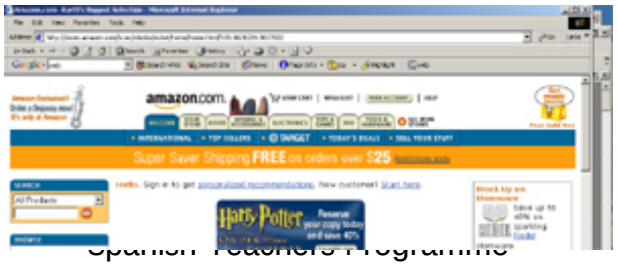
GRID e Informática Científica
Al servicio de la Ciencia

Miguel Angel Marquina
IT Department - CERN

What is the Grid?

- The World Wide Web provides seamless access to information that is stored in many millions of different geographical locations

- In contrast, the Grid is an emerging infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe.



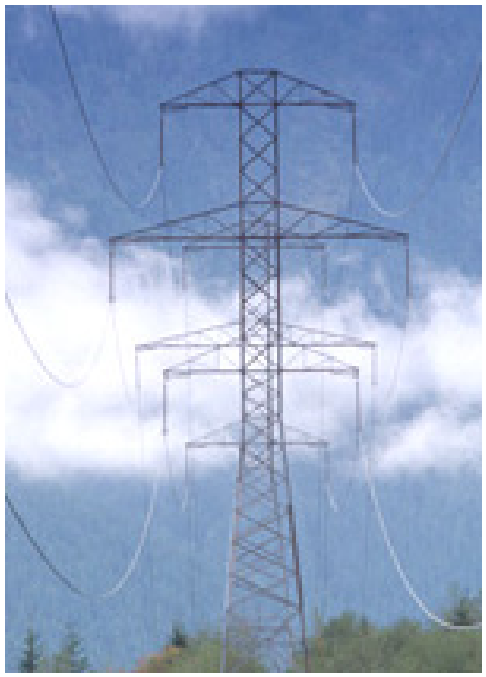
What is the Grid?

• The term was coined by Ian Foster and Carl Kesselman (Grid bible “The Grid: blueprint for a new computing infrastructure”).

• The name is chosen by analogy to the electric power grid: plug-in to computing power without worrying where it comes from, like a toaster.

• The idea has been around under other names for a while (distributed computing, metacomputing, …).

•This time, technology is in place to realise the dream on a global scale.



How will it work?

• It relies on advanced software, called middleware, which ensures seamless communication between different computers and different parts of the world

• A search engine will not only find the data the scientist needs, but also the data processing techniques and the computing power to carry them out

• It will ship the computing task to wherever in the world there is spare capacity, and send the result back to the scientist

How will it work?

The GRID middleware:

- Finds convenient places for the scientists “job” (computing task) to be run
- Optimises use of the widely dispersed resources
- Organises efficient access to scientific data
- Deals with authentication to the different sites that the scientists will be using
- Interfaces to local site authorisation and resource allocation policies
- Runs the jobs
- Monitors progress
- Recovers from problems

… and ….

Tells you when the work is complete and transfers the result back!

What are the challenges?

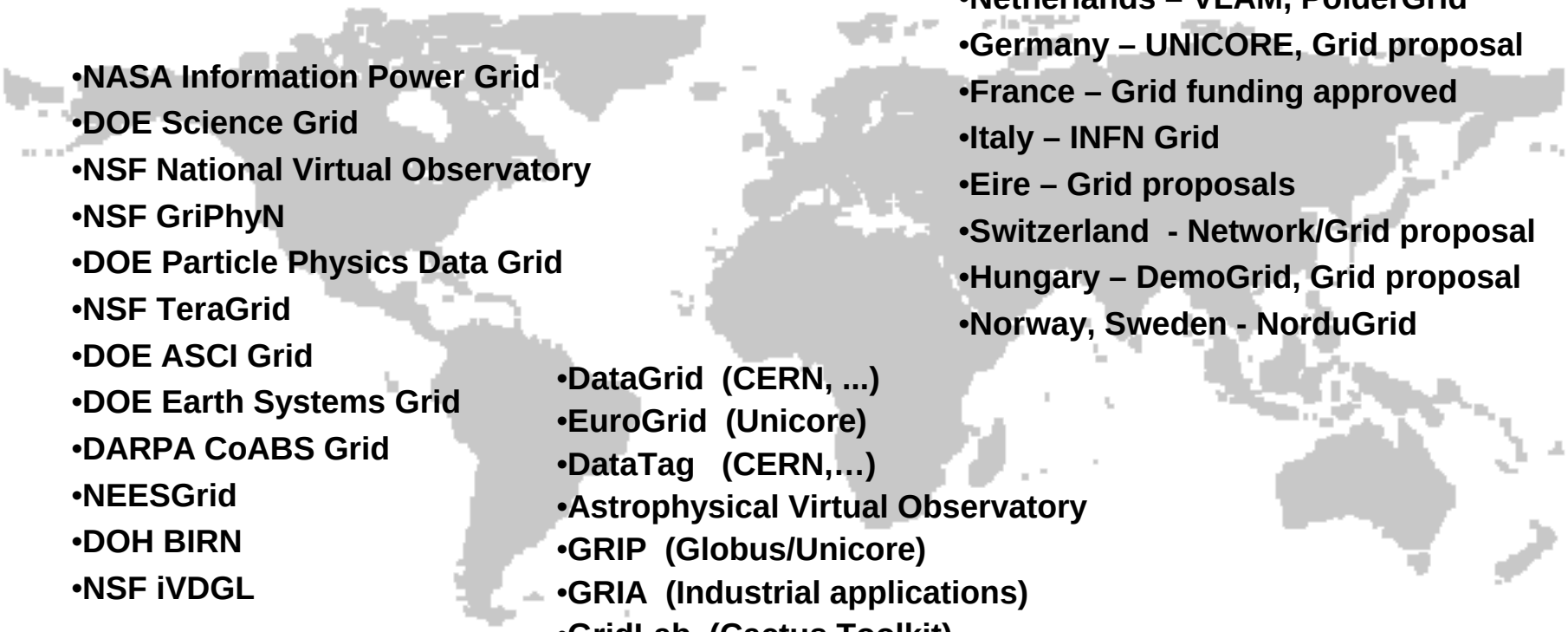
The Grid must:

- share data between thousands of scientists with multiple interests
- link major computer centres, not just PCs
- ensure all data accessible anywhere, anytime
- grow rapidly, yet remain reliable for more than a decade
- cope with different management policies of different centres
- ensure data security: more is at stake than just money!
- and at CERN, up and running by 2009 ☺☺☺

Benefits for Science

- More effective and seamless collaboration of dispersed communities, both scientific and commercial
- Ability to run large-scale applications comprising thousands of computers, for wide range of applications
- Transparent access to distributed resources from your desktop, or even your mobile phone
- The term **“e-Science”** has been coined to express these benefits

Grid projects in the world



- **NASA Information Power Grid**
- **DOE Science Grid**
- **NSF National Virtual Observatory**
- **NSF GriPhyN**
- **DOE Particle Physics Data Grid**
- **NSF TeraGrid**
- **DOE ASCI Grid**
- **DOE Earth Systems Grid**
- **DARPA CoABS Grid**
- **NEESGrid**
- **DOH BIRN**
- **NSF iVDGL**

- **UK e-Science Grid**
- **Netherlands – VLAM, PolderGrid**
- **Germany – UNICORE, Grid proposal**
- **France – Grid funding approved**
- **Italy – INFN Grid**
- **Eire – Grid proposals**
- **Switzerland - Network/Grid proposal**
- **Hungary – DemoGrid, Grid proposal**
- **Norway, Sweden - NorduGrid**

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

Grid Applications for Science

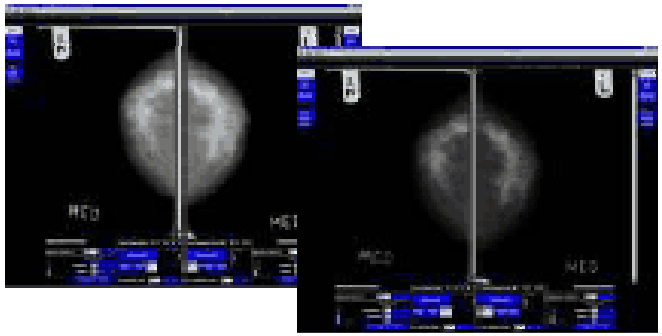
- Medical/Healthcare *(imaging, diagnosis and treatment)*
- Bioinformatics *(study of the human genome and proteome to understand genetic diseases)*
- Nanotechnology *(design of new materials from the molecular scale)*
- Engineering *(design optimization, simulation, failure analysis and remote Instrument access and control)*
- Natural Resources and the Environment *(weather forecasting, earth observation, modeling and prediction of complex systems)*

Medical/Healthcare Applications

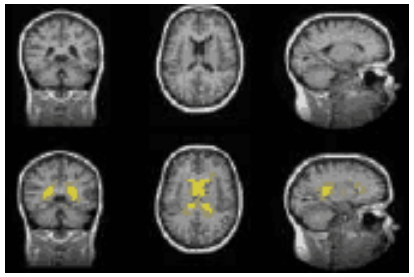
- Digital image archives
- Collaborative virtual environments
- On-line clinical conferences

"The Grid will enable a standardized, distributed digital mammography resource for improving diagnostic confidence"

"The Grid makes it possible to use large collections of images in new, dynamic ways, including medical diagnosis."



"The ability to visualise 3D medical images is key to the diagnosis of pathologies and pre-surgical planning"



Quotes from: http://gridoutreach.org.uk

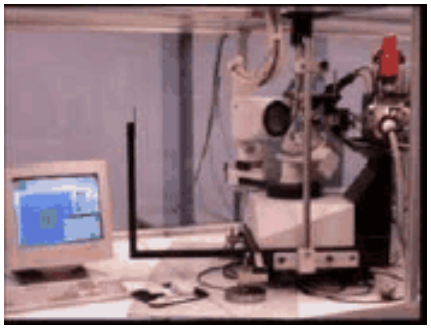
Bioinformatics

- Capturing the complex and evolving patterns of genetic information, determining the development of an embryo
- Understanding the genetic interactions that underlie the processes of life-form development, disease and evolution.

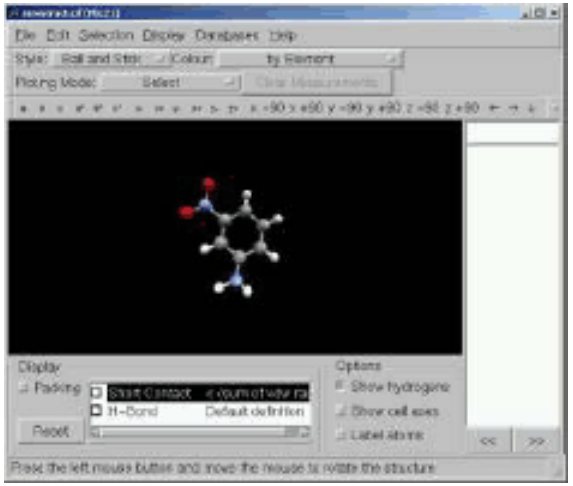
"Every time a new genome is sequenced the result is compared in a variety of ways with other genomes. Each code is made of 3.5 billion pairs of chemicals…"

Nanotechnology

- New and 'better' materials
- Benefits in pharmaceuticals, agrochemicals, food production, electronics manufacture from the faster, cheaper discovery of new catalysts, metals, polymers, organic and inorganic materials



"The Grid has the potential to store and analyze data on a scale that will support faster, cheaper synthesis of a whole range of new materials."



Quotes from: http://gridoutreach.org.uk

Natural Resources/Environment

- Modeling and prediction of earthquakes
- Climate change studies and weather forecast
- Pollution control
- Socio-economic growth planning, financial modeling and performance optimization

"Federations of heterogeneous databases can be exploited through the Grid to solve complex questions about global issues such as biodiversity."

Quotes from: http://gridoutreach.org.uk

Precursors of the Grid

- SETI@home: sharing of spare PC processing power to analyze radio signals

- Napster: sharing of data (music) between computers

- Entropia DCGrid: commercial solution for sharing workstations within a company

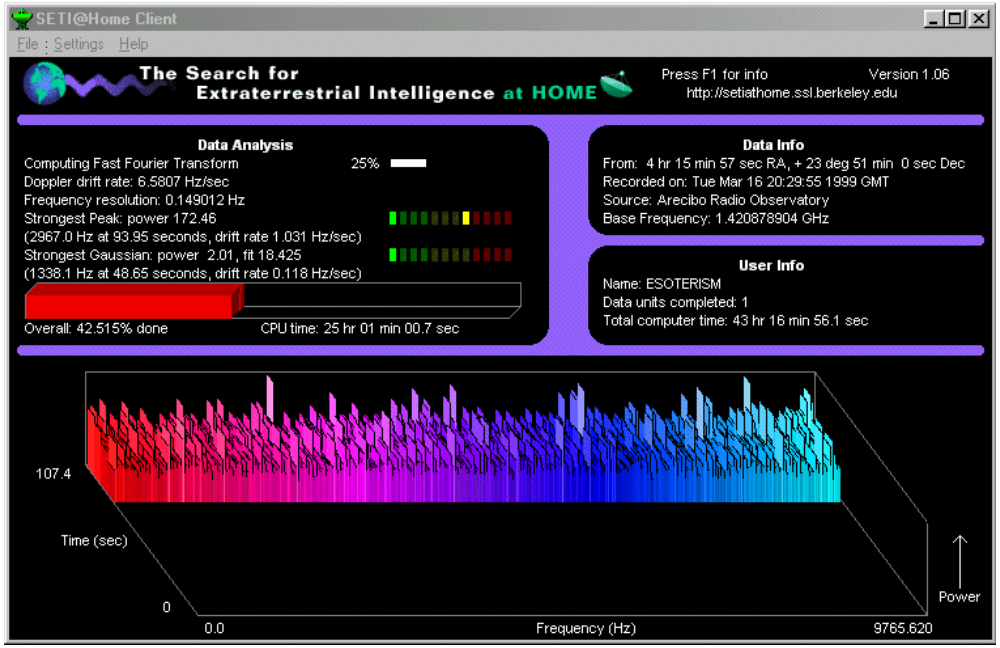
The difference:
The Grid CERN is developing will combine resources at major computer centers, and require dedicated equipment and sophisticated middleware to monitor and allocate resources

SETI@home: a grassroots Grid

>1 million years of computer processing time

>3.5 million have downloaded the screensaver

>30 Teraflops rating (ASCI White = 12 Teraflops)



Spinoff from SETI@home

Spawned a cottage industry

Xpulsar@home, Genome@home, Folding@home, evolutionary@home, FightAIDS@home, SARS@home...

Spawned a real industry

Entropia, United Devices, Popular Power...

Major limitations:

Only suitable for massively parallel problems

Cycle scavenging relies on goodwill

Who will use Grids?

- Computational Scientists & Engineers: large scale modeling of complex structures
- Experimental Scientists: storing and analyzing large data sets
- Collaborations: large scale multi-institutional projects
- Corporations: global enterprises and industrial partnership
- Environmentalists: climate monitoring and modeling
- Training & Education: virtual learning rooms and laboratories