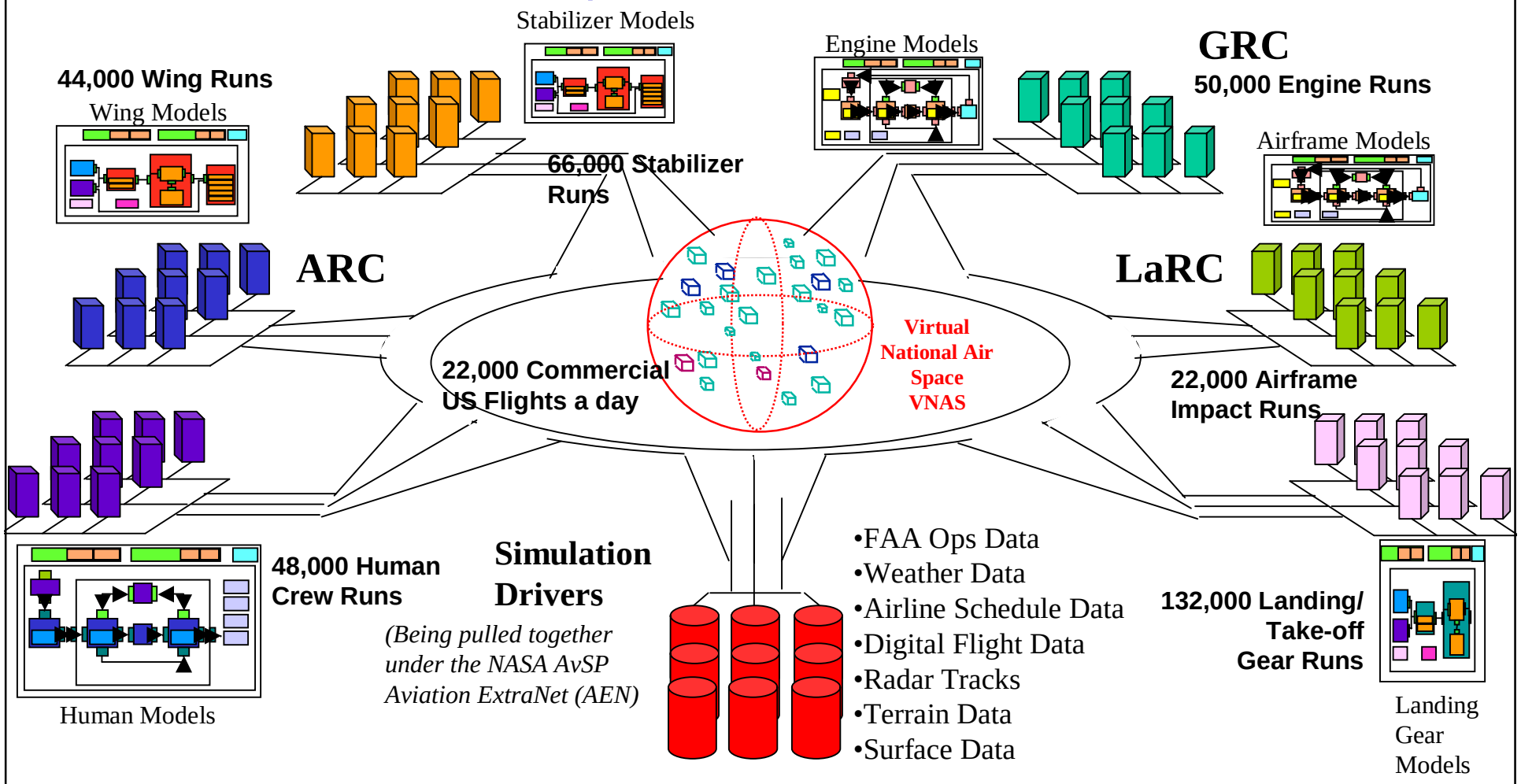


- **FAA Ops Data**
- **Weather Data**
- **Airline Schedule Data**
- **Digital Flight Data**
- **Radar Tracks**
- **Terrain Data**
- **Surface Data**

Simulated aircraft are inserted into a realistic environment, which requires adding many types of operations data to the systems simulation

Multi-disciplinary Simulations

National Air Space Simulation Environment



Many aircraft, flight paths, airport operations, and the environment are combined to get a virtual national airspace

US Grid Projects (1)

- NASA Information Power Grid
- DARPA CoABS Grid
- DOE Science Grid
- NSF National Virtual Observatory
- NSF GriPhyN
- DOE Particle Physics Data Grid
- NSF DTF TeraGrid
- DOE ASCI DISCOM Grid

US Grid Projects (2)

- DOE Earth Systems Grid
- DOE FusionGrid
- NEES Grid
- NIH BIRN
- NSF iVDGL
- NSF GEON
- NSF NEON
- …..

EU Grid Projects (1)

- DataGrid (CERN, ..)
- EuroGrid (Unicore)
- Damien (Metacomputing)
- DataTag (TransAtlanticTestbed, …)
- Astrophysical Virtual Observatory
- GRIP (Globus/Unicore)
- GRIA (Industrial applications)
- GridLab (Cactus Toolkit, ..)
- CrossGrid (Infrastructure Components)
- EGSO (Solar Physics)

EU Grid Projects (2)

- GridStart (Coordination, …)
- FLOWGRID (CFD VO)
- OPENMOL (Chemistry, Pharma, ..)
- GRACE (Distributed Search, …)
- COG (Industrial ontolgies, ..)
- MOSES (Knowledge Grid)
- BIOGRID (Biotechnology industry)
- GEMSS (e-Healthcare, ..)
- SELENE (Metadata, P2P, ..)

National Grid Projects (1)

- UK - e-Science Grid
- Netherlands – VLAM-G, DutchGrid
- Germany – UNICORE Grid, D-Grid
- France – Etoile Grid
- Italy – INFN Grid
- Eire – Grid-Ireland
- Scandinavia - NorduGrid
- Poland – PIONIER Grid
- Hungary – DemoGrid
- ….

National Grid Projects (2)

- Japan – JpGrid, ITBL
- South Korea – N*Grid
- Australia – Nimrod-G, ….
- Thailand
- Singapore
- AsiaPacific Grid ?
- Pragma?
- …

UK e-Science Funding

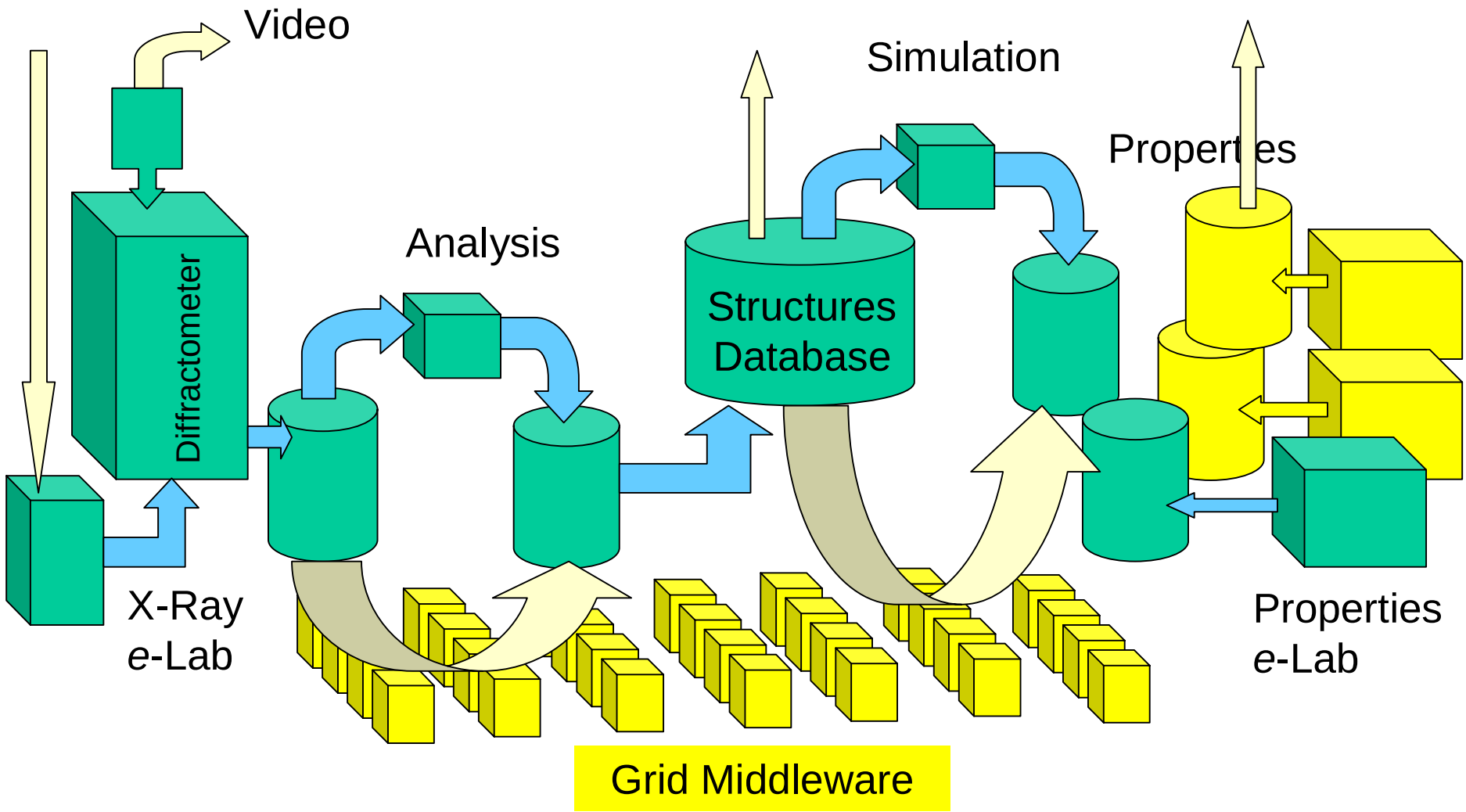
First Phase: 2001 –2004

- Application Projects
  - £74M
  - All areas of science and engineering
- Core Programme
  - £35M
  - Collaborative industrial projects

Second Phase: 2003 –2006

- Application Projects
  - £96M
  - All areas of science and engineering
- Core Programme
  - £16M + £25M (?)
  - Core Grid Middleware

Comb-*e*-Chem Project



National Crystallographic Service Workflow

Send sample material to NCS service

Collaborate in e-Lab experiment and obtain structure

Search materials database and predict properties using Grid computations

Download full data on materials of interest

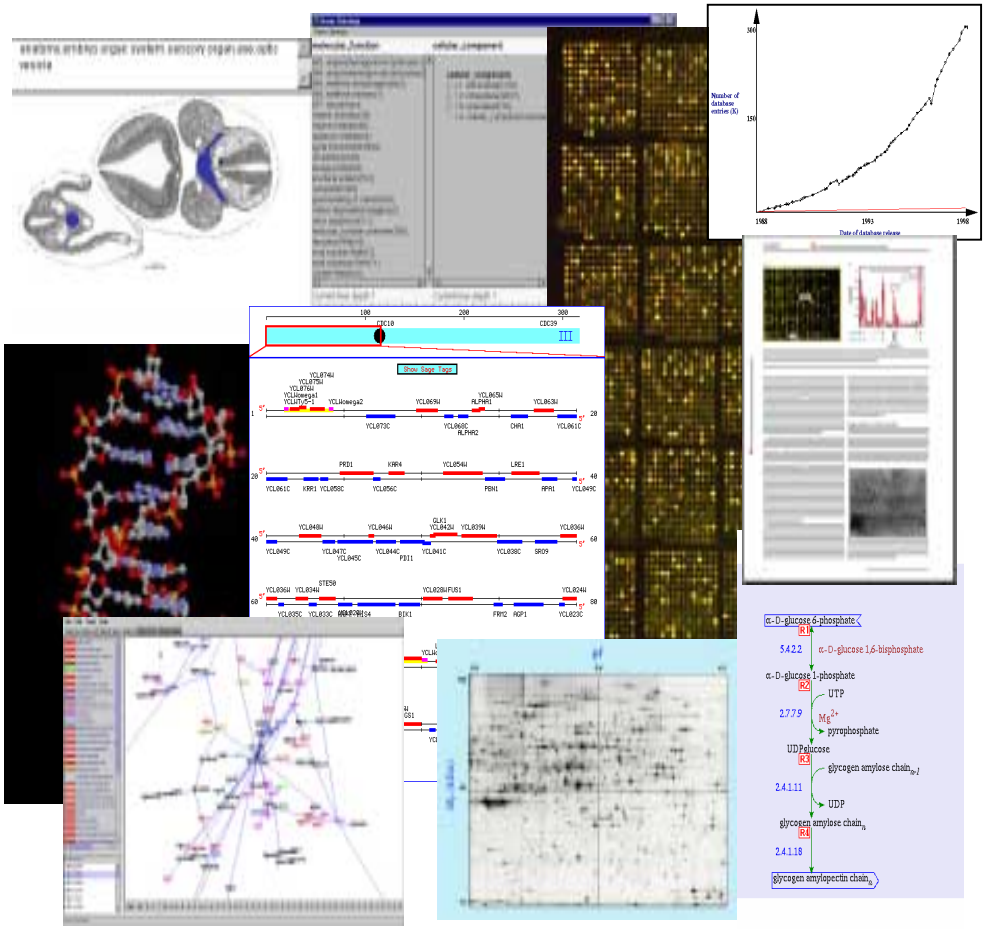
X-Ray e-Laboratory

Structures Database

Computation Service

myGrid Project

- Imminent ‘deluge’ of data
- Highly heterogeneous
- Highly complex and inter-related
- Convergence of data and literature archives



Discovery Net Project

In Real Time

Scientific Information

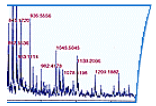
Scientific Discovery

Literature

Databases

Operational Data

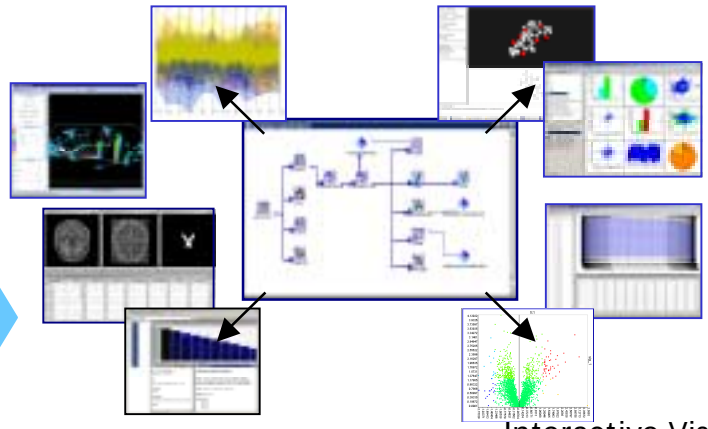
Images



Instrument Data

Real Time Integration

Workflow Construction

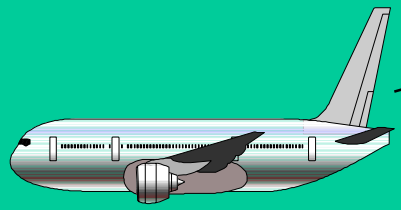


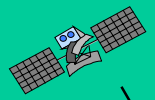
Dynamic Application Integration

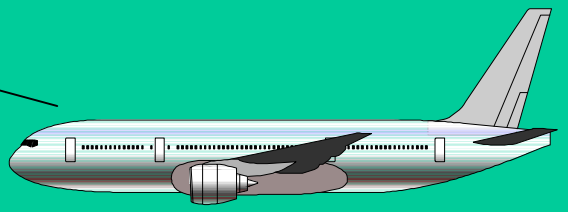
Interactive Visual Analysis

Using Distributed Resources

DAME Project







In flight data

Global Network
eg: SITA

Airline

Ground
Station

DS&S Engine Health Center

Internet, e-mail, pager

Maintenance Centre

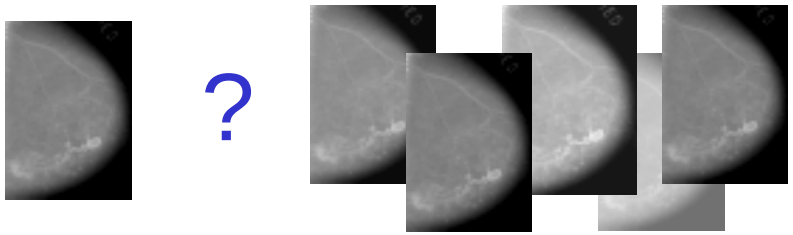
Data centre

eDiamond

Applications of SMF

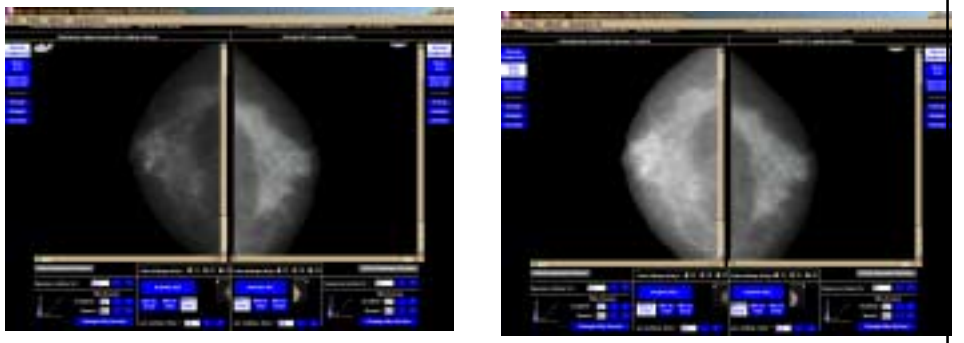
Training and Differential Diagnosis

“Find one like it”



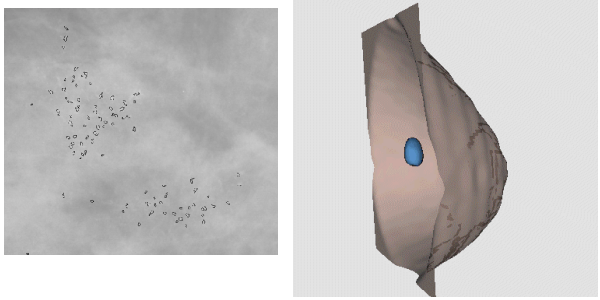
Teleradiology and QC

VirtualMammo



Advanced CAD

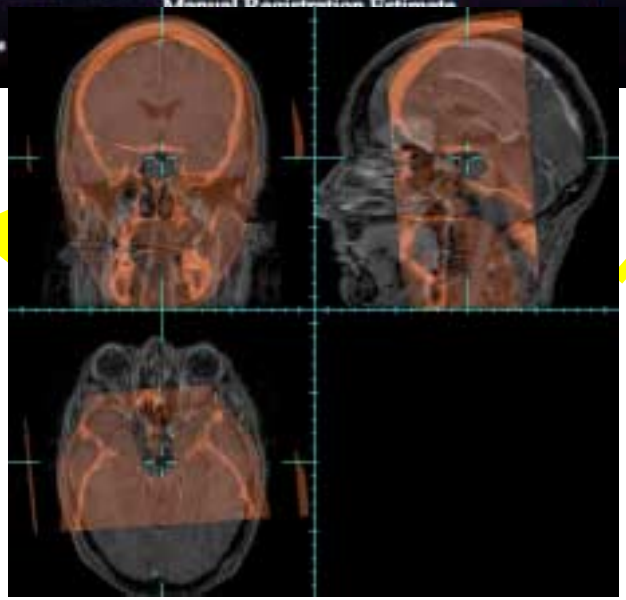
SMF-CAD workstation

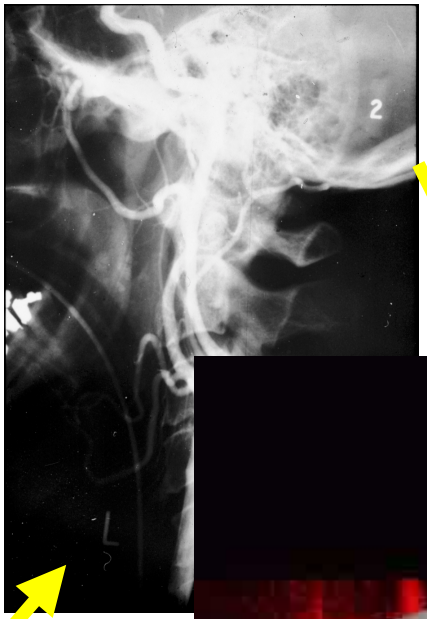


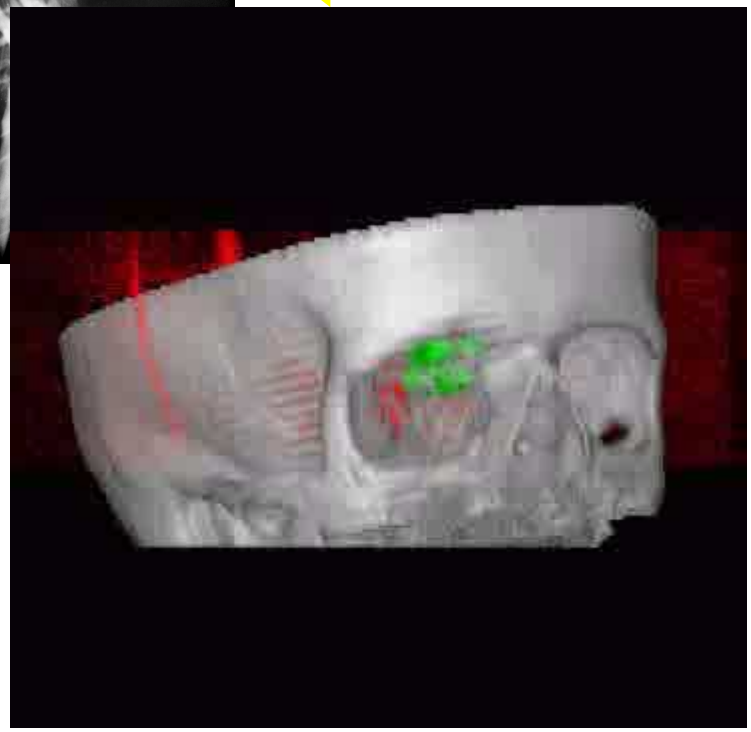
Epidemiology

SMFcomputed breast density

Image guided interventions





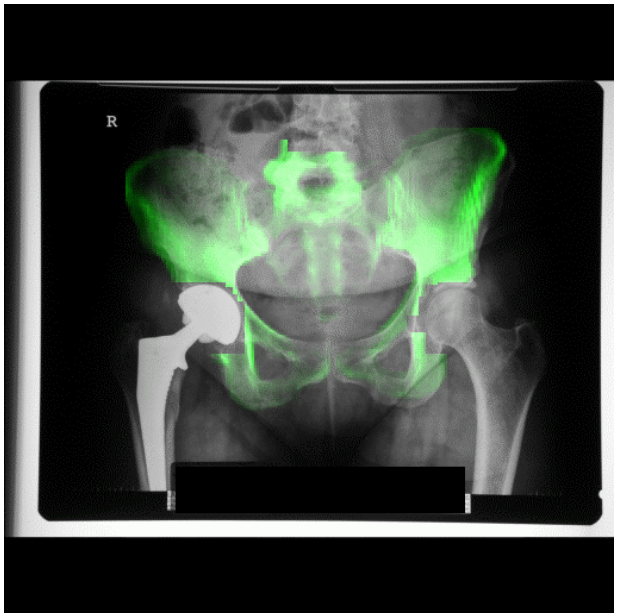


Images Courtesy
Derek Hill
Guy's Hospital

Surgical verification

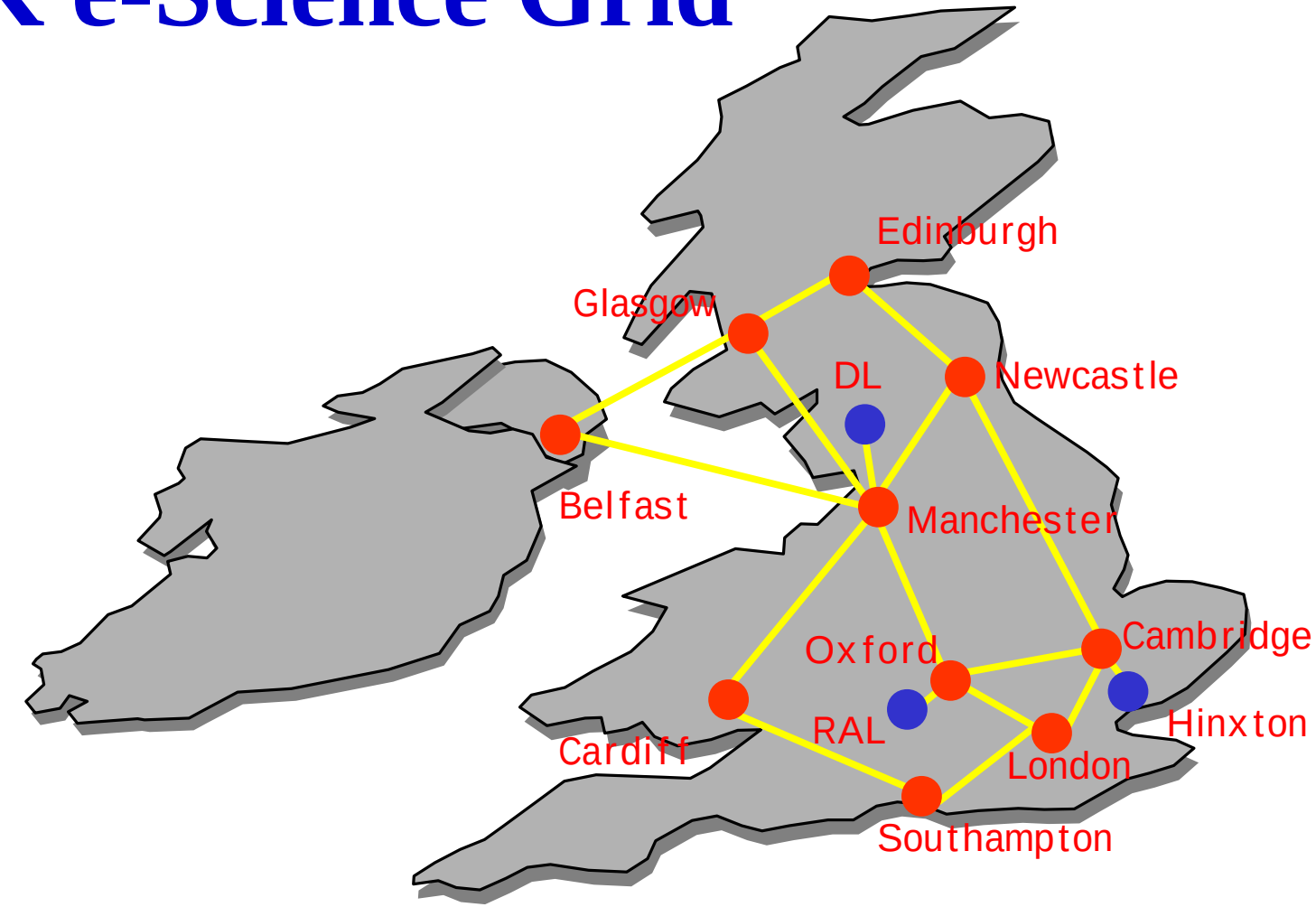
Accuracy of surgical placement against plan

- Surgeon plans on X-ray or CT, uses database of prostheses
- Operation takes place using plan as guidance
- Post operative X-ray evaluated for accuracy of placement
- Data stored and used for short term assessment and long term evaluation studies



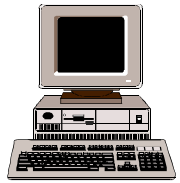
Courtesy of
Ian Revie
Depuy International

UK e-Science Grid

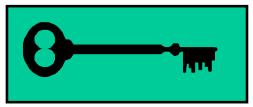


Single Sign-On and Digital Certificates

1. Scientist wishes to access a resource, so he sends a copy of the certificate to the resource



2. Resource says: prove it's your certificate



Private Key

Challenge

Response

3. Scientist proves that he has the corresponding private key

4. Resource is convinced that scientist is who he claims to be and decides to give him access

Access Grid – Group Conferencing

Multi-site group-to-group conferencing system

Continuous audio and video contact with all participants

Globally deployed

All UK e-Science Centres have AG rooms

Widely used for technical and management meetings

The UK Grid Experience

- UK Programme on Grids for e-Science
  - £75M for e-Science Applications
- UK Grid Core Programme for Industry
  - £35M for collaborative industrial R&D

➢ Over 80 UK companies participating

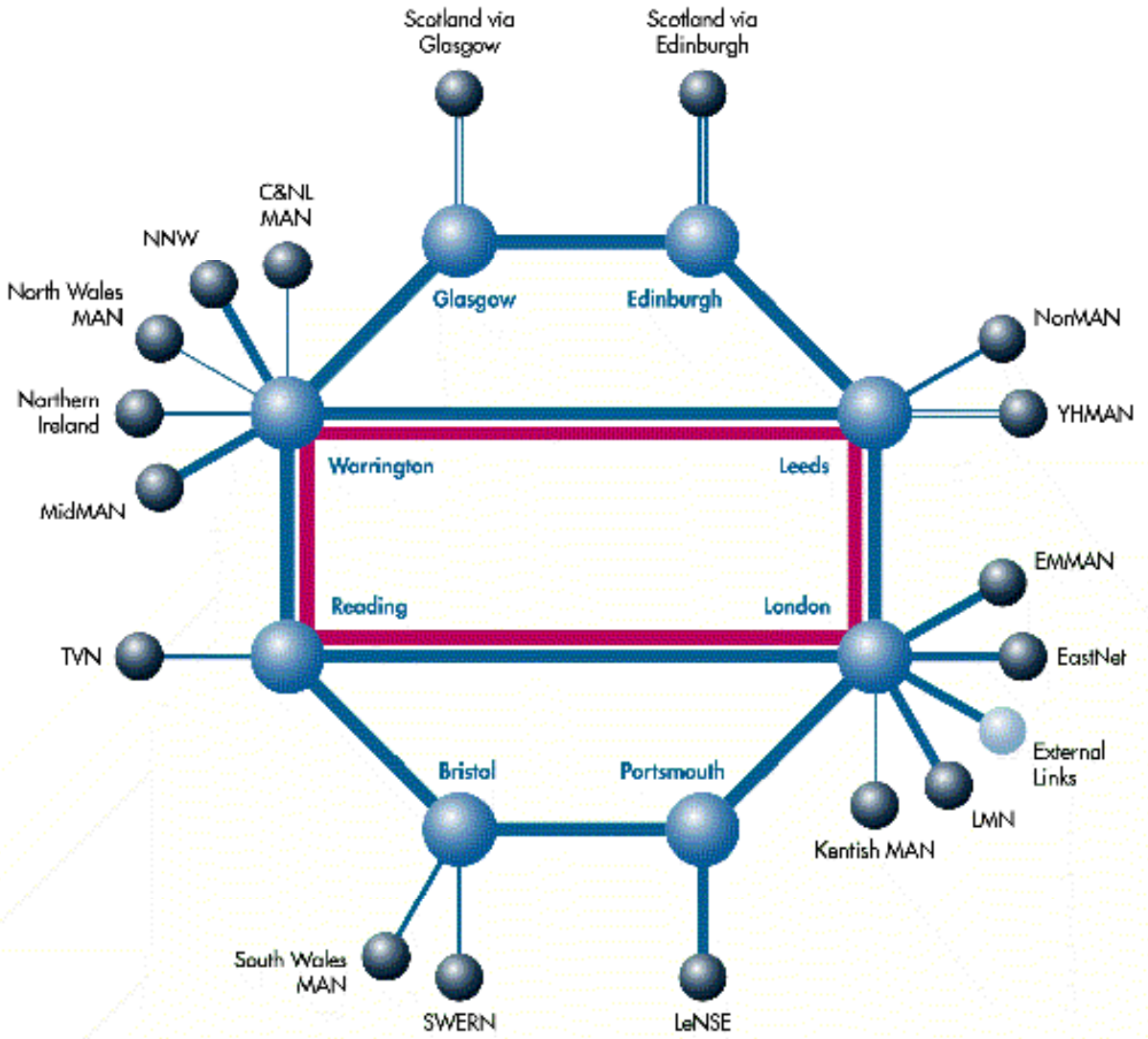
➢ Over £30M industrial contributions

- Engineering, Pharmaceutical, Petrochemical
- IT companies, Commerce, Media

Subset of Industrial Involvement

- **IT Companies**
  - Sun, IBM, Intel, Microsoft
  - SGI, HP, Fujitsu, Cisco, …
- **Major End User Companies**
  - Rolls Royce, Data Systems and Solutions, BAESystems, Shell, Siemens
  - GSK, Astra-Zeneca, Pfizer, Merck, Schlumberger, BT, …
- **SMEs**
  - NAG, Cybula, Compusys, Mesophotonics, Fluent, Epistemics, Mirada, ….

SuperJANET4



Research Prototype Middleware to Production Quality

- Research projects are not funded to do the regression testing, configuration and QA required to produce production quality middleware
- Common rule of thumb is that it requires at least 10 times more effort to take 'proof of concept' research software to production quality

➢ Key issue for UK e-Science projects is to ensure that there is some documented, maintainable, robust grid middleware by the end of the 5 year £250M initiative

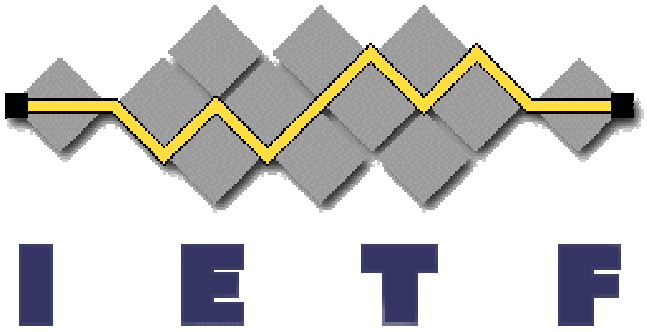
Open Grid Services Architecture

- Development of Web Services
- OGSA will provide

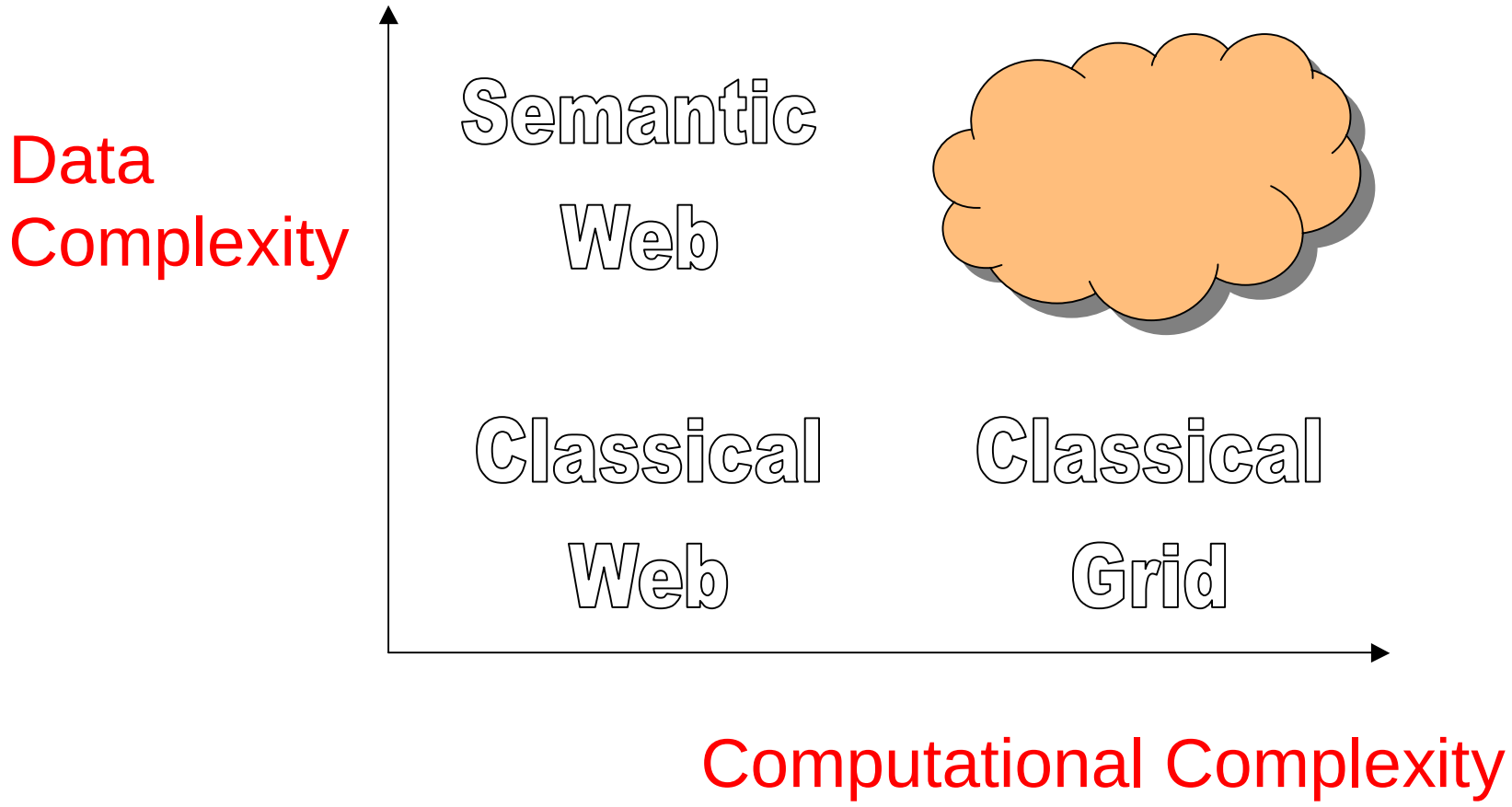
  Naming /Authorization / Security / Privacy/…

➢ Projects looking at higher level services: Workflow, Transactions, DataMining, Knowledge Discovery…

➢ Exploit Synergy: Commercial Internet with Grid Services



Databases in the Grid



OGSA – DAI Project

- Key middleware project for UK Program
  - Phase 1 funding £3M
  - Phase 2 funding £1.5M

- Three Centres involved:
  - Edinburgh, Manchester and Newcastle

- Industrial partners:
  - IBM US, IBM Hursley and Oracle UK

➢ Develop high-quality data-centric middleware

OGSA – DAI Project

- Beta versions released April 2003
  - XML Database Interface
  - Relational Database Interface
- Prototype
  - Distributed Query Service
- Final versions to be delivered July 2003
  - Integrate release with Globus GT3
- OGSA-DAI 2 Project now approved (£1.5M)
  - Continued development and more functionality

Core ‘e-Science’ Middleware

- Need to develop open source, open standard compliant, Grid Middleware stack that will integrate and federate with industrial solutions
- Software Engineering focus

  Aim is to produce robust, well-documented, re-usable software that is maintainable and can evolve to embrace emerging Grid Service standards

➢ Major focus of Phase 2 of the UK e-Science Programme - £20M ‘OMII’ initiative

A UK Open Middleware Infrastructure Institute

- Repository for UK-developed Open Source ‘e-Science/Cyber-infrastructure’ Middleware
- Compliance testing for GGF/WS standards
- Documentation, specification and QA
- Fund work to bring ‘research project’ software up to ‘production strength’
- Fund Middleware projects for identified ‘gaps’
- Work with US NMI, EU Projects and others
- Support from major IT companies

A Global Open Middleware Infrastructure Institute?

- Repository for global Open Source ‘e-Science/Cyber-infrastructure’ Middleware
- Compliance testing for GGF/WS standards
- Documentation, specification and QA
- Fund work to bring ‘research project’ software up to ‘production strength’
- Fund Middleware projects for identified ‘gaps’
- Federated ‘Global OMII’ structure of EU, US and A-P middleware engineering activities
- Supported by all major IT companies

National Data Curation Centre

- In next 5 years e-Science projects will produce more scientific data than has been collected in the whole of human history
- In 20 years can guarantee that the operating and spreadsheet program and the hardware used to store data will not exist

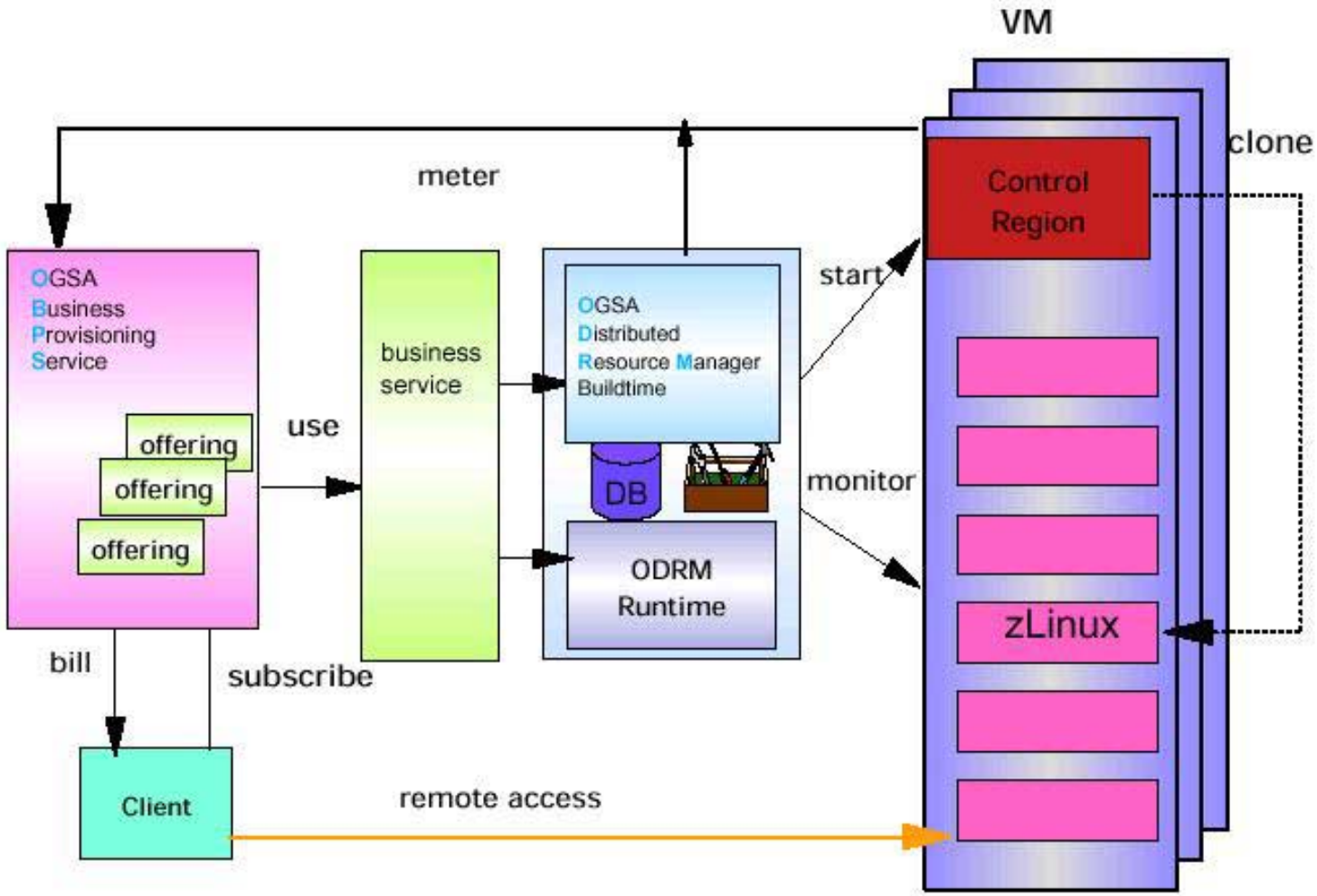
➢ Need to research and develop technologies and best practice for curating digital data

➢ Need to liaise closely with individual research communities and data archive centres

IBM : Interview with Irving Wladawsky-Berger (August 2001)

- ‘Grid computing is a set of research management services that sit on top of the OS to link different systems together’
- ‘We will work with the Globus community to build this layer of software to help share resources’
- ‘All of our systems will be enabled to work with the grid, and all of our middleware will integrate with the software’

Capacity on Demand Proof of Concept: zLinux-on-Demand

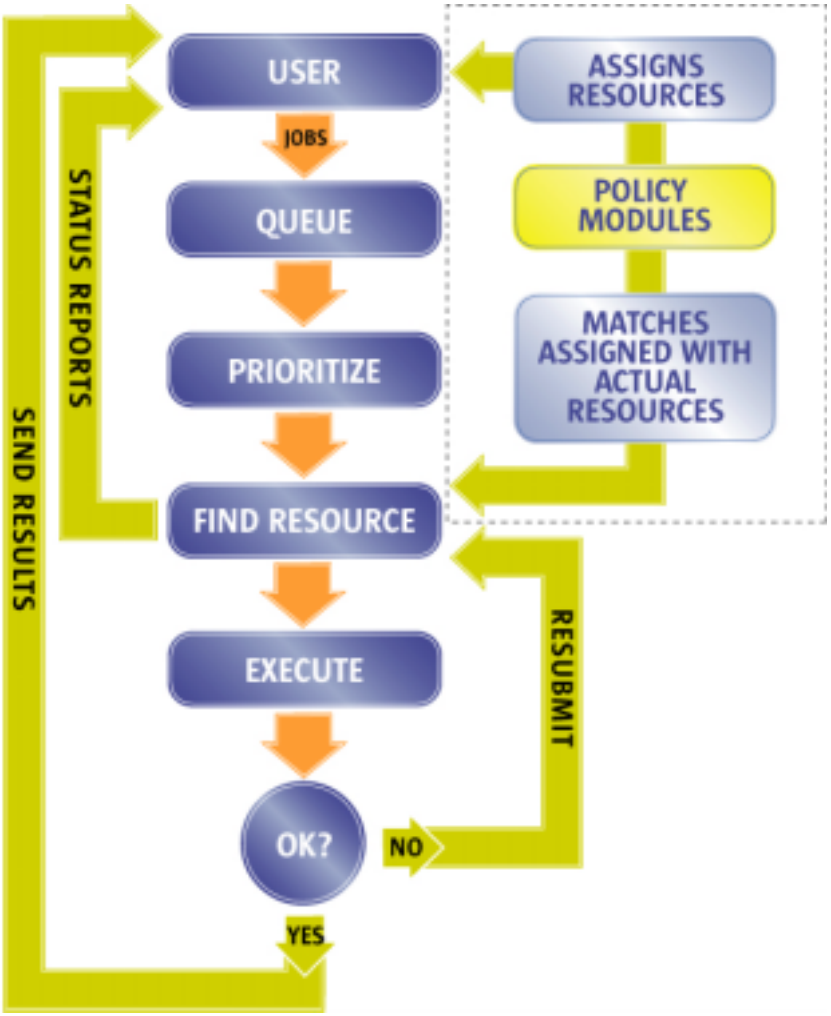


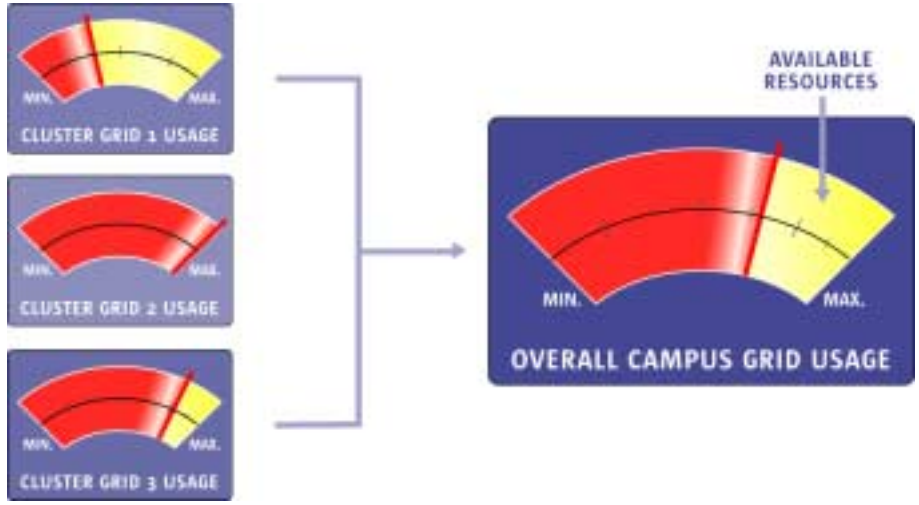
Sun and the Grid

‘Grid Computing is one of the three next big things for Sun and our customers’

Ed Zander, COO

SGE Enterprise Edition





Microsoft and the Grid:

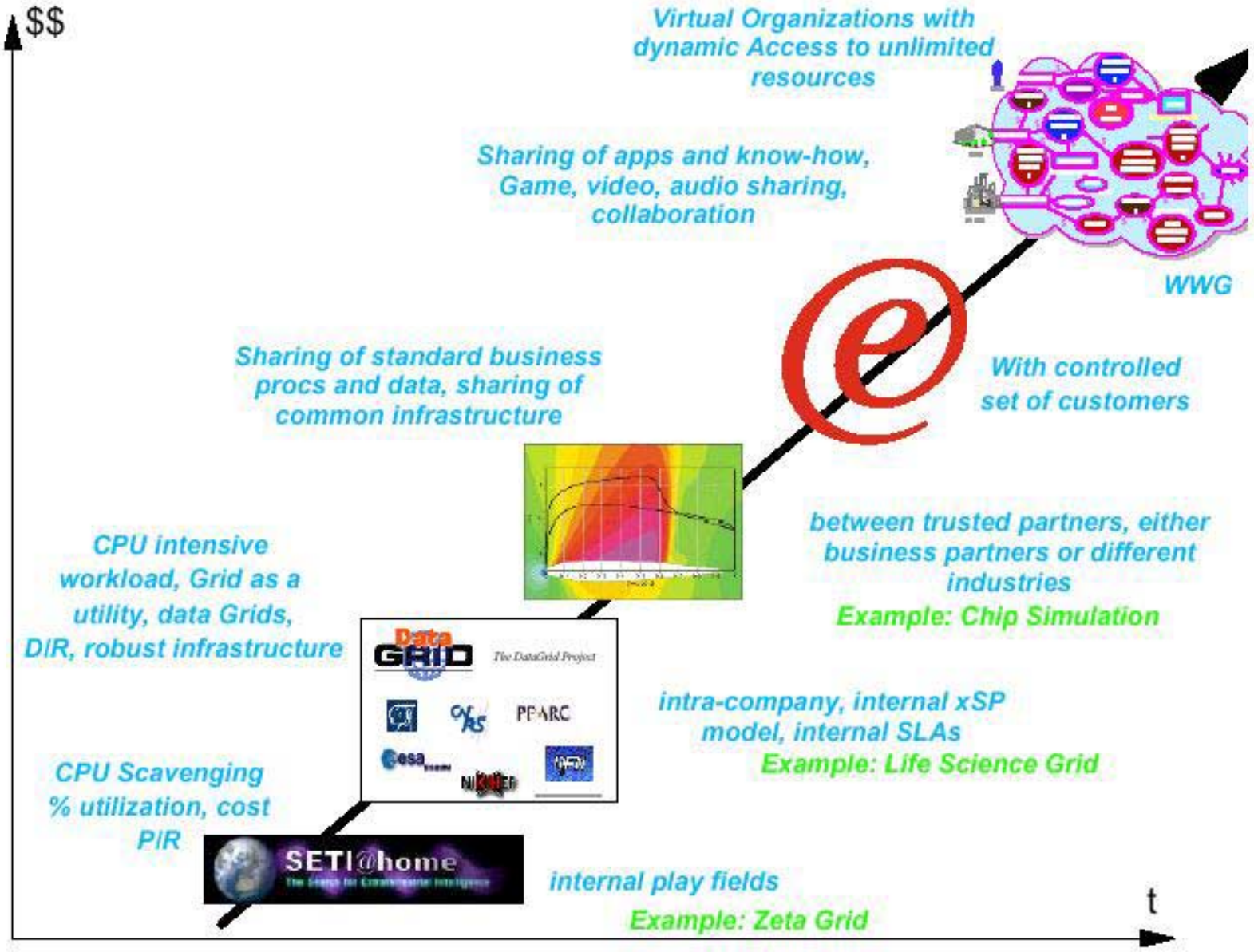
‘The alignment of OGSA with XML Web services is important because it will make Internet-scale, distributed Grid Computing possible’

Robert Wahbe,

General Manager of Web Services

- .NET Grid Project with Edinburgh

The Steps of Grid Computing

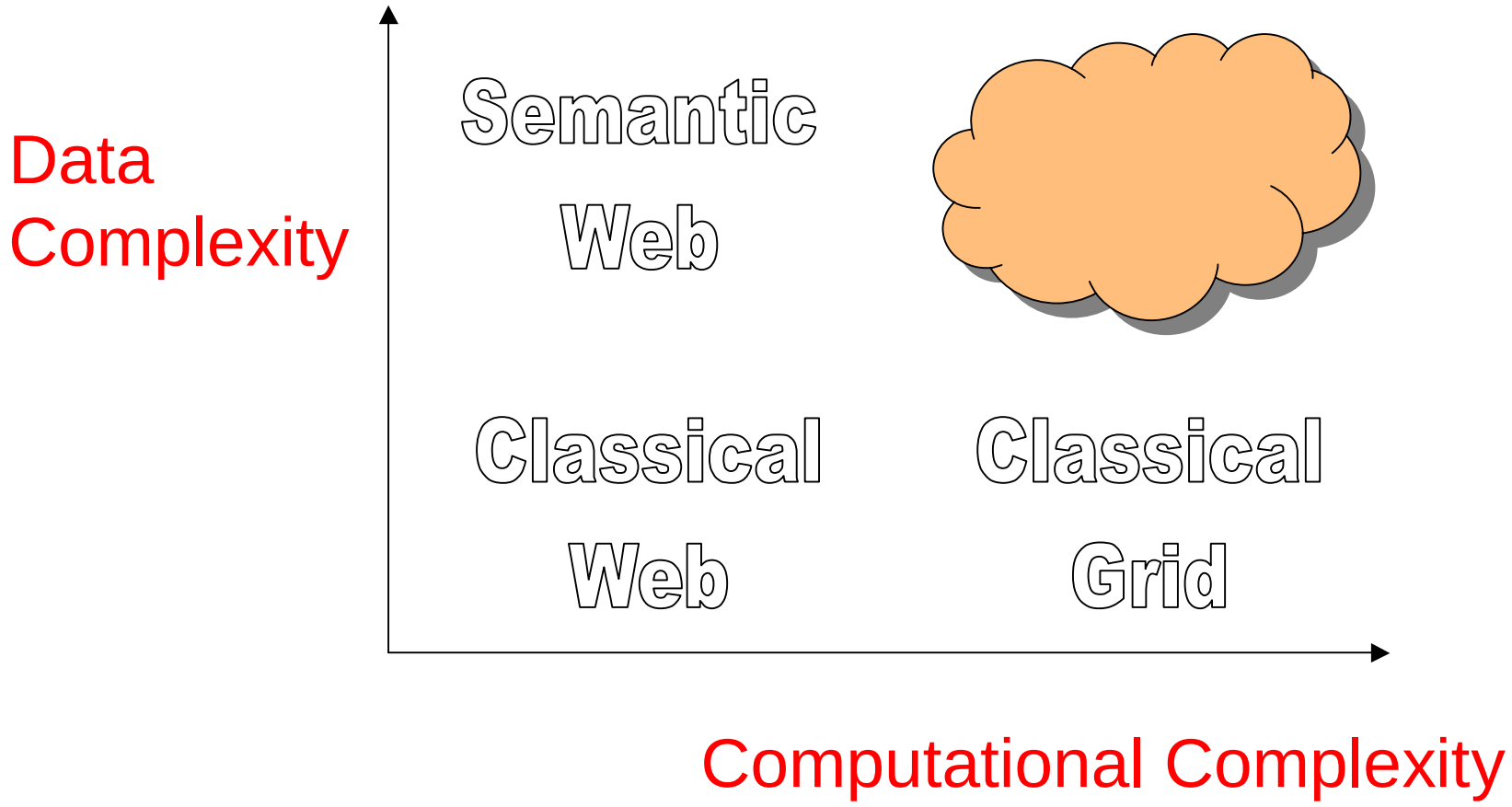


Status of the Grid

- **Today -** mainly a research tool, used by scientists - just like the early days of the Web
- **Tomorrow -** sophisticated combinations of services to locate data and information, applications to process the data, and computer systems to run them on

➢ The Grid infrastructure of the future will support Virtual Organisations, c-ommerce and e-Utilities

Databases in the Grid



e-Science and Grid Middleware

‘e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.’

John Taylor

- High value Grid applications of the future will involve interoperability of Grid and Web Data Services and not just cluster or cycle-stealing Grid computing

e-Government and the Grid

‘[The Grid] intends to make access to computing power, scientific data repositories and experimental facilities as easy as the Web makes access to information.’

Tony Blair, 2002

Acknowledgements

With thanks to:

Gerd Breiter, Phillipe Bricard, David Boyd, Jens Jensen, Mike Brady, Derek Hill, Carole Goble, Yike Guo, Jeremy Frey, Bill Johnston, Ray Browne, Jim Fleming, Anne Trefethen and others