Applications and the Grid

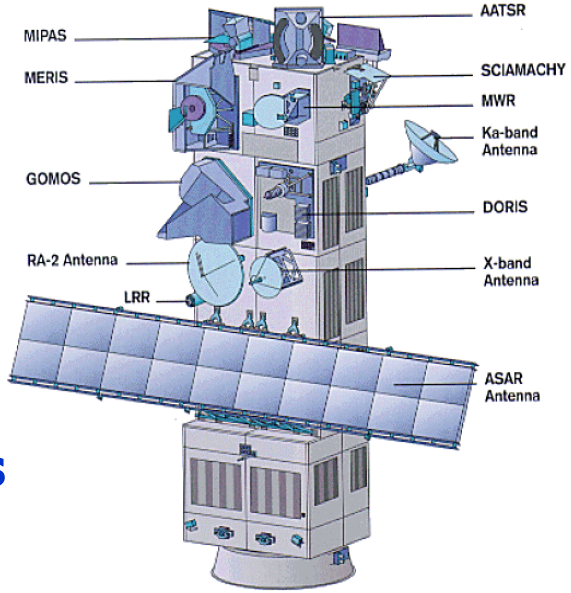
The European DataGrid Project Team

http://www.eu-datagrid.org

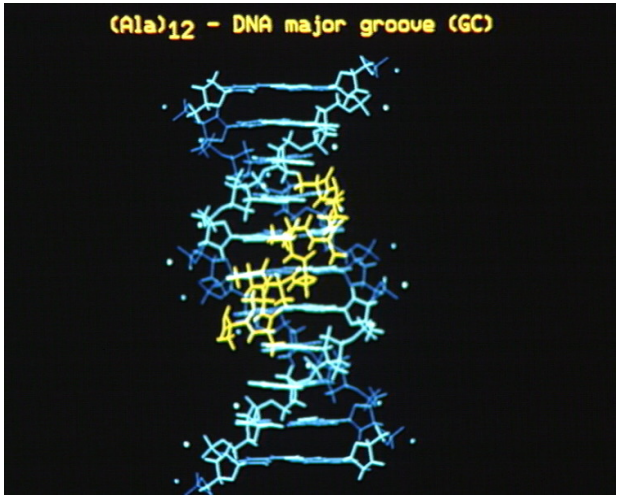
EDG Application Areas

High Energy Physics

Earth Observation Science Applications



Biomedical Applications

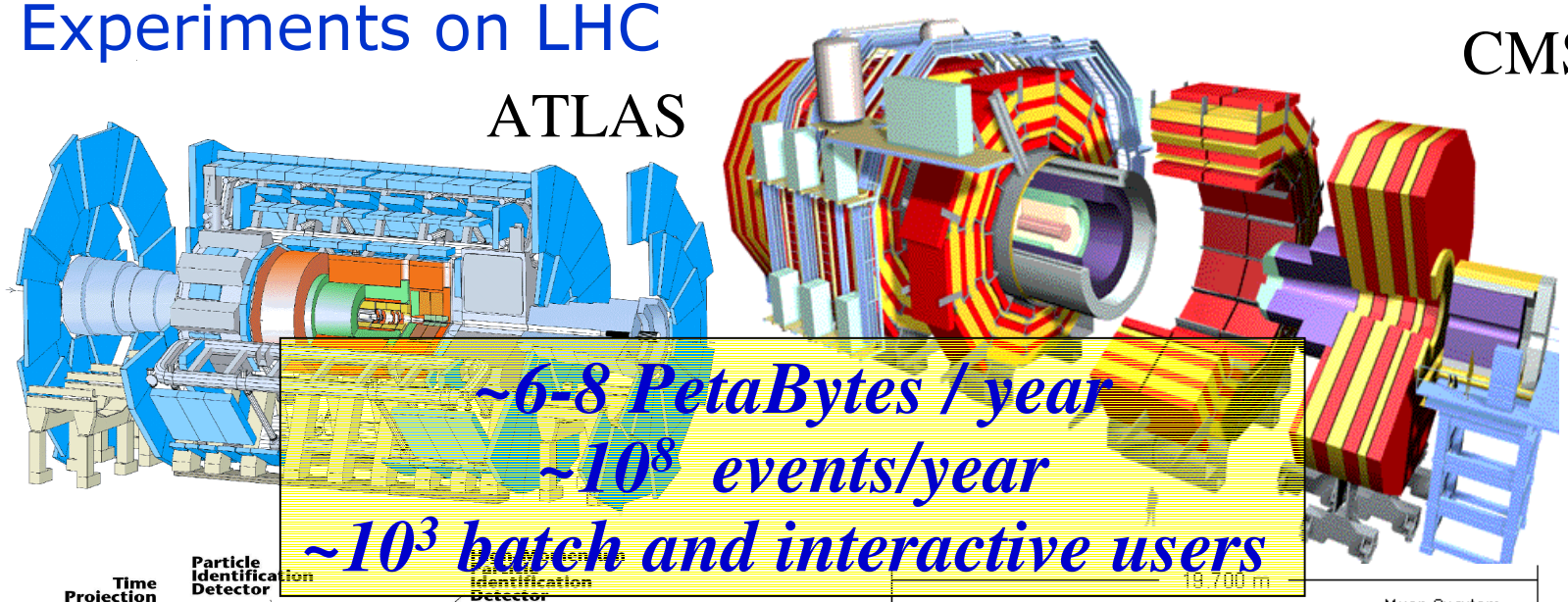


High Energy Physics

4 Experiments on LHC

ATLAS

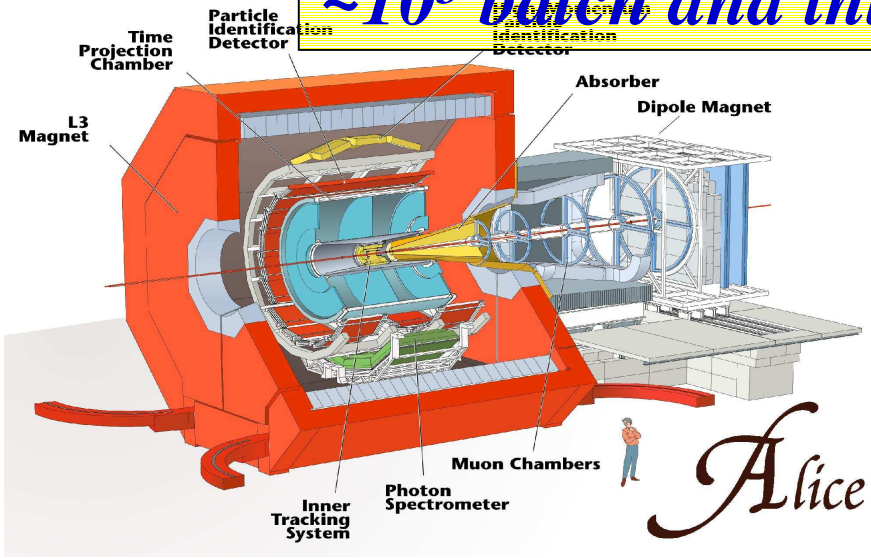
CMS

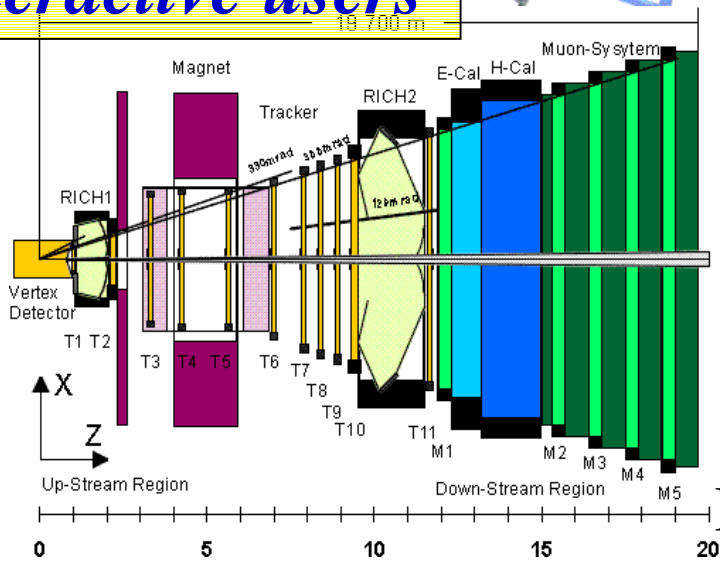


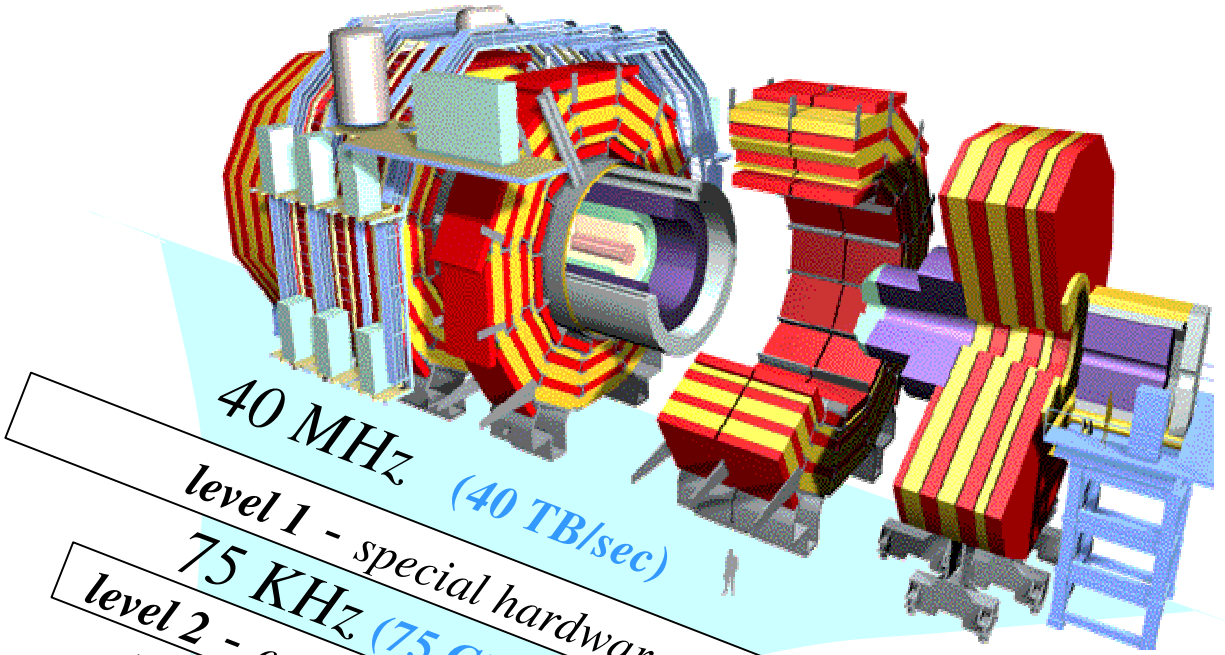
~6-8 PetaBytes / year

~10^8 events/year

~10^3 batch and interactive users





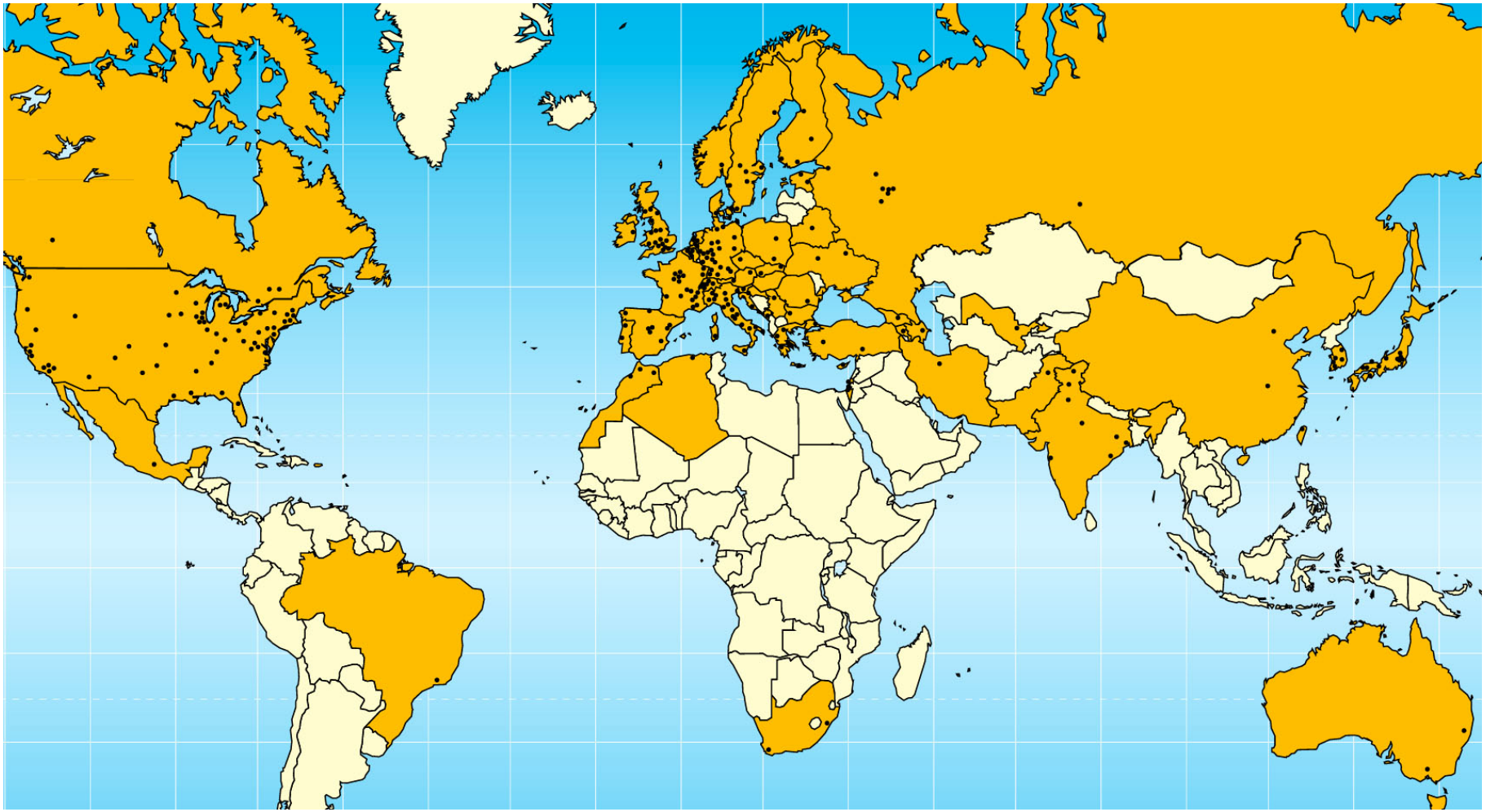


online system

multi-level trigger
filter out background
reduce data volume

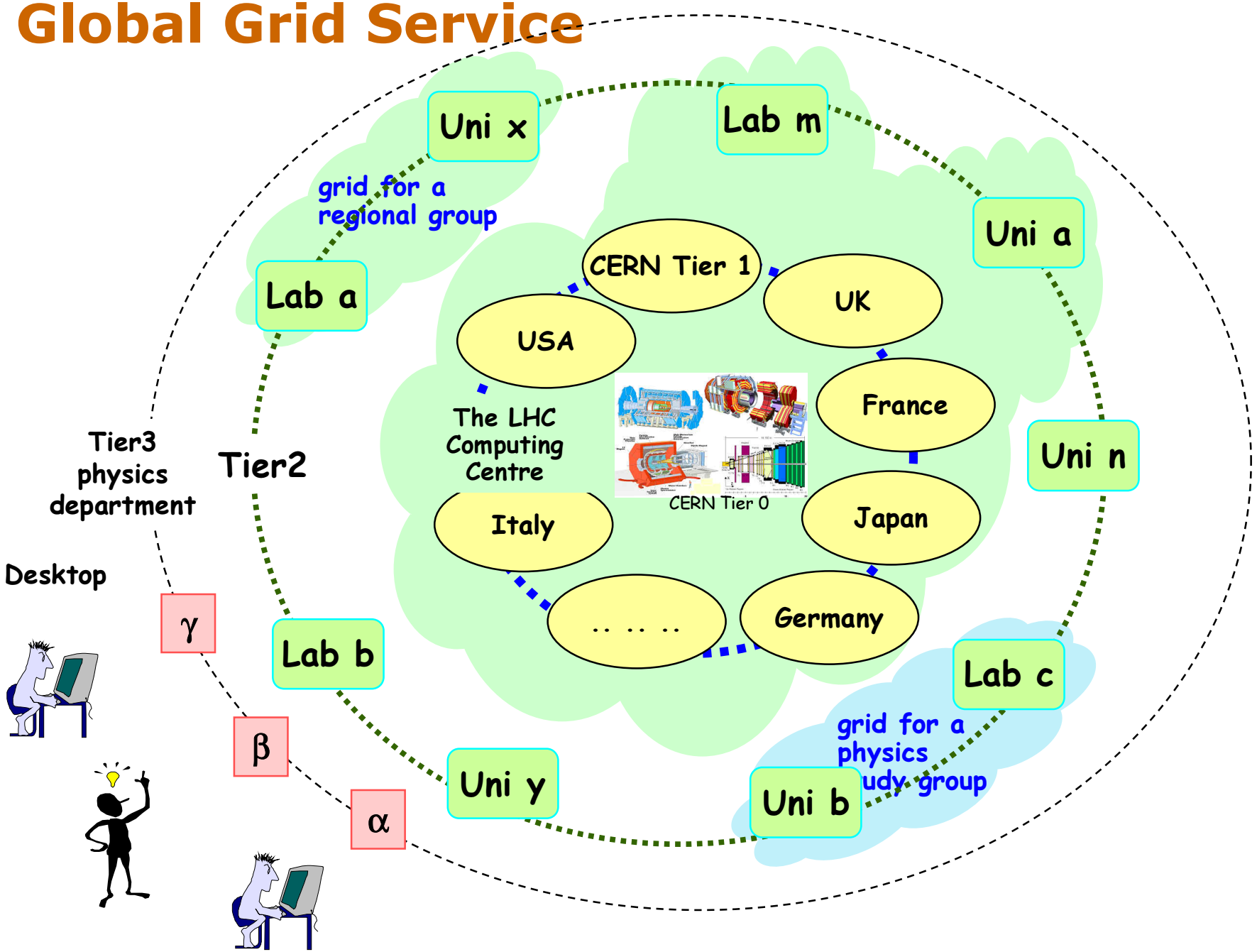
les.robertson@cern.ch

CERN's Network in the World



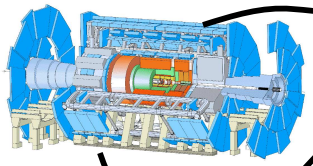
Europe: 267 institutes, 4603 users
Elsewhere: 208 institutes, 1632 users

Deploying the LHC Global Grid Service



les.robertson@cern.ch

Data Handling and Computation for Physics Analysis



detector

event filter (selection & reconstruction)

reconstruction

raw data

event summary data

processed data

event reprocessing

batch physics analysis

analysis

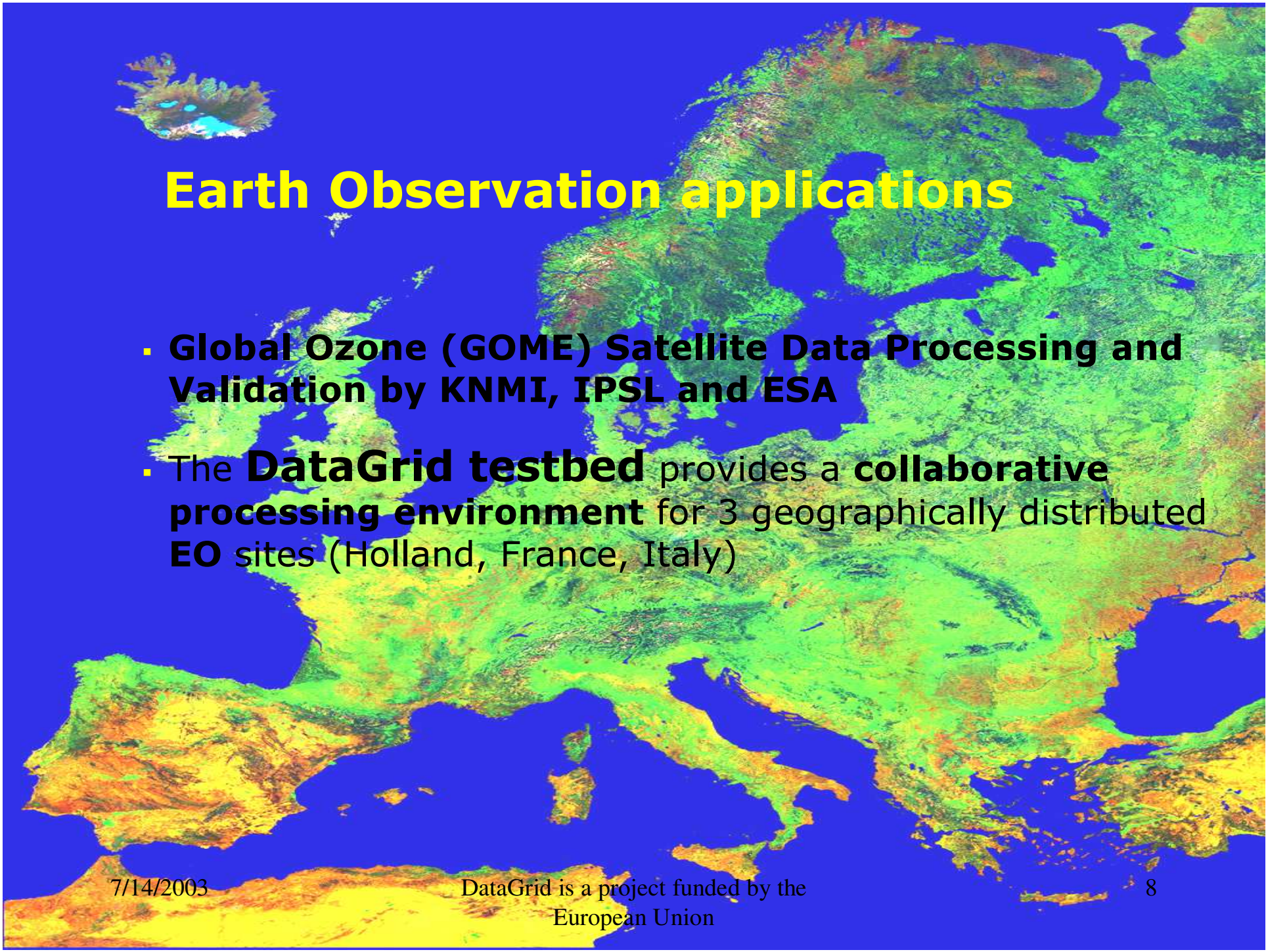
analysis objects
(extracted by physics topic)

event simulation

simulation

interactive physics analysis

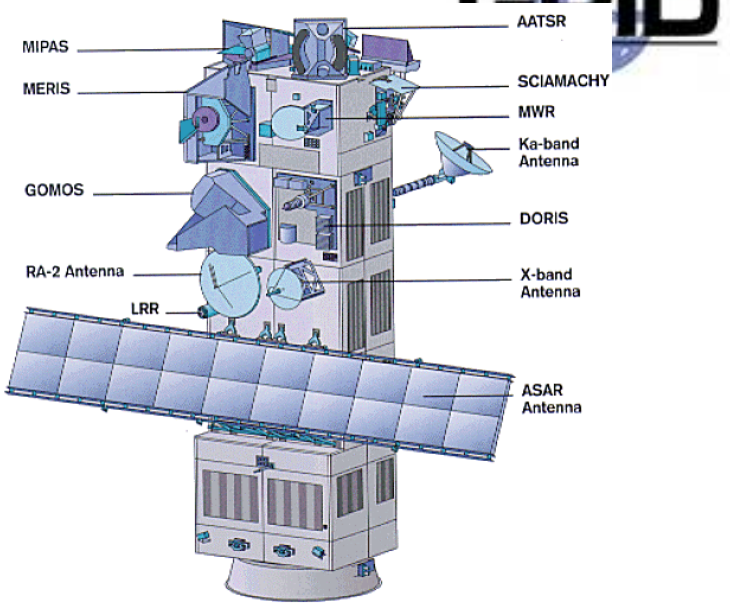
les.robertson@cern.ch



Earth Observation applications

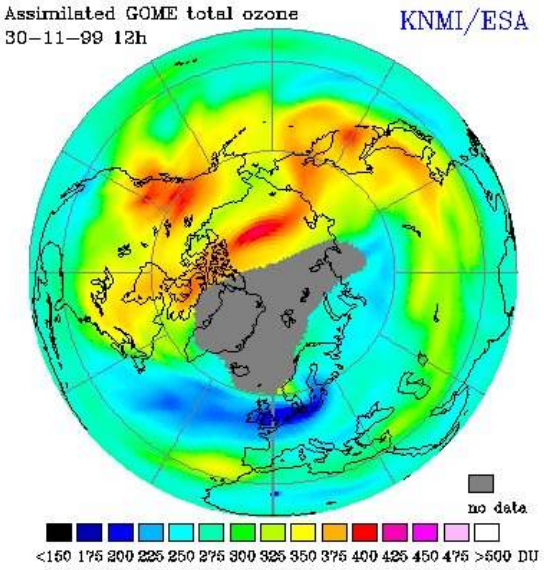
- **Global Ozone (GOME) Satellite Data Processing and Validation by KNMI, IPSL and ESA**
- The **DataGrid testbed** provides a **collaborative processing environment** for 3 geographically distributed **EO** sites (Holland, France, Italy)

Earth Observation



ESA missions:

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

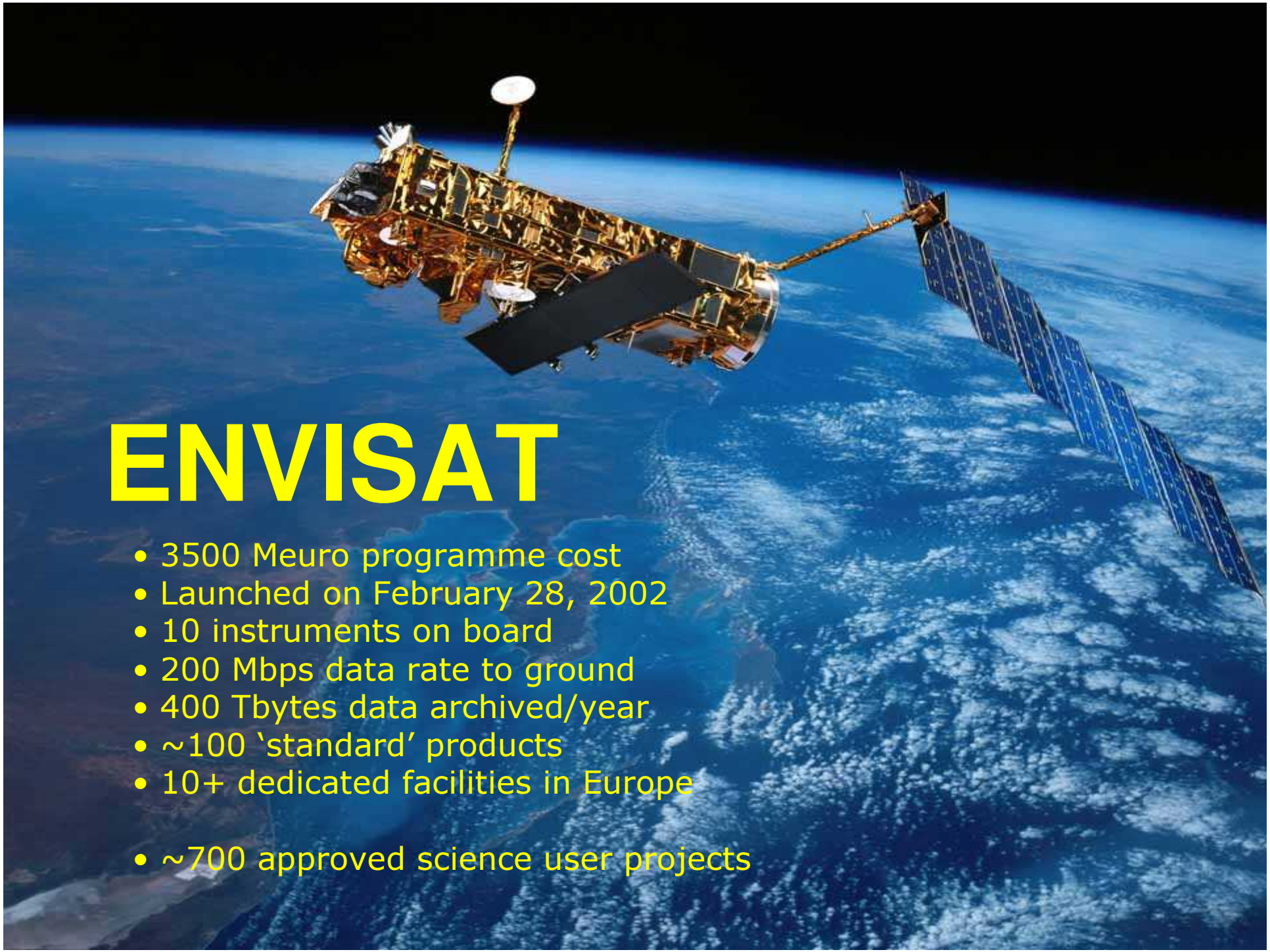


DataGrid contribute to EO:

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

Source: L. Fusco, June 2001

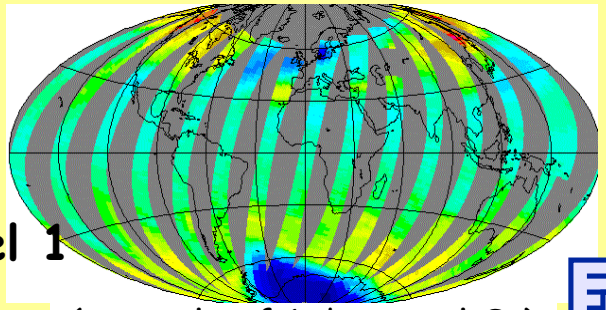
Federico.Carminati , EU review presentation, 1 March 2002


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

EDG EO challenge: Processing / validation of 1y of GOME data

Raw satellite data from the GOME instrument (~75 GB - ~5000 orbits/y)

Level 1



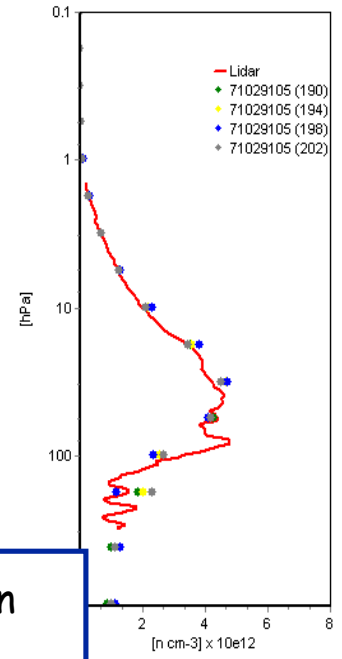
(example of 1 day total O_3)

ESA(IT) KNMI(NL)
Processing of raw GOME data to ozone profiles.
2 alternative algorithms
~28000 profiles/day

Level 2

LIDAR data (7 stations, 2.5MB per month)

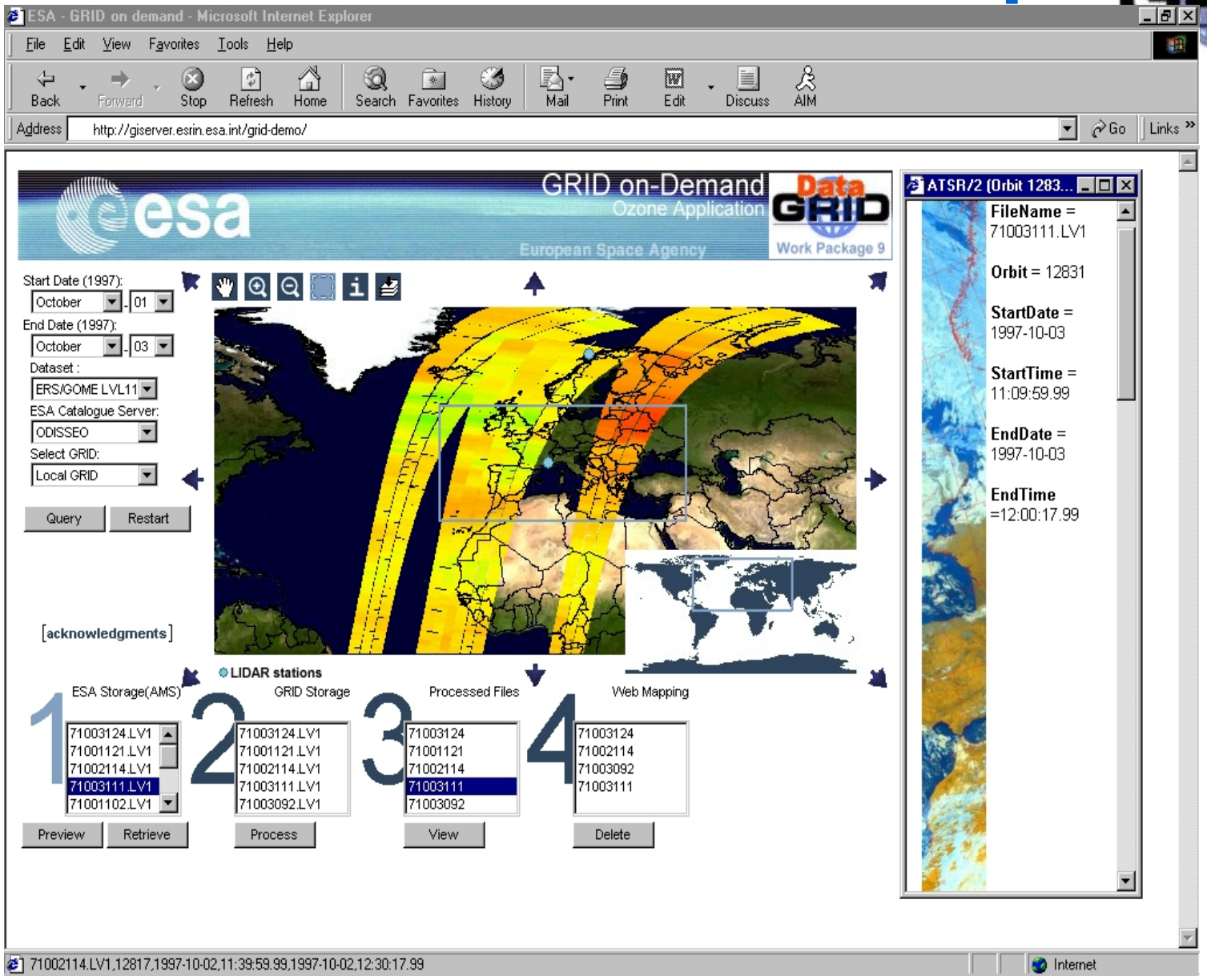
IPSL(FR)
Validate some of the GOME ozone profiles (~10^6/y)
Coincident in space and time with
Ground-Based measurements



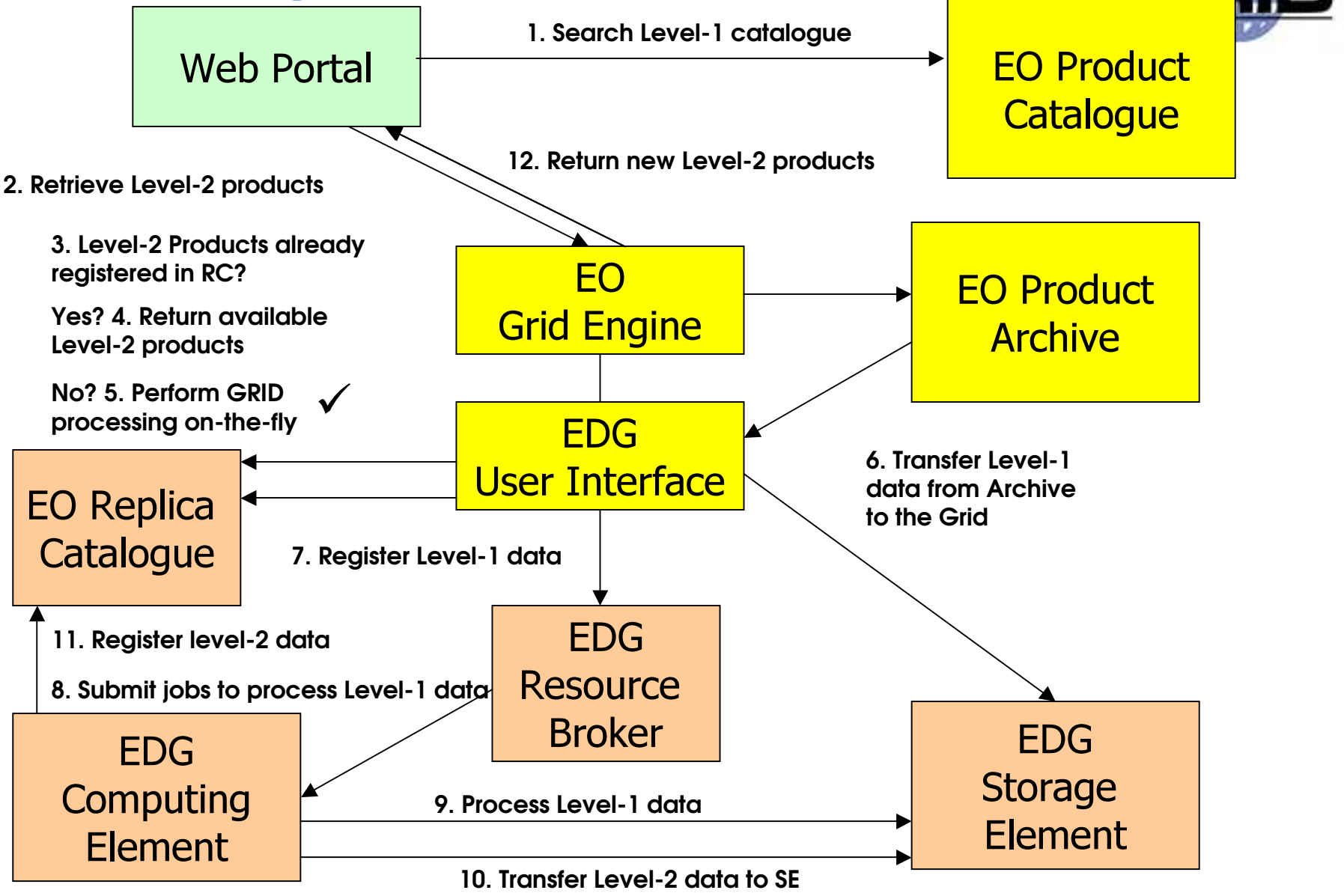
DataGrid environment

Visualization & Analyze

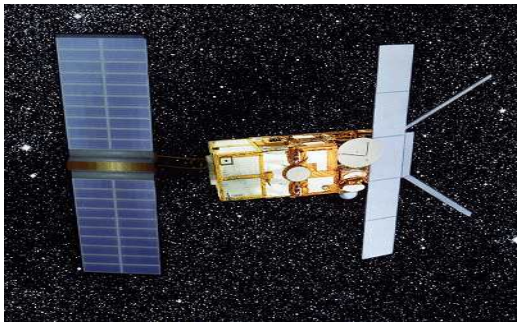
EO WebMap Portal



Processing Sequence



GOME Ozone Profile Validation



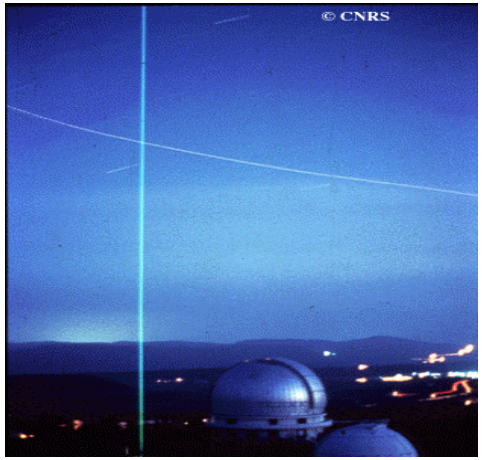
ERS/GOME satellite

50 km

10 km

OZONE LAYER

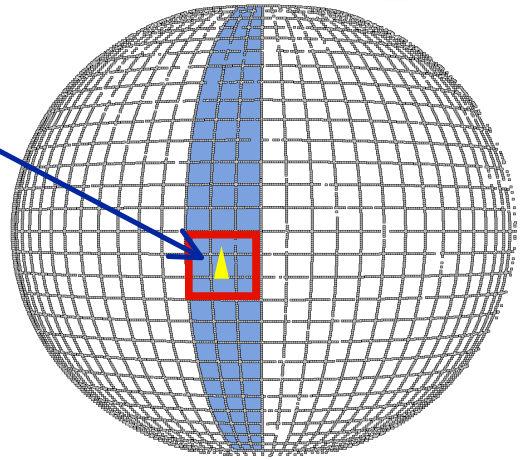
- ✓ Goals of the DataGrid application
  validate satellite data with all ground based data available in an easy way:
  - ➢ Comparison of ozone profiles provided by satellite with lidar data in different locations and times (see the web portal)
  - ➢ Statistical comparison and analysis in order to improve algorithms.



Lidar at the Haute Provence Observatory

Validation Processing Sequence

Satellite data validation



Level 2 Catalogue

2

Level 2 Catalogue

Queries and data information retrieval from the Gome Level 2 orbit or pixel metadata catalogues

3

Submission of the Job in the GRID

GRID Portal

Queries and data information retrieval from the Lidar metadata catalogue

1

Lidar data catalogue

GRID

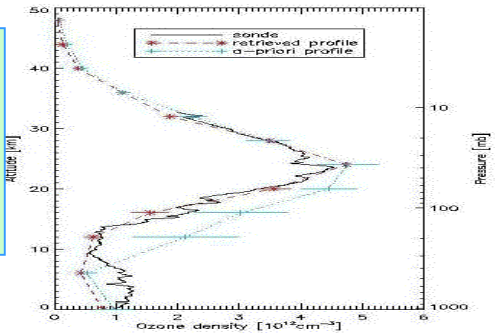
Storage Elements with Lidar data

Computing Element

Storage Elements with Gome L2 data

4

When completed comparison between lidar and satellite ozone profiles



Validation Output

Figure 1:

Estimation of the bias between Gome and Lidar using one month of data.

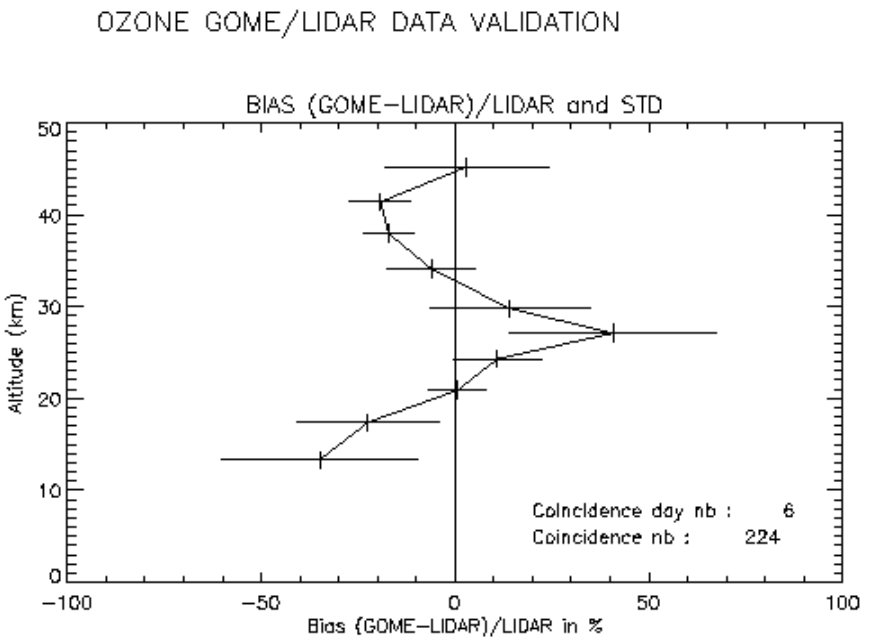
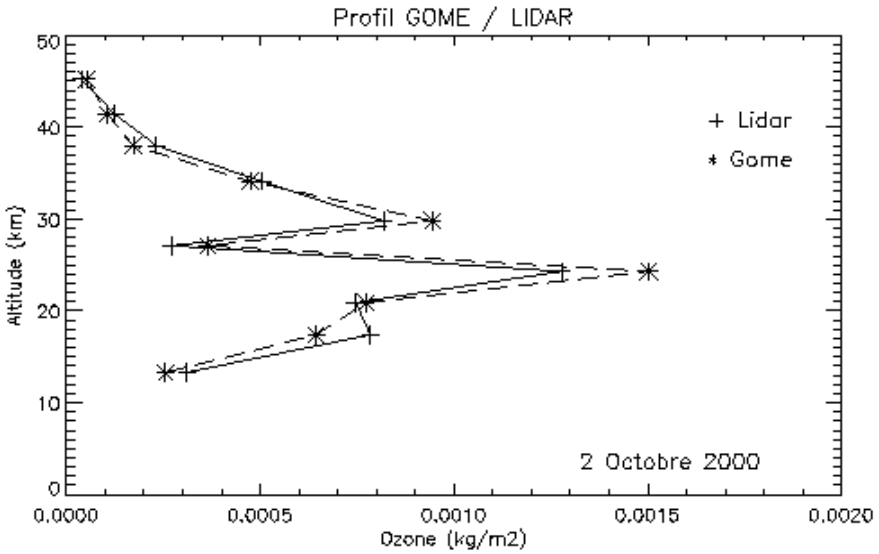


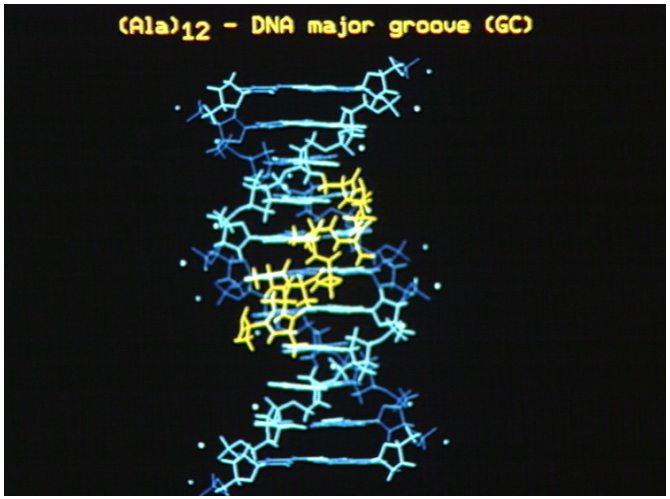
Figure 2 :

example of 2 profiles : Comparison between Gome profile and lidar profile for the 2nd October 2000.



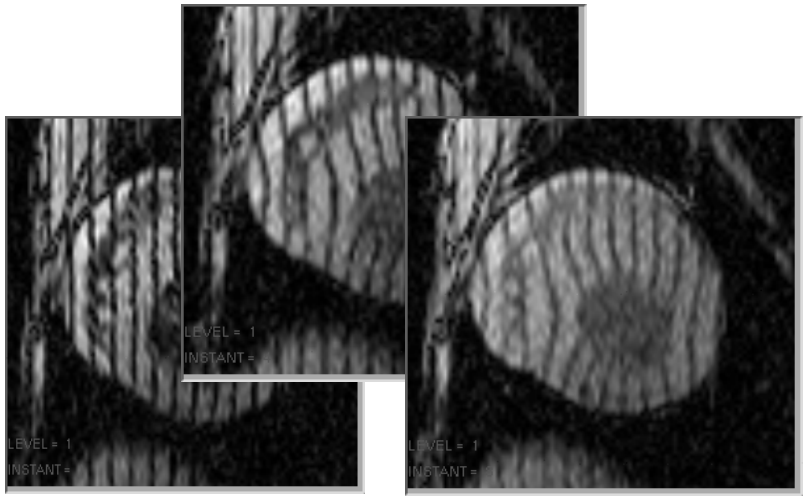
Biomedical Applications

Genomics, post-genomics, and proteomics



Explore strategies that facilitate the sharing of genomic databases and test grid-aware algorithms for comparative genomics

Medical images analysis



Process the huge amount of data produced by digital imagers in hospitals.

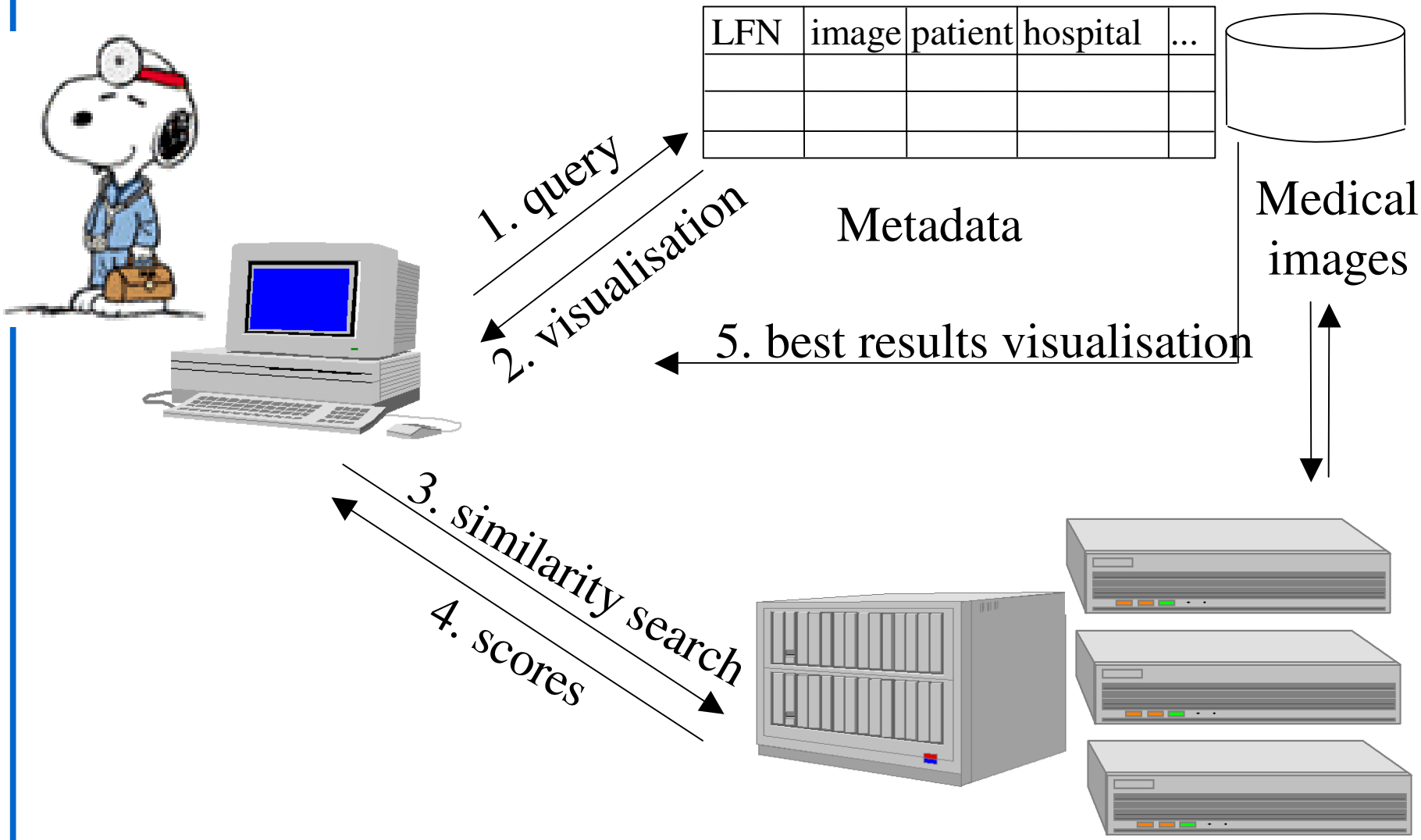
Biomedical Applications

Applications deployed
Applications tested on EDG
Applications under preparation

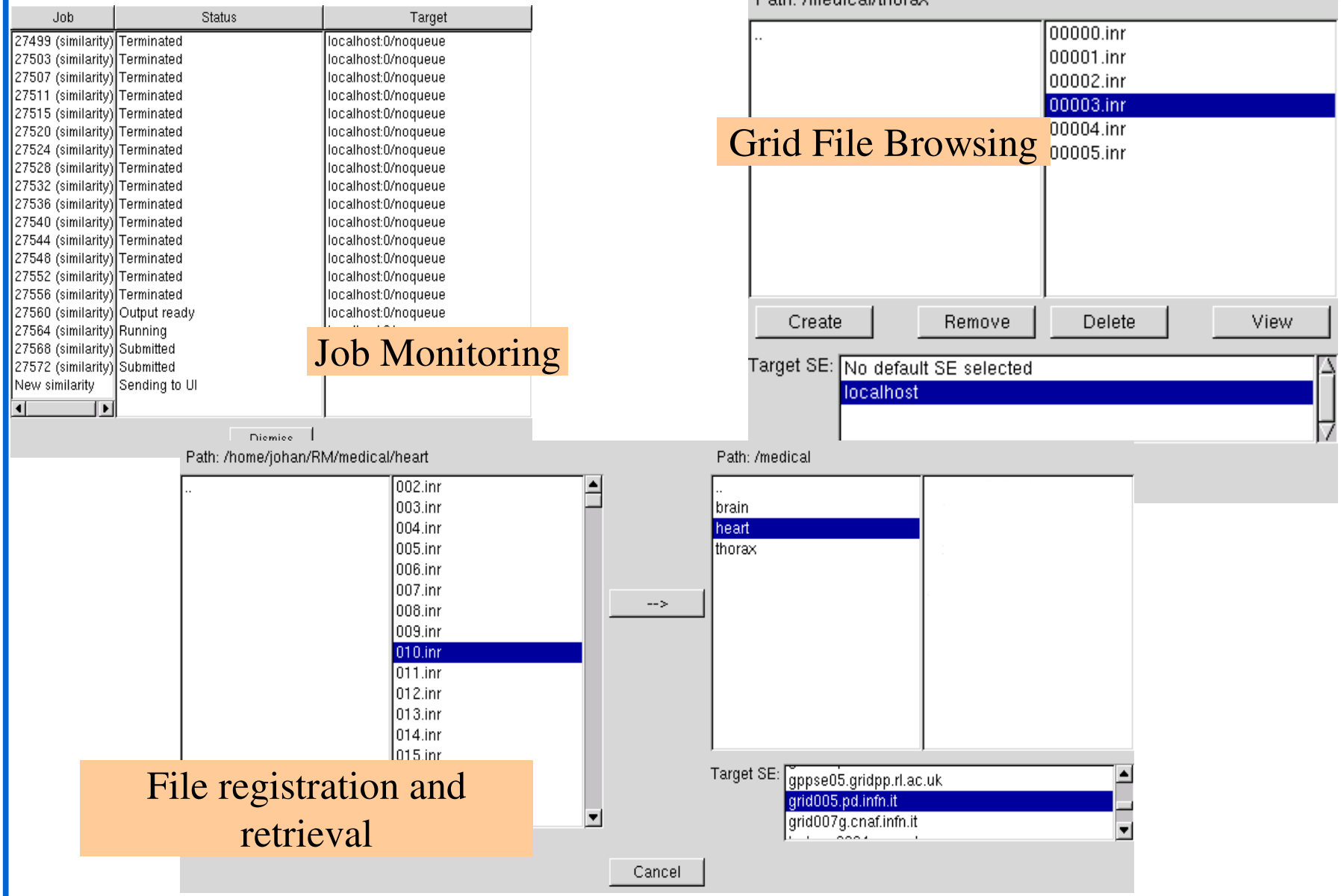
- **Bio-informatics**
  - Phylogenetics : BBE Lyon (T. Sylvestre)
  - Search for primers : Centrale Paris (K. Kurata)
  - Statistical genetics : CNG Evry (N. Margetic)
  - Bio-informatics web portal : IBCP (C. Blanchet)
  - Parasitology : LBP Clermont, Univ B. Pascal (N. Jacq)
  - Data-mining on DNA chips : Karolinska (R. Médina, R. Martinez)
  - Geometrical protein comparison : Univ. Padova (C. Ferrari)
- **Medical imaging**
  - MR image simulation : CREATIS (H. Benoit-Cattin)
  - Medical data and metadata management : CREATIS (J. Montagnat)
  - Mammographies analysis ERIC/Lyon 2 (S. Miguet, T. Tweed)
  - Simulation platform for PET/SPECT based on Geant4 : GATE collaboration (L. Maigne)

Medical Imaging

H

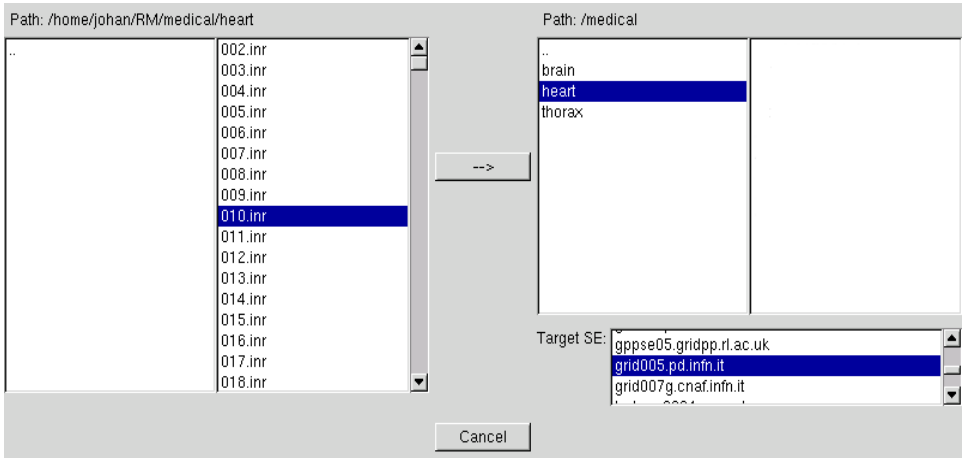


Graphic layer



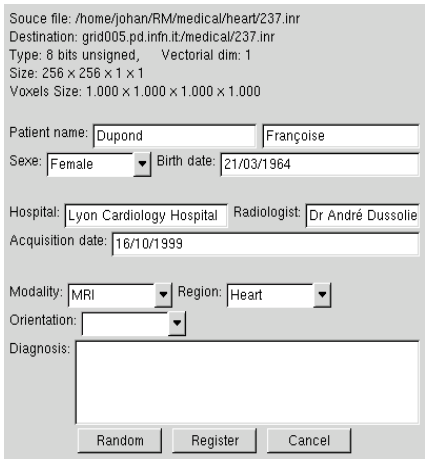
Graphical Interfaces

Image registration



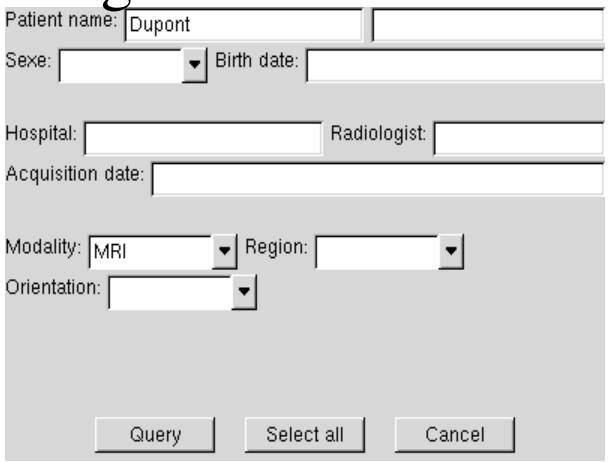
Local files

Grid files

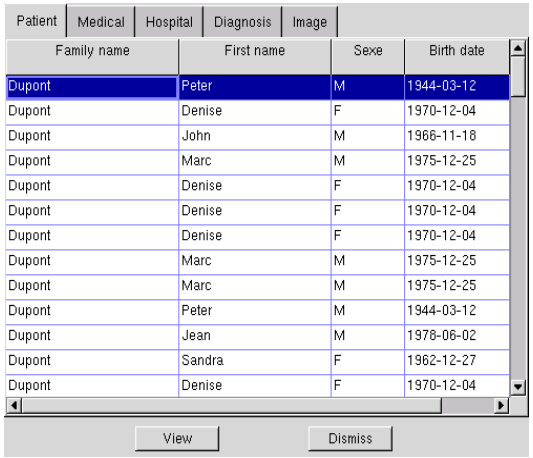


Metadata

Image retrieval

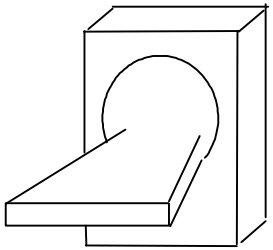


Query over metadata

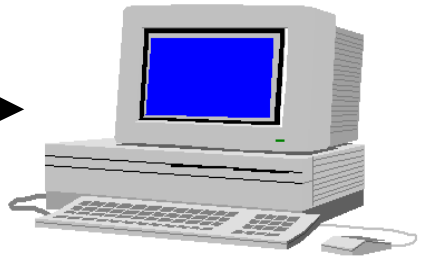


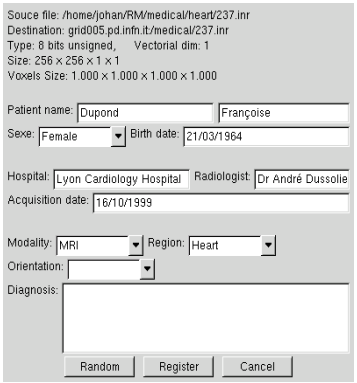
Query result

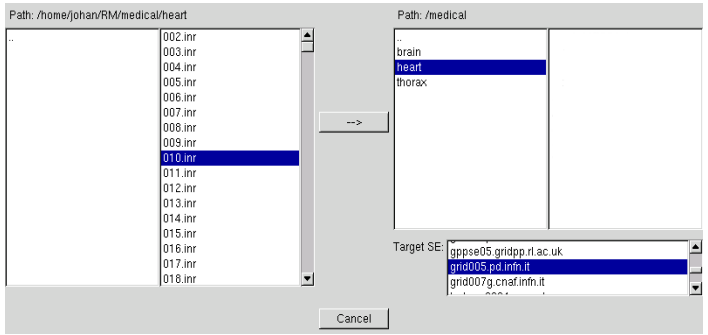
Image Registration



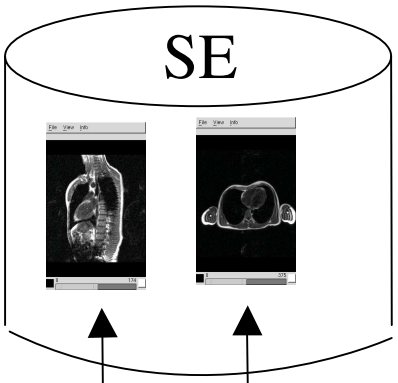
Imager





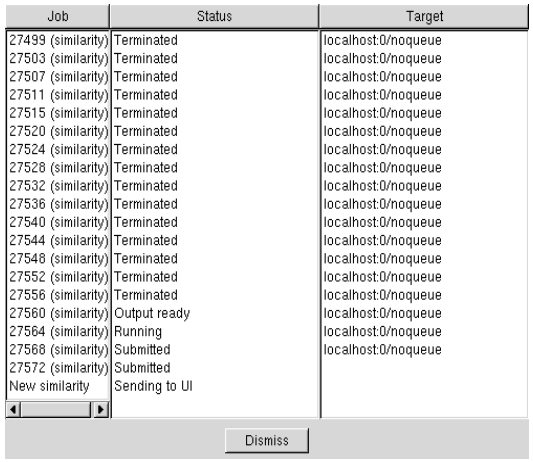


LFN	image	patient	hospital	...

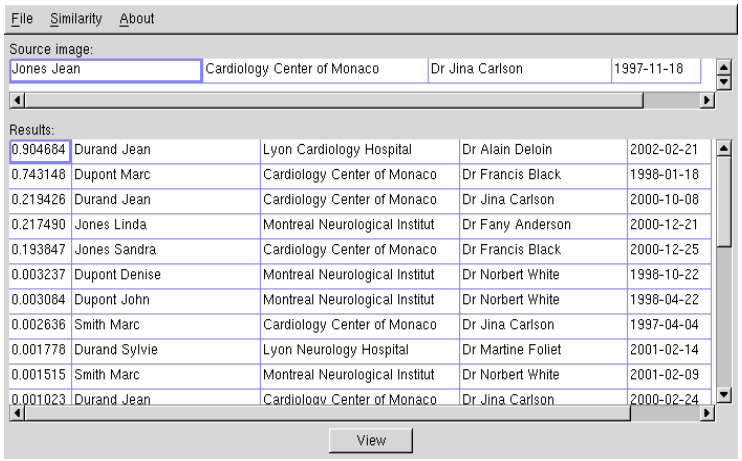


Similarity search

Similarity computation

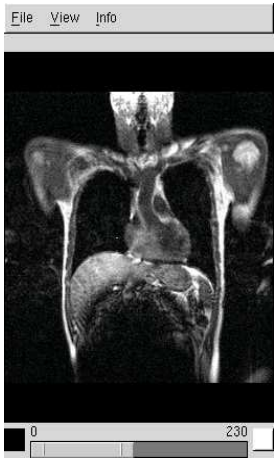


Job monitoring



Ranked list of images

Results visualization

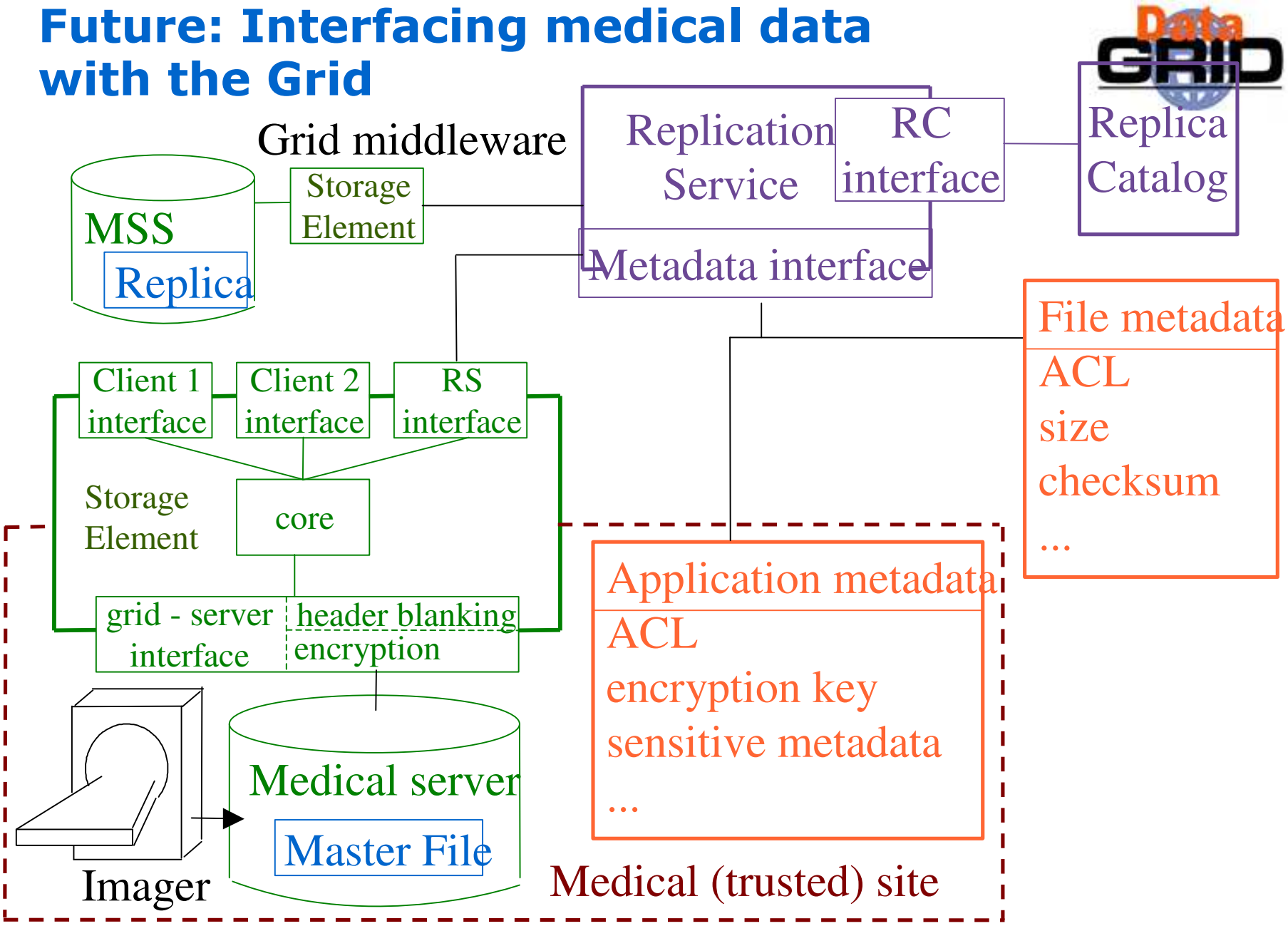


Source image

Most similar images

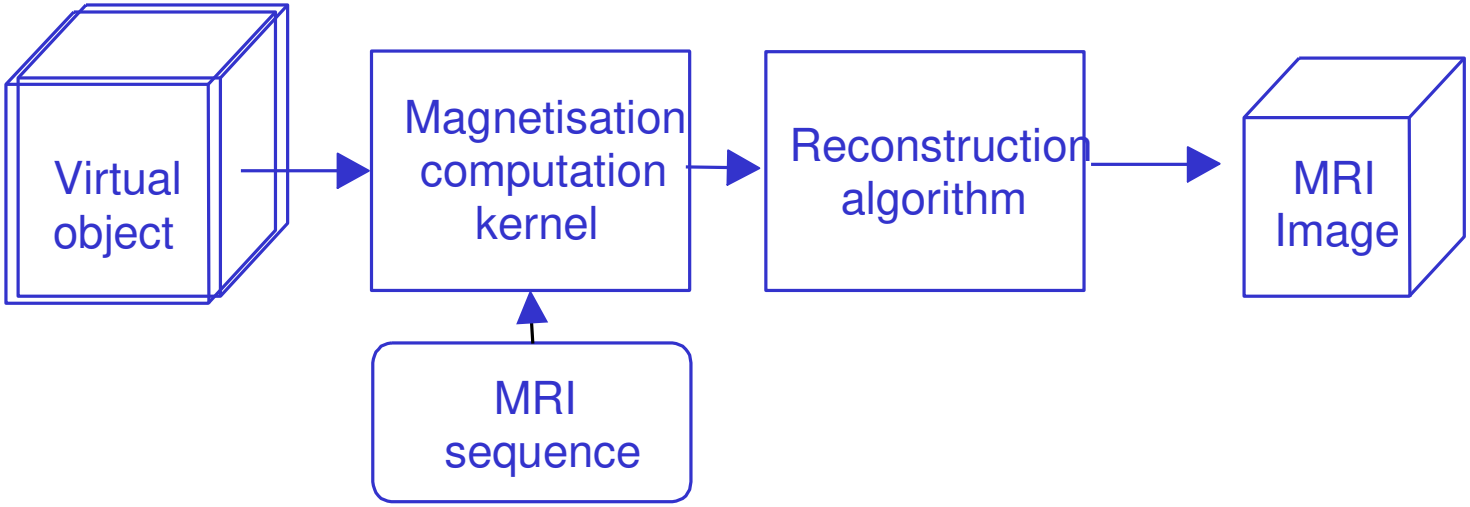
Low score images

Future: Interfacing medical data with the Grid



Parallel Processing

- Magnetic Resonance Images simulation using the grid



- 3 levels of parallelism:
  - Parallel isochromat computations
  - Multi-slice MRI computation
  - Parallel magnetization kernel

Summary

- Use Cases
  - High Energy Physics
  - Earth Observation
  - Biomedical Applications

Further Information

- High Energy Physics

  http://datagrid-wp8.web.cern.ch/DataGrid-WP8/

- Bio-Informatics

  http://marianne.in2p3.fr/datagrid/wp10/index.html

- Earth Observation

  http://styx.esrin.esa.it/grid/