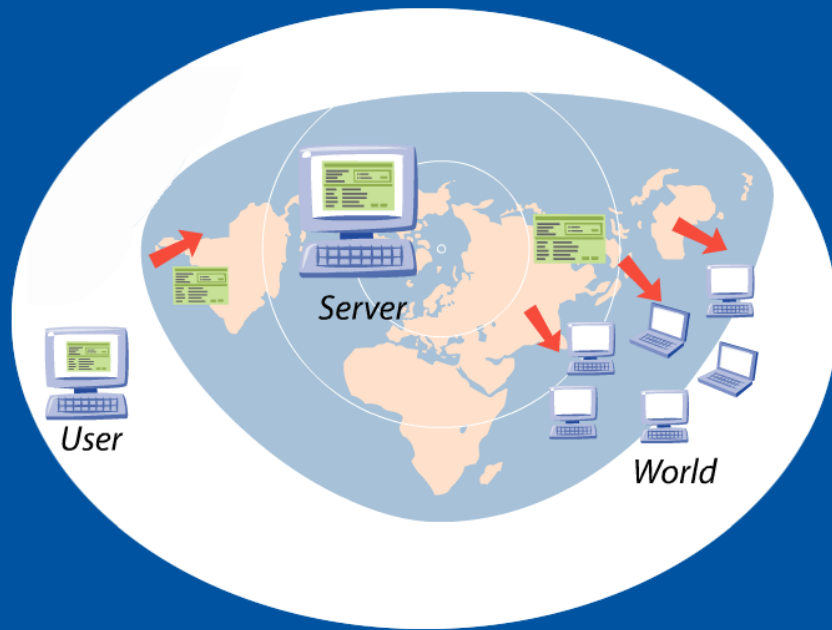


*From the  
Web*



*to the  
Grid*



*Dr. Francois Grey  
IT Department, CERN*



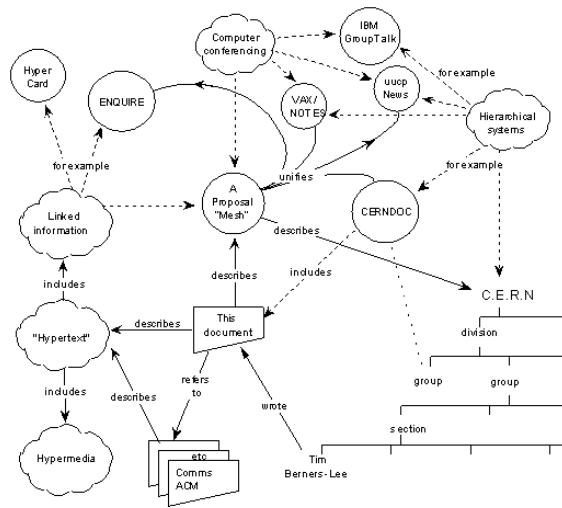
# Why was the Web invented at CERN?

- Science depends on free **access to information** and exchange of ideas. CERN is the hub of a worldwide community of **6500 scientists** in **80 countries**.
- CERN has a long history of being at the forefront of **scientific computing** and **networking** (first lab on Internet outside the US).
- During the preparation of the previous large project LEP, the need to share documents in a global way became vital.

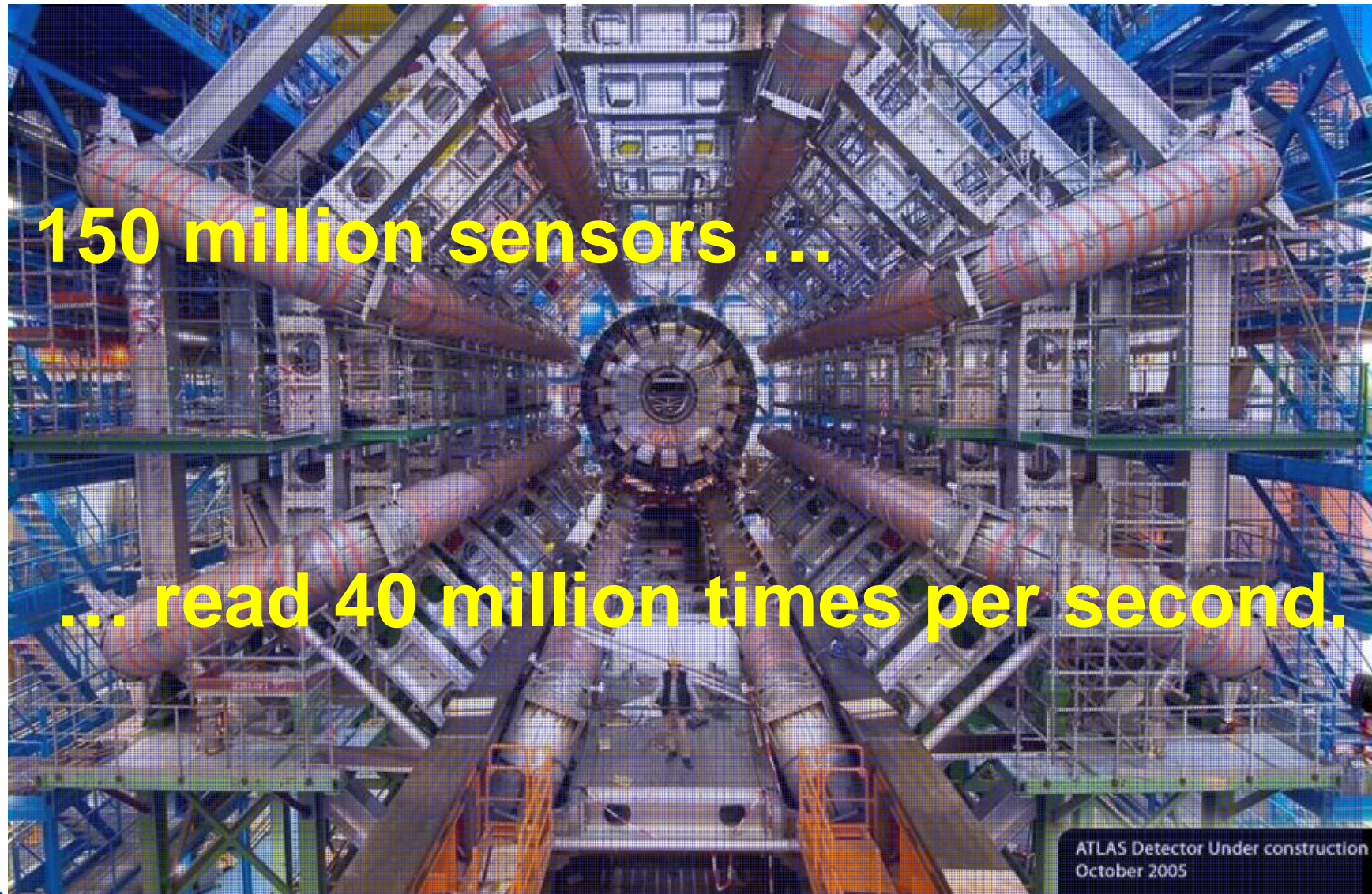


# How did the Web start?

- 1989: **Tim Berners-Lee** circulates “Information Management: A proposal” to help with future Large Hadron Collider project.
- 1991: Early **www system released** to high energy physics via the CERN program library. First web servers located in European physics laboratories.
- 1993: First Mosaic browser; web reaches 500 servers and 1% of Internet traffic; CERN places **WWW in the public domain**.



## View of the ATLAS detector (under construction)



# Petabyte science

- Detectors read 40 million times per second
- After filtering, 100 collisions of interest per second
- > 1 Megabyte of data per collision  
recording rate > 1 Gigabyte/sec
- $10^{10}$  collisions recorded each year  
stored data > 10 Petabytes/year

**1 Megabyte (1MB)**

*A digital photo*

**1 Gigabyte (1GB)**

**= 1000MB**

**5GB = A DVD movie**

**1 Terabyte (1TB)**

**= 1000GB**

**World annual  
book production**

**1 Petabyte (1PB)**

**= 1000TB**

**Annual production of  
one LHC experiment**

**1 Exabyte (1EB)**

**= 1000 PB**

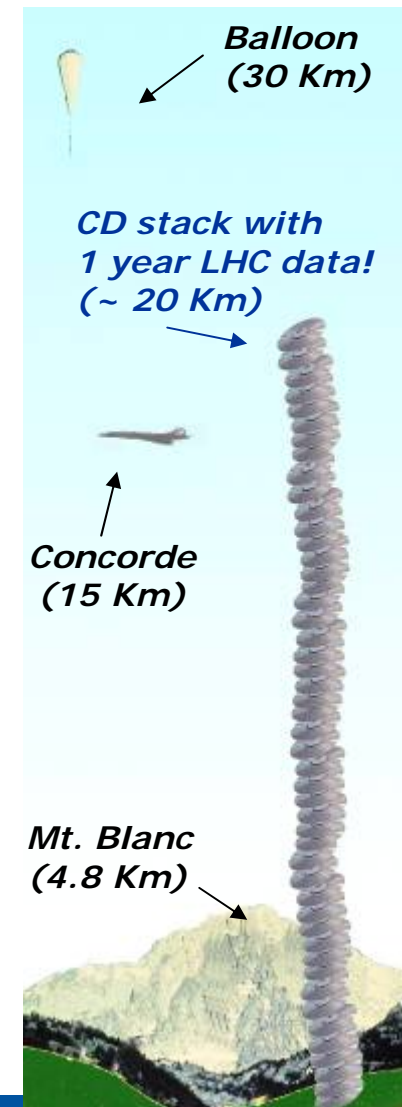
**3EB = World annual  
information production**



# Data Storage for the LHC

- LHC data correspond to about 20 million CDs each year!

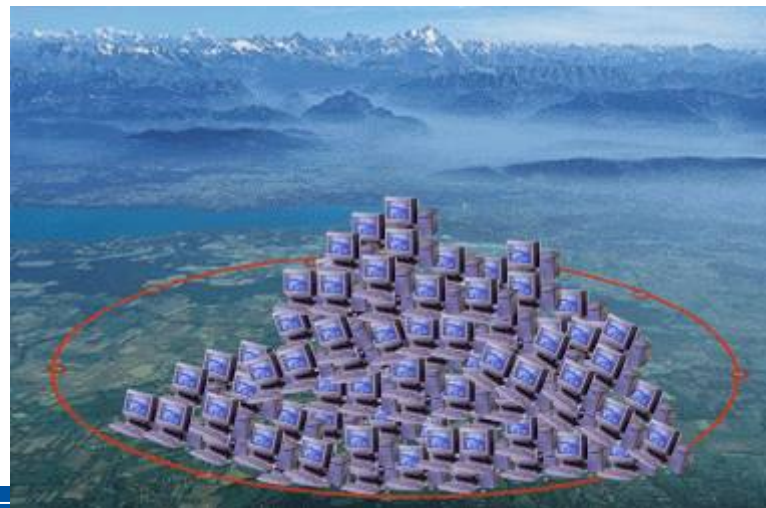
**Where will the experiments store all of these data?**



# Data Processing for the LHC

- LHC data analysis requires a computing power equivalent to ~ 100,000 of today's PC processors!

**Where will the experiments find such a computing power?**





# Computing at CERN

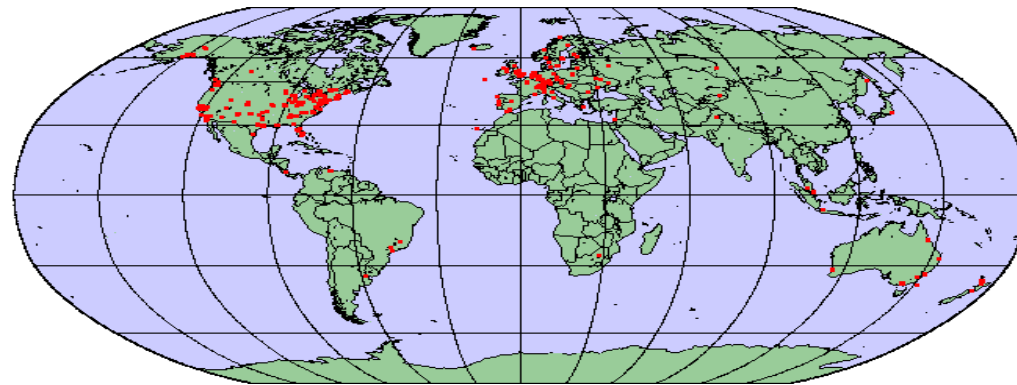
- High-throughput computing based on reliable “commodity” technology
- 2000 dual processor PCs
- 3 Petabytes of data on disk and tapes

**Nowhere near enough!**



# Computing for LHC

- Problem: even with Computer Centre upgrade, CERN can provide only a fraction of the necessary resources.
- Solution: CERN has over 250 partner institutes in Europe, over 200 in rest of the world. Most have significant computing resources. Build a Grid that **unites these computing resources.**



# Grid history

- Name “Grid” chosen by analogy with electric power grid (Foster and Kesselman 1997)
- Vision: plug-in computer for processing power just like plugging in toaster for electricity.
- Concept has been around for decades (distributed computing, metacomputing)
- Key difference with the Grid is to realise the vision on a global scale.

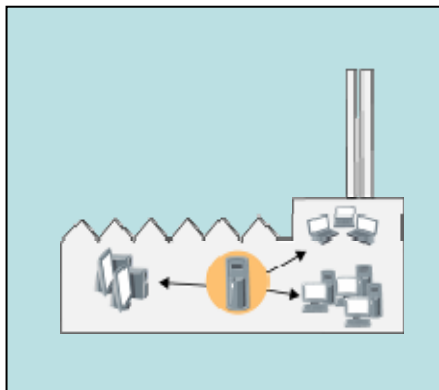


# Different Grids for different needs

There is as yet no unified Grid, like there is a single web.

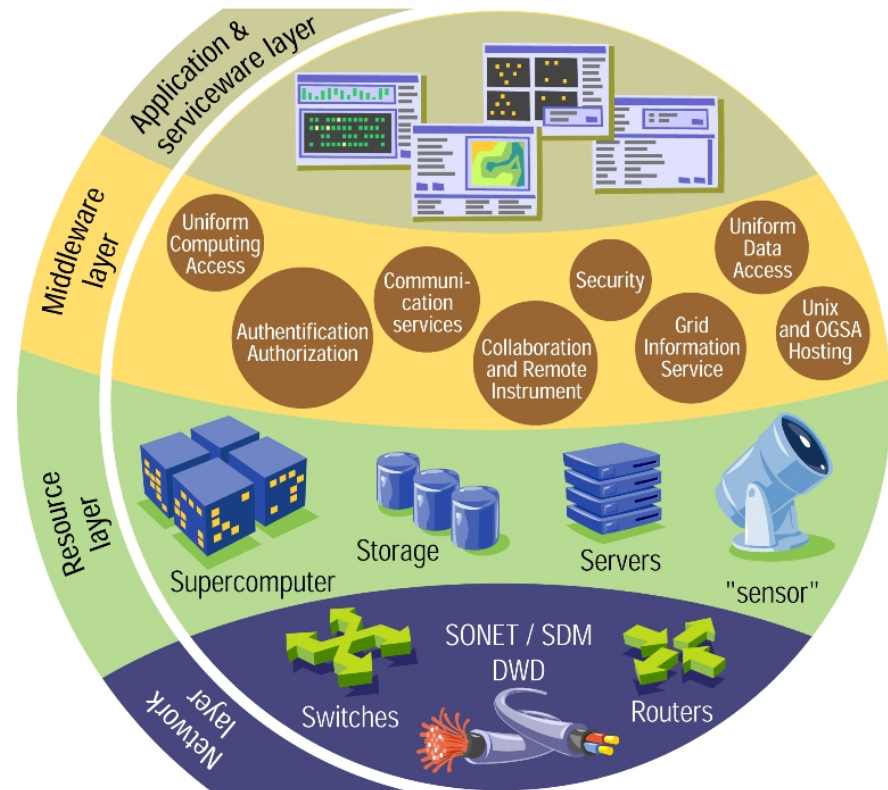
Rather **there are many Grids** for many applications:

- **Enterprise Grids** link together PCs within one company.
- **Volunteer computing** links together public computers.
- **Scientific Grids** link together major computing centres.
- Latest trend **federates national Grids** into global Grid infrastructure.
- High Energy Physics is a driving force for this.



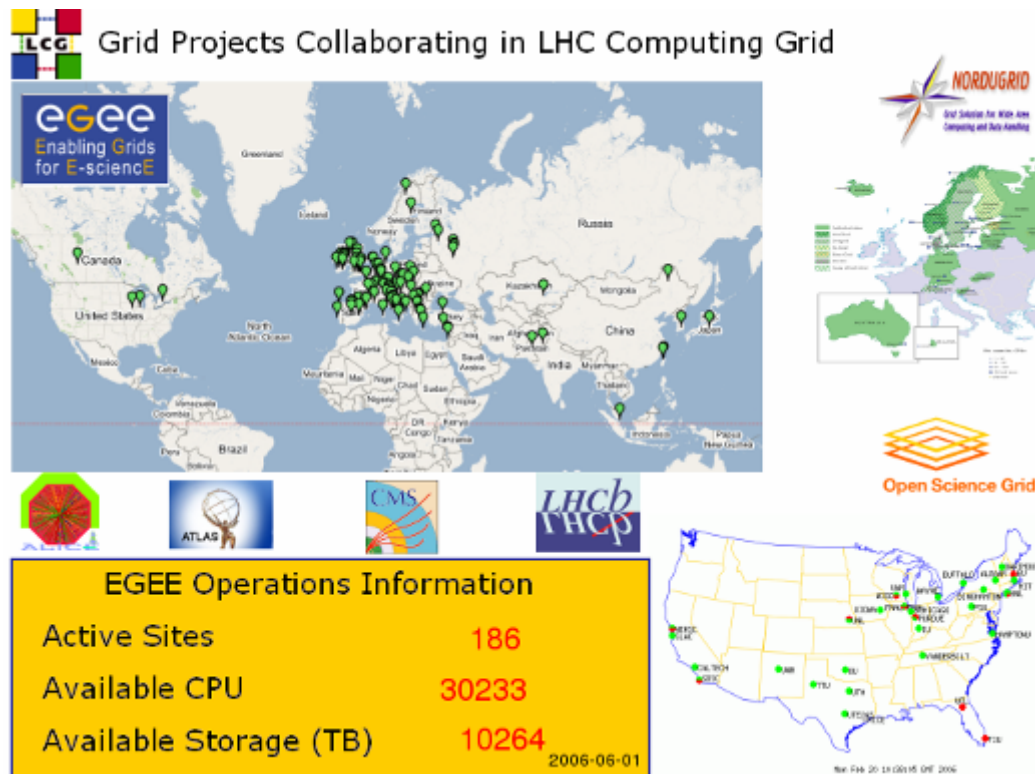
# How does the Grid work?

- It relies on advanced software, called **middleware**.
- Middleware automatically finds the **data** the scientist needs, and the **computing power** to analyse it.
- Middleware balances the load on different resources. It also handles **security, accounting, monitoring** and much more.

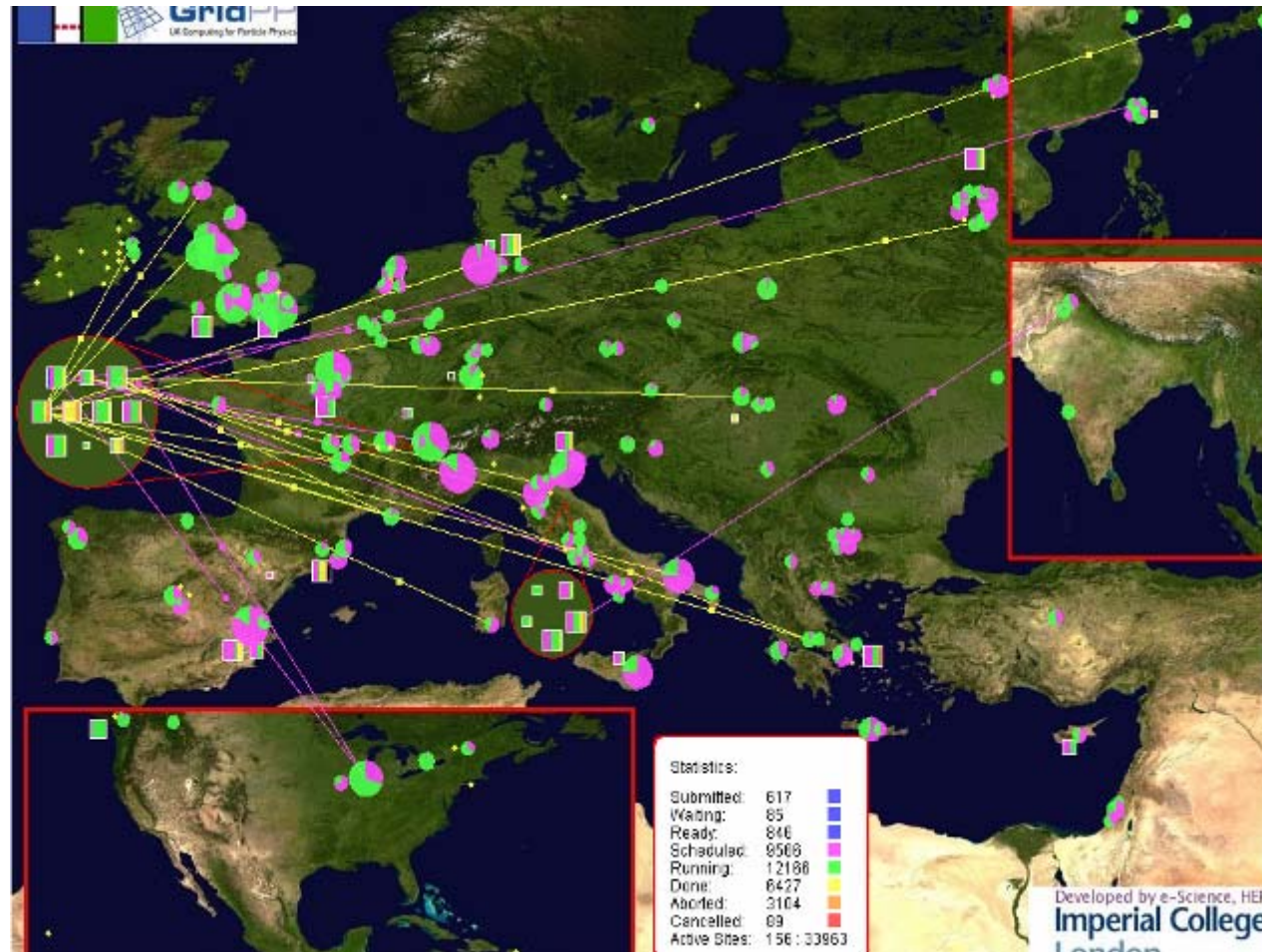


# LCG: LHC Computing Grid project

- The LHC Computing Grid project launched a **service** with 12 sites in 2003. Today 200 sites in 30 countries with 16,000 PCs
- LCG **service** based on three Grids: EGEE, NorduGrid, OSG
- LCG supports **virtual organisations** for LHC experiments



# LHC Computing Grid is working now!



# EGEE: Enabling Grids for E-scienceE



**EGEE: a multi-science Grid infrastructure**

**91 institutional partners**

**> 180 sites, 40 countries**

**> 24,000 processors**

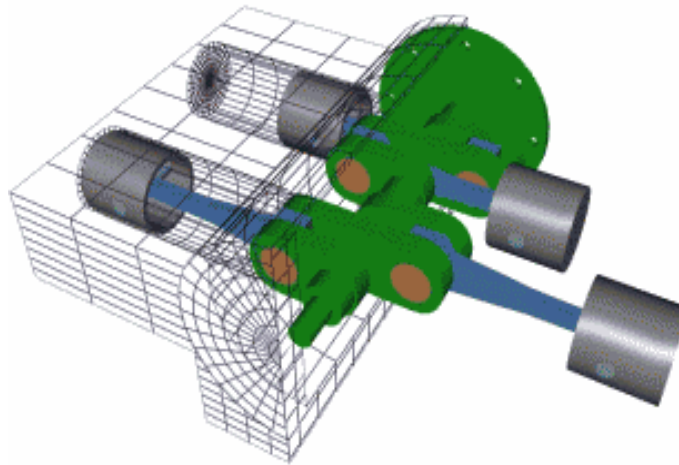
**~ 5 PB storage**

country	sites	country	sites	country	sites
Austria	2	India	2	Russia	12
Belgium	3	Ireland	15	Serbia	1
Bulgaria	4	Israel	3	Singapore	1
Canada	7	Italy	25	Slovakia	4
China	3	Japan	1	Slovenia	1
Croatia	1	Korea	1	Spain	13
Cyprus	1	Netherlands	3	Sweden	4
Czech Republic	2	Macedonia	1	Switzerland	1
Denmark	1	Pakistan	2	Taipei	4
France	8	Poland	5	Turkey	1
Germany	10	Portugal	1	UK	22
Greece	6	Puerto Rico	1	USA	4
Hungary	1	Romania	1	CERN	1



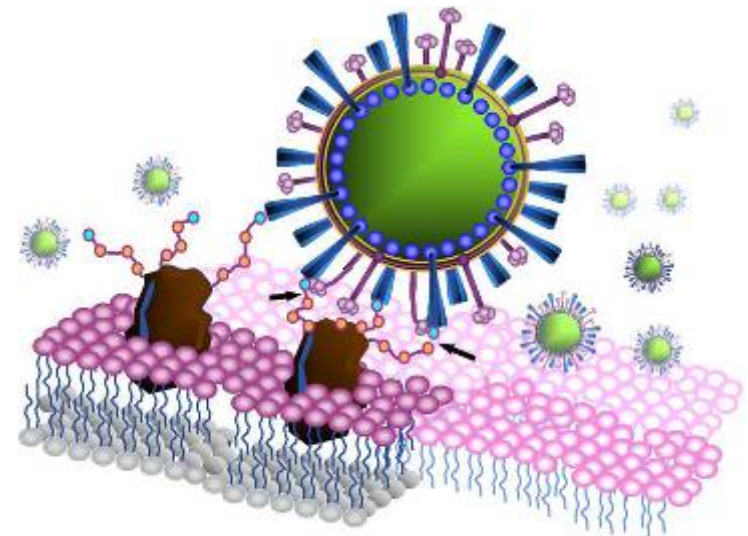
# Possible applications of EGEE

- Government: Help government agencies share data and forecasting resources (earthquakes, floods)
- Education: Help students access large computing resources for educational purposes (shrink digital divide)
- Business: Help small companies access large computing resources for drug design, materials simulation etc.



# Example: EGEE Attacks Avian Flu

- EGEE used to analyse 300,000 possible potential drug compounds against bird flu virus, H5N1.
- 2000 computers at 60 computer centres in Europe, Russia, Taiwan, Israel ran during four weeks in April - the equivalent of 100 years on a single computer.
- Potential drug compounds now being identified and ranked



*Neuraminidase, one of the two major surface proteins of influenza viruses, facilitating the release of virions from infected cells. Image Courtesy Ying-Ta Wu, AcademiaSinica.*

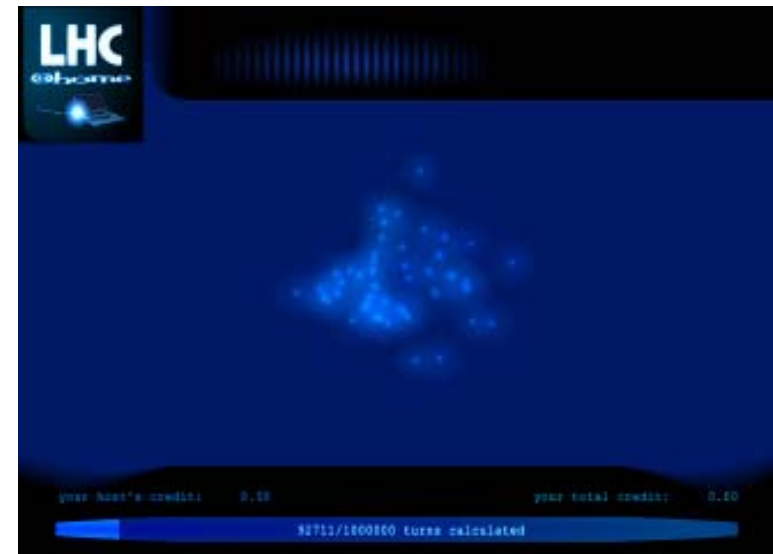
# LHC@home: volunteer computing

- Volunteer computing project based on BOINC platform (used by SETI@home)
- Calculates stability of LHC proton orbits
- Resources (PCs, laptops) donated by volunteers around the world
  - more than 30 000 users, over 1500 CPU years of processing
  - Note: SETI@home downloaded on more than 5m machines!
- Objectives: extra CPU power and raising public awareness of CERN and the LHC.
- Started as an outreach project
- UK is now a key partner (GridPP w. Queen Mary and Imperial College)



# The @home community

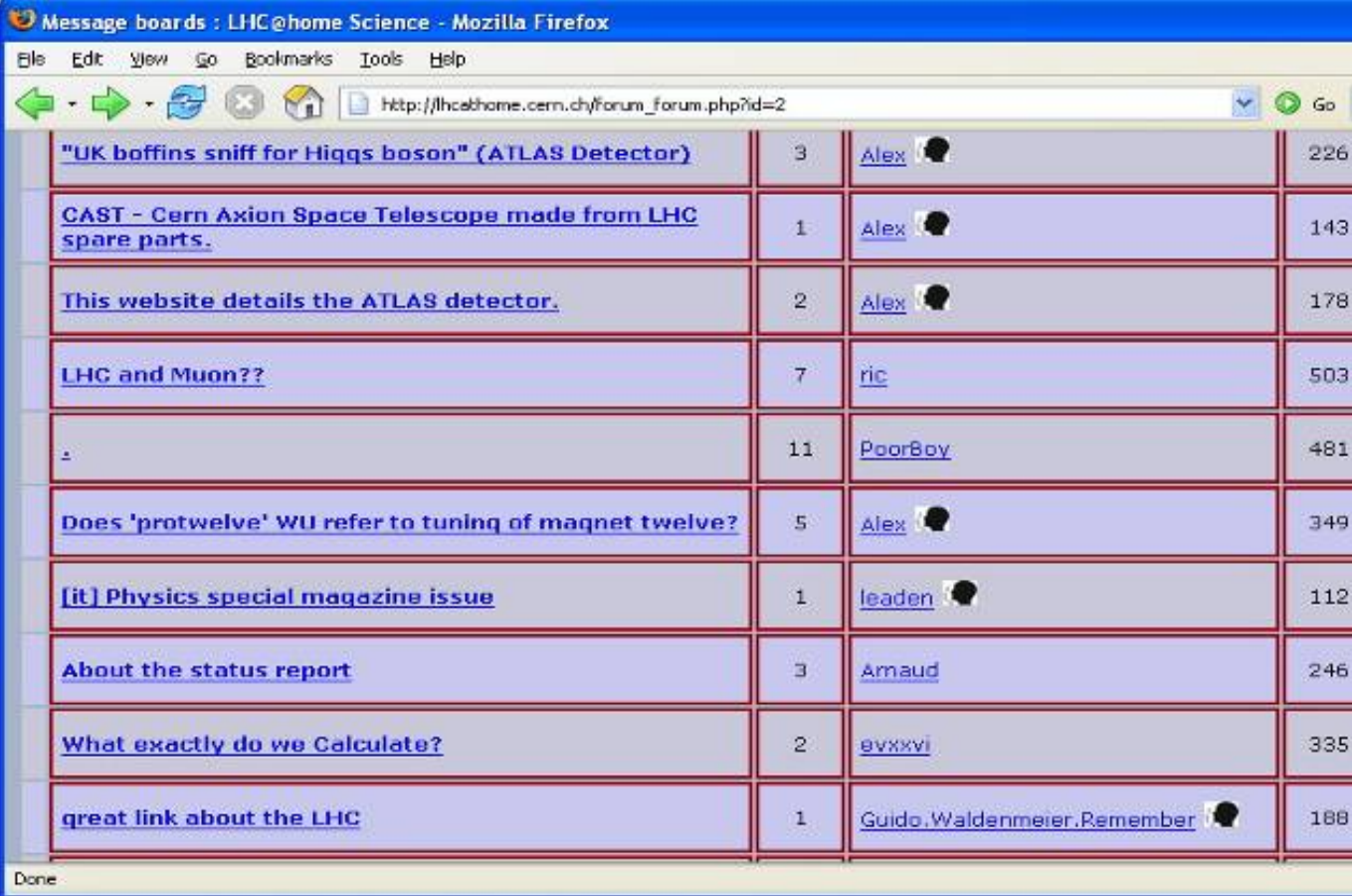
- Competition between individuals and teams for “credit”.
- Websites and regular updates on status of project by scientists.
- Forums for users to discuss the science behind the project.
- The volunteers show great interest in CERN and the LHC.
- Supply each other with scientific information.



LHC@home screensaver



## Example of one of the discussion forums on LHC@home



<a href="#">"UK boffins sniff for Higgs boson" (ATLAS Detector)</a>	3	<a href="#">Alex</a>	226
<a href="#">CAST - Cern Axion Space Telescope made from LHC spare parts.</a>	1	<a href="#">Alex</a>	143
<a href="#">This website details the ATLAS detector.</a>	2	<a href="#">Alex</a>	178
<a href="#">LHC and Muon??</a>	7	<a href="#">ric</a>	503
<a href="#">.</a>	11	<a href="#">PoorBoy</a>	481
<a href="#">Does 'protwelve' WU refer to tuning of magnet twelve?</a>	5	<a href="#">Alex</a>	349
<a href="#">[it] Physics special magazine issue</a>	1	<a href="#">leaden</a>	112
<a href="#">About the status report</a>	3	<a href="#">Amaud</a>	246
<a href="#">What exactly do we Calculate?</a>	2	<a href="#">evxxvi</a>	335
<a href="#">great link about the LHC</a>	1	<a href="#">Guido, Waldenmeier, Remember</a>	188

Done



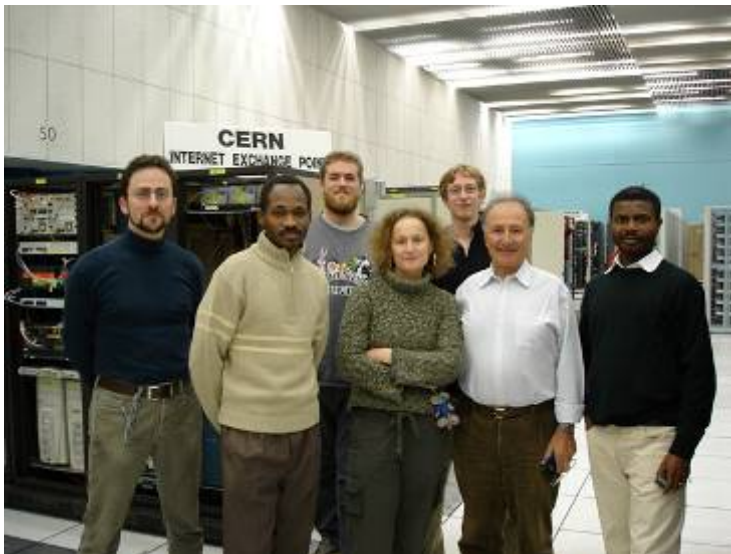
# Volunteer computing for participative science

- SETI@home Extraterrestrial radio signal search
  - FightAids@home AIDS drugs design
  - Einstein@home Gravitational wave search
  - Rosetta@home Protein shapes for disease cure
  - Climateprediction.net Climate modelling
  - Compute against Cancer Study effects of chemotherapy
  - Evolution@home Genetic causes of extinction
  - Folding@home Protein folding for disease cure
  - QMC@home Quantum simulation of molecules
  - GIMPS Search for large prime numbers
  - ChessBrain Distributed chess supercomputer
  - ...and many more!
- 
- Helen McGlone will explain how to join a project and create a team tomorrow in the workshop on Grids.



# Africa@home: dealing with the digital divide

- 3 month student project
- 500 PCs joined in 2 days
- >100 CPU-years in 2 months
- Demoed at WSIS Tunis
- Went public in July



# Conclusions and Speculations

- Scientific Grids are already working!
- *But the Grid today is where the Web was 10 years ago*
  
- The Grid is not (yet) generally accessible
- *But '@home' projects let people anywhere participate in real science.*
  
- The Grid is helping to globalise science
- *This may help to narrow the digital divide*
  
- There is a lot of hype about the Grid
- *Nano, Grid, e-Science are just words...*
- *...but words can make young people dream!*





For more information about the Grid...



***Thank you!***

