

Grids and Grid Applications

C. Loomis (LAL-Orsay)

EGEE Induction (Clermont-Ferrand)

March 22, 2005

- **Based on previous presentations by:**
 - Dave Berry (NeSC) & David Fergusson
- **Contains material from:**
 - Andrew Grimshaw (Univ. of Virginia)
 - Bob Jones (EGEE Tech. Director)
 - Mark Parsons (EPCC)
 - EDG Training Team
 - Roberto Barbera (INFN)
 - Ian Foster (Argonne National Laboratories)
 - Jeffrey Grethe (SDSC)
 - The National e-Science Centre
 - M. Petitdidier (EGAPP presentation)
 - O. Gervasi (EGAPP presentation)

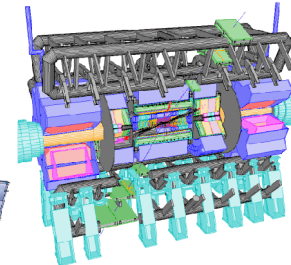
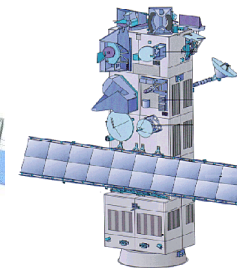
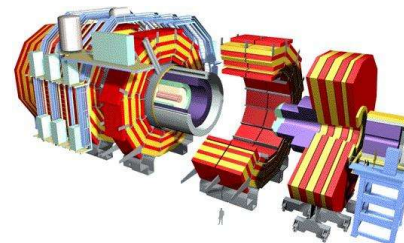
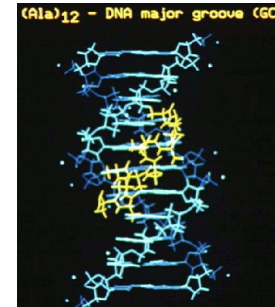
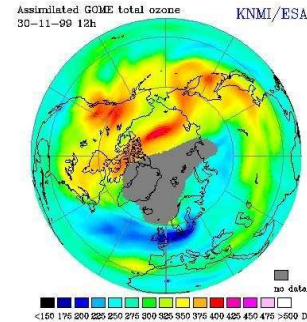
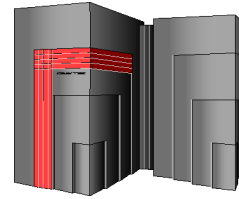
- **EGEE: Enabling *Grids* for *E-science***
- **Introduction to Grid Computing**
 - Motivation
 - Expectations & Constraints
 - Historical Perspective
 - Grid Architectures
 - Converging Technologies
- **Grid Applications (e-Science)**
 - Characteristics of e-Science
 - EGEE application areas
 - Typical Scenarios
- **Summary/Questions**

- **Goal in one sentence:**
 - Allow scientists from multiple domains to use, share, and manage geographically distributed resources transparently.
- **Simple statement, many consequences:**
 - Not specific to a particular application.
 - Jobs, policies cross administrative & political domains.
 - Sharing requires a means for accounting.
 - Transparency implies standardized services & APIs.
 - Access control for data and services.
 - Dynamic and heterogeneous resources.

- **Users**
 - Scientists with tasks requiring computational resources.
- **Virtual Organizations**
 - People from different institutions with common goals.
 - Share computational resources to achieve those goals.
- **System Administrators**
 - People responsible for keeping an institute's resources running.
 - Ensuring efficient and correct use of available resources.
- **Real Organizations**
 - Institutes, funding agencies, governments, ...
- **Standards Bodies**
 - OASIS, GGF, W3C, IETF, ...

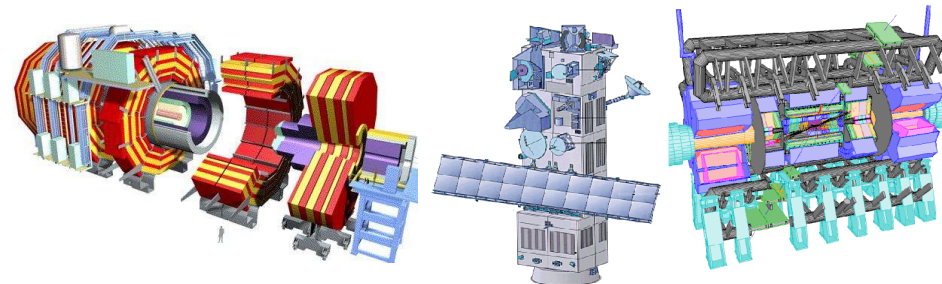
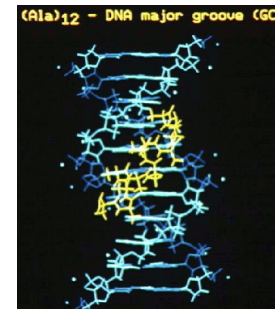
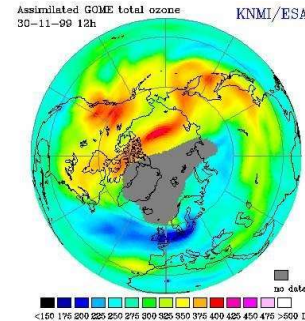
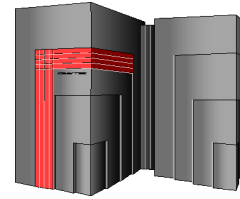


Grid “Middleware”





Grid “Middleware”



Grid technology allow scientists:

- access resources universally
- interact with colleagues
- analyze voluminous data
- share results



Grid "Middleware"

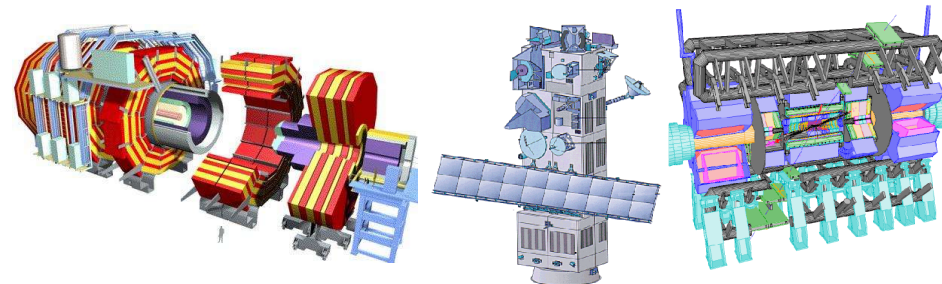
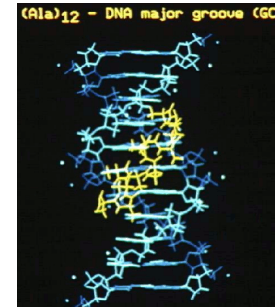
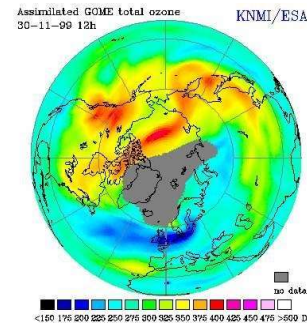


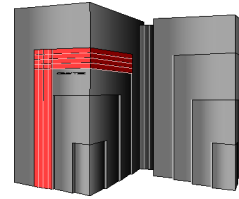
Includes traditional resources:

- raw compute power
- storage (disk, tape, ...)
- network connectivity

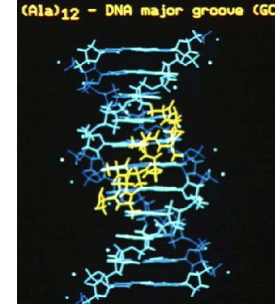
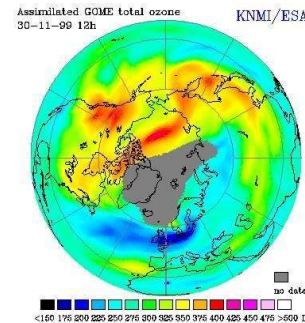
Resources are:

- heterogeneous
- dynamic





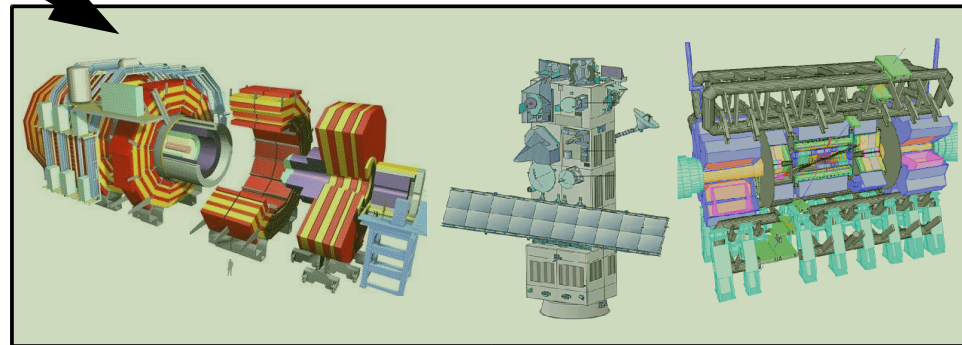
Grid "Middleware"

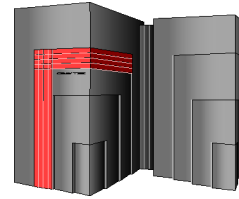


Detectors produce huge amounts of data for analysis.

Non-traditional resources:

- scientific instruments
- conferencing technologies
 - video
 - audio
 - chat





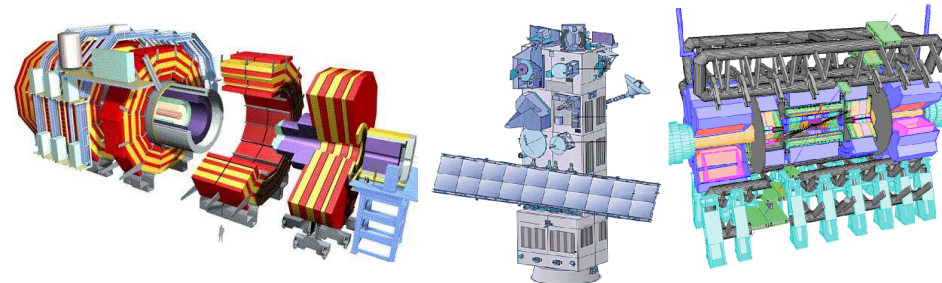
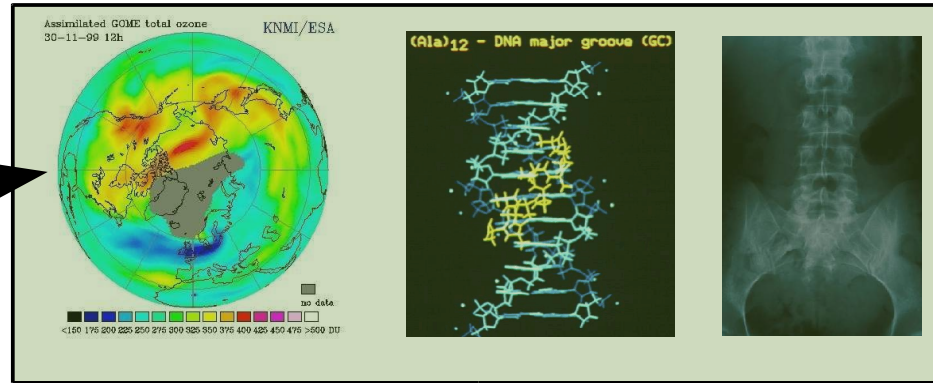
Grid "Middleware"

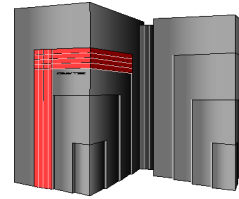
Access to data:

- data files and datasets
- databases
- replica metadata
- application metadata

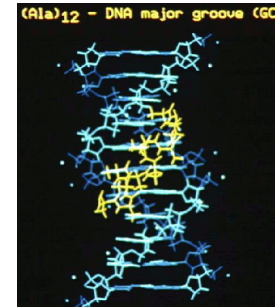
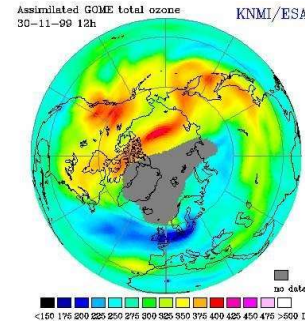
Manage data:

- transfer and copy data
- locate relevant data



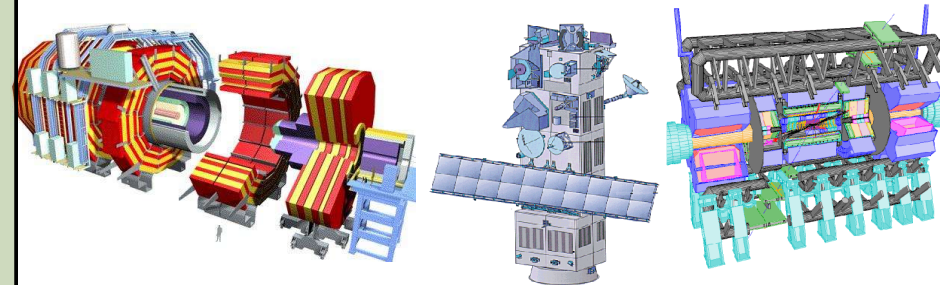


Grid "Middleware"



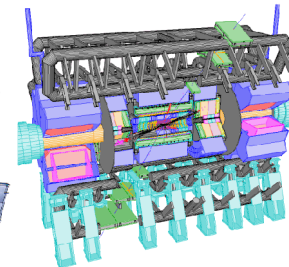
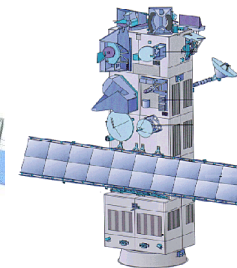
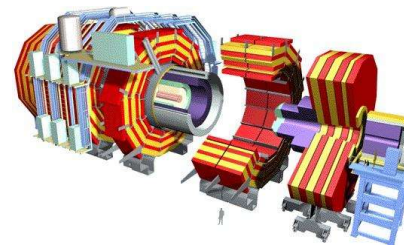
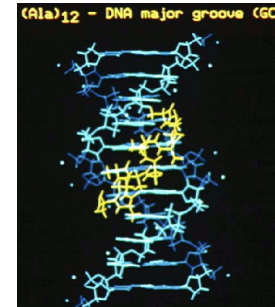
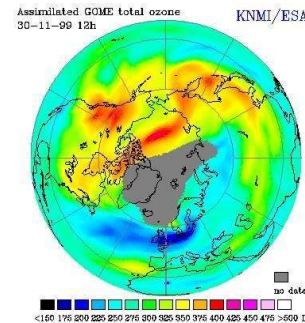
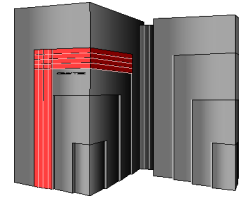
Services:

- high-level services to facilitate use of grid
 - e.g. job brokering
- application-specific services
 - e.g. portals





Grid “Middleware”



What is the grid?

- **Middleware:**
 - service interoperability
 - high-level services
- **Resources:**
 - provided by participants
 - shared for efficient use

- **Avoid reinventing the wheel:**
 - Many computational tasks are common.
 - High-level, standardized services avoid duplication.
 - Scientists concentrate on results rather than tools.
- **Resource needs grow with time:**
 - Start small for testing.
 - Push limits for ultimate sensitivity.
 - Grid APIs make finding and using additional resources easier.
- **Data access:**
 - Find and access existing data more easily.
 - Share results for others to build upon.

- **Use of computing resources varies with time.**
 - Analysis rush before major conferences.
 - End-of-quarter financial analyzes.
 - July and August holidays.
- **Current solutions:**
 - Buy peak needed capacity; idle in non-peak periods.
 - Buy average capacity; delay results.
- **Grid solution:**
 - Share resources to time-shift availability.
 - Buy average capacity but get timely results!
 - Improve reliability with automatic failover.

Access Grid
DISCOM
DOE Science Grid
Condor
ESG (Earth System Grid)
Fusion Collaboratory
Globus
GrADSoft (Grid Application Development Software)
Grid Canada
GRIDS (Grid Research Integration Development & Support Center)
GriPhyN (Grid Physics Network)
iVDGL (International Virtual Data Grid Laboratory)
Music Grid
NASA Information Power Grid
NCSA Alliance Access Grid

AstroGrid
AVO (Astrophysical Virtual Observatory)
Comb-e-chem
CrossGrid
DAME (Distributed Aircraft Maintenance Environment)
DAMIEN (Distributed Applications and Middleware for Industrial Networks)
DataTAG
Discovery Net
DutchGrid
EDG (European DataGrid)
EGSO (European Grid of Solar Observations)
GEODISE (Grid Enabled Optimisation & Design Search for Engineering)

GRIA (Grid Resources for Industrial Applications)
Grid-Ireland
GridLab (Grid Application Toolkit and Testbed)
GridPP
LCG (LHC Computing Grid)
MyGrid
NGIL (National Grid for Learning Scotland)
NorduGrid (Nordic Testbed for Wide Area Computing and Data Handling)
PIONIER Grid
Reality Grid
ScotGrid

ApGrid
ApBioNet
Grid Forum Korea
PRAGMA (Rim Applications and Grid Middleware Assembly)
Grid Datafarm for Petascale Data Intensive Computing
Gridbus Project

- **European Funded**

- European DataGrid
- CrossGrid
- DataTAG
- LHC Computing Grid
- GridLab
- EUROGRID
- DEISA
- EGEE



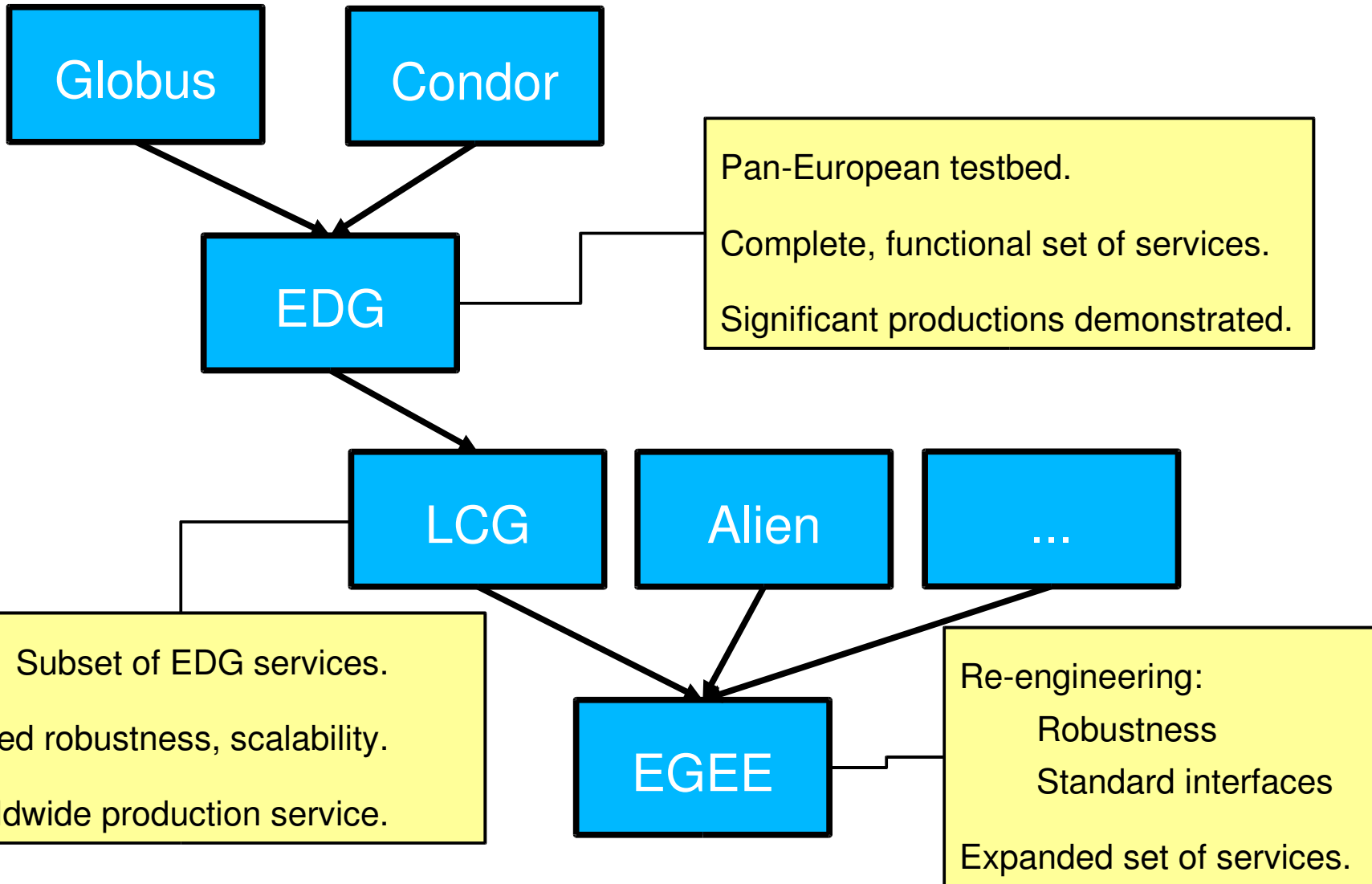
crossgrid



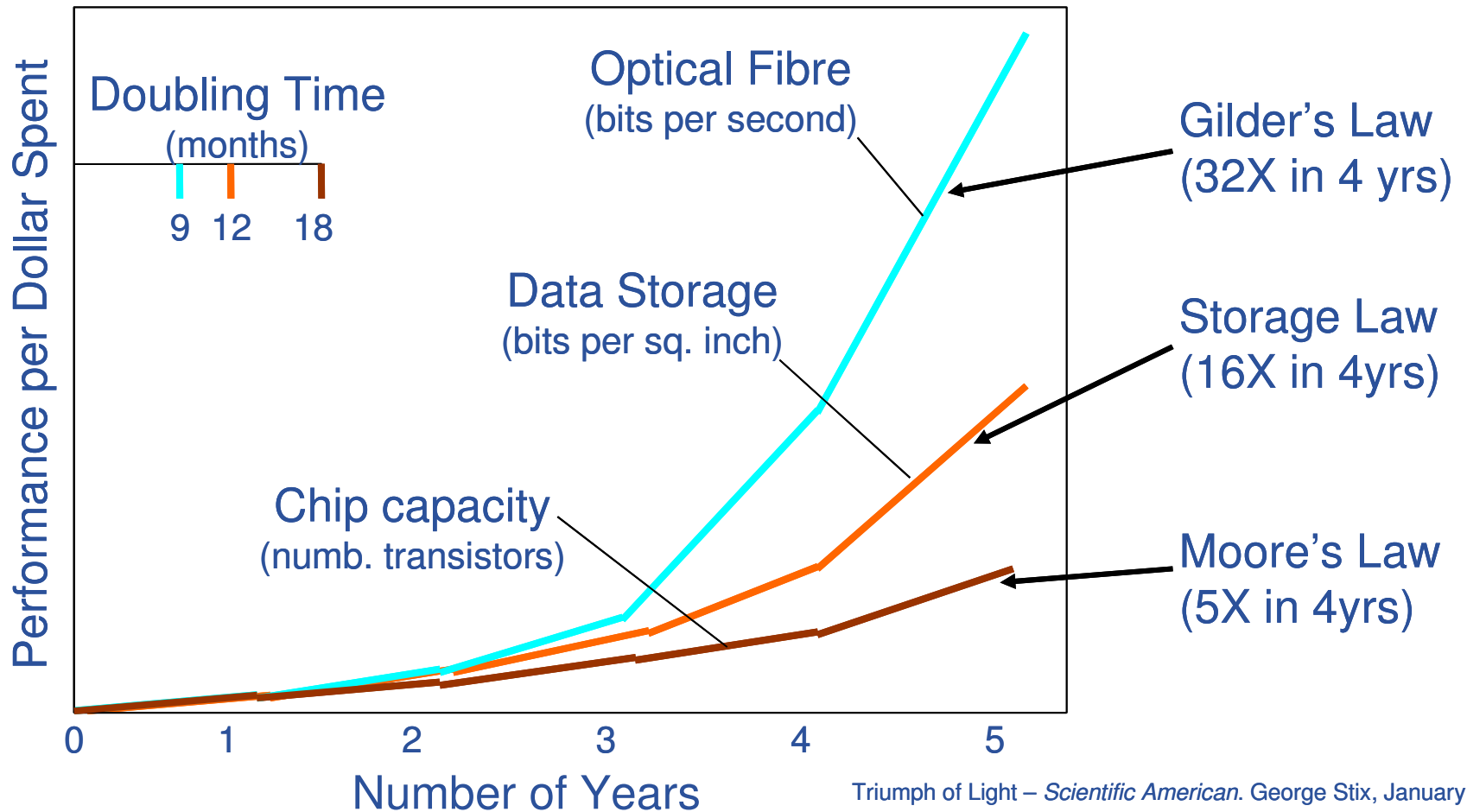
- **National Grid Efforts**

- INFN Grid
- NorduGrid
- UK e-Science Programme



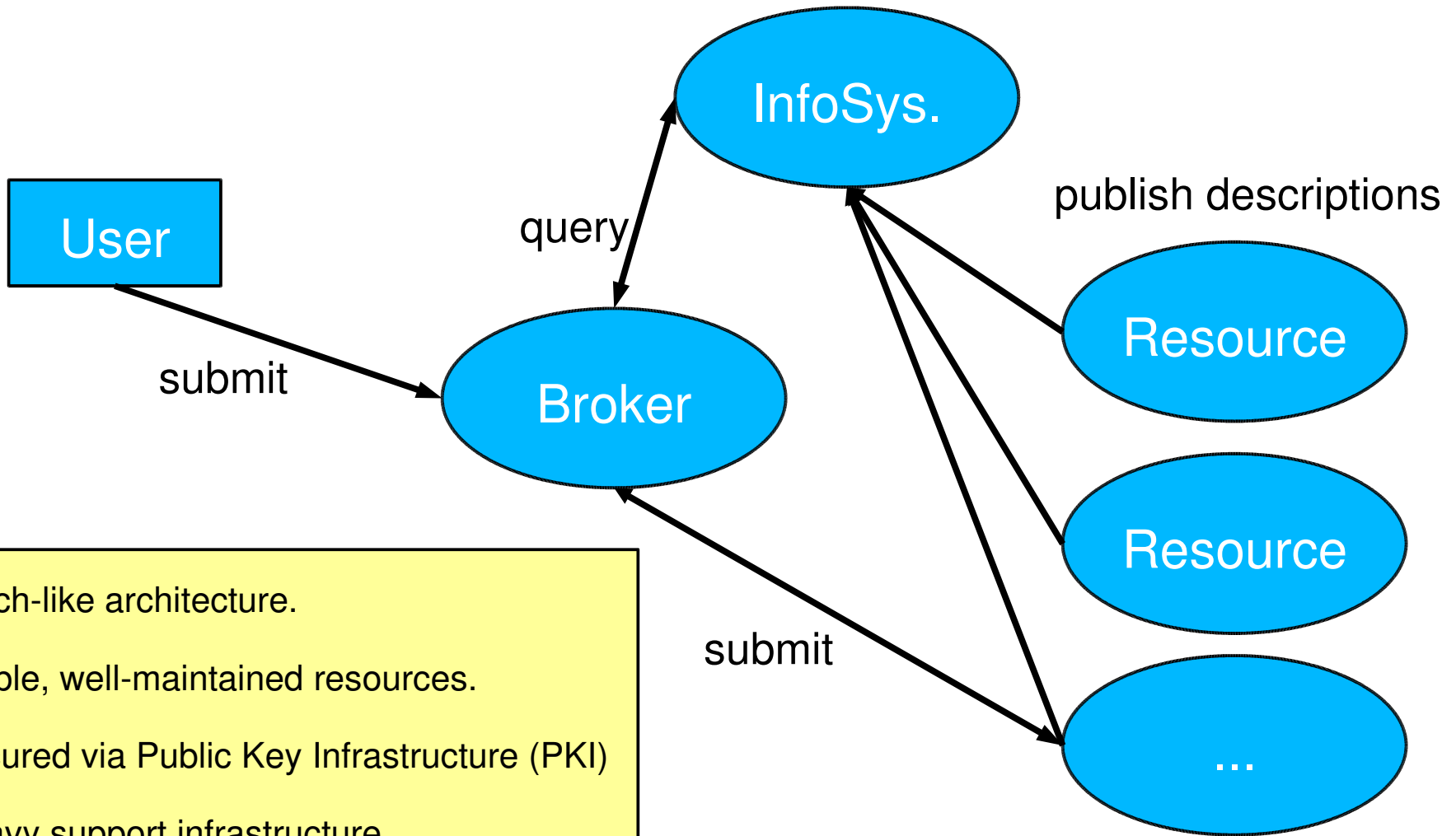


- Relative CPU, storage, and network capability impacts computing architecture.

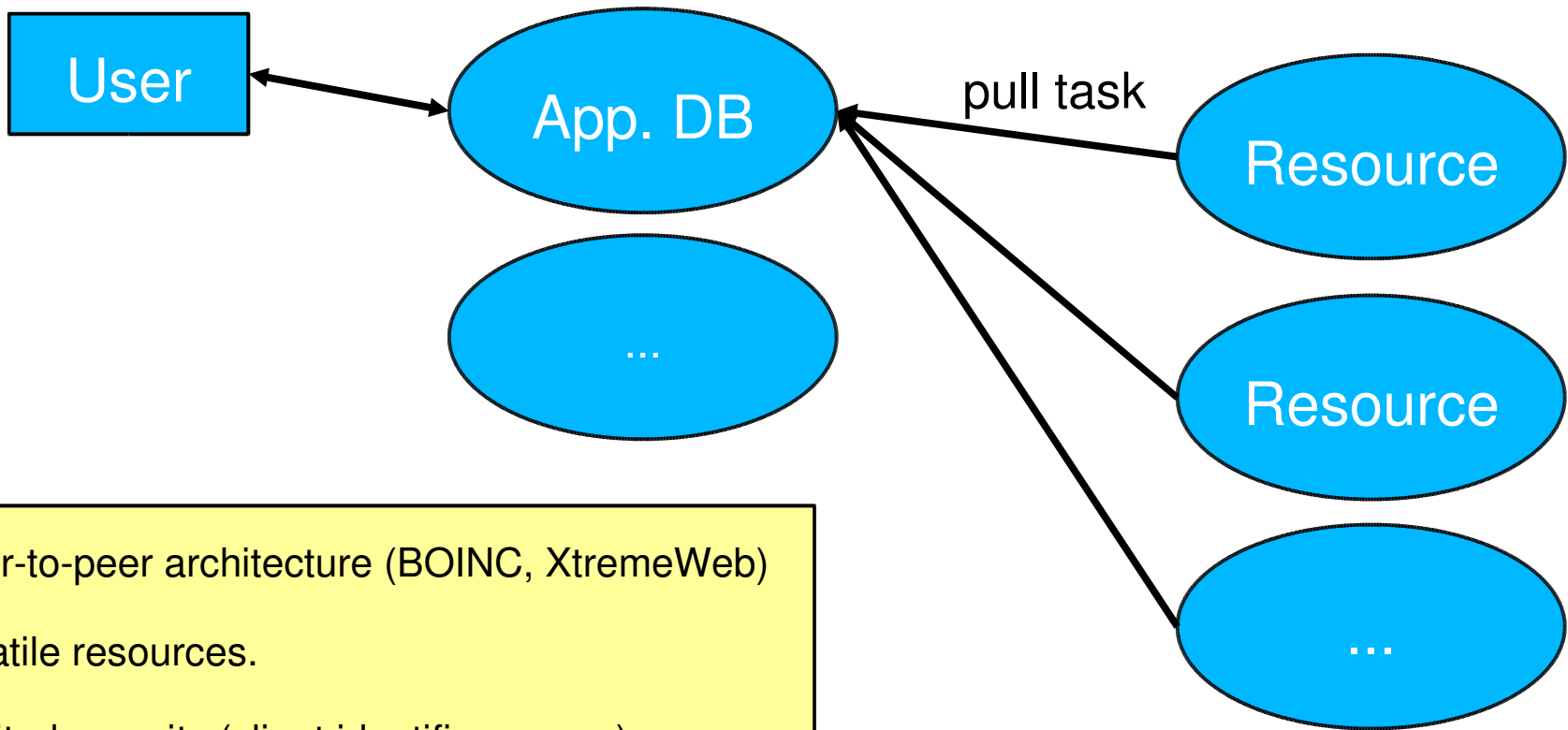


Triumph of Light - Scientific American. George Stix, January 2001

- **Local Computing**
 - All computing resources at single site.
 - People move to resources to work.
- **Remote Computing**
 - Resources accessible from distance.
 - All significant resources still centralized.
- **Distributed Computing**
 - Resources geographically distributed.
 - Specialized access; largely data transfers.
- **Grid Computing**
 - Resources and services geographically distributed.
 - Standard interfaces; transfers of computations and data.



- Batch-like architecture.
- Stable, well-maintained resources.
- Secured via Public Key Infrastructure (PKI)
- Heavy support infrastructure.
- Can handle large range of resources.



Peer-to-peer architecture (BOINC, XtremeWeb)

Volatile resources.

Limited security (client identifies server).

Lightweight infrastructure.

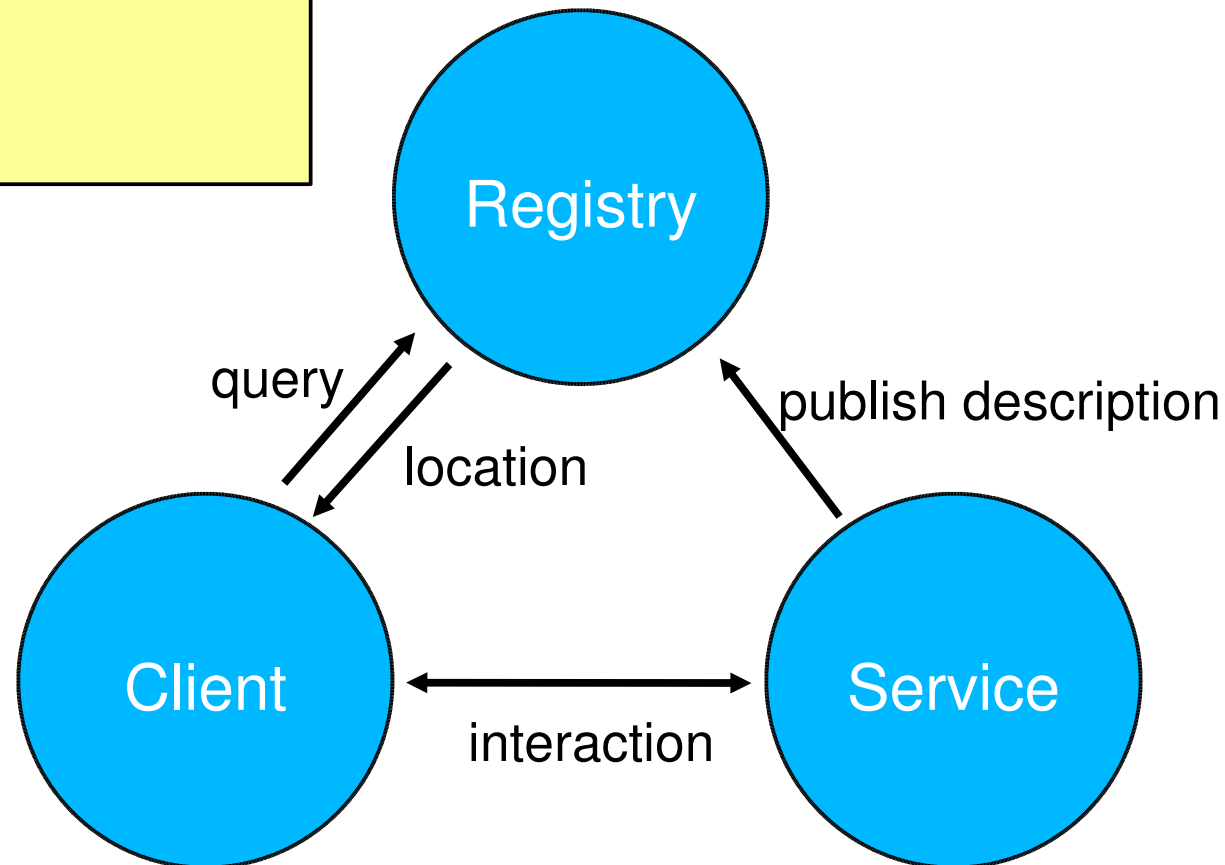
Handles limited types of resources.

Existing LCG system is largely service-oriented.

EGEE evolving to a clean SOA:

standard interfaces

standard technologies



- **Web Services**

- Clean, complete specification of service APIs.
- Supported technology:
 - Good support within commercial sector.
 - Adequate support within open-source community.
- Very active → proposed standards rapidly evolving.

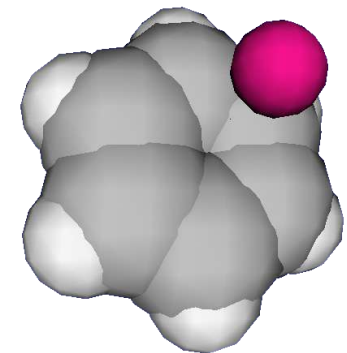
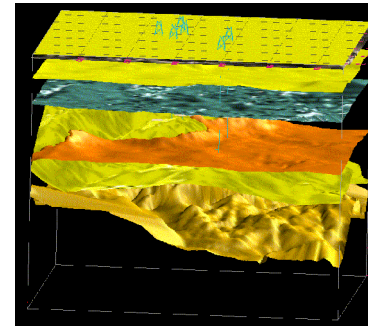
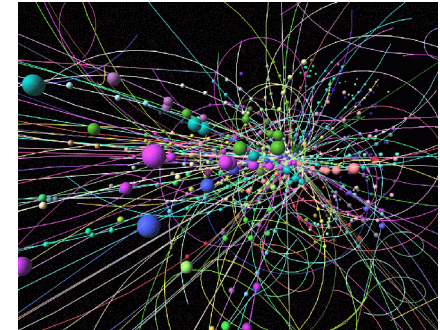
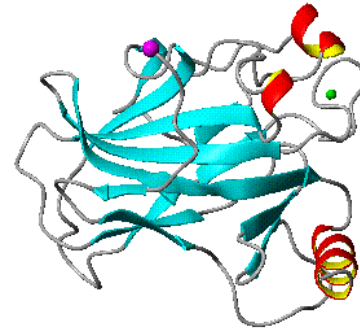
- **EGEE Service Evolution**

- Plain web services:
 - Avoid “proprietary” protocols and interfaces.
 - Fairly stable, will ease further evolution.
- Adopt WSRF and/or WS-* standards as appropriate.

- **Expect user-visible changes in APIs.**

- ***e-Science***: Pushing frontiers of scientific discovery by exploiting advanced computational methods.
- **Use grid technology to:**
 - Generate, curate, and analyze research data.
 - Develop and explore models and simulations.
 - Facilitate sharing of data, results, and resources.

- **Biomedical Applications**
 - imaging, diagnosis, treatment
 - genome and protein studies
- **High-Energy Physics**
 - simulation of particle interactions
 - analysis of detector data
- **Earth Science**
 - observing terrestrial conditions
 - natural resources
- **Computational Chemistry**
 - simulation of chemical properties



- **Batch Use**
 - Use the grid as a huge computational resource.
 - Simulation plays an important role in nearly all fields.
- **Portal Use**
 - Use grid for load balancing of standardized applications.
 - Provides easy interface which hides grid complexities.
- **Agent Use**
 - Centralized control and monitoring of a large production.
 - “Agent” jobs contact central database when started.
- **Interactive Use**
 - Provide improved response time for intensive calculations.
 - Debugging of applications *in situ*.

- **Multi-site parallel jobs:**
 - Can't guarantee simultaneous start of all processes.
 - Can't easily control bandwidth and latencies between processes.
 - Can expose MPI-enabled site as a grid resource.
- **Tasks requiring real-time response.**
 - Same as above: start up and response latencies, bandwidth.
 - If start up latency OK, can use “agent” model.

Grid technology attractive to many scientific endeavors:

- **Provides means of sharing resources to:**
 - reduce overall hardware cost
 - reduce response times
 - improve reliability
- **Standardized, high-level APIs:**
 - allow services to inter-operate effectively
 - allow scientists to concentrate on science rather than tools
- **EGEE:**
 - improving the technology
 - deploying a powerful, worldwide grid

- **Questions?**