

# Grid Applications and Projects



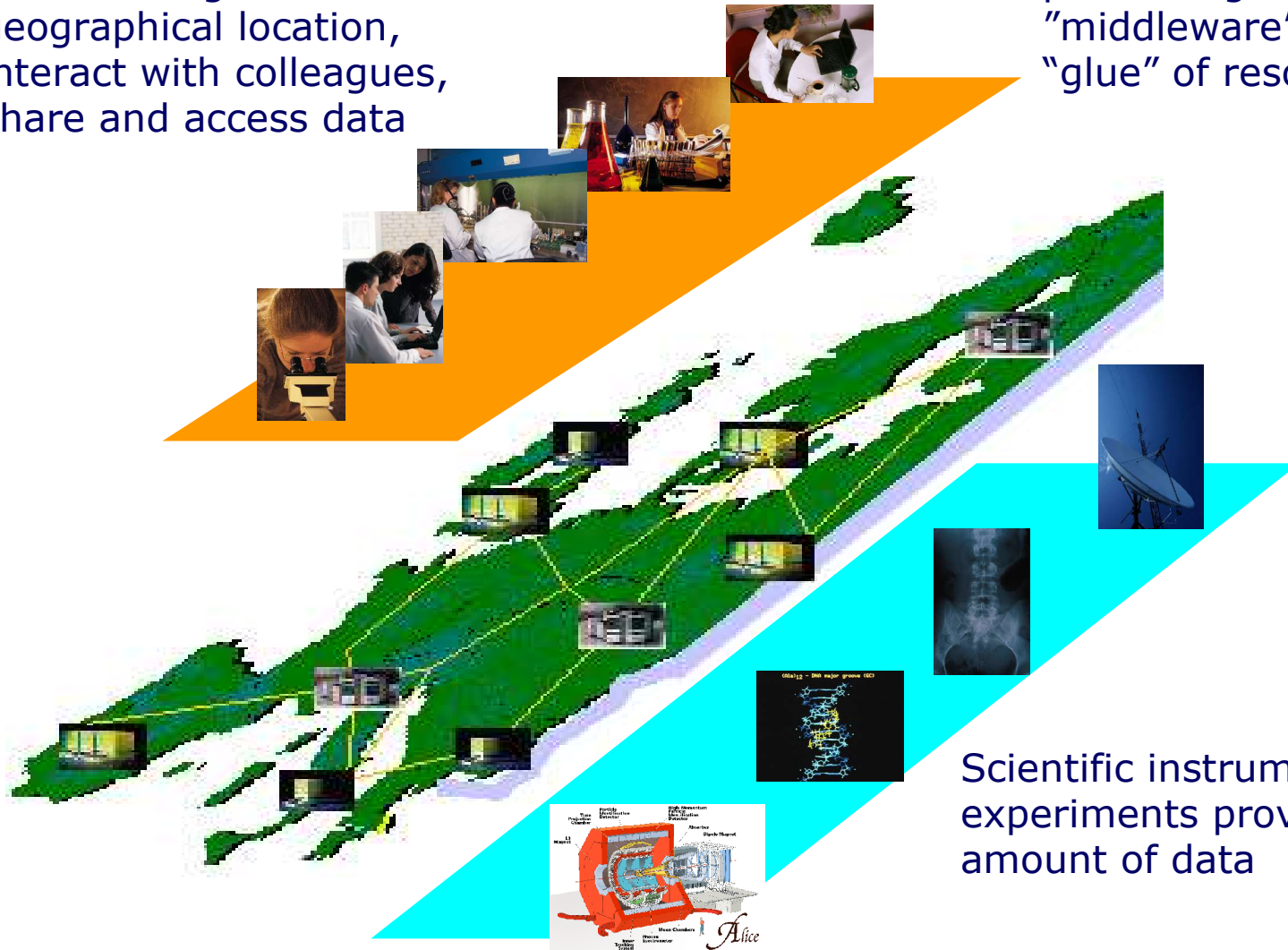
*Dr. Rüdiger Berlich, Christopher Jung  
Forschungszentrum Karlsruhe*

*DESY, Hamburg / Germany, 24.3.05*

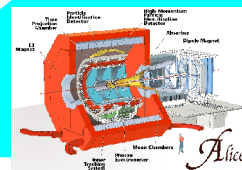
*Slides contributed by R.Barbera and  
EGEE teams*

Researchers perform their activities regardless of geographical location, interact with colleagues, share and access data

The Grid: networked data processing centres and "middleware" software as the "glue" of resources.

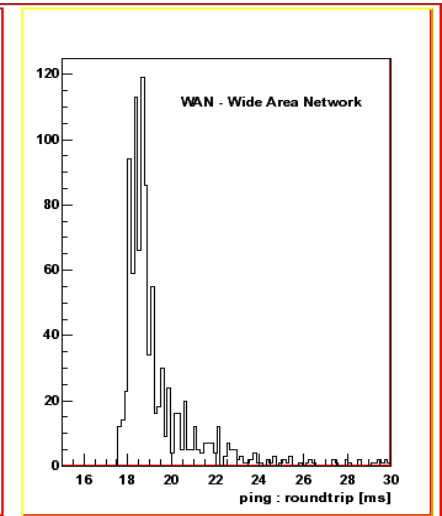
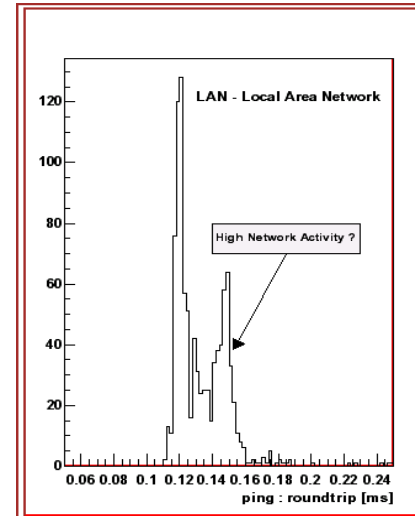
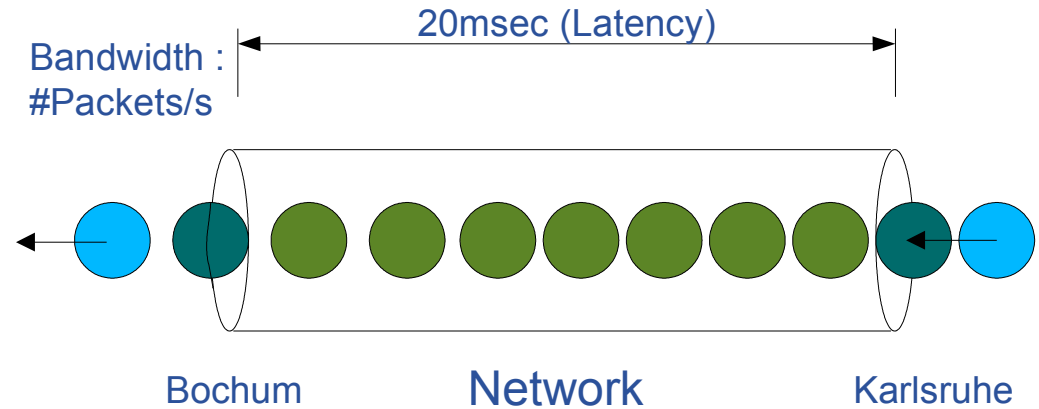


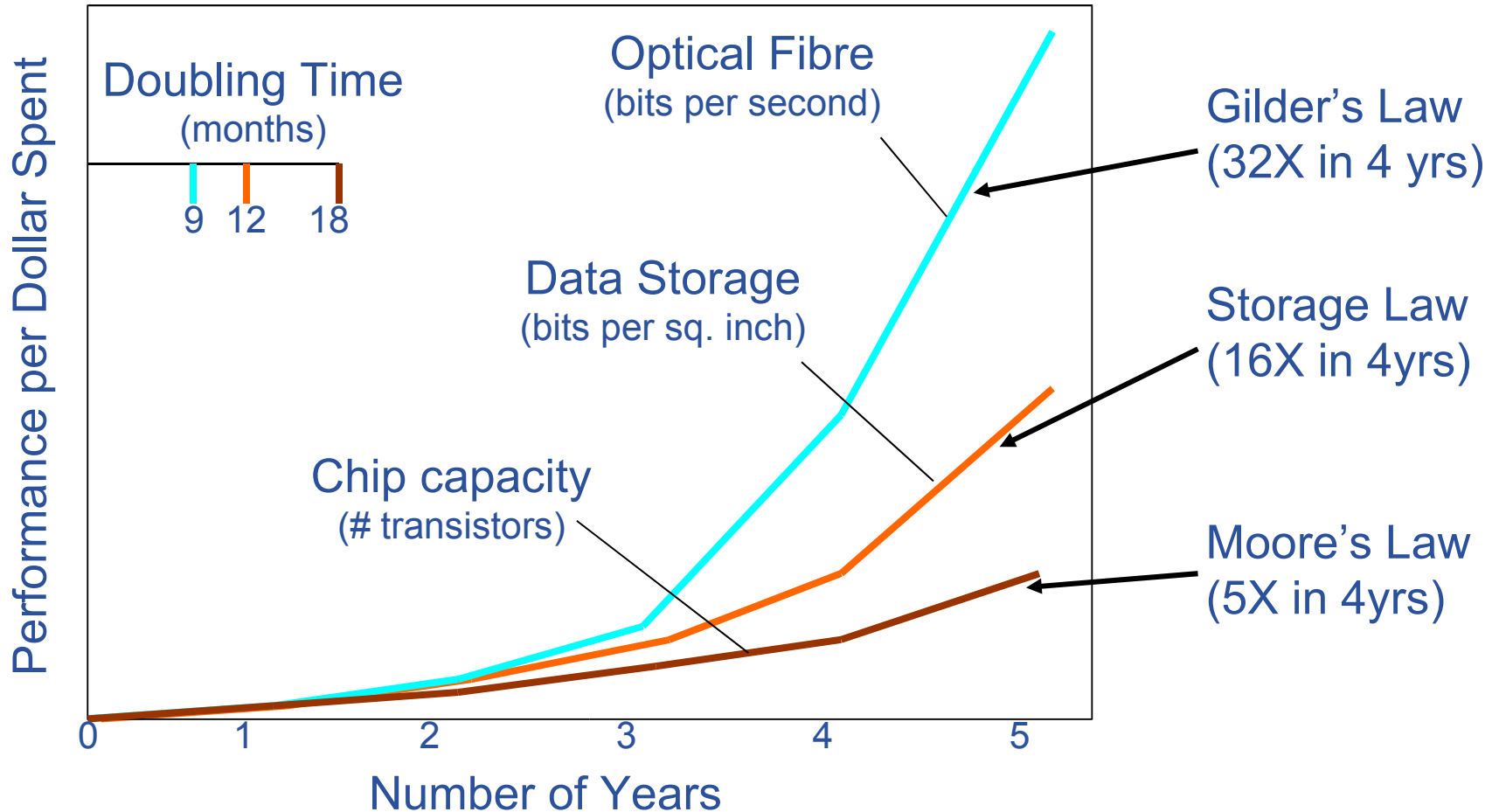
Scientific instruments and experiments provide huge amount of data



# *Technical and Monetary Constraints*

- „Speed“ of a network consists of two components
- Bandwidth (scales to any number)
- Latency (doesn't scale)
- Possible application types in a Grid are limited by latency
- Latency comparable to mean access times of old MFM harddrive

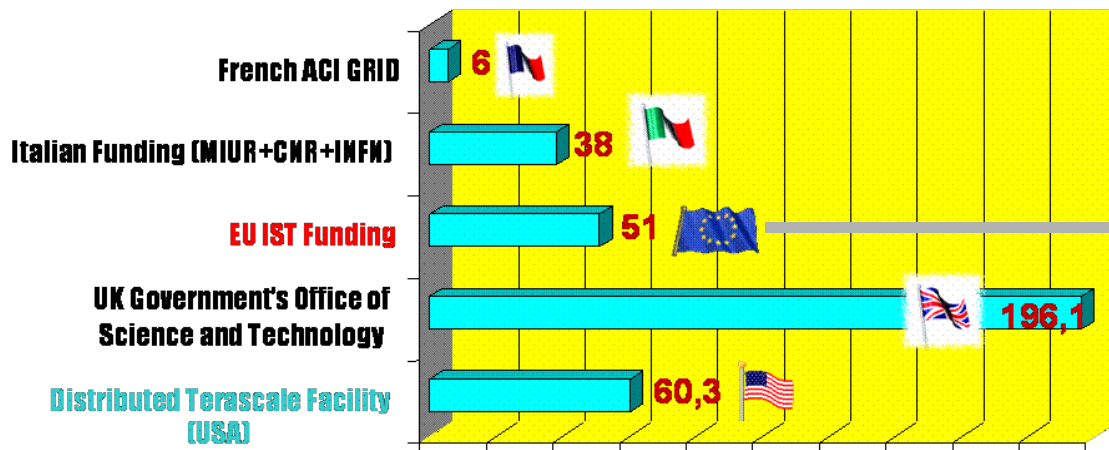




Triumph of Light – *Scientific American*. George Stix, January 2001

## Present Grid investments in EU/US

**Million €**



+ D-Grid Initiative  
(just forming :  
300 Million Euro  
expected funding)

**Future figures:**

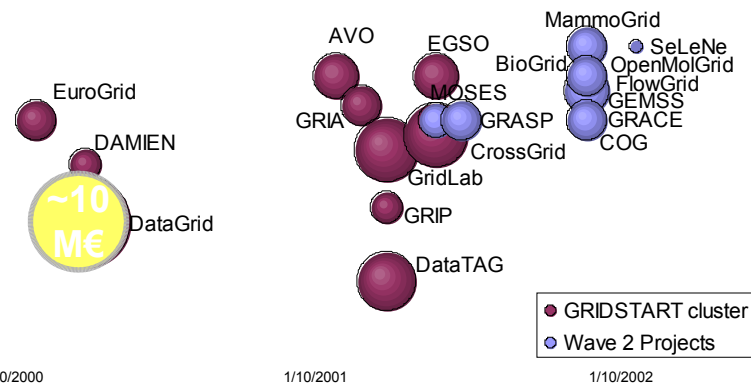
*US Cyber Infrastructure:*

*1020 M\$*

*Japan (A-P) Grid:*

*~500 M\$*

### IST Grid Projects

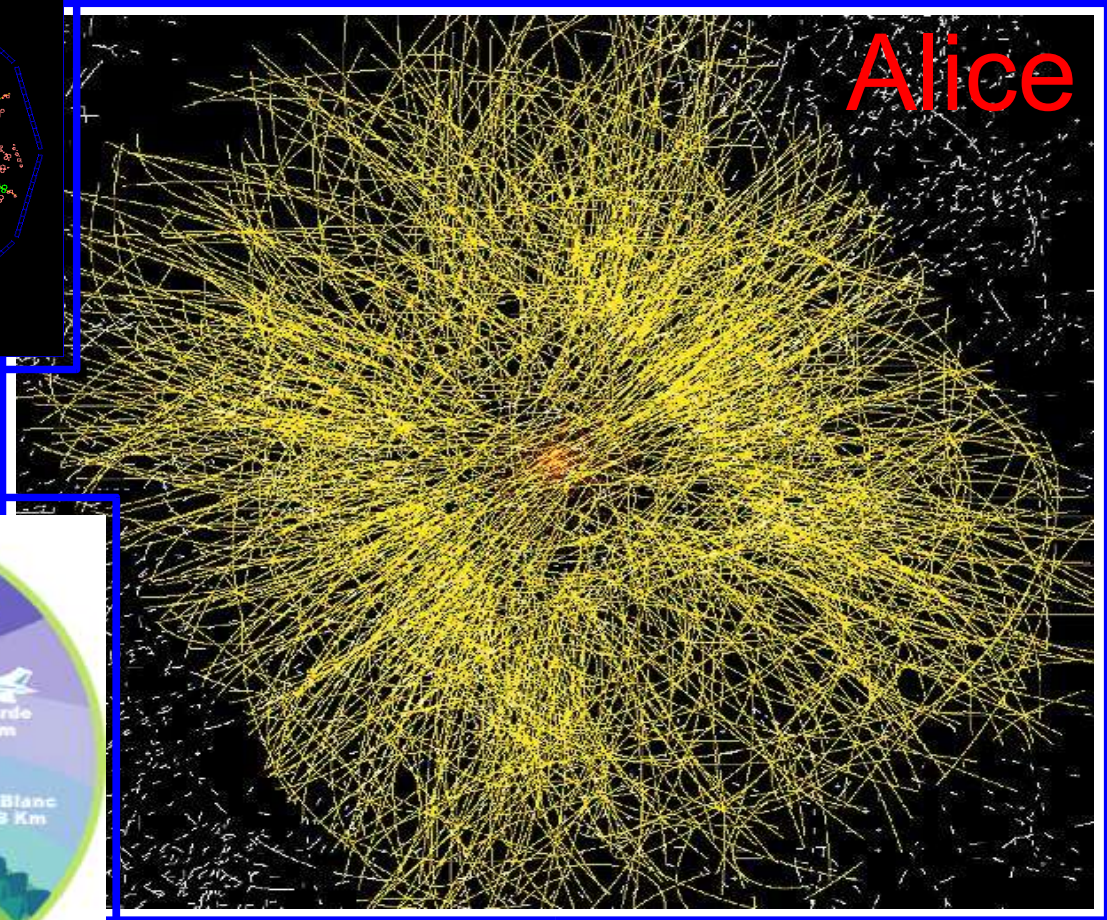
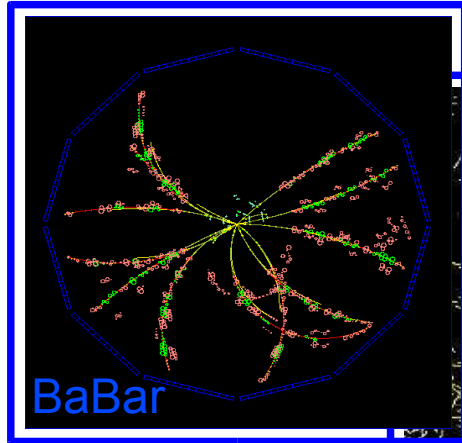


# ***“Typical” application areas***

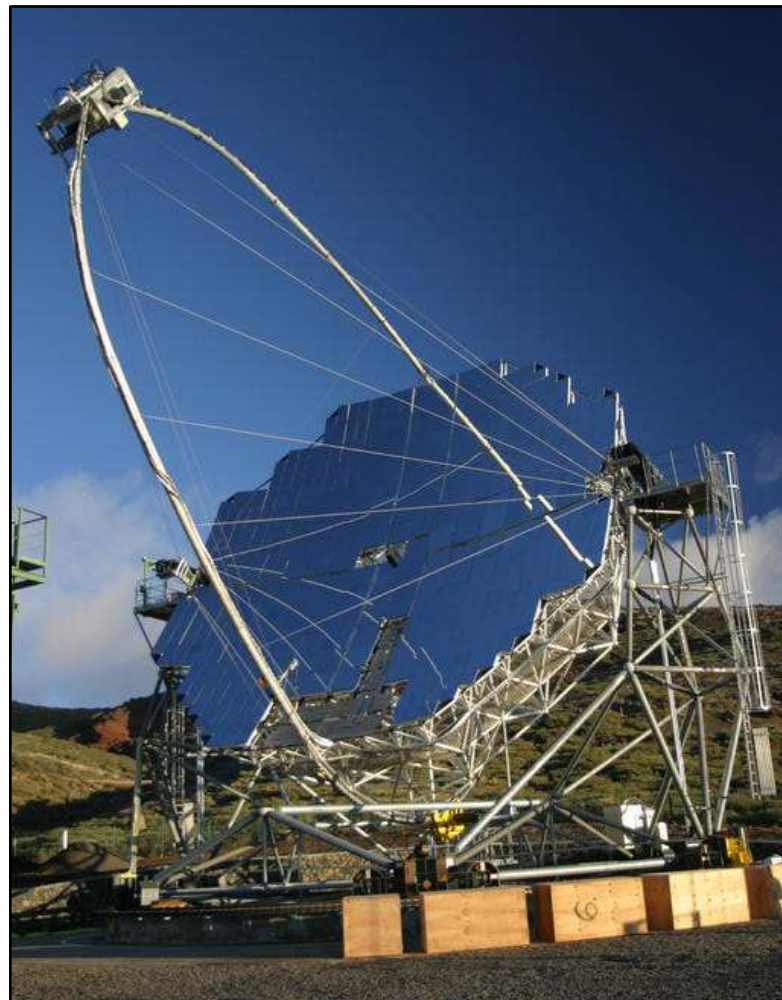
- **Medical/Healthcare** (*imaging, diagnosis and treatment*)
- **Bioinformatics** (*study of the human genome and proteome to understand genetic diseases*)
- **Nanotechnology** (*design of new materials from the molecular scale*)
- **Engineering** (*design optimization, simulation, failure analysis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



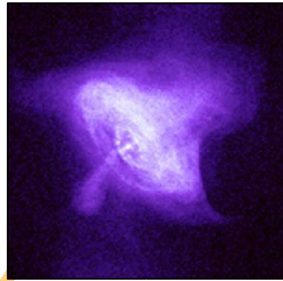




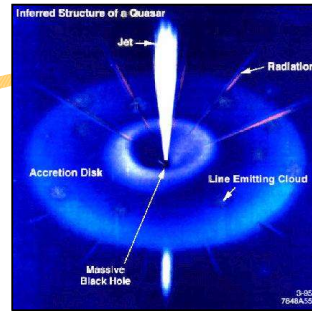
- **Largest Imaging Air Cherenkov Telescope**  
(17 m mirror dish)
- Located on Canary Island **La Palma** (@ 2200 m asl)
- Lowest **energy threshold** ever obtained with a Cherenkov telescope
- Aim: detect  **$\gamma$ -ray sources** in the unexplored energy range:  
**30 (10)-> 300 GeV**



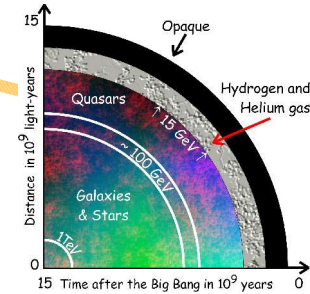




■ Pulsars

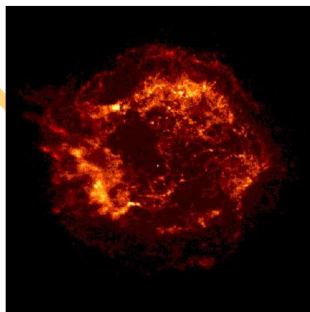


■ AGNs

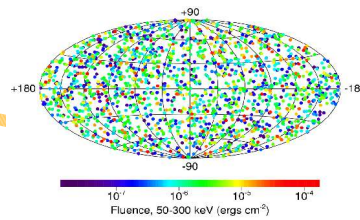


■ Cosmological  $\gamma$ -Ray Horizon

■ Origin of Cosmic Rays

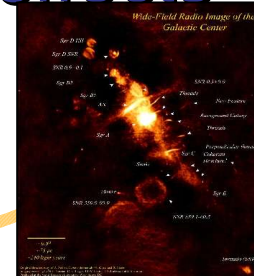


■ SNRs



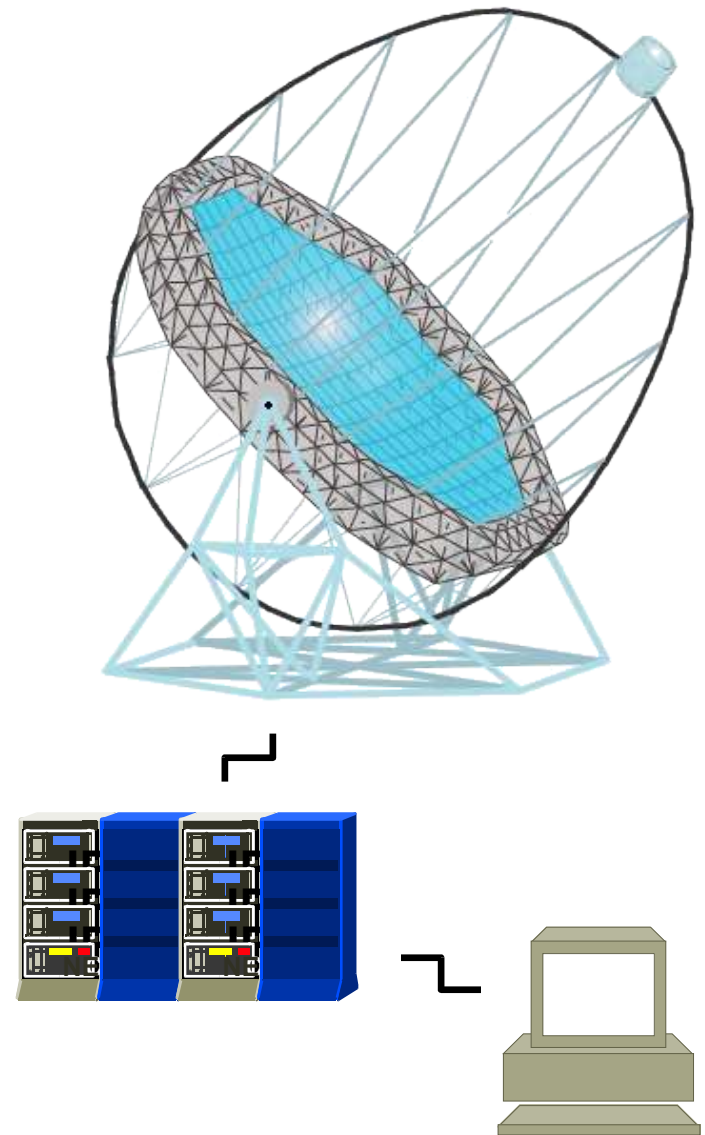
■ GRBs

■ Tests of Quantum Gravity effects



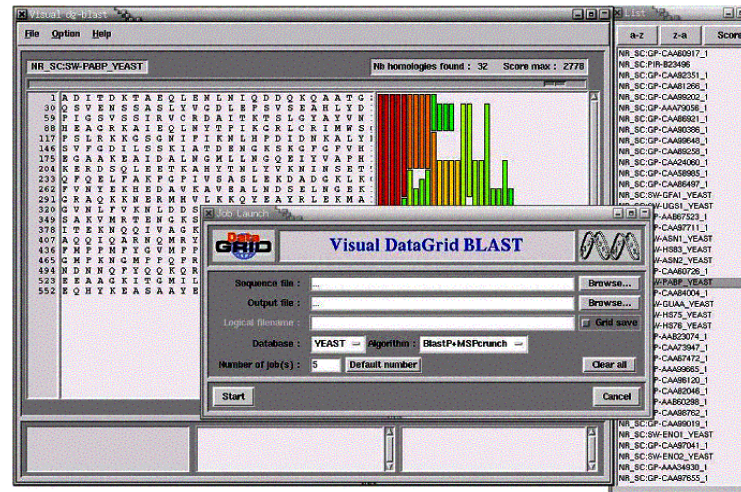
■ Cold Dark Matter

- **Event Size:**
  - 577 PM x 1 Byte x 30 samples  
⇒ ~ 20 kByte/event
- **Data Acquisition Rate:**
  - 500 Hz typical trigger rate  
⇒ ~ 10 MByte/sec
- **Data Storage Requirements:**
  - ~ 1000 h / year  
useful moonless observation time  
⇒ ~ 36 TByte/year

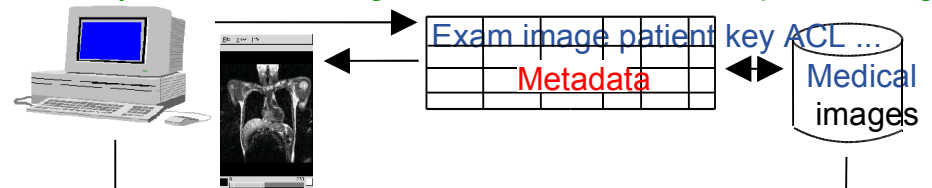


- **Bio-informatics**
  - Phylogenetics
  - Search for primers
  - Statistical genetics
  - Bio-informatics web portal
  - Parasitology
  - Data-mining on DNA chips
  - Geometrical protein comparison
- **Medical imaging**
  - MR image simulation
  - Medical data and metadata management
  - Mammographies analysis
  - Simulation platform for PET/SPECT

|  |                          |
|--|--------------------------|
|  | Applications deployed    |
|  | Applications tested      |
|  | Applications under prep. |



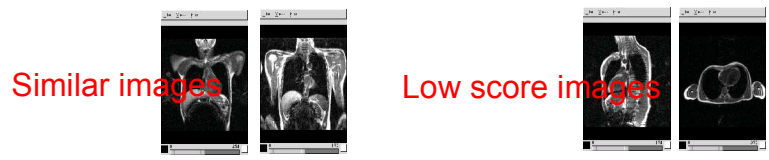
1. Query the medical image database and retrieve a patient image



2. Compute similarity measures over the database images

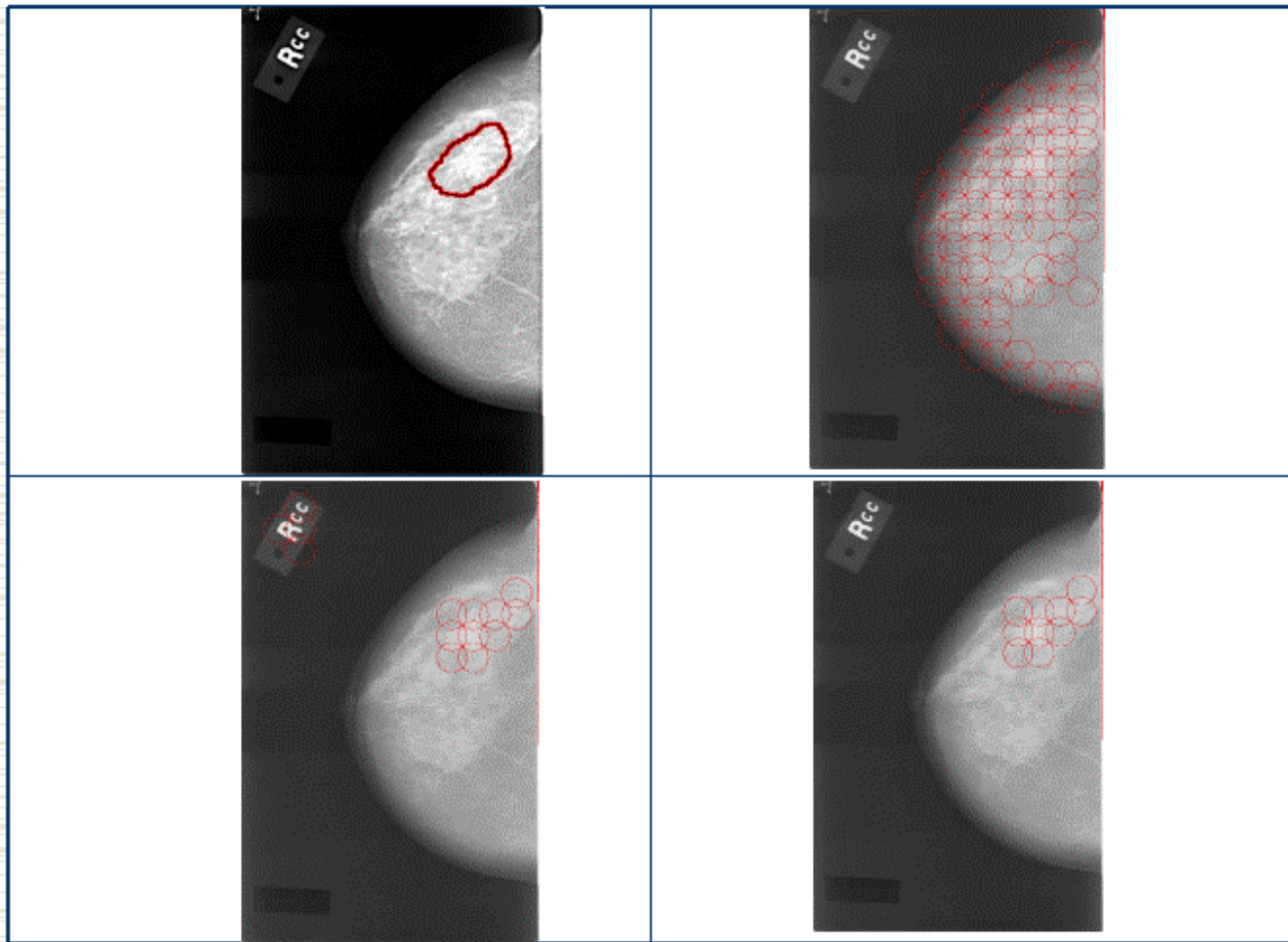


3. Retrieve most similar cases



## Phylogenetic - Wikipedia.org:

A phylogeny (or phylogenesis) is the origin and evolution of a set of organisms, usually of a species. A major task of systematics is to **determine the ancestral relationships among known species** (both living and extinct), and the most commonly used methods to infer phylogenies include cladistics and phenetics.



September 21st, 2001

*Datagrid Meeting, Lyon*

15

## Mammogrid -> AliEn



## Similarity computation

| Job                | Status        | Target              |
|--------------------|---------------|---------------------|
| 27499 (similarity) | Terminated    | localhost:0/noqueue |
| 27503 (similarity) | Terminated    | localhost:0/noqueue |
| 27507 (similarity) | Terminated    | localhost:0/noqueue |
| 27511 (similarity) | Terminated    | localhost:0/noqueue |
| 27515 (similarity) | Terminated    | localhost:0/noqueue |
| 27520 (similarity) | Terminated    | localhost:0/noqueue |
| 27524 (similarity) | Terminated    | localhost:0/noqueue |
| 27528 (similarity) | Terminated    | localhost:0/noqueue |
| 27532 (similarity) | Terminated    | localhost:0/noqueue |
| 27536 (similarity) | Terminated    | localhost:0/noqueue |
| 27540 (similarity) | Terminated    | localhost:0/noqueue |
| 27544 (similarity) | Terminated    | localhost:0/noqueue |
| 27548 (similarity) | Terminated    | localhost:0/noqueue |
| 27552 (similarity) | Terminated    | localhost:0/noqueue |
| 27556 (similarity) | Terminated    | localhost:0/noqueue |
| 27560 (similarity) | Output ready  | localhost:0/noqueue |
| 27564 (similarity) | Running       | localhost:0/noqueue |
| 27568 (similarity) | Submitted     | localhost:0/noqueue |
| 27572 (similarity) | Submitted     | localhost:0/noqueue |
| New similarity     | Sending to UI |                     |

Dismiss

Job monitoring

| File Similarity About   |               |                                |                   |            |
|---|---------------|--------------------------------|-------------------|------------|
| Source image:<br>Jones Jean    Cardiology Center of Monaco    Dr Jina Carlson    1997-11-18 |               |                                |                   |            |
| Results:  |               |                                |                   |            |
| 0.904684  | Durand Jean   | Lyon Cardiology Hospital       | Dr Alain Deloin   | 2002-02-21 |
| 0.743148  | Dupont Marc   | Cardiology Center of Monaco    | Dr Francis Black  | 1998-01-18 |
| 0.219426  | Durand Jean   | Cardiology Center of Monaco    | Dr Jina Carlson   | 2000-10-08 |
| 0.217490  | Jones Linda   | Montreal Neurological Institut | Dr Fany Anderson  | 2000-12-21 |
| 0.193947  | Jones Sandra  | Cardiology Center of Monaco    | Dr Francis Black  | 2000-12-25 |
| 0.003237  | Dupont Denise | Montreal Neurological Institut | Dr Norbert White  | 1998-10-22 |
| 0.003084  | Dupont John   | Montreal Neurological Institut | Dr Norbert White  | 1998-04-22 |
| 0.002636  | Smith Marc    | Cardiology Center of Monaco    | Dr Jina Carlson   | 1997-04-04 |
| 0.001778  | Durand Sylvie | Lyon Neurology Hospital        | Dr Martine Follet | 2001-02-14 |
| 0.001515  | Smith Marc    | Montreal Neurological Institut | Dr Norbert White  | 2001-02-09 |
| 0.001023  | Durand Jean   | Cardiology Center of Monaco    | Dr Jina Carlson   | 2000-02-24 |

View

Ranked list of images

### Results visualization



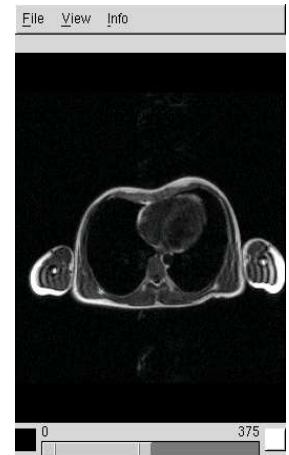
Source image



Most similar images



Low score images





### 3.3 Heart Modeling

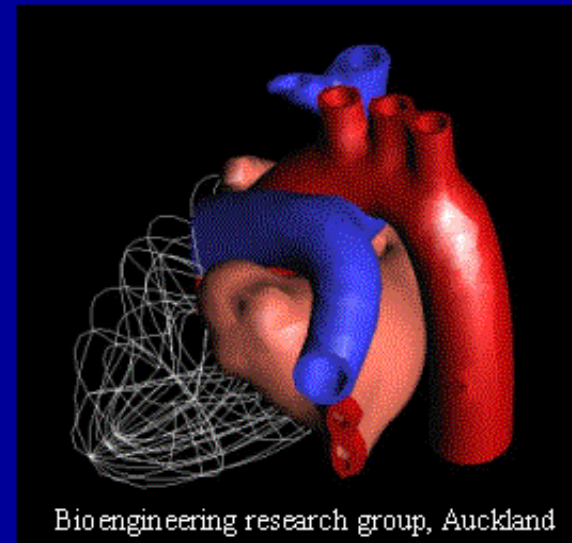
- **Objectives: modeling heart anatomy, dynamics and physiology for heart image processing**

- bio-mecanical model**

- electrical model**

- very complex structure**

- biological scale out of range**



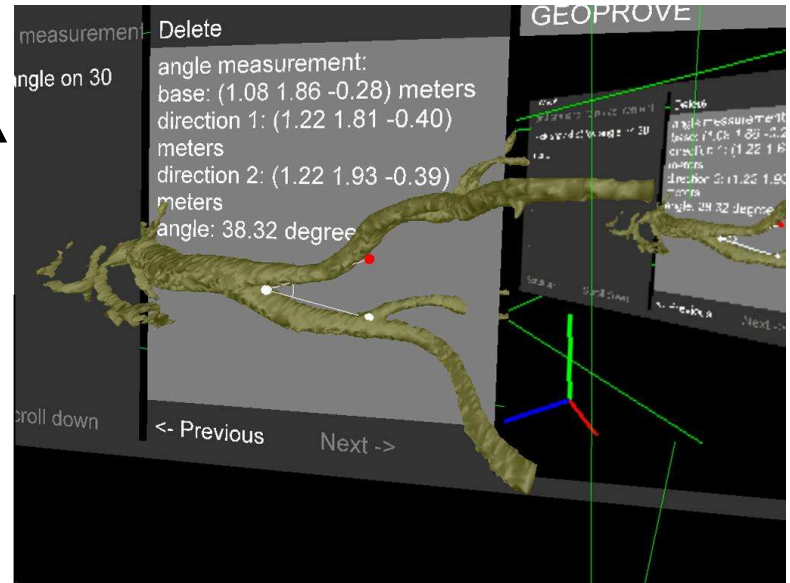
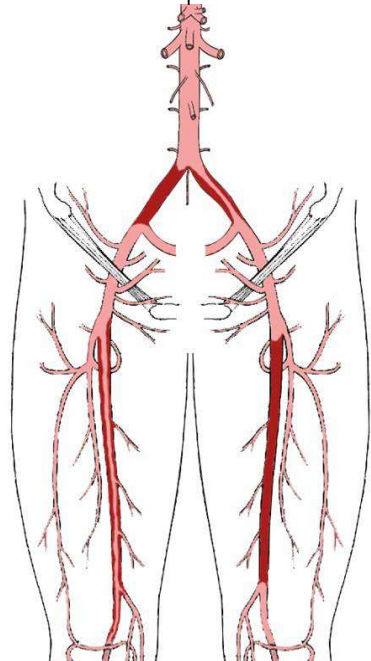
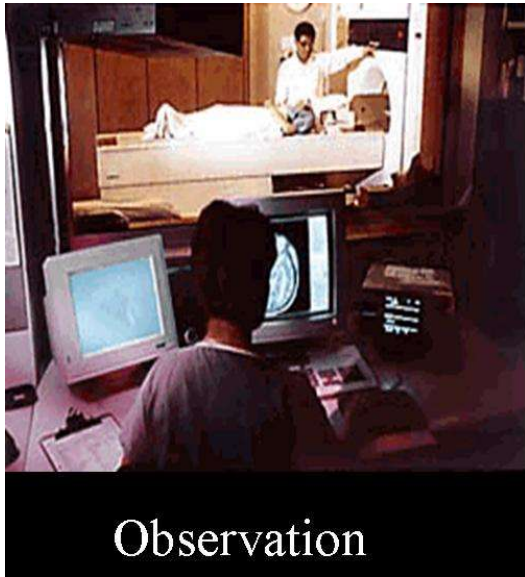
Bioengineering research group, Auckland

- **Finite Element modeling**

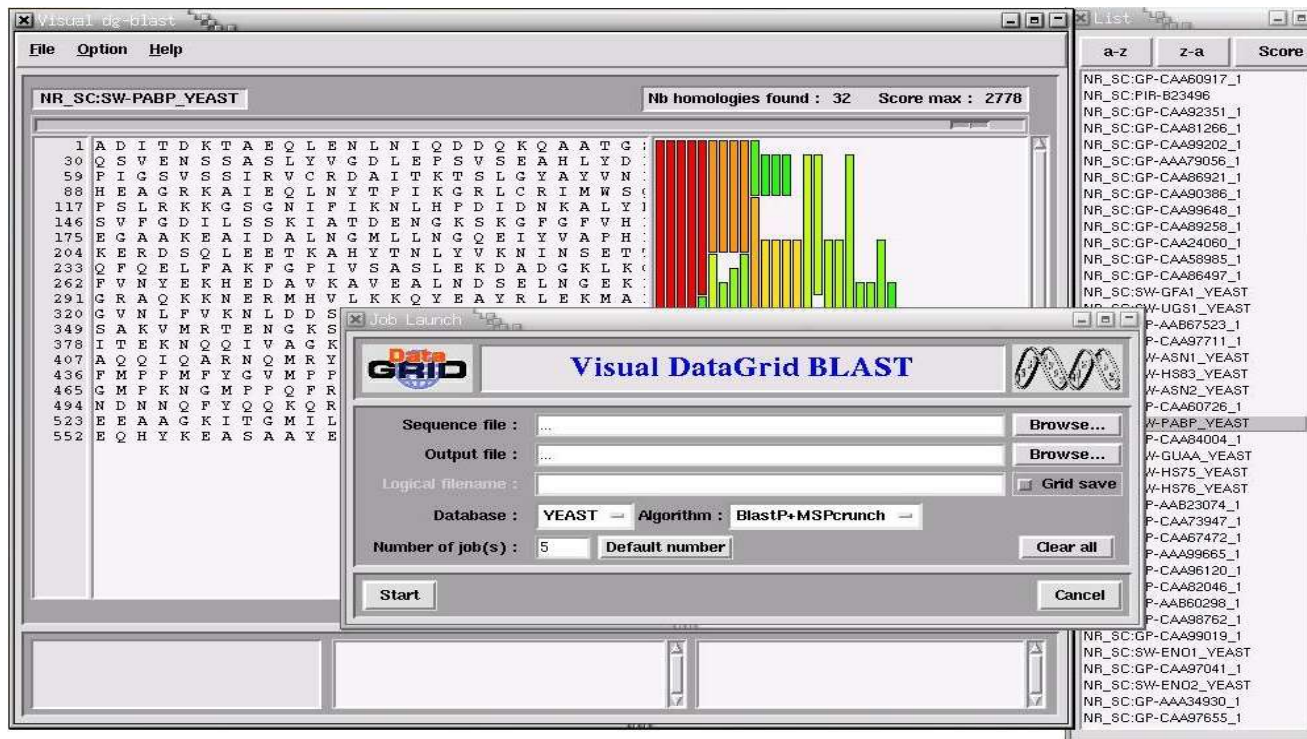
- elements oriented in heart fibers direction: fine resolution**

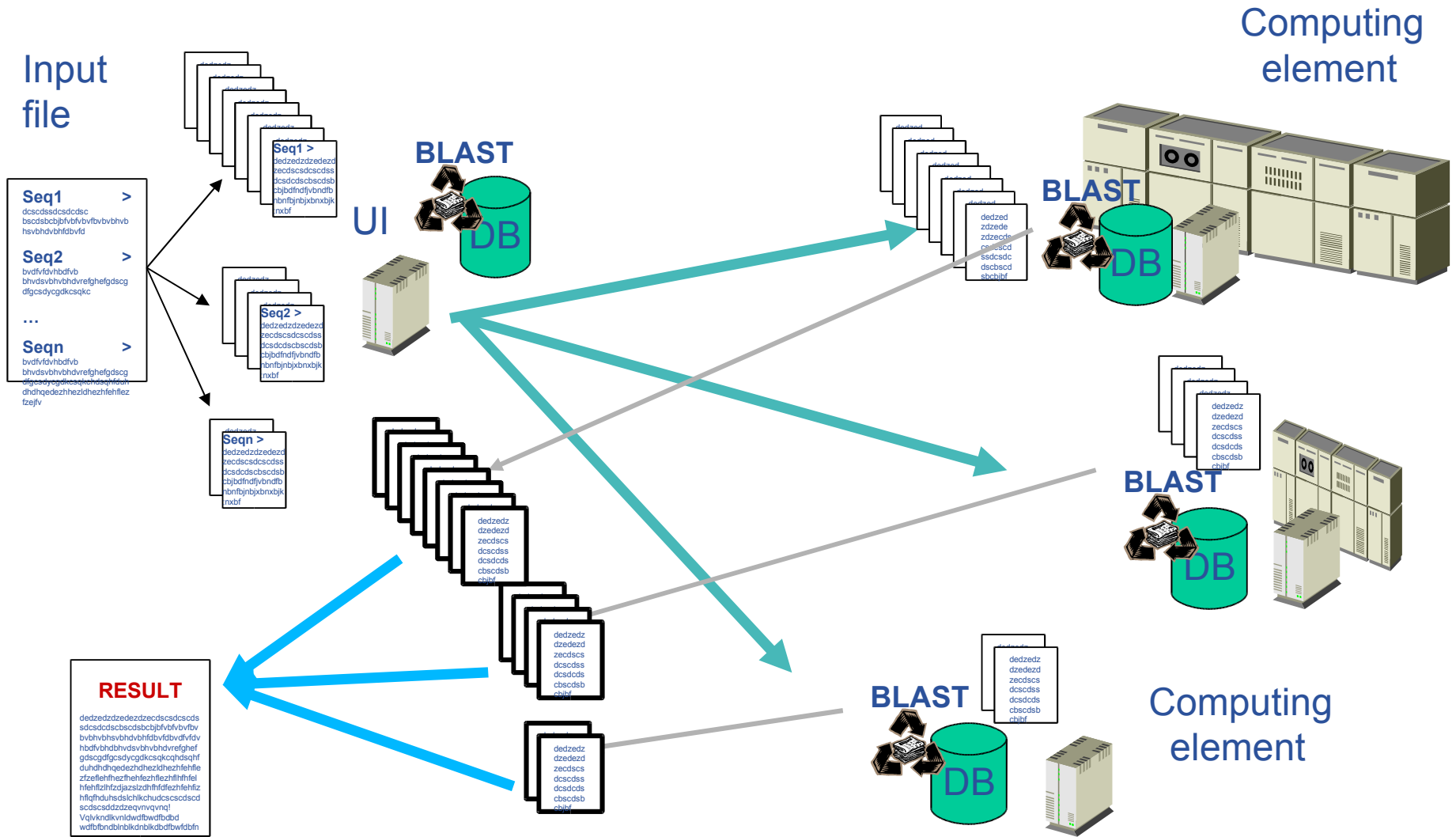
- electrical propagation model based on bidomain theory**

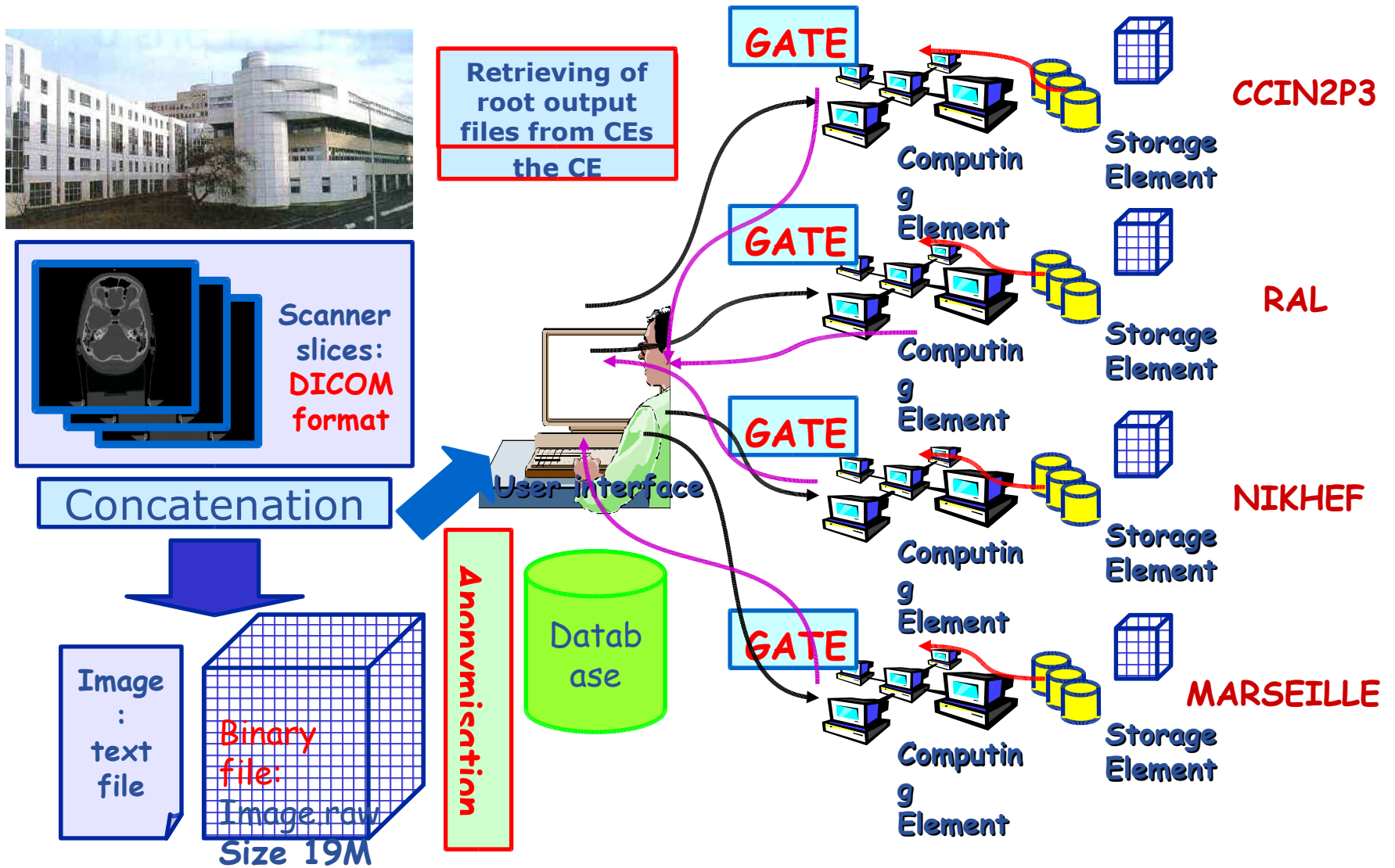
- 4D model (3D+T)**



- BLAST is the first step for analysing new sequences: to compare DNA or protein sequences to other ones stored in personal or public databases. Ideal as a grid application.**
  - Requires resources to store databases and run algorithms
  - Can compare one or several sequence against a database in parallel
  - Large user community





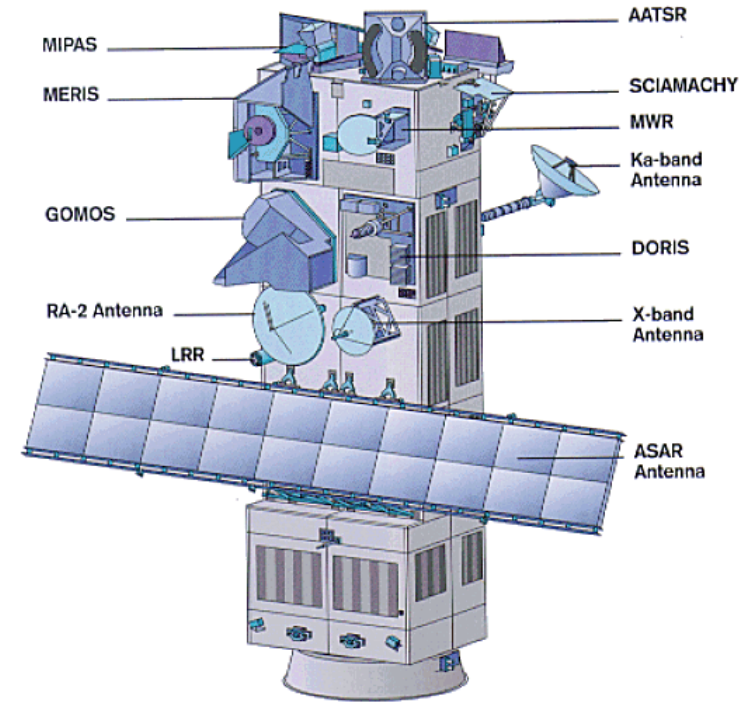
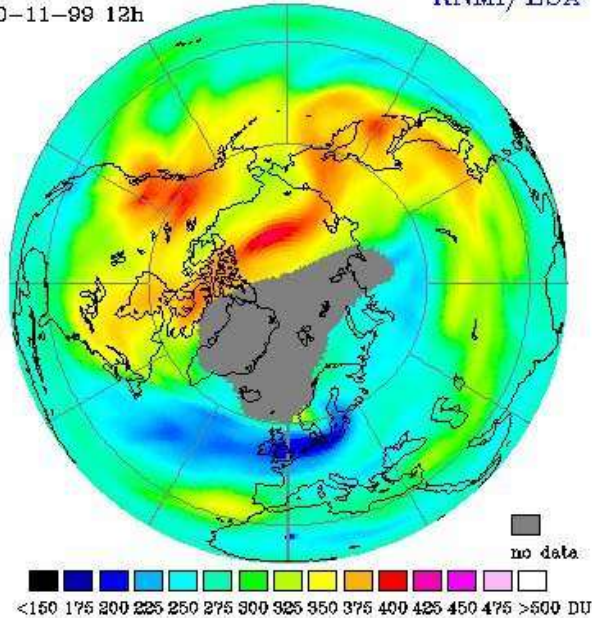




## ESA missions:

- about 100 Gbytes of data per day (ERS 1/2)
- 500 Gbytes, for the ENVISAT mission (2002).

Assimilated GOME total ozone  
30-11-99 12h  
KNMI/ESA



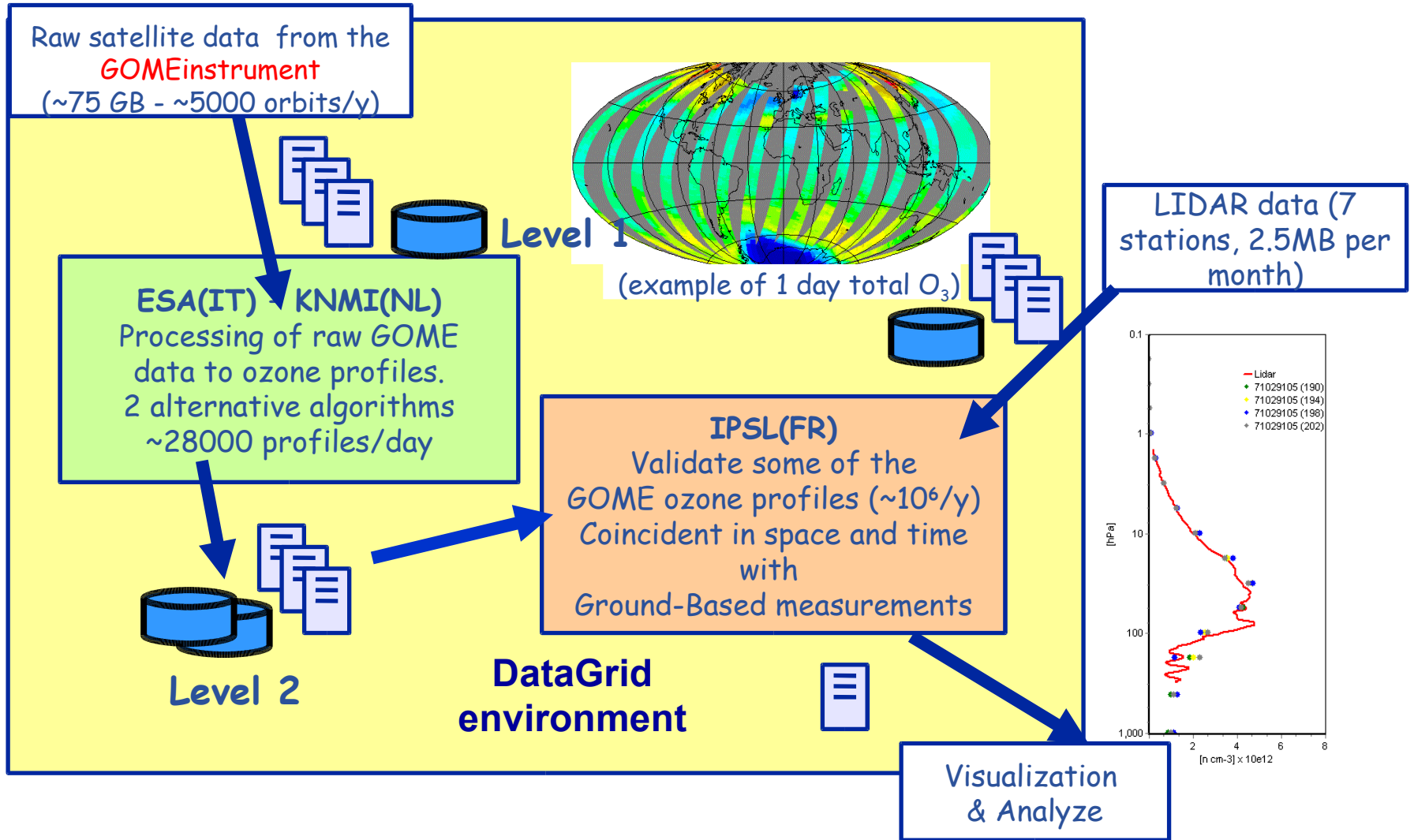
## Grid contribute to EO:

- enhance the ability to access high level products
- allow reprocessing of large historical archives
- improve Earth science complex applications (data fusion, data mining, modelling ...)

A photograph of the ENVISAT satellite in orbit above Earth. The satellite is a complex, gold-colored structure with various instruments and antennas. A long, thin solar panel array extends from the main body. The Earth's surface is visible below, showing blue oceans and white clouds. The word "ENVISAT" is overlaid in large, bold, yellow letters on the left side of the image.

# ENVISAT

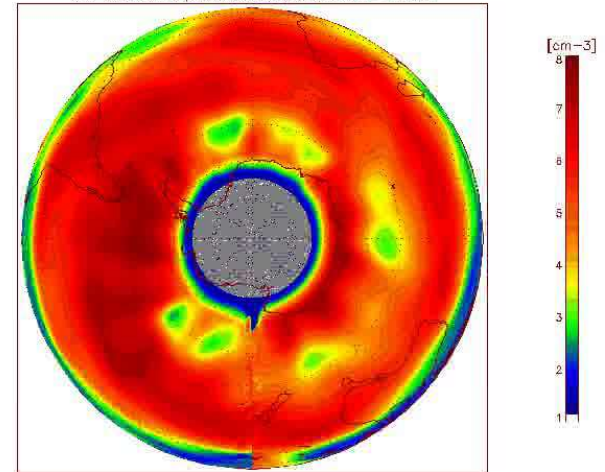
- 3500 Meuro programme cost
- Launched on February 28, 2002
- 10 instruments on board
- 200 Mbps data rate to ground
- 400 Tbytes data archived/year
- ~100 'standard' products
- 10+ dedicated facilities in Europe
- ~700 approved science user projects



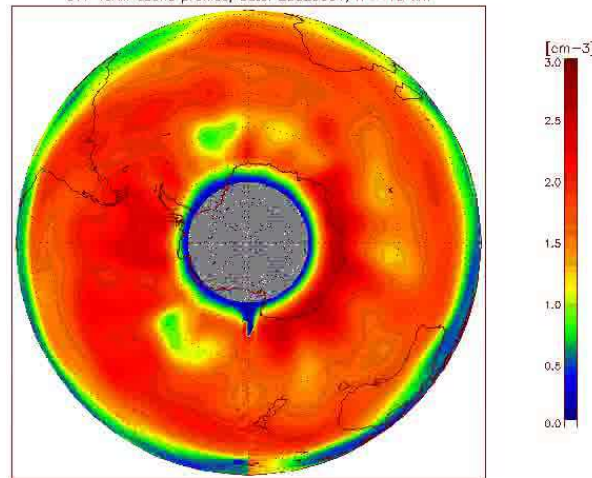


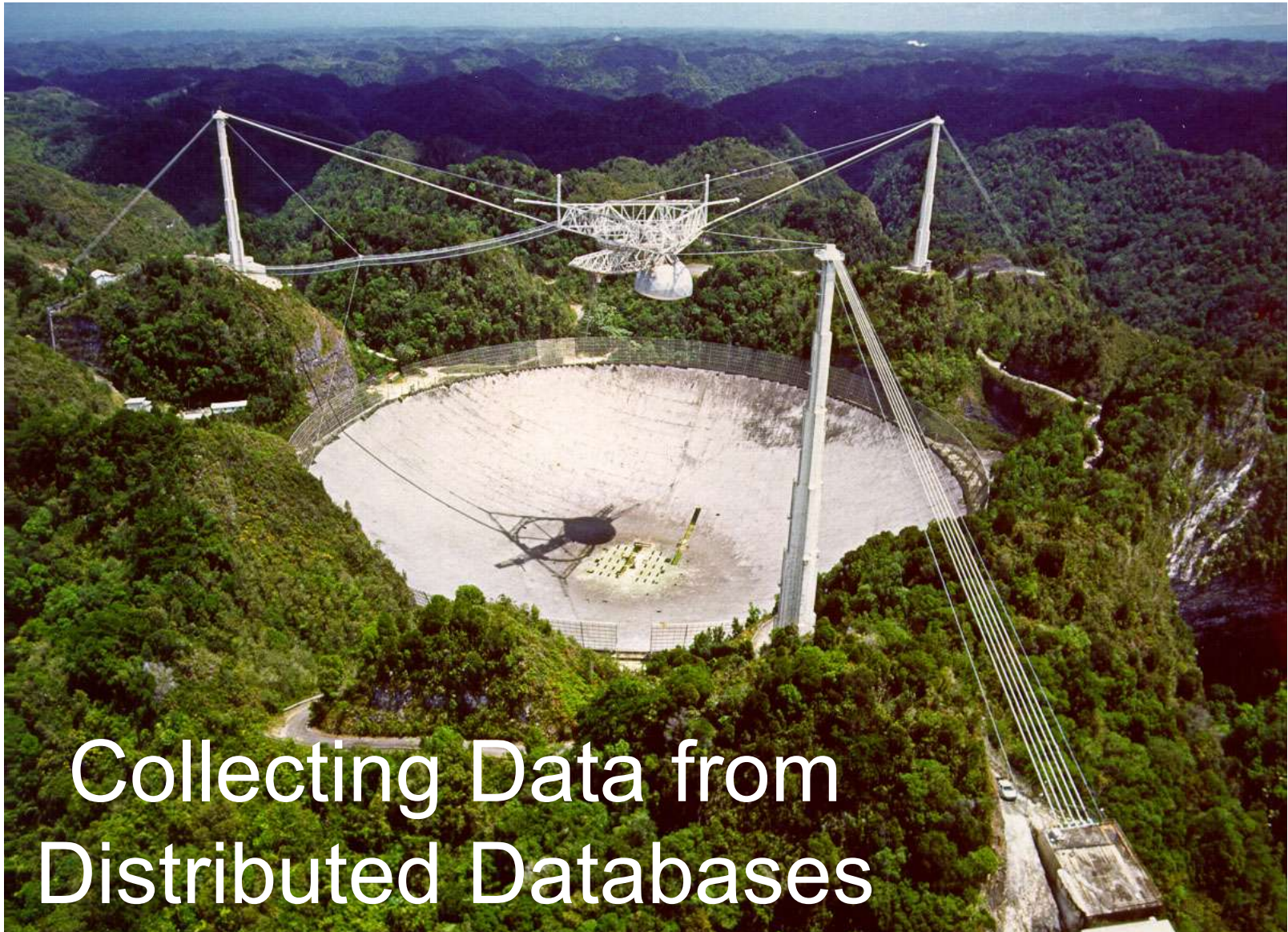
S. Casadio - ESA ESRIN  
 (GOME 3D Ozone volume over Antarctica -  
 Sept 02, NNO Level 2 products generate  
 in EDG)

UTV-IGAM ozone profiles, date: 20020901, h = 16 km



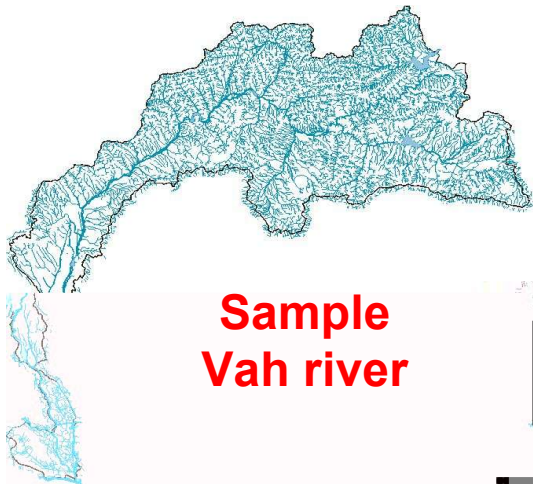
UTV-IGAM ozone profiles, date: 20020901, h = 10 km



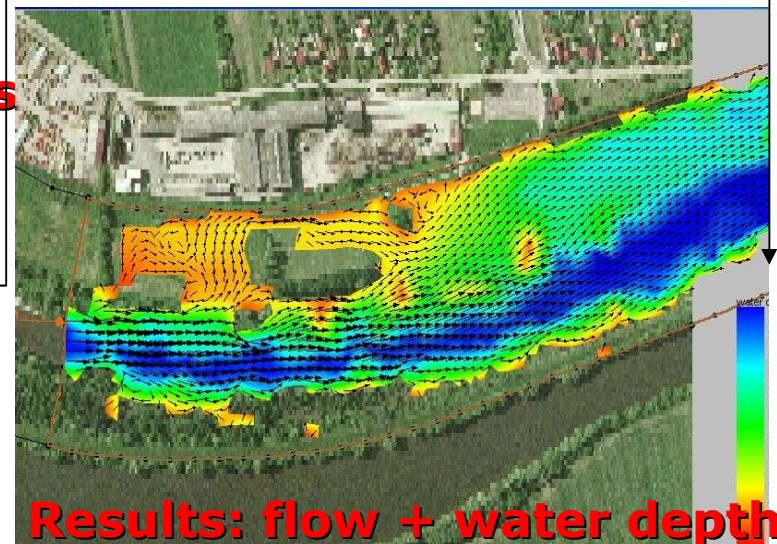


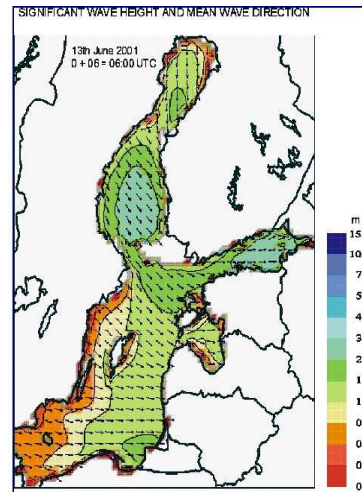
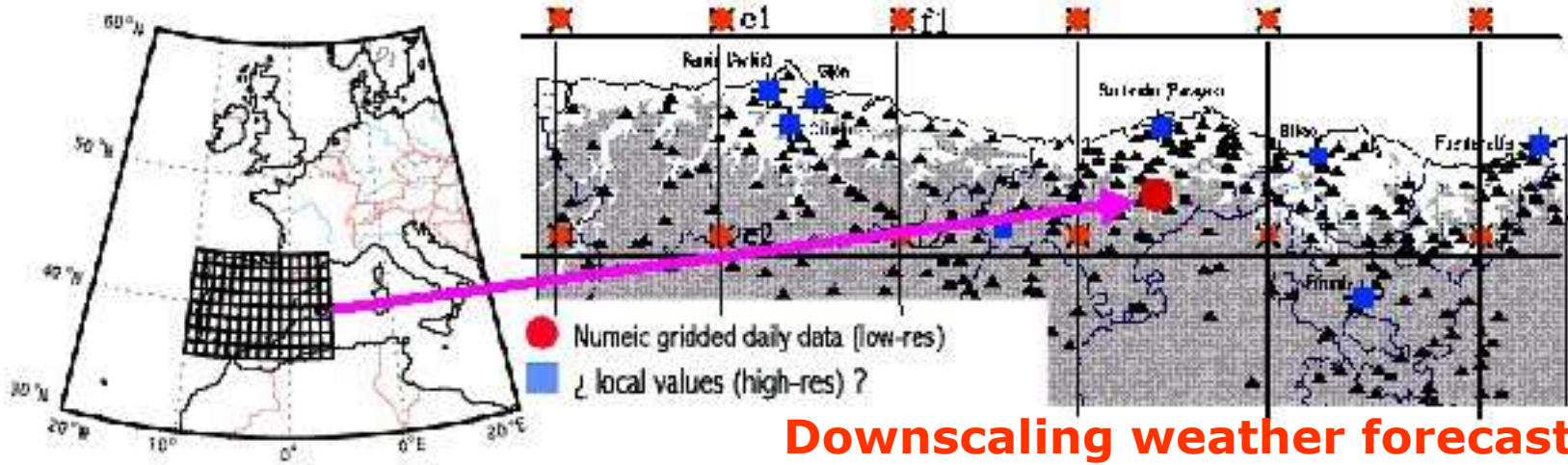
Collecting Data from  
Distributed Databases



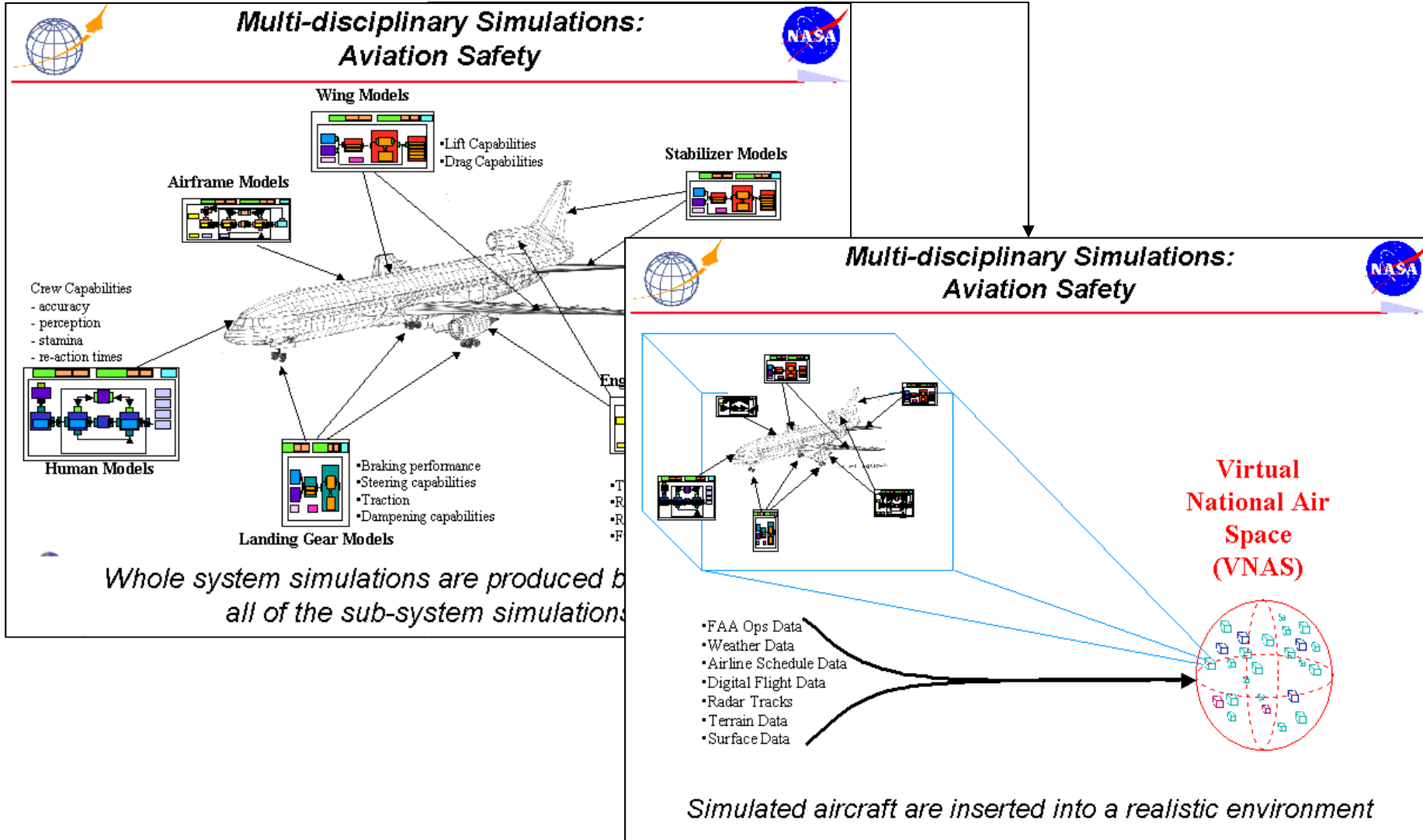


**Geographical Information Systems**



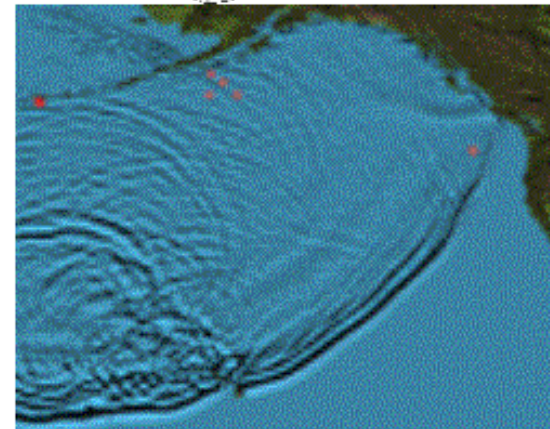
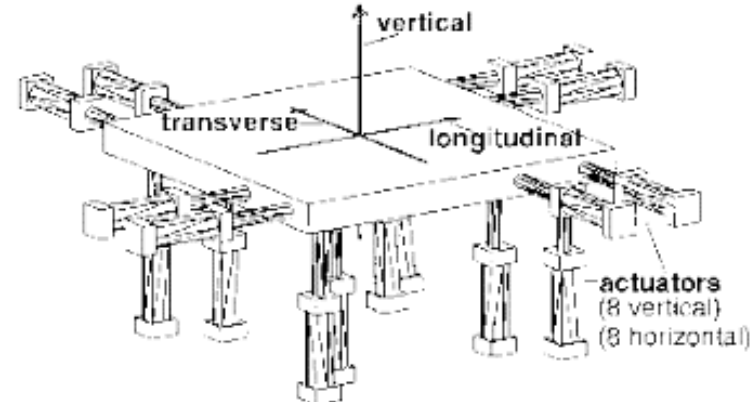


**Significant wave height and mean wave direction**



## Network for Earthquake Engineering Simulation

- NEESgrid: national infrastructure to couple earthquake engineers with experimental facilities, databases, computers, & each other
- On-demand access to experiments, data streams, computing, archives, collaboration



NEESgrid: Argonne, Michigan, NCSA, UIUC, USC



Paintings are being scanned in at  
30 GB each  
in the EU CRISATEL Project



Museo Virtual de Artes El Pais (MUVA)  
<http://www3.diarioelpais.com/muva/>.

Books are being scanned in at  
767 MB per page  
1/2 Terabyte for Gutenberg Bible



# ***Grid and Grid Business***



Ian Foster at Sun booth during SC2001, Denver



Irving Wladawsky-Berger at GGF4 in Toronto - \$ 4 billion IBM investment

- High entry level
- Large scale enterprises (??)



- Selling compute cycles
- Sun: 1\$ / CPU-hour

- Pushing Tin, selling computing equipment
  - large and small companies
- “GSP” (Grid Service Provide)
  - Webservices, Storage Service Provider
- Training, Support and Consulting
  - Grid Computing is a new technology, people need training
- Decentralization

## “Embarrassingly parallel” applications

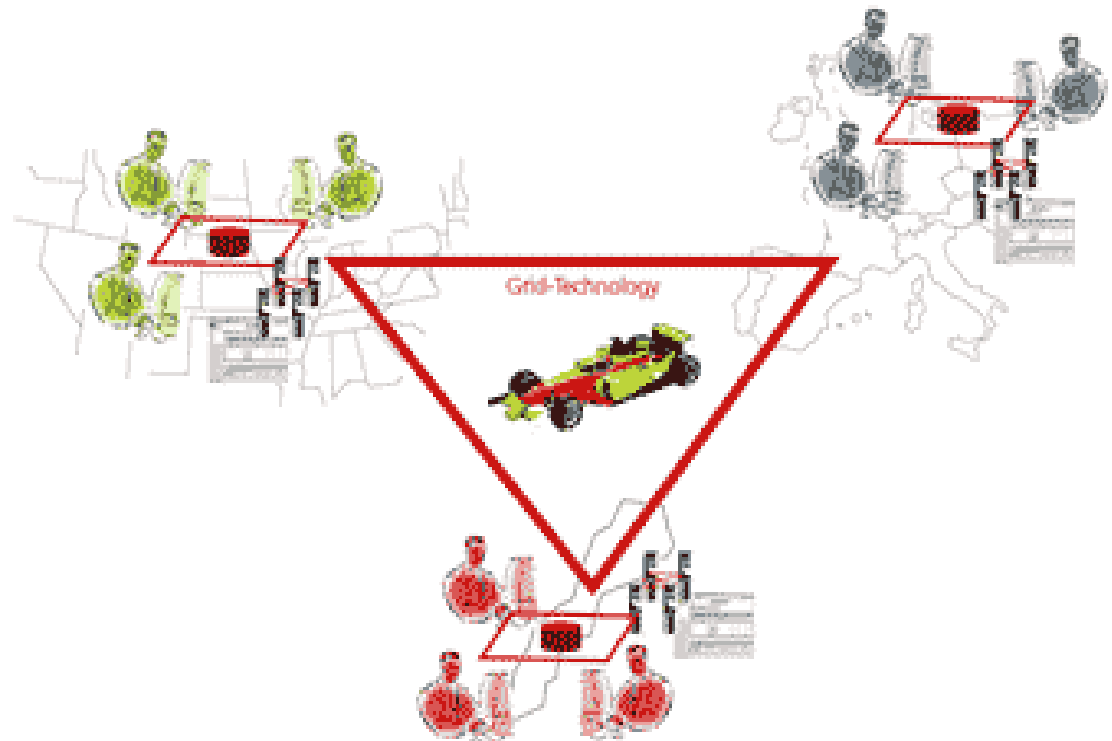
- Film industry (“Titanic” style);  
**rendering (Maja)**
- Some types of Monte Carlo simulations, e.g financial risk analysis





## Integration, Virtualisation

- SIMDAT, InGRID
- Collaboration between several product development teams using grid technology by federating product development environments.
- Thin clients for Grid resources ?



## Services

- Google, eBay, data mining, **spell checking** ...

## Data collection, remote access

- DAME (**Sheffield University**)
- sensor grids



## Video conferencing

- Access Grid



Remote video

Visualisation

Microphones

Cameras

## Now different notion of “Grid Computing”

- Oracle: Mostly high-availability cluster
- Sun Grid Engine: Batch Submission
- IBM Discovery Link: Federating databases
- **Plus:** Webservices and Grid Computing are moving towards each other



# *Other Grid Initiatives, New EU Projects in FP6*

- US counterpart to EDG / EGEE
- Deployment of an international Data Grid with dozens of sites and thousands of processors.
- Operated jointly by the U.S. Grid projects iVDGL, GriPhyN and PPDG, and the U.S. participants in the LHC experiments ATLAS and CMS.

In the summer of 2004 a total the European Commission launched 12 research projects in the context of FP6 in the area of Grid technologies (see press release) that will receive EUR 52 million of EU funding. The bulk of the EU funding is going to 4 projects - SIMDAT, NextGRID, Akogrimo and CoreGRID with 9 Mio. Euro each.

- **NextGRID:** Architecture for Next Generation Grids
- **SimDat:** Industry focus; development of fundamental Grid techniques
- **Akogrimo:** Access to Knowledge through the Grid in a mobile world; “pervasiveness of Grid computing”
- **CoreGrid:** aims at strengthening and advancing scientific and technological excellence
- DataMiningGrid, GridCoord, HPC4U, K-WF Grid, OntoGrid, Provenance, UniGridS

- **CrossGrid**: Interactive use of the Grid (MPI, etc.)
- Builds on EDG
- Project was finished this month
- **CrossGrid-2 ??**
- **Condor**: Batch submission. Condor-G: Globus-enabled
- NIMROD: Parametric optimisation studies across a Grid



- National Grid deployment projects (D-Grid, E-Science, Business Grid)
- Focus on integration and production quality infrastructure, mostly not Grid research
- Large funding
- Industry involvement
- TeraGrid, Geant: network infrastructure (our backbone)

# *Summary / Conclusion*

- We are not alone -> Integration !
- Many Grid Applications already available (although not in the mainstream)
- Typical Areas:
  - „Embarrassingly Parallel“ Applications (or „nicely parallel“)
  - Collaboration
  - Distributed Databases
  - Some real parallel processing

# Thanks for listening !

**We'd like to thank the German Federal Ministry of Education and Research, BMB+F, as well as Forschungszentrum Karlsruhe / Germany for their continuous interest and support !**



**bmb+f** - Förderschwerpunkt  
Hadronen -  
und Kernphysik  
Großgeräte der physikalischen  
Grundlagenforschung