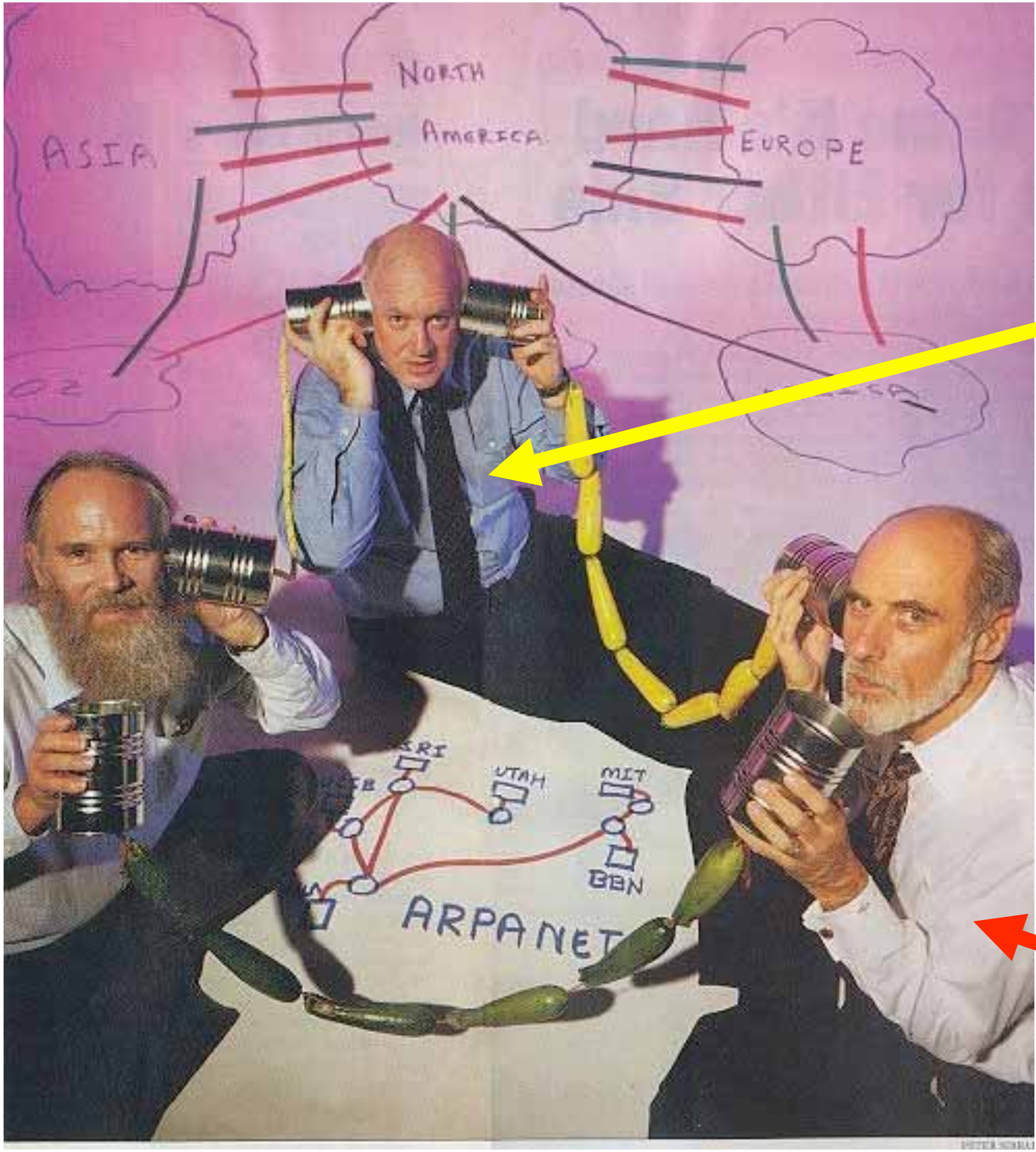
Grid Computing Evolution and Challenges for Resilience, Performance and Scalability

Luca Simoncini

University of Pisa, Italy

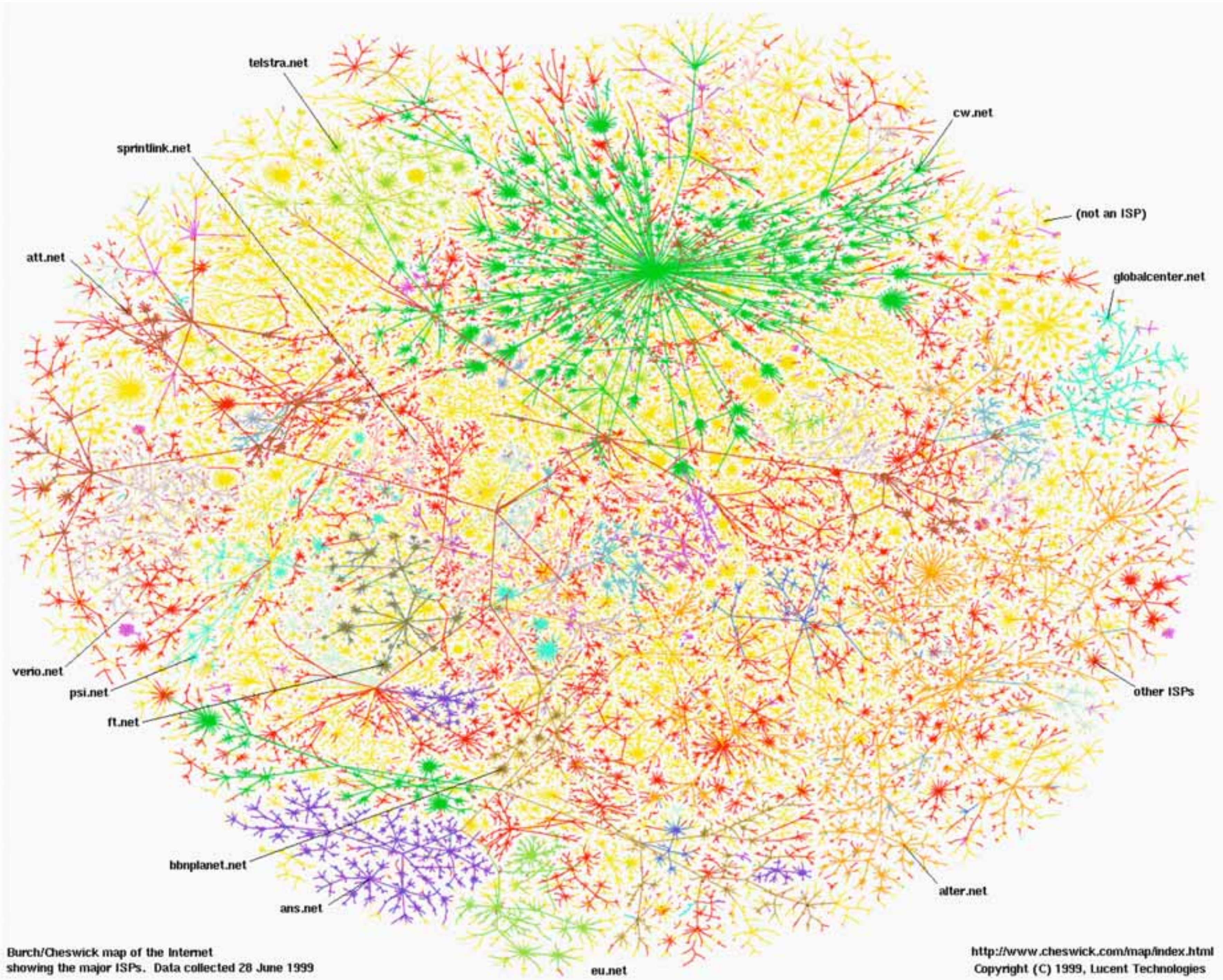


This photo was published in the August 8, 1994 issue of Newsweek and commemorates the 25th anniversary of the ARPANET. **Jon Postel**, **Steve Crocker** and I spent hours helping the photographer prepare for this shot.
Jon drew all the pictures, Steve and I strung the zucchini and the yellow squash. I think we must have collectively spent about 8 hours on this.

Note that this network can't work - there is no mouth/ear link anywhere!!!

Such was the state of networking in the primitive 1960s...

*Picture from **Vint Cerf***



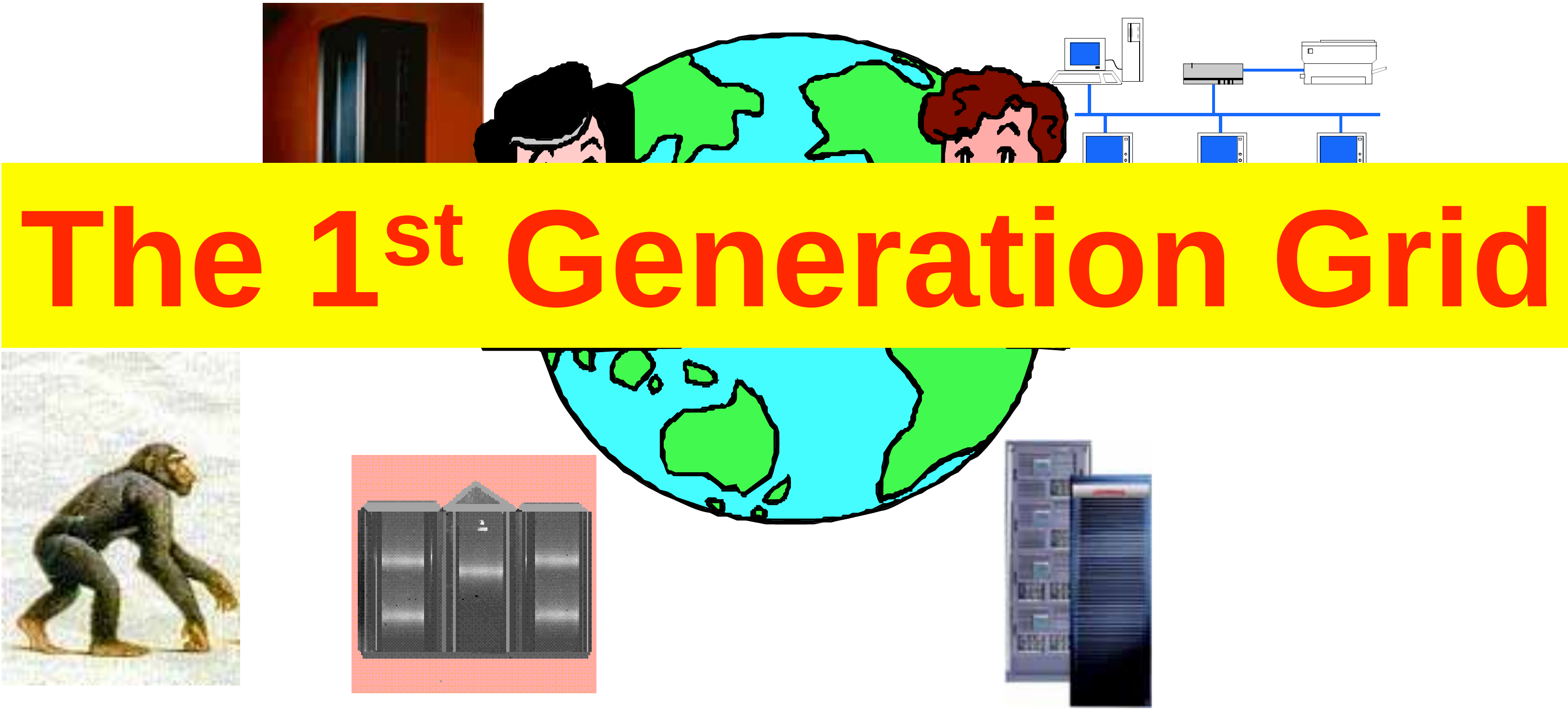
Burch/Cheswick map of the Internet
showing the major ISPs. Data collected 28 June 1999

http://www.cheswick.com/map/index.html

The term **“Grid”** means different things to different users groups and application domains.

- **Virtual organizations**. The Grid is seen as the collection of enabling technologies for building virtual organizations over the Internet.
- **Integration of resources.** The Grid is about building large-scale, distributed applications from distributed resources using a standard implementation-independent infrastructure.
- **Universal computer**. According to some (e.g., IBM-GRID25), the Grid is in effect a universal computer with memory, data storage, processing units, etc. that are distributed and are used transparently from applications.
- **Supercomputer interconnection**. The Grid is the result of interconnecting supercomputer centers together and enabling large-scale, long-running scientific computations with a very high demand regarding all kinds of computational, communication, and storage resources.
- **Distribution of computations**. Finally, there are those who see cycle-stealing applications, such as SETI@HOME, as typical Grid applications without any requirements for additional, underlying technologies.

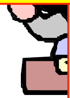
Grid Evolution - Metacomputing

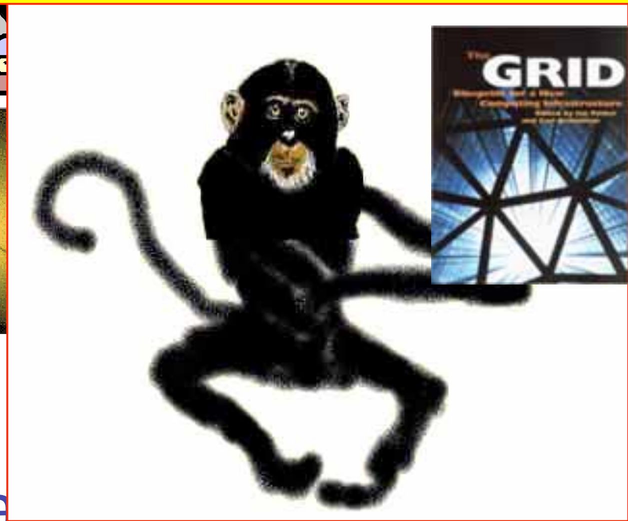


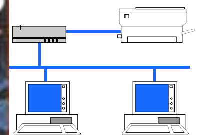
- ❑ Different Supercomputing Resourses
  - ❖ geographically distributed
  - ❖ used as a single **powerful** parallel machine (clear, High-Performance orientation)

Grid Evolution

The 2nd Generation Grid

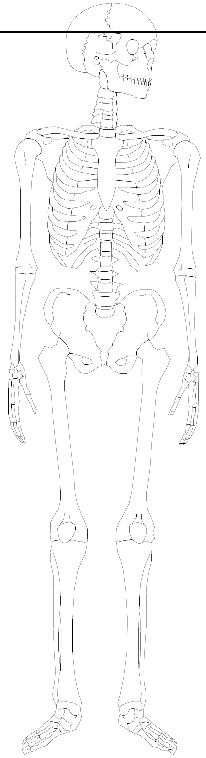


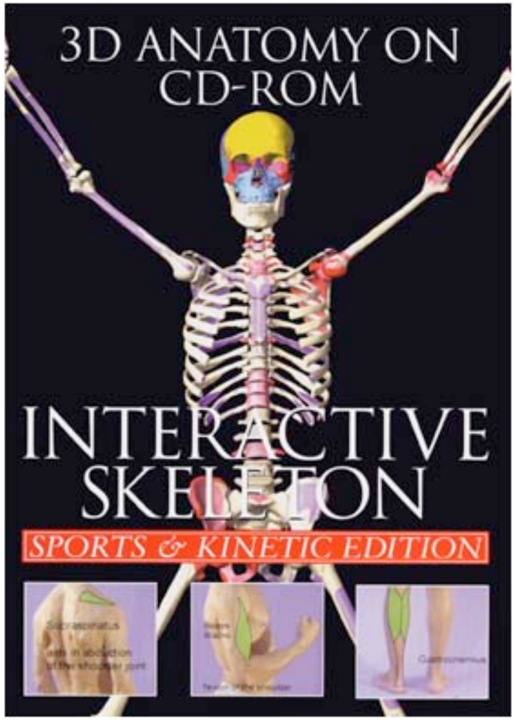


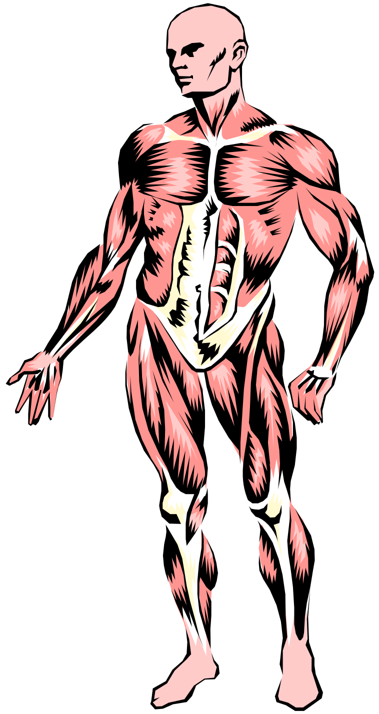


Grid computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation.

The Anatomy of the Grid:
Enabling Scalable Virtual Organizations







By Ian Foster, Carl Kesselman, and Steven Tuecke

The International Journal of High Performance Computing Applications
Volume 15, number 3, pages 200–222, Fall 2001

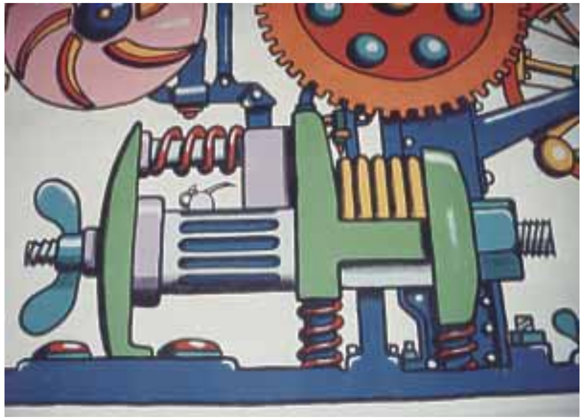
Open Question

Is the far-reaching vision offered by Grid Computing

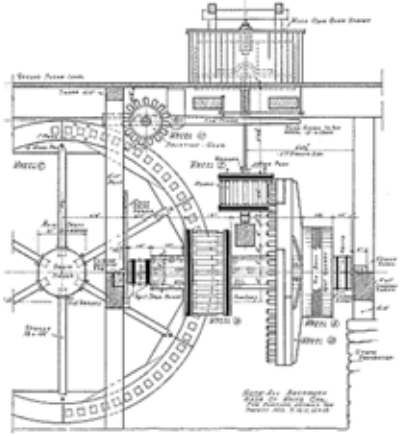
obscured by the

lack of interoperability standards

among Grid technologies ?



Interoperability



- Describes whether or not two components of a system that were developed with different tools or different vendor products can work together

hp invent

IBM

Platform

Microsoft

The Bottom Line in Distributed & Grid Computing

AVAKI

ORACLE

NEC

Sun microsystems

ENTROPIA PC GRID COMPUTING

FUJITSU

sgi

intel

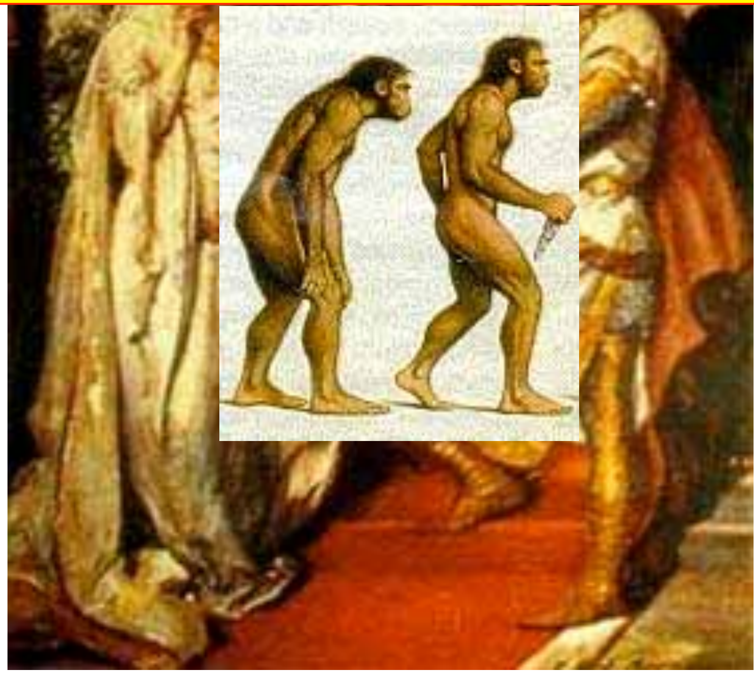
How to guarantee interoperability among Grids ?

Grid Standards & Alliances

- GGF
  - Research and Industry, use cases, architectures and specifications (OGSA, OGSI/WSRF)
- DMTF
  - Distributed Mgt. standards and models (CIM)
- OASIS
  - eBusiness & Web Services Management (WS-RF, WS-Notification, WSDM, …)
- EGA
  - Promote and grow Enterprise grid computing
- IETF
  - Internet architectures & specifications (SNMP, SMI)
- W3C
  - Web Services architectures and specifications
- SNIA
  - Advance the adoption of storage networks as complete and trusted solutions"

Grid Evolution

The 3rd Generation Grid



The marriage of the **Web technology** with the **2nd Generation Grid technology** led to new and generic Grid Services

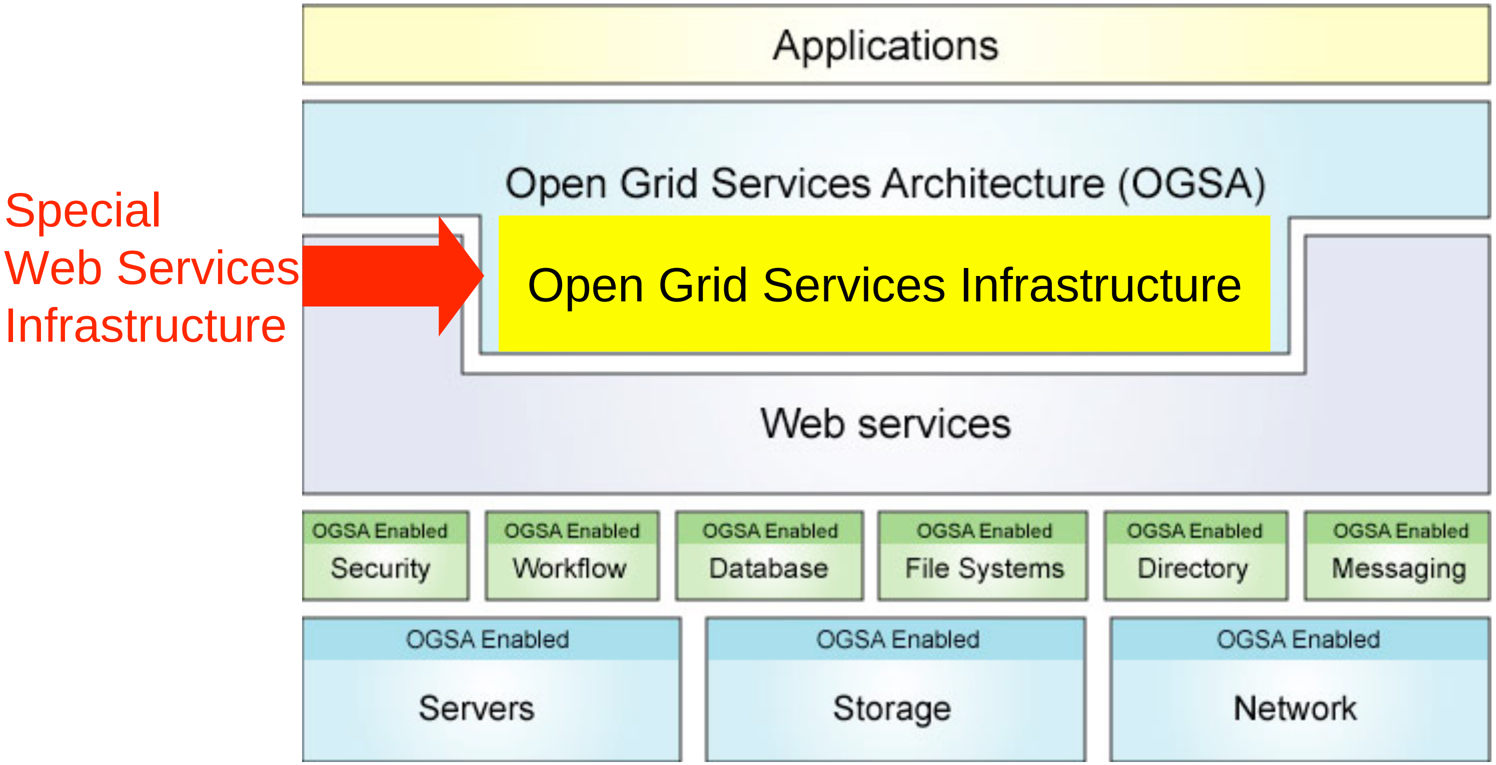
The Physiology of the Grid

An Open Grid Services Architecture for Distributed Systems Integration

I. Foster, C. Kesselman, J. Nick, S. Tuecke, January, 2002

http://www.globus.org/research/papers/ogsa.pdf

OGSA - OGSI


Special
Web Services
Infrastructure
Applications
Open Grid Services Architecture (OGSA)
Open Grid Services Infrastructure
Web services
OGSA Enabled
Security
OGSA Enabled
Workflow
OGSA Enabled
Database
OGSA Enabled
File Systems
OGSA Enabled
Directory
OGSA Enabled
Messaging
OGSA Enabled
Servers
OGSA Enabled
Storage
OGSA Enabled
Network

Hot News From

January 20, 2004

globusWORLD™
www.globusworld.org

Major Grid Services News:

The Globus Alliance and IBM in conjunction with HP announced details of the new:

WS-Resource Framework

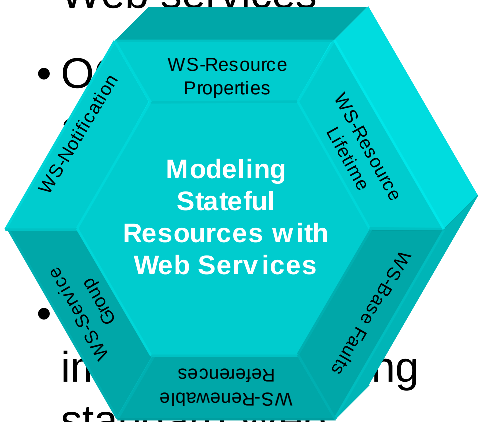
a further convergence of Grid services and Web services.

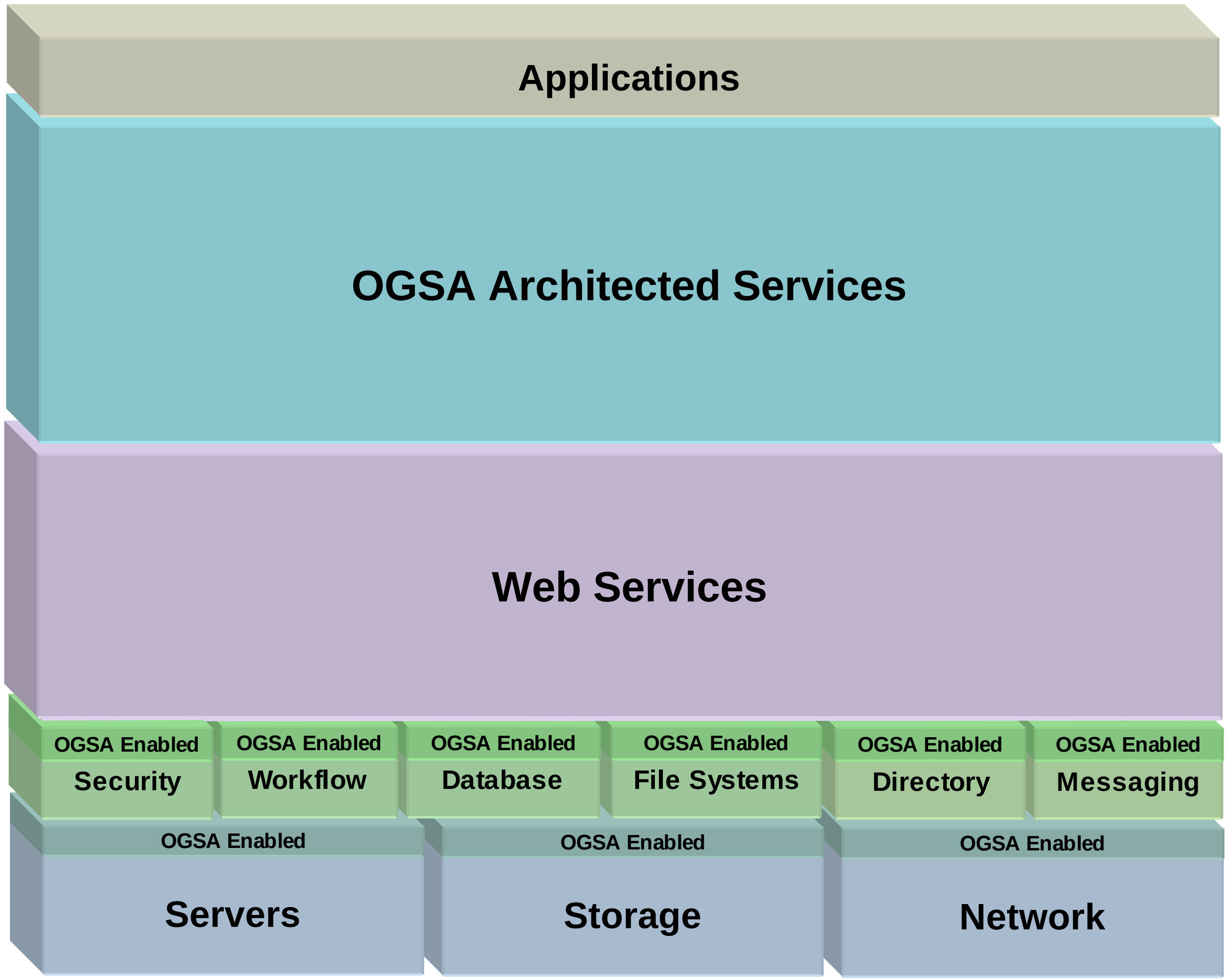
See: presentations by Daniel Sabbah of IBM and Ian Foster of the Globus Alliance for details.

How these proposals relate to OGSA

WS-Resource Framework & WS-Notification are an evolution of OGSI

- OGSA Services can be defined and implemented as Web services
- O
- in ng standard Web services development tools
- Grid applications will NOT require special Web services infrastructure





the globus consortium

January 24, 2005

About The Globus Consortium

- **The Globus Consortium - Bringing Open Source Grid Technology to the Enterprise**
  The Globus Consortium is the world's leading organization championing open source Grid technologies in the enterprise. With the support of industry leaders IBM, Intel, HP, and Sun Microsystems, the Globus Consortium draws together the vast resources of IT industry vendors, enterprise IT groups, and a vital open source developer community to advance use of the Globus Toolkit in the enterprise.

- The **Globus Toolkit** is the de facto standard for Grid infrastructure enabling IT managers to view all of their distributed computing resources around the world as a unified virtual datacenter. By giving enterprises access to computing resources as they need it, IT costs can go up and down as business demands. An open Grid infrastructure is the pre-requisite to fulfilling the promise of utility computing.

Sponsor-level members:

hp invent

IBM

intel

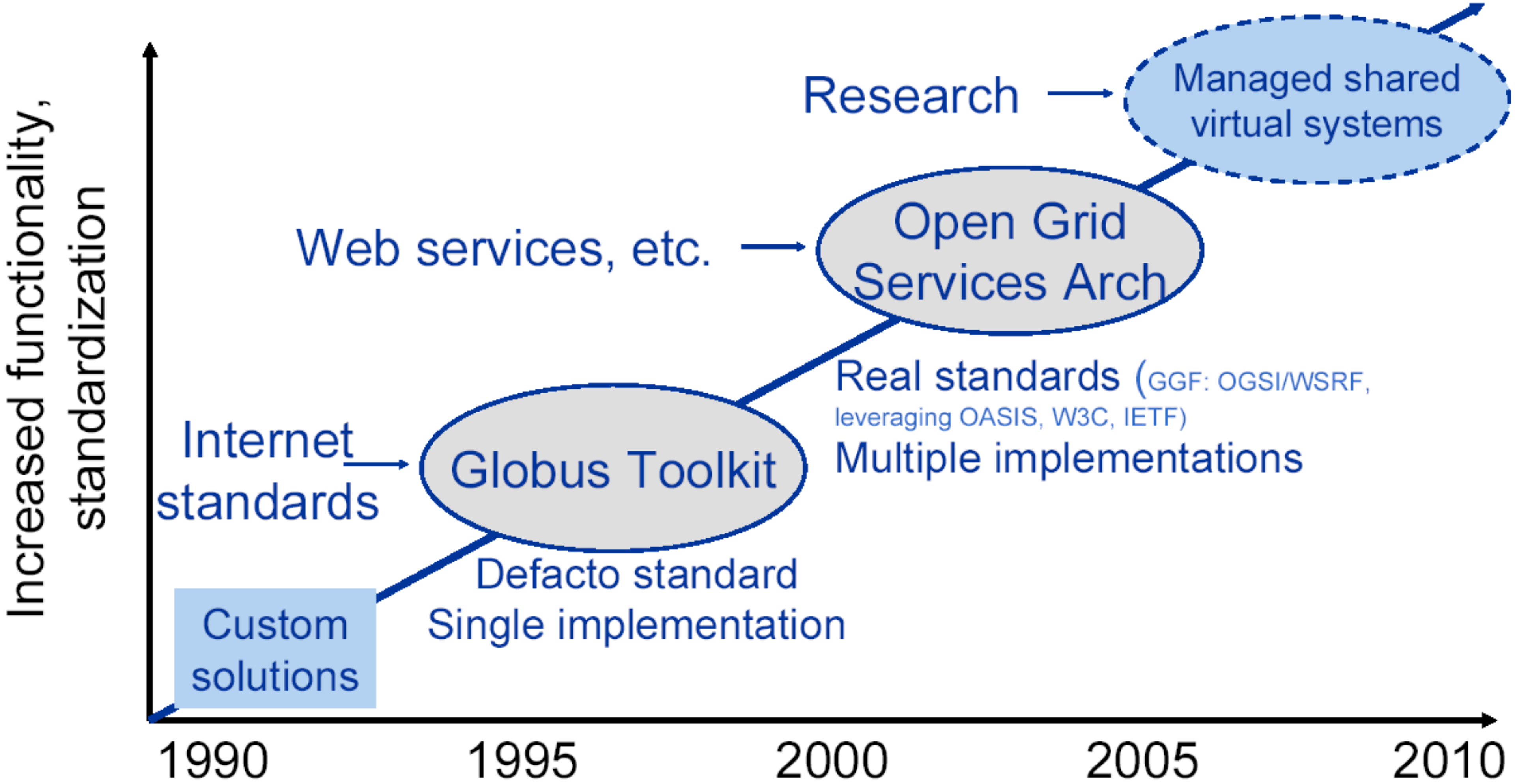
Sun microsystems

Contributor-level members:

NORTEL

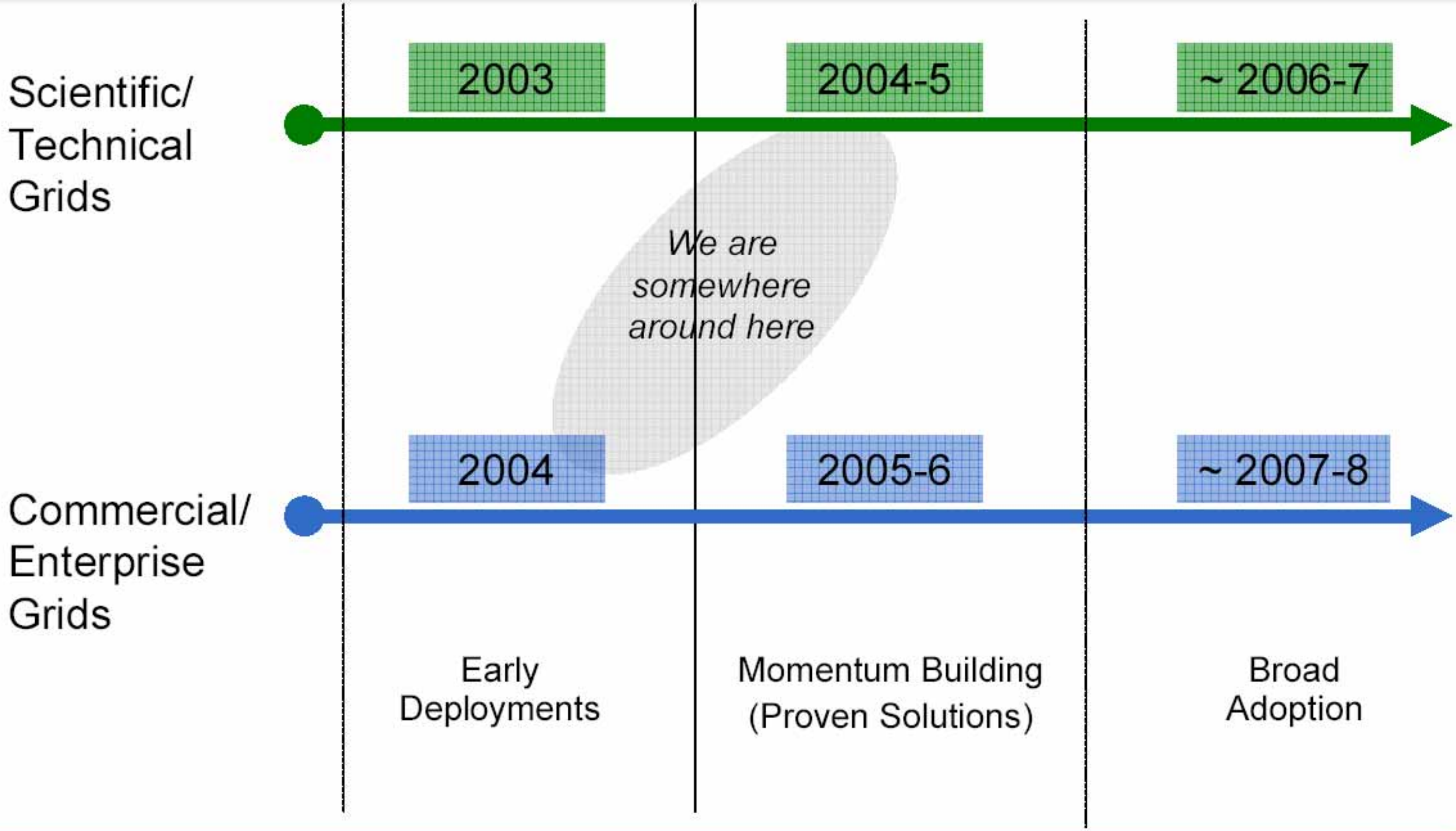
Univa

Developing Grid Standards



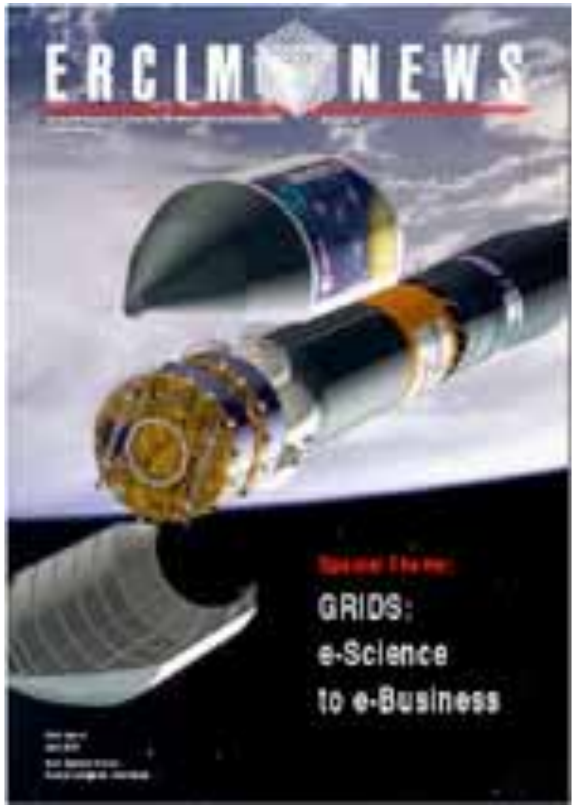
Source: Ian Foster - foster@mcs.anl.gov

Realistic Expectations



What is boiling in the (European) pot?

ERCIM News No.59, October 2004



ERCIM News No.45, April 2001

NGG1 and NGG2

Terms of reference

- Identify Research Priorities
  - 5 to 7 year timeframe
  - Include implementation strategies
- Propose an Implementation Roadmap
- Align Priorities with the European Research Agenda
- Network and Liaise with the Grid Community
- Propose actions to Improve International Collaboration

New Grid Research Projects

Total EU Funding:
52 M€

Start:
Mid 2004

grid@asia

GRIDCOORD
Building the ERA
in Grid research

Grid-based generic enabling application technologies to facilitate solution of industrial problems
SIMDAT

EU_driven Grid services architecture for business and industry
NEXTGRID

Mobile Grid architecture and services for dynamic virtual Organisations
AKOGRIMO

European_wide virtual laboratory for longer term Grid research _ foundation for next generation Grids
COREGRID

K-WF Grid
Knowledge based workflow & collaboration

UniGridS
Extended OGSA Implementation based on UNICORE

HPC4U
Fault tolerance, dependability for Grid

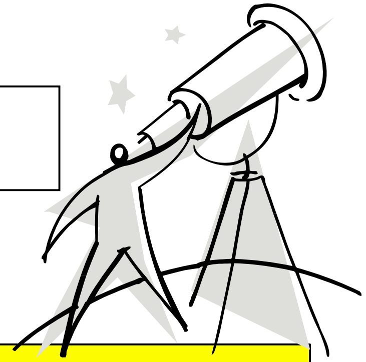
inteliGRID
Semantic Grid based virtual organisations

OntoGrid
Knowledge Services for the semantic Grid

DataminingGrid
Datamining tools & services

Provenance
Provenance for Grids

NGG from 3 Different Perspectives



The end users perspective

How the Grid might be
in everyday life,
ness drives Grid

The Grid as a structural entity with a
collection of capabilities and pro
Critical for an indication of the s
term of numbers, geography an
administrative domains.

What will it be like to
program the Grid?
What constraints
have to be observed
when developing
Grids?

The architectural perspective

NGG: The Wish List

❖ **Transparent and reliable**

❖ **Open to wide user and provider communities**

❖ **Pervasive and ubiquitous**

❖ **Secure and provides trust**
- ➢ Across multiple administrative domains

❖ **Easy to use and to program**

❖ **Persistent**
- ➢ Local and personal persistence as well as global persistence
- ➢ Strict reproducibility

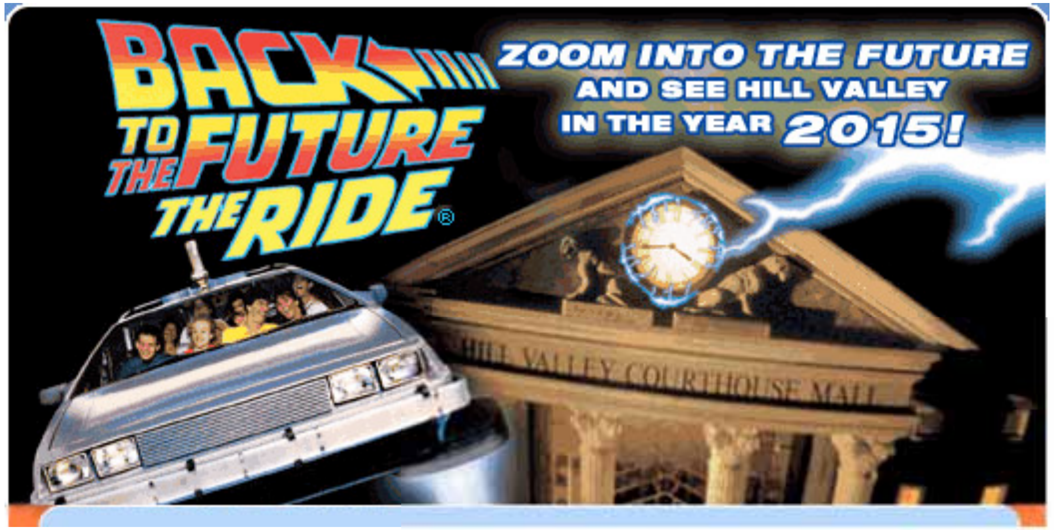
❖ **Person-centric**

❖ **Scalable and Scale Independent**

❖ **Easy to configure and manage**
- – Self managing

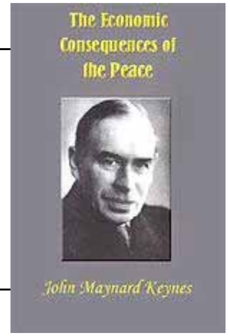
❖ **Based on standards for software and protocols**

GGF OASIS IETF
W3C WORLD WIDE WEB consortium
OMG OBJECT MANAGEMENT GROUP
WS-I WEB SERVICES INTEROPERABILITY ORGANIZATION
CCA Common Component Architecture



Looking into the Future

From e-Science to €-Business



- Towards the realisation of the "invisible Grid", offering key features for A Service-oriented Knowledge Utility
  - a new paradigm for software and service delivery, for the next decade.
- **Next Generation Grids 2 - Expert Group Report**
  - **http://www.cordis.lu/ist/grids/index.htm**
  - **ftp://ftp.cordis.lu/pub/ist/docs/ngg2_eg_final.pdf**

Service-Oriented architecture (SOA)

Definition

http://www.service-architecture.com/web-services/articles/service-oriented_architecture_soa_definition.html

- A service-oriented architecture is essentially a collection of services.
- A service is a function that is well-defined, self-contained, and does not depend on the context or state of other services.
- These services communicate with each other.
- The communication can involve either simple data passing or it could involve two or more services coordinating some activity.

Service-Oriented architecture (SOA)

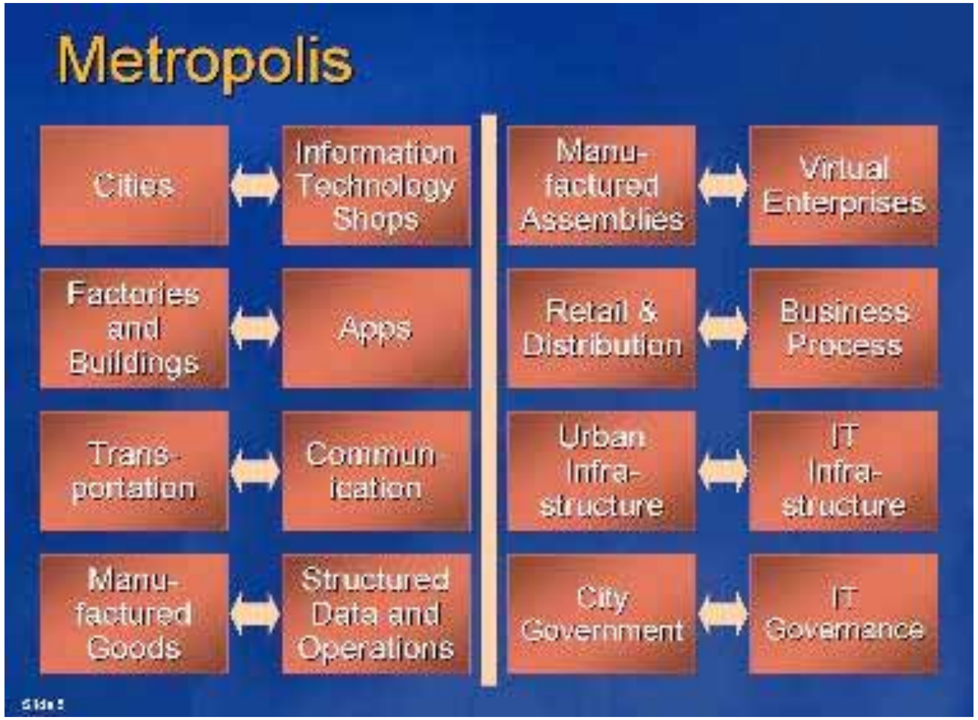
Definition

http://msdn.microsoft.com/architecture/soa/default.aspx

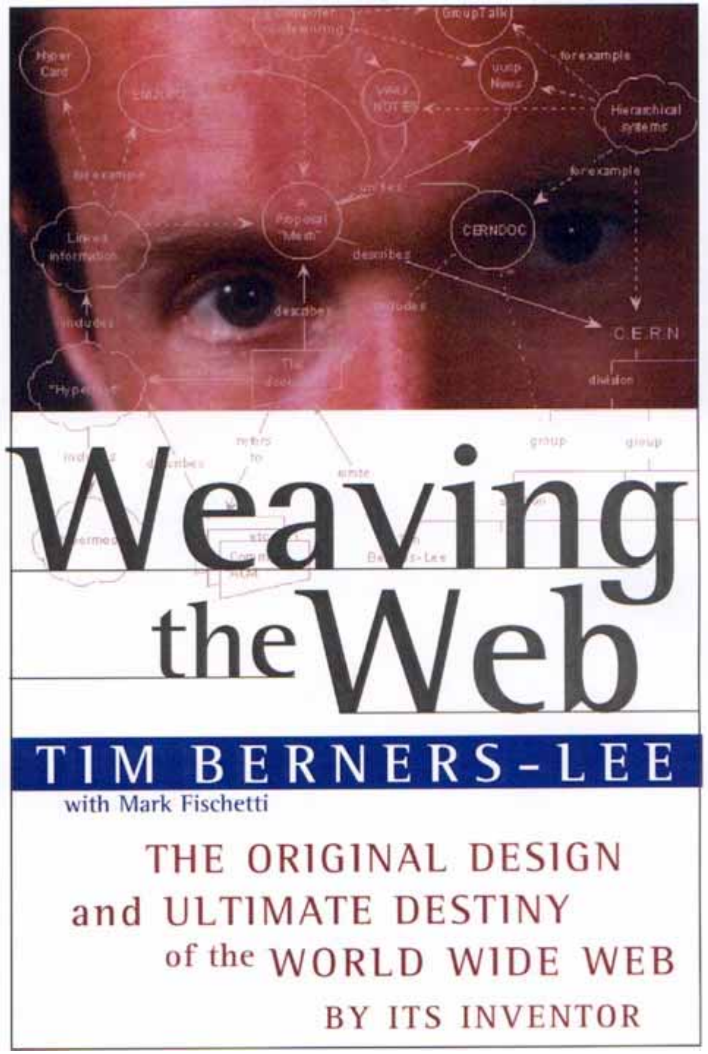
- The goal for Service Oriented Architecture (SOA) is a world-wide mesh of collaborating services that are published and available for invocation on a Service Bus.
- Adopting SOA is essential to delivering the business agility and IT flexibility promised by Web Services.
- These benefits are delivered not just by viewing service architecture from a technology perspective or by adopting Web Service protocols, but also by requiring the creation of a Service Oriented Environment that is based on specific key principles.

Metropolis : Envisioning the Service-Oriented Enterprise

http://msdn.microsoft.com/seminar/shared/asp/view.asp?url=/architecture/media/en/metrov2_part1/manifest.xml



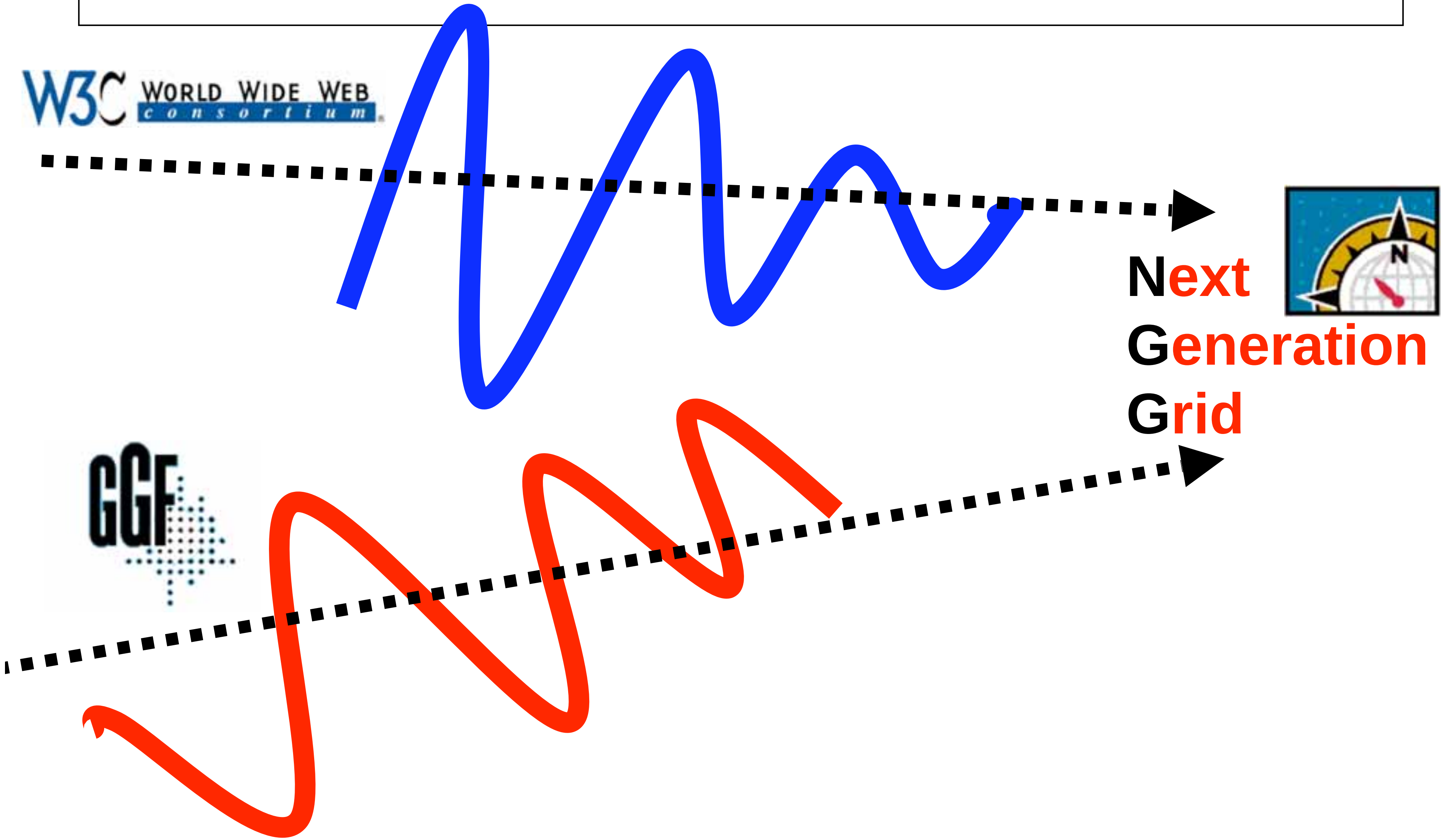
Semantic Web



1999

- "In the first part, the Web becomes a much more powerful means for collaboration between people …In the second part of the dream, collaborations extend to computers . ….
- A 'Semantic Web' which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives will be handled by machines talking to machines, leaving humans to provide the inspiration and intuition. . . The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form."

Convergence of Interests



Convergence is a need !

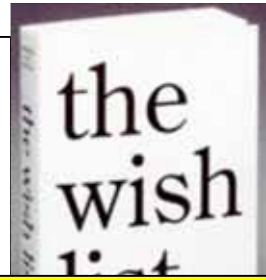
The long-awaited technologic convergence of the communications industry.

Mandatory

No Standard… ?
No Industrial/
Business Interest !

GGF
OASIS
IETF
W3C WORLD WIDE WEB consortium
OMG OBJECT MANAGEMENT GROUP
WS-I WEB SERVICES INTEROPERABILITY ORGANIZATION
CCA Common Component Architecture

Next Generation Grid Properties

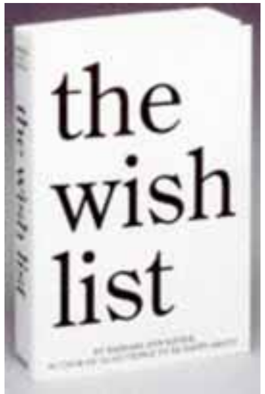


The current Grid implementations DO NOT individually possess all of these properties

Future Grids NOT possessing these properties are unlikely to be of significant use and, therefore, inadequate from business perspectives

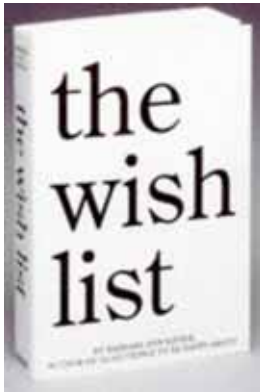
Performance and Dependability are key properties for NGG, but they are perceived as contrasting properties:

1)Long periods of grid services unavailability impact on performance
2)Techniques for resiliency may introduce overheads



Performability of grids is a holistic approach that has to include also security and business concerns

Challenges for performable grid systems and services

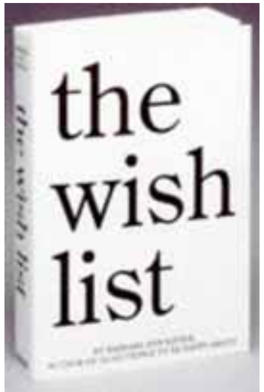


1. Standardization

- Definition of standards for metrics, models, modeling languages and formalisms
- Definition of benchmarks
- Independent approaches determine different means and tools for metrics and models
- Dominant projects that dictate standards, not necessarily have the best approach to performance and dependability

- Role of GGF W3C WORLD WIDE WEB consortium

and of the other standard bodies

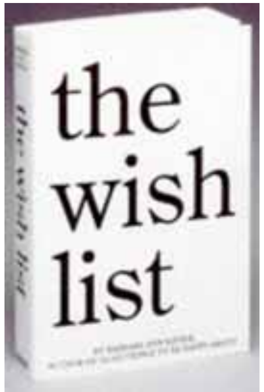


2. Virtualization

Virtualization enables a service to be offered seamlessly without awareness of what underlying services are used, their location, who provides them and if are used by others:

Hierarchy of services that can be managed as atomic entities, but introduce many problems from a modeling and measurement point of view:

- It is impossible to determine what resources are being used; different uses of the same service can be made by distinct sets of resources
- If a resources is overused, a task can be migrated to an alternative with different non-functional properties
- Different services may employ the same set of underlying services, becoming correlated and affected by common mode failures
  - this is a problem in both analysis and in design for deciding where and when using resilience techniques
- Difficult prediction of resource's workload
  - on-line monitoring of resources but role of interdependencies
- Complexity of models of system behavior
- Little work on this issue

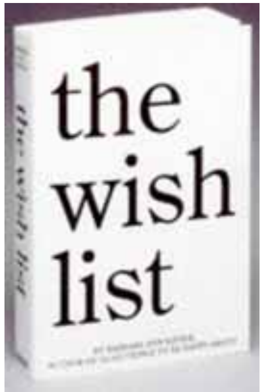


3. Measurement of complex systems

The size of grid systems, their heterogeneity and dynamicity create problems for performability analysis.

- **What to measure and where to measure**
- **Model-based evaluation of large complex systems will have to cope with large state spaces**
- **Simulation will have unacceptable run times**
- **Analytical models of complex systems, if available, are very costly to solve**

- **Need of techniques for efficient solutions of large models and for finding simple approximations**
- **Production of trustworthy approximations and verifiable techniques for model simplification**

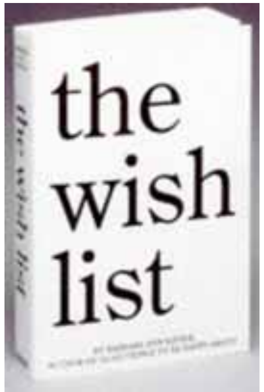


4. Resource management

Effective management of resources is a key part for providing QoS to customers; managing performability requires up to date knowledge of the state of the system operation:

- Being entirely up to date is unreasonable
- Performance may be increased if the choice of where directing a particular request is based on the best information available
- Predictive mechanisms:
  - efficient decomposition techniques
  - accurate approximations
  - scenario specific heuristics

- Identification of quasi-optimal policies and their evaluation
- Application oriented easily usable mechanisms



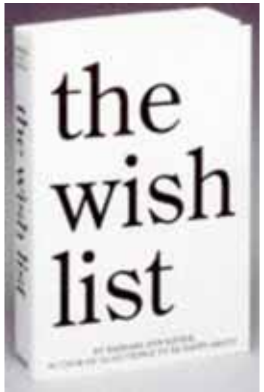
5. Realistic parameterization of systems

Performability models are only as good as the data that is used to populate them. If performance or availability is predicted on a conservative estimate for user demand then the system may have too little capacity and a far poorer expected performability

It is important to have accurate information *on demand* and for proposed models to be accurately verified against real data

Quite apart some work on grid scheduling, still much is to be done for:

- providing the right level of information across a wide range of systems in an accurate and timely manner
- providing new applications with accurate historical data from similar applications to be able to make accurate performability predictions



6. Business metrics

- ❖ Real metrics of interest are financial
- ❖ Increasing performability introduces costs

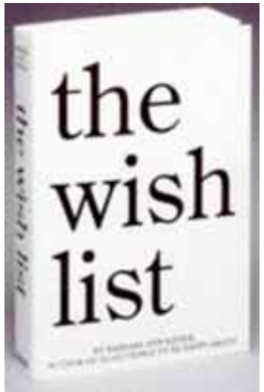
there is a need for a trade-off

- ➢ Grid systems are not simply a technical solution, but rather a different way of organizing business

- ➢ The core model is going to be a business process model and the technical models are going to be add-ons to this

- ➢ Need of understanding of charging models and their impact on user behavior

The relationship between charging and performability is very complex



7. Performance and security

- ❖ Grid systems involve sharing of large set of personal data some of which very valuable
- ❖ Protection of data is a key issue
- ❖ Making open systems secure is difficult and can introduce large unwanted overheads
- ❖ Some users may privilege performance over security and decide to turn off security measure
- ❖ Even if security developers do not consider performability as orthogonal to security, for sure, it is a secondary consideration for them.

Much work has to be done:

- ➢ to define acceptable trade-offs between security and performability
- ➢ to identify accurate even if approximate measures of security

More Research is needed…

- introduction of performability services
- understanding, integration of all these viewpoints and their absorption into standards

More international cooperation is needed….