

e-Science & the GRID

Where we are; Where we're going; and What it means to IT Support

Matthew J. Dovey
Technical Manager
Oxford e-Science Centre
OUCS
(matthew.dovey@oucs.ox.ac.uk)

What is e-Science?

e-Science GRID GRID

In the GRID below you will find four common phrases associated with the "grid".

First Prize: £3.5M!

E	A	G	E	Q	M	N	K
H	X	H	X	T	P	O	K
Y	Q	C	C	P	P	W	L
E	X	C	I	T	I	N	G
G	Z	G	T	T	F	F	K
J	N	O	I	D	I	A	U
C	Y	D	N	Z	Z	N	Z
H	Z	H	G	L	F	R	G
Z	Z	Z	Z	Z	Z	Z	Z
W	H	A	T	I	S	I	T

From "OUCS Nudes", 1 April 2003

What is e-Science?

e-Science GRID GRID

In the GRID below you will find four common phrases associated with the "grid".

First Prize: £3.5M!

From "OUCS Nudes", 1 April 2003

E	A	G	E	Q	M	N	K
H	X	H	X	T	P	O	K
Y	Q	C	C	P	P	W	L
E	X	C	I	T	I	N	G
G	Z	G	T	T	F	F	K
J	N	O	I	D	I	A	U
C	Y	D	N	Z	Z	N	Z
H	Z	H	G	L	F	R	G
Z	Z	Z	Z	Z	Z	Z	Z
W	H	A	T	I	S	I	T

What really is e-Science?

“In the future, e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists.”

<http://www.research-councils.ac.uk/escience/>

What really is e-Science?

“In the future, e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists.”

<http://www.research-councils.ac.uk/escience/>

What *is* e-Science, now?

- “e-Science means science increasingly done through distributed global collaborations enabled by the Internet, using very large data collections, terascale computing resources and high performance visualisation”
(John Taylor, Director General of the Research Councils, OST)
- “e-Science will change the dynamic of the way science is undertaken”

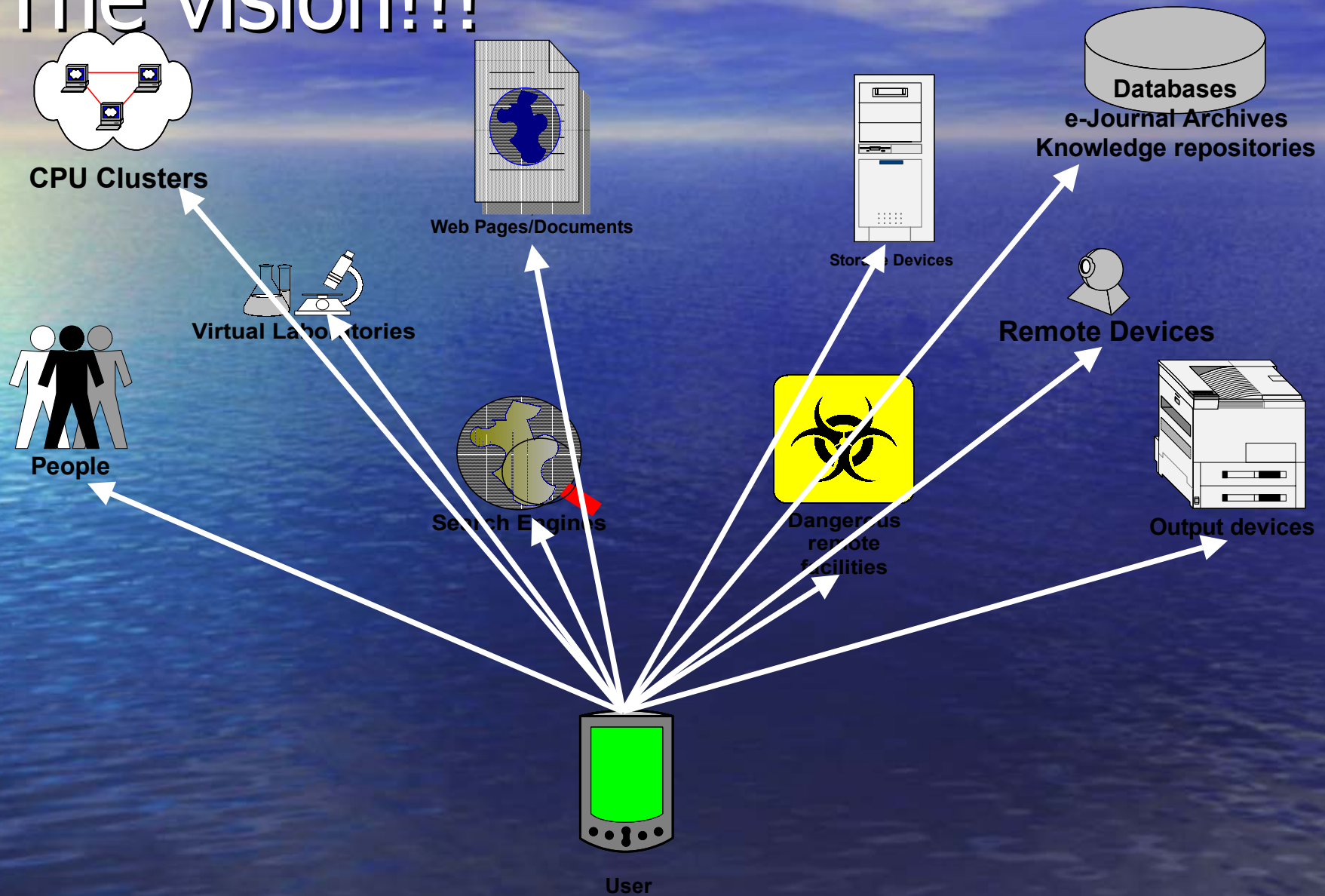
Errm, and that means?

- e-Science is basically "Science"
 - (and the deplorable trend of putting "e-" in front of perfectly good English words)
- The "e-" refers to using electronic communications, the internet etc. for enabling collaborative and distributed research.
- Oh, and its "big" (big data, big processing)!

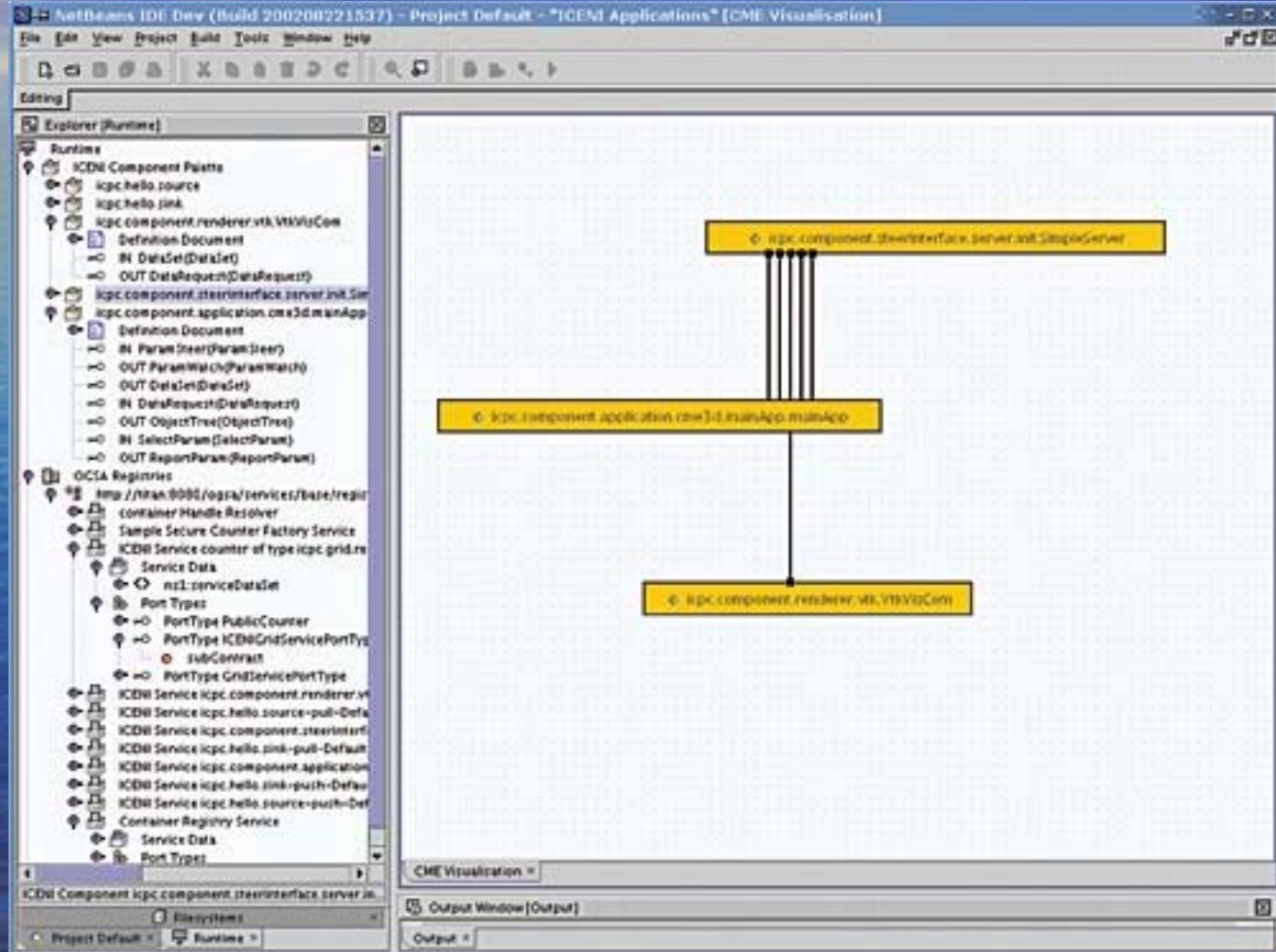
So e-Science is e-mail?

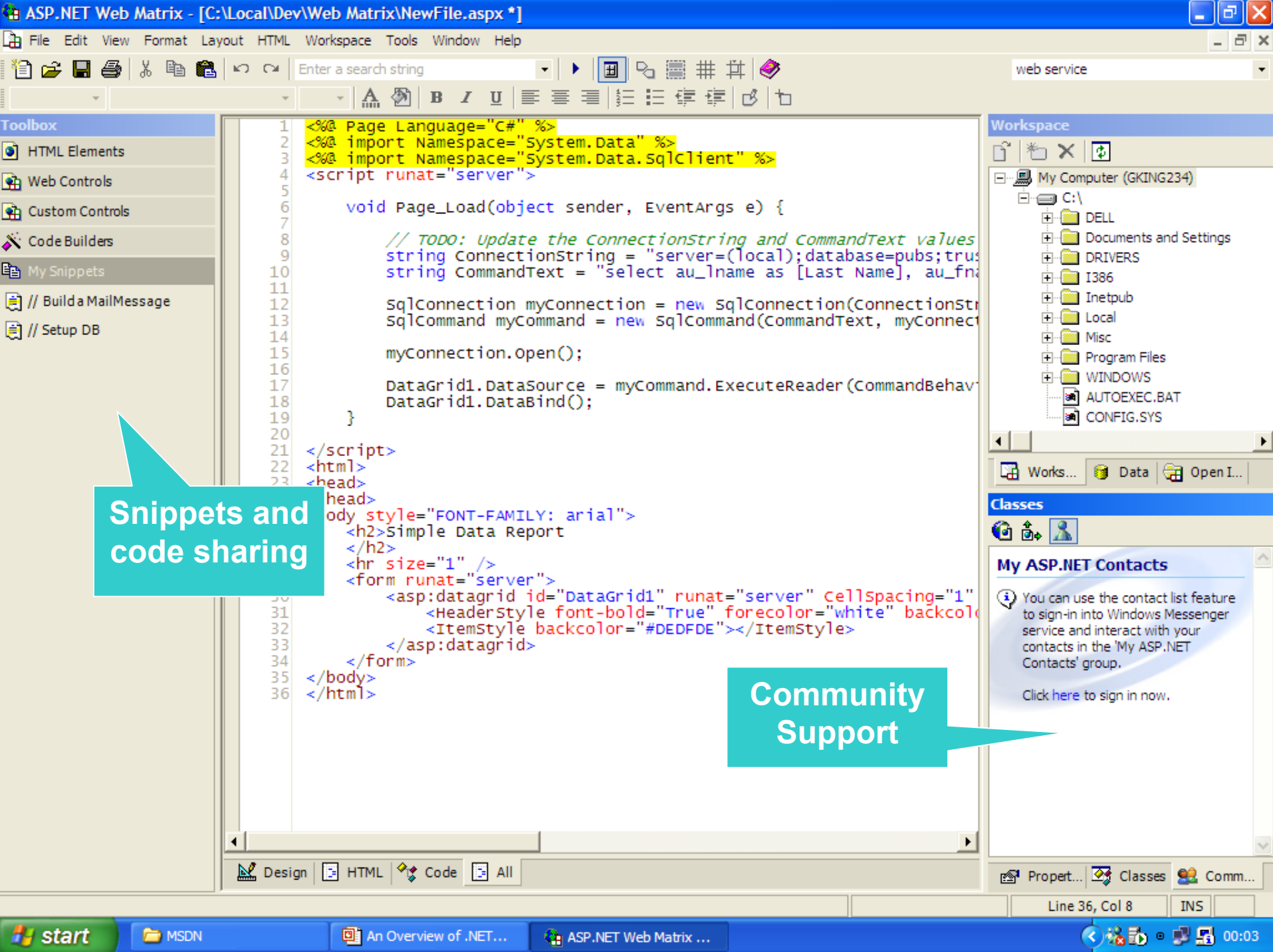
- Successful science using electronic means or internet is “e-Science”
- So a group of physicists who would not normally work together, communicating by e-mail is a minimum metric.
- But the e-Science vision is far more...

The vision!!!



The Vision (2)





Snippets and code sharing

Community Support

What is this GRID, then?

- “The Grid is a software infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources.” (*The Grid*, eds. Foster & Kesselman)
- “The Grid is an emergent infrastructure capable of delivering dependable, pervasive and uniform access to a set of globally distributed, dynamic and heterogeneous resources. It brings challenges of scalability, interoperability, fault tolerance, resource management and security.” (UK e-Science Director)

Errr, and *that* means?

- The GRID is to Computing Resources as the Web is to Documents
- The GRID is vision of a collection of resources (people, processors, storage, scientific devices, information, knowledge, etc.) and the mechanisms to access them online and remotely
- “GRID” is often used to refer to the vision, the system and also enabling software

What is *a* GRID?

- GRID is an enabling technology for online collaboration and discovery and access to remote resources
- GRID is middleware
- A number of GRID middleware technologies:
 - Globus
 - Condor
 - SRB
 - Sun GRID Engine
 - etc.

GRIDs Beyond e-Science

- Apart from the depressing use of "e-"
 - e-social-science
 - e-humanities
 - e-arts
 - e-theology (deus ex machina?)
- GRID is an enabling technology for collaborative research
- GRID is an enabling technology for resource discovery, access, use
- GRID is an enabling technology for ubiquitous computing
- GRID has applications outside of just the science arena

A Brief History of GRIDs

- Ian Foster “The Grid: Blueprint for a New Computing Infrastructure” publish 1998
- Series of GRID (or GRID-like) Technologies
 - Globus
 - Condor
 - JINI/JavaSpaces
 - SRB
 - AccessGRID
- No commonality; No common infrastructure; No common vocabulary (some weren’t aware of “GRID”); No common interfaces; No intercommunications

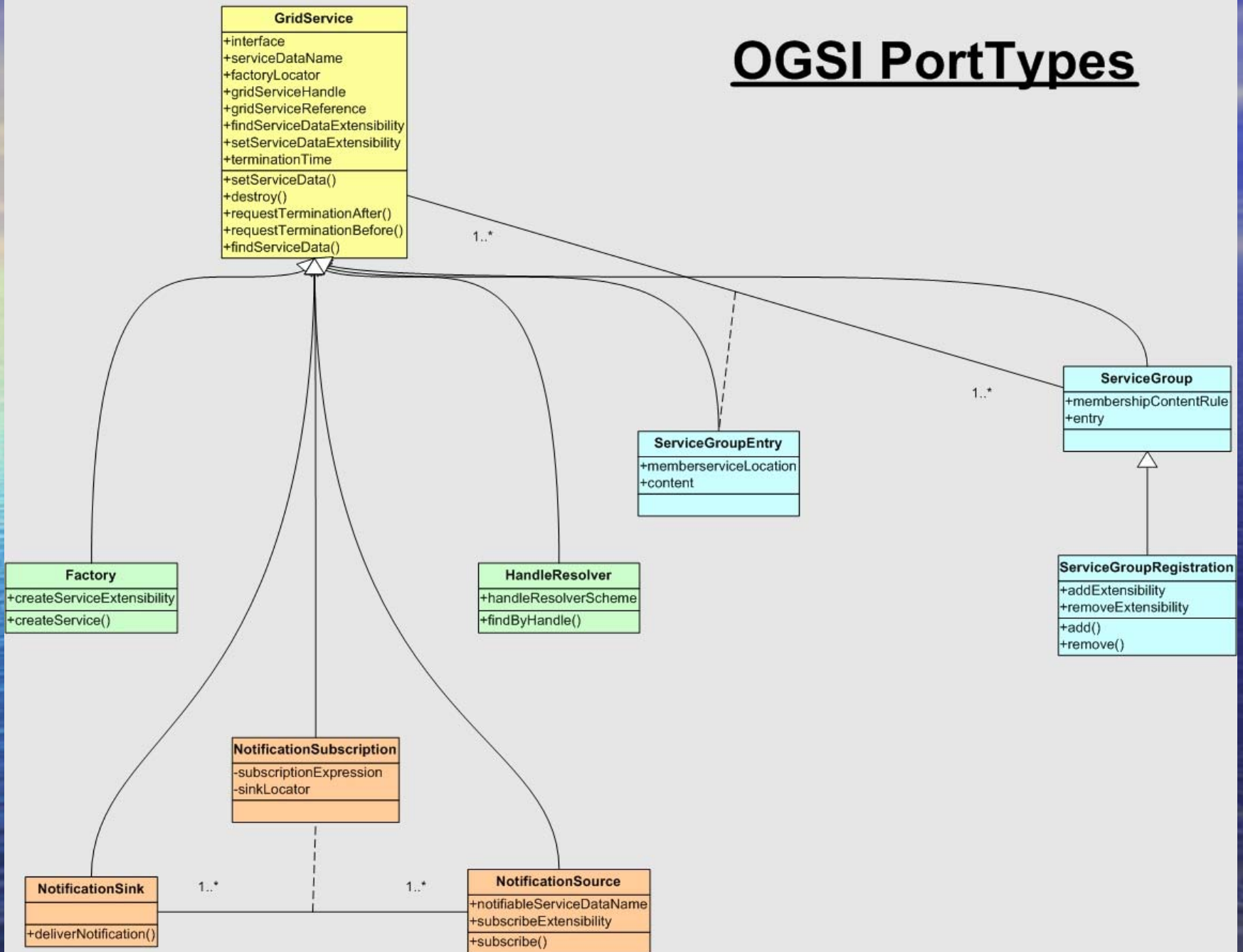
Enter the Open GRID

- To date different GRID technologies use different tools, API's etc.
- GRID's not interoperable!
- Global GRID Forum (cf W3C)
 - Open GRID Service Architecture
 - Open GRID Service Infrastructure
 - Based on WebServices
 - Adds O-O capabilities and foundation service (port)types
 - Globus 3 will be reference implementation

OGSI

- Open GRID Service Infrastructure
- Web Service based
- Common vocabulary/concepts
- Base “Object Oriented” types for GRIDServices

OGSI PortTypes

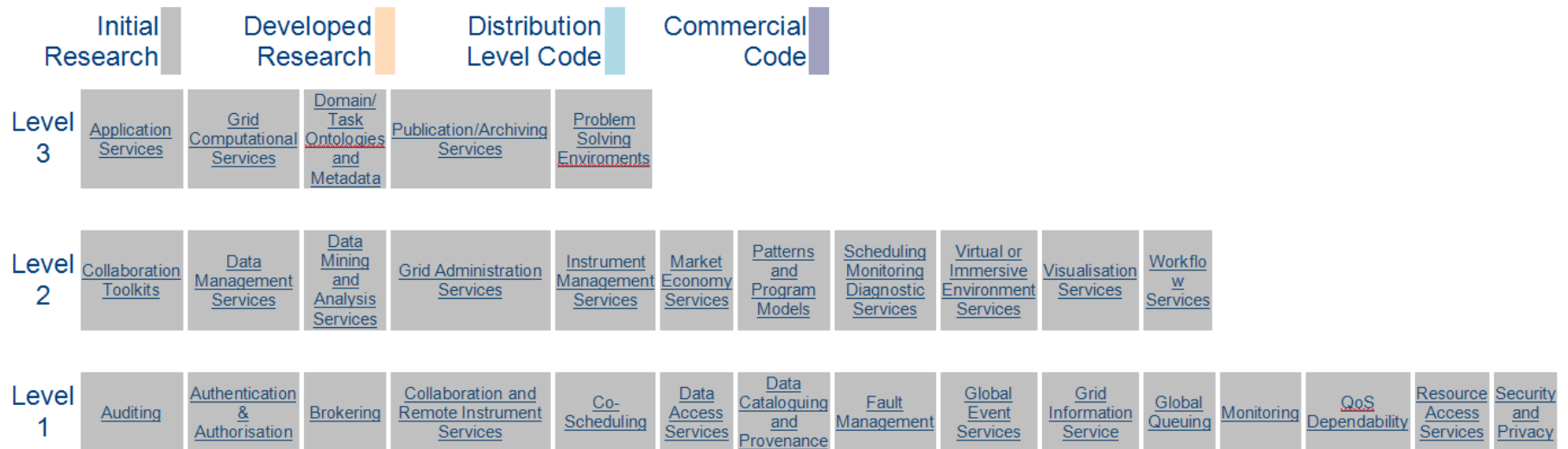


OGSA

- Open GRID Service Architecture
- Builds on OGSI
- Defines APIs/GRID Services for
 - Scheduling
 - Management/Monitoring
 - Database Access
 - Etc.

Middleware Components

Components



Current Implementations

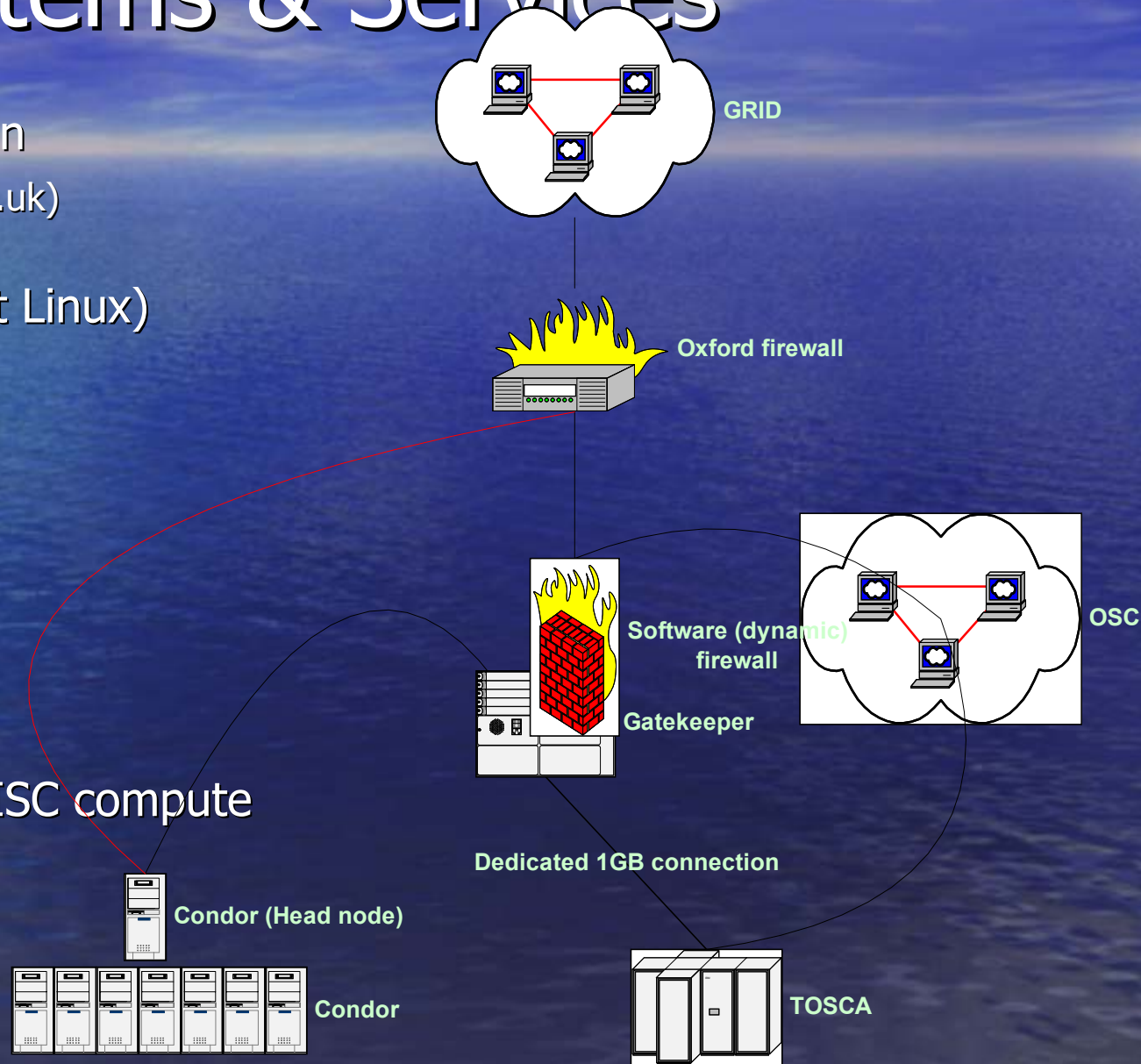
- All beta's and previews!
- Globus (Java and C)
- GridLite (Java)
- .Net (Two: Virginia and Edinburgh)
- Even Perl

E-Science and GRIDs in Oxford

- Oxford e-Science Centre based within OUCS and Comlab
 - OUCS - Developing the infrastructure for e-Science in Oxford
 - Comlab – Applications/Software Engineering
- Based within RTS
 - Develop the use of e-Science as a research technology
 - Develop the use of GRID as an enabling technology for research

OeSC Systems & Services

- Certificate Verification (<http://ca.grid-support.ac.uk>)
- Gatekeepers (Redhat Linux)
- Sun GRIDEngine
- Condor cluster
- Link to OSC
- Bidding for a large JISC compute cluster



OeSC Technical Expertise

- Globus
- X509
- MetaDirectory Services
- Authentication
- Firewalls
- WebServices & .Net
- Java & Java COG
- Sun GRIDEngine
- Portals
- Service/resource metadata and discovery
- AccessGRID – real time communications

Current OeSC Projects

- VideoWorks
- Remote Microscopy
- GRID Workload Management
- Climate Predication
- GeoVis/GeoDise
- DAME
- Reality GRID
- e-Diamond
- Biomolecular simulations
- Structural Biology
- Security for the EU DataGrid
- DCOCE

OeSC Within OUCS/RTS

- Authentication Developments
 - DCOCE
- Portal Developments
 - Need a portal to bring together the virtual project teams
 - Based on uPortal (UK expertise in developing GRID/Certificate based services for uPortal)
- Collaborative working
 - Access GRID (node at Comlab; nodes planned at Churchill and Begbroke; have desktop version running)
- Virtual Teams
 - Management and communications across distributed teams (OII)
- Resource Discovery
 - GRID as a new tool
 - GGF GIR Working Group (OULS/SERS)
- Support for OpenSource and Open Standards
 - GTR, GGF, OASIS, etc.

Futures – Interdisciplinary Hub

- e-Science Hub physical location being created in the University
 - Nerve centre for e-Science
 - 'Independent', not a separate department...
 - Located in close proximity to Computing Laboratory
 - Place to build teams, receive visitors, give lectures, overflow
 - Hub for nurturing new collaborations
 - Form new interdisciplinary centre for the University
- Two dedicated academic staff:
 - Dual role: core staff for Hub, complement expertise of e-Science team
 - Professor – database expert; established through Computing Laboratory
 - UL – Web services; Software Engineering
 - Aim for external funding (Development Office); strong support from Comlab
- Externally funded joint staff
 - EG NERC fellowship; department/OeSC shared post
 - Joint with CCLRC?

OeSC in the UK GRID and beyond

- Work on the “Level 2 GRID”
 - Lead for GRID Security within the ETF
 - Organised workshop on firewalls
 - Proposed various solutions: host database, dynamic firewall, VPN
 - Developed prototype dynamic firewall script
 - Developing a prototype trusted host database system
 - Helped develop integration test scripts
- Globus 3 and the OGSA GRID
 - Town Meeting on requirements and plans for the OGSA GRID
 - Working on Web Service definitions for upcoming projects
 - Working on Web Service to GRID Service migration paths
 - Perceived to be easier than GT2 to GT3 migration
 - Working with the OASIS WSDM TC
 - Working with RAL and Manchester on UDDI registries for GRID
 - Working with OASIS UDDI TC on GRID Requirements

UK National GRIDs

	GT2	OGSA
Level 1 Enthusiasts Prototyping	2002	Current Phase
Level 2 Functional with Some Real Applications	April 2003	??
Level 3 Useable GRID	??	??
...

Impact on IT Support – Hand Holding

– Getting and using certificates

- Getting the certificates isn't easy (but improving)
- Getting the certificates into tools isn't easy (but improving)
- Certificates don't roam easily

– Client certificate authentication

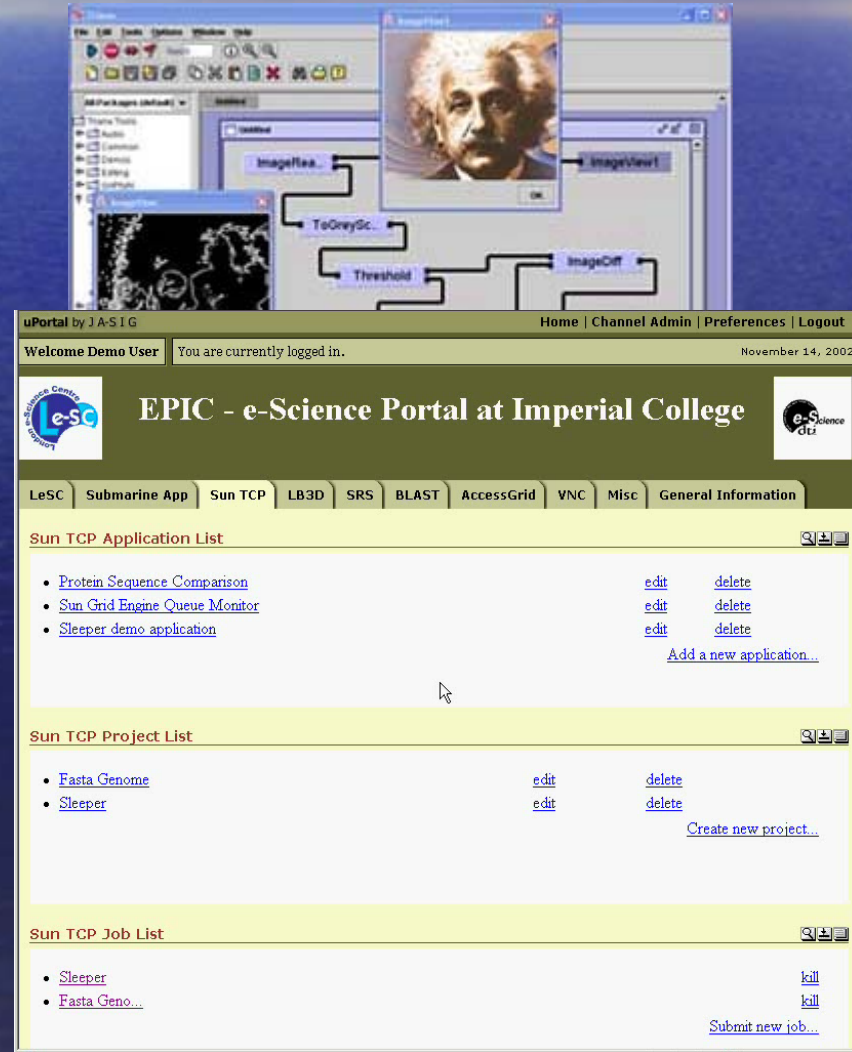
- Easier on some platforms than others

Impact on IT – Network Bandwidth

- E-Science Potentially involved large amounts of data
- Both data and jobs can move so some optimisation
- Many projects think their data is “big” but actual bandwidth used is not too bad (e.g 6GB per day)

Impact on IT Support – Locating and Accessing Resources

- Currently a programmers field
 - Some very nice technologies in WebServices (Visual Studio .Net, WebServices for Office, JBuilder, Elipse etc.)
 - GRIDService technologies very immature
- The Future
 - Dedicated workbenches
 - Drag and Drop interfaces
 - Portals (see portal workshop mid July – <http://www.nesc.ac.uk>)



The screenshot displays a web portal interface for EPIC - e-Science Portal at Imperial College. The page features a navigation menu with various application categories: LeSC, Submarine App, Sun TCP, LB3D, SRS, BLAST, AccessGrid, VNC, Misc, and General Information. Below the menu, there are three main sections: Sun TCP Application List, Sun TCP Project List, and Sun TCP Job List. Each section contains a list of items with 'edit' and 'delete' links. The top of the page has a header with 'uPortal by JA-SIG', 'Home | Channel Admin | Preferences | Logout', and a date 'November 14, 2002'. There is also a small image of Albert Einstein in the background of the top part of the screenshot.

Impact on IT Support – Security

- Globus 2 uses port mapping technologies
 - OeSC within ETF looking at potential solutions using VPNs, dynamic application aware firewalls etc.
 - GRID/Web Service over https will hopefully be a solution
- Some GRID infrastructures use Peer 2 Peer technologies
- Collaborative Virtual Organisations are based on realtime chat, file sharing, messenger applications etc.
- Transmission of jobs – mobile code
 - Potential use of JVM style sandbox technologies

Impact on IT Support - AccessGRID

- Large scale multicast video-conferencing
 - AccessGRID 1.x
 - Little flakey
 - AccessGRID 2.x
 - More robust – but certificate based



•Impacts

- AccessGRID rooms need planning (location of screens, projectors, microphones)
- Network bandwidth (support of multicast)
- Personal AccessGRID nodes will need hand-holding

Assistance to IT Support – GRID Monitoring & Management

- We have GITS (GRID Integration Test Scripts) for testing point to point interaction of GRID Services
- Planned UK Open Middleware Infrastructure Institute will provide interoperable testbeds
- GRID Monitoring systems for monitoring performance, network bandwidth etc.



[InfoPortal Home](#) - [MDS Browser](#) - [Map](#) - [About](#) - [CLRC e-Science](#) - [HPCPortal](#) - [Registration](#) - [Support](#) - [Projects](#) - [Systems](#) - [Applications](#) - [Users](#) - [Mailto](#)

	atlantis iridis	beowulf1 dl	bezier man	blue02 iridis	blue05 iridis	blue07 iridis	bouscat ca
columbus0 rl	course0 esc	course1 esc	course2 esc	course3 esc	course4 esc	esc dl	esc4 dl
esc5 dl	escpc15 esc	fermat cfb	frik mvc	gatekeeper ox	gatekeeper2 ox	gilmour nesc	ginfo grid-support
green cfb	grid-ireland ca	grid2 esc	grid4 esc	grid9 esc	herschel amtp	hrothgar esc	inca cf
ironwood sucx	i101 hpcx	login hpcx	loki	loki dl	mercury0 doc	metropolis sucx	neptune dcs
node00 cluster	node01 cluster	node02 cluster	node03 cluster	node04 cluster	node05 cluster	node06 cluster	node07 cluster
node08 cluster	node09 cluster	node10 cluster	node11 cluster	node12 cluster	node13 cluster	node14 cluster	node15 cluster
node16 cluster	node17 cluster	node18 cluster	node19 cluster	node20 cluster	node21 cluster	node22 cluster	node23 cluster
node24 cluster	node25 cluster	node26 cluster	node27 cluster	node28 cluster	node29 cluster	node30 cluster	node31 cluster
node32 cluster	pacifica iridis	parsifal cpc	pioneer-i0 pioneer	ramshope ncl	rhea lesc	saturn icpc	silica esc
splozin1 dl	sunresearch qub	tcisp213 hpc	thor-i0 thor	titan doc	vermont mvc	vikings-i00 vikings	vikings-i01 vikings
vikings-i02 vikings	wallace mvc	wesc6-comac grid	wesc7-comac grid	wesc8-comac grid	wesc9-comac grid	wk-pc1 dl	wren cfb

UK e-Science MDS - Last Update: Jun 26, at 10:45:35 BST

Machine Information

local MDS

beowulf1.dl.ac.uk Daresbury
Pentium Beowulf

node20.cluster.cpc.wmin.ac.uk

Beowulf at Daresbury

Old Linux Beowulf head node.



Plots of Ping Monitoring Data from icfamom.rl.ac.uk

Remote Host is **grid001f.cnaf.infn.it**

packet size is 100

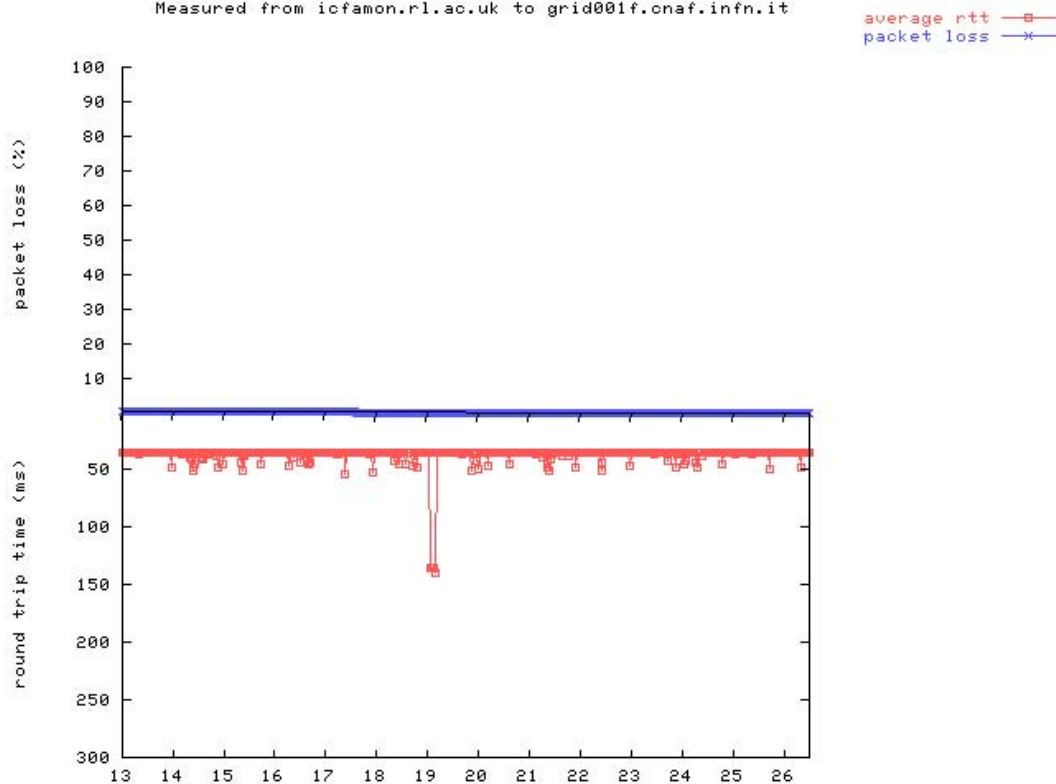
from Fri Jun 13 0:15:00 2003 to Thu Jun 26 11:45:01 2003

There is a time interval every day

View rtt frequency plot of this data [here](#)

Average Round Trip Time and Packet Loss (100 byte packets)

Measured from icfamom.rl.ac.uk to grid001f.cnaf.infn.it



Futures of Monitoring and Management – OASIS WSDM

- New OASIS TC (started up last month)
- Based on previous floundered OASIS TC
- Defining management of distributed resources USING Web services
- Defining management OF Web services operations and WSDL.
- Collaborate with W3C, GGF, DMTF, OASIS
- Chairs: Heather Kreger (IBM, Chair of WSA MTF) & Winston Bumpus (Novell, DMTF President)

Web Service Architecture – MTF

(<http://www.w3c.org>)

- Defining the manageability characteristics of the architectural elements of the Web Services architecture, i.e.:
 - Identification - data that uniquely identifies the element
 - Status - information about operational state of a element (up: busy/idle; down: stopped/saturated/crashed)
 - Configuration - a collection of behavioural properties which may be changed (persistent over instances)
 - Metrics - raw atomic, unambiguous information for managment purposes e.g. response times
 - Operations - methods that control or help manage the entity (instance specific)
 - Events - changes in the state of the entity e.g a lifecycle state change, or a state change.

DMTF

(<http://www.dmtf.org>)

Models real world managed objects (WBEM, CIM).

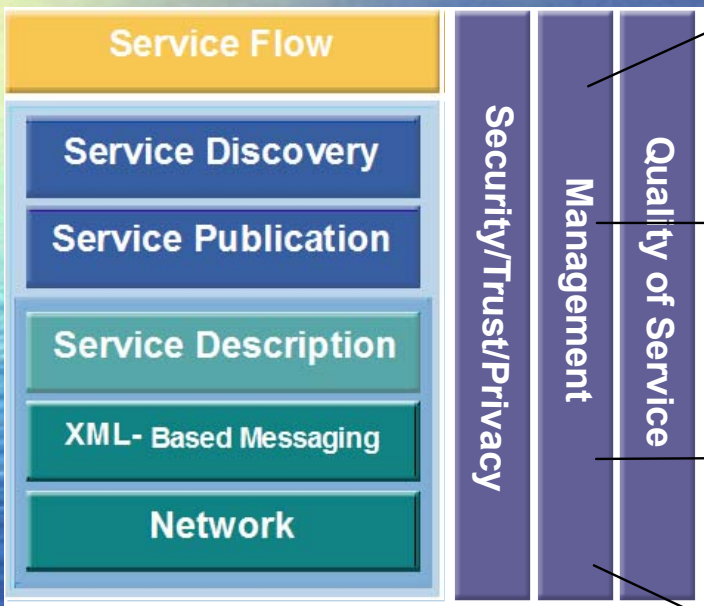
Large existing model (not in web/grid format/granularity)

- Application Working Group
 - Intends to model management of web services
- Interoperability Working Group
 - Defining a CIM/SOAP protocol in WSDL: CIM/Ops as WSDL operations and xmlCIM as the body of SOAP messages over HTTP

GRID

- OGSA Working Group on GRID Management and Monitoring
 - Object Oriented Inheritance
 - GRIDServices will be self describing
 - GRIDServices will be self managed

Web Services stack



Management requirements for a Manageable Web Services Architecture
W3C WS Arch WG, Management TF

Mangeability of Web Services
OASIS WSDM

Web service based access to management data -
OASIS WSDM, GGF CMM

Manageability portTypes -
OASIS WSDM, GGF CMM, DMTF

Questions & Discussion

