EDG Applications

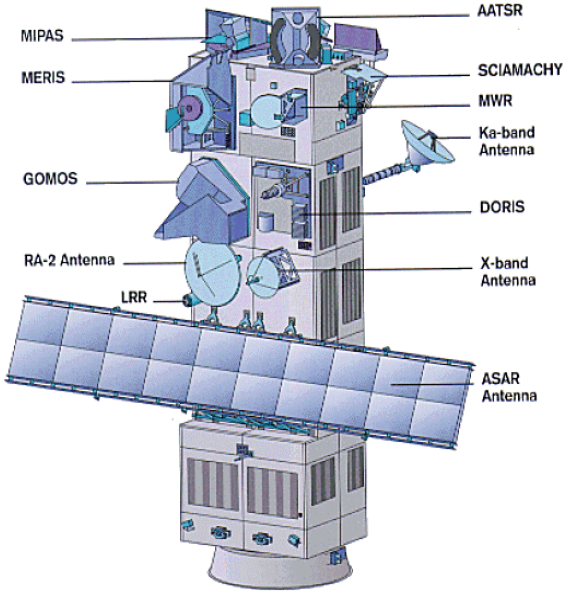
The European DataGrid Project Team

http://www.eu-datagrid.org

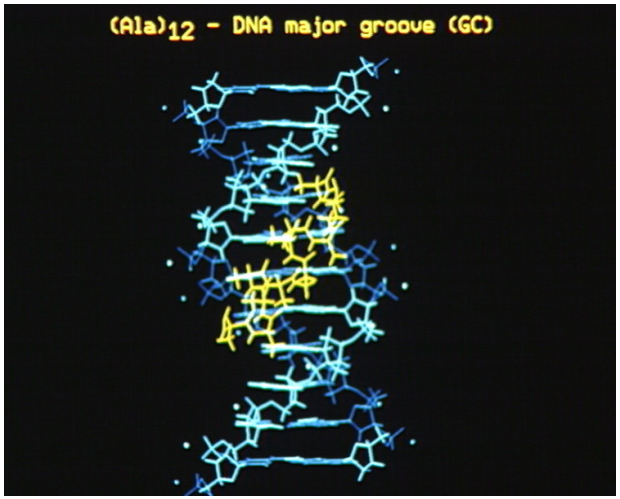
EDG Application Areas

High Energy Physics

Earth Observation Science Applications



Biomedical Applications

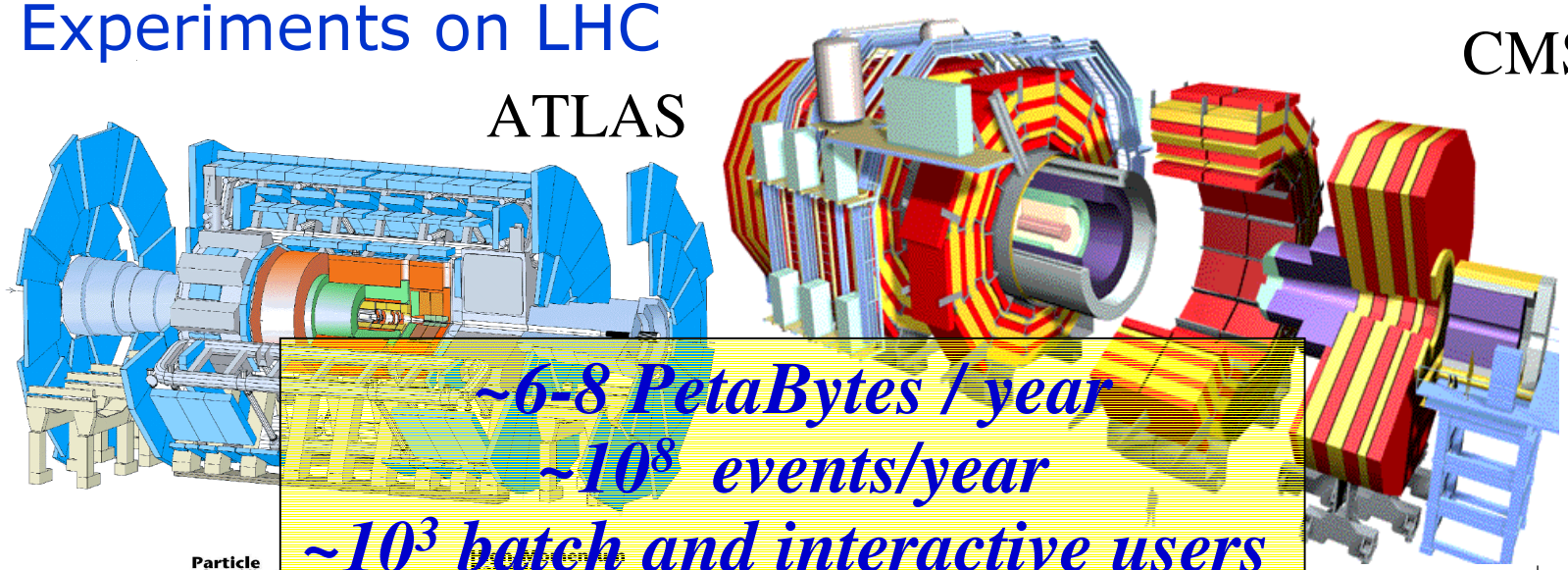


High Energy Physics

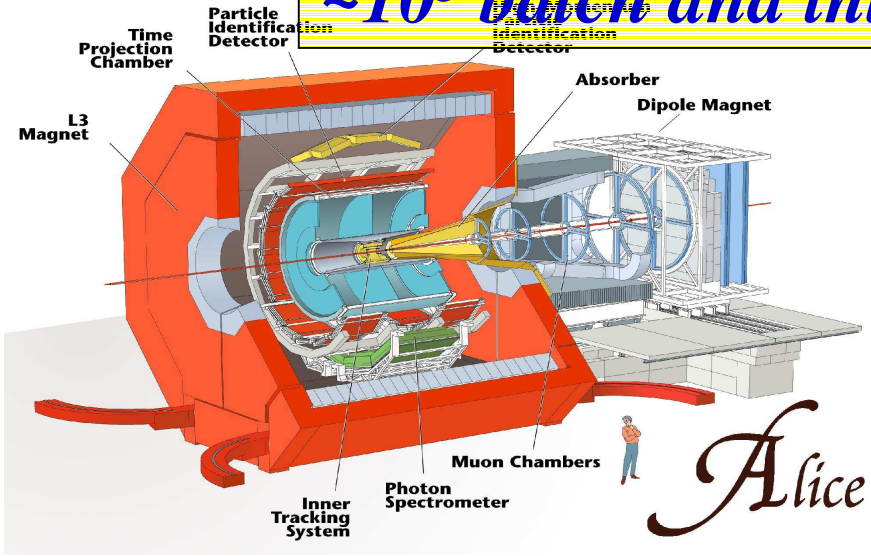
4 Experiments on LHC

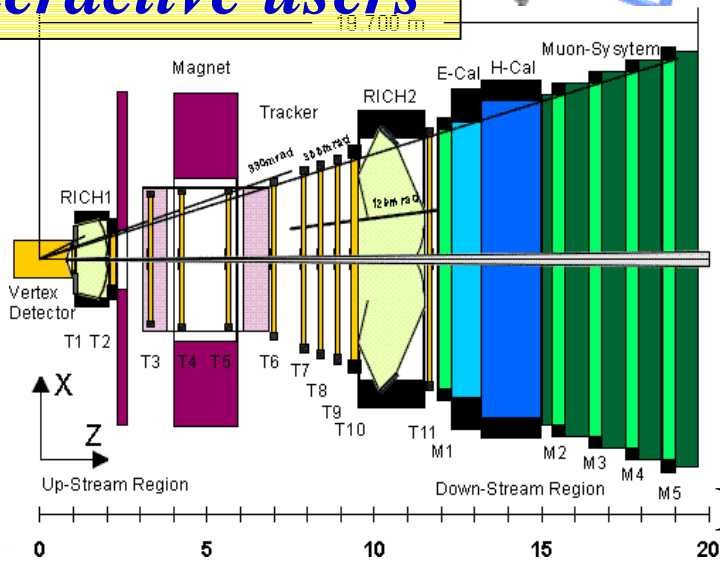
ATLAS

CMS

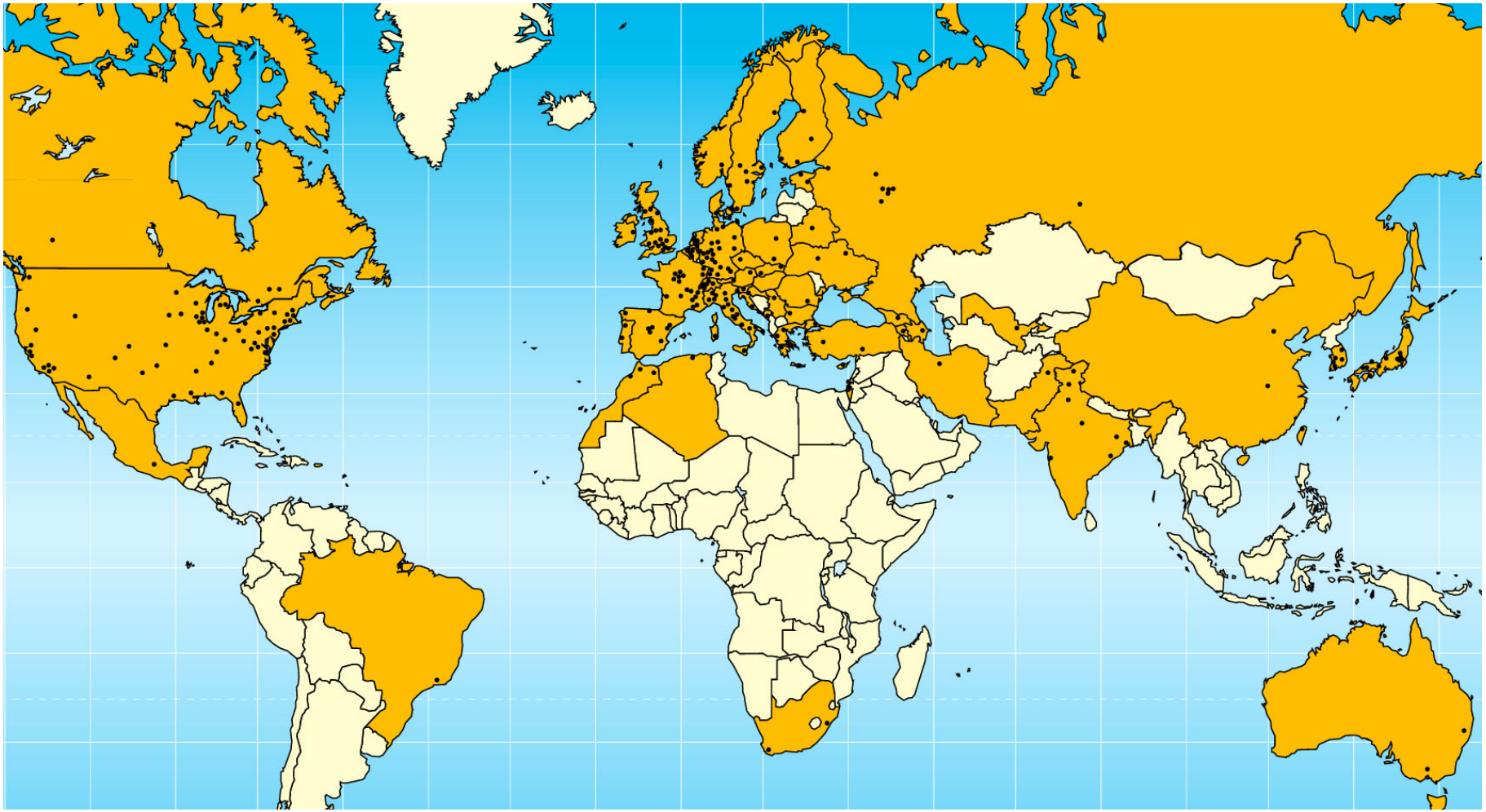


~6-8 PetaBytes / year
~10^8 events/year
~10^3 batch and interactive users



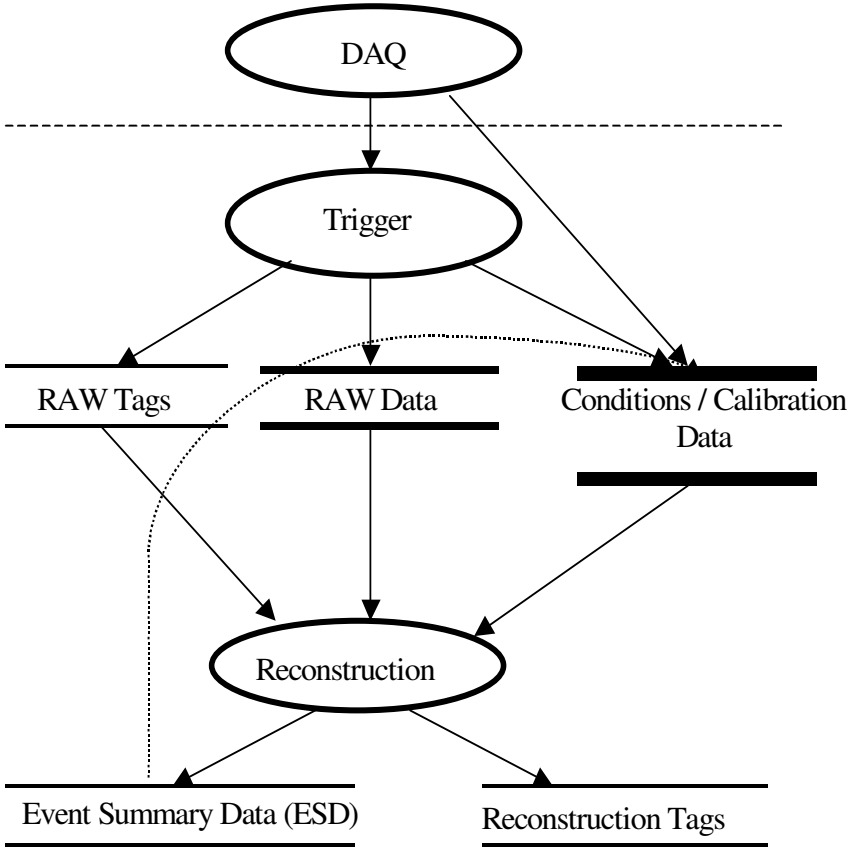


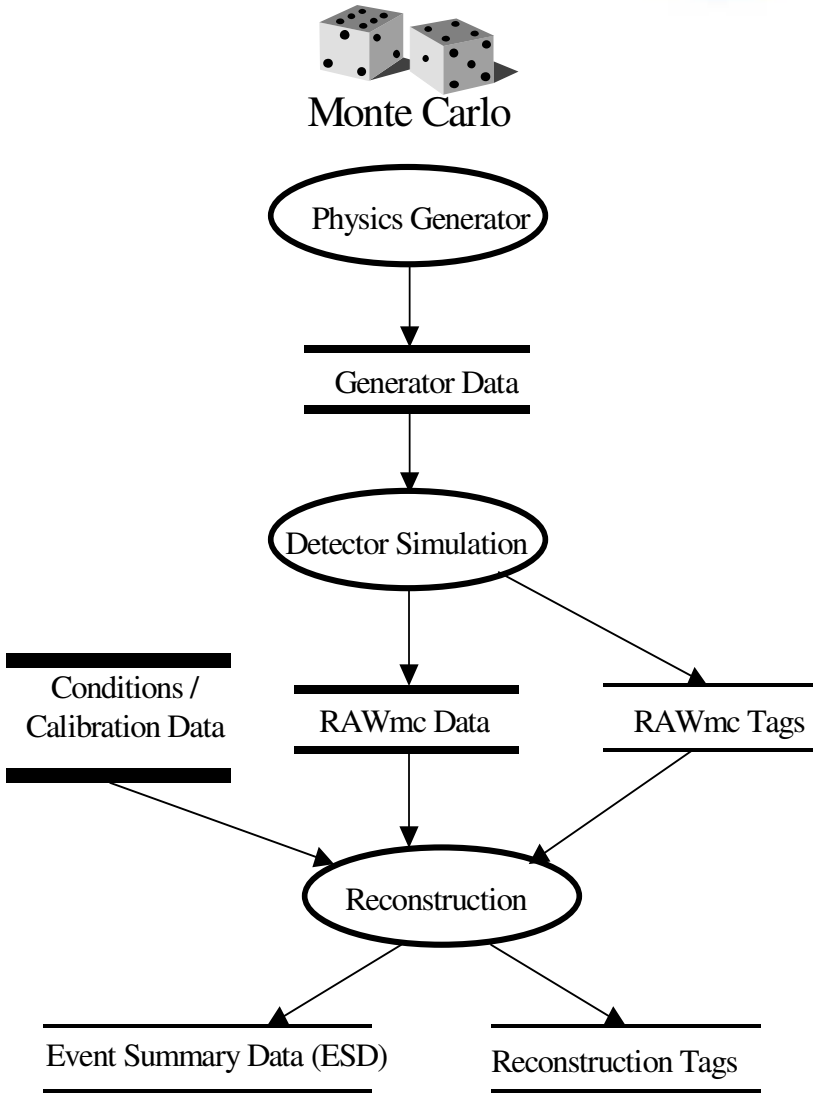
CERN's Network in the World



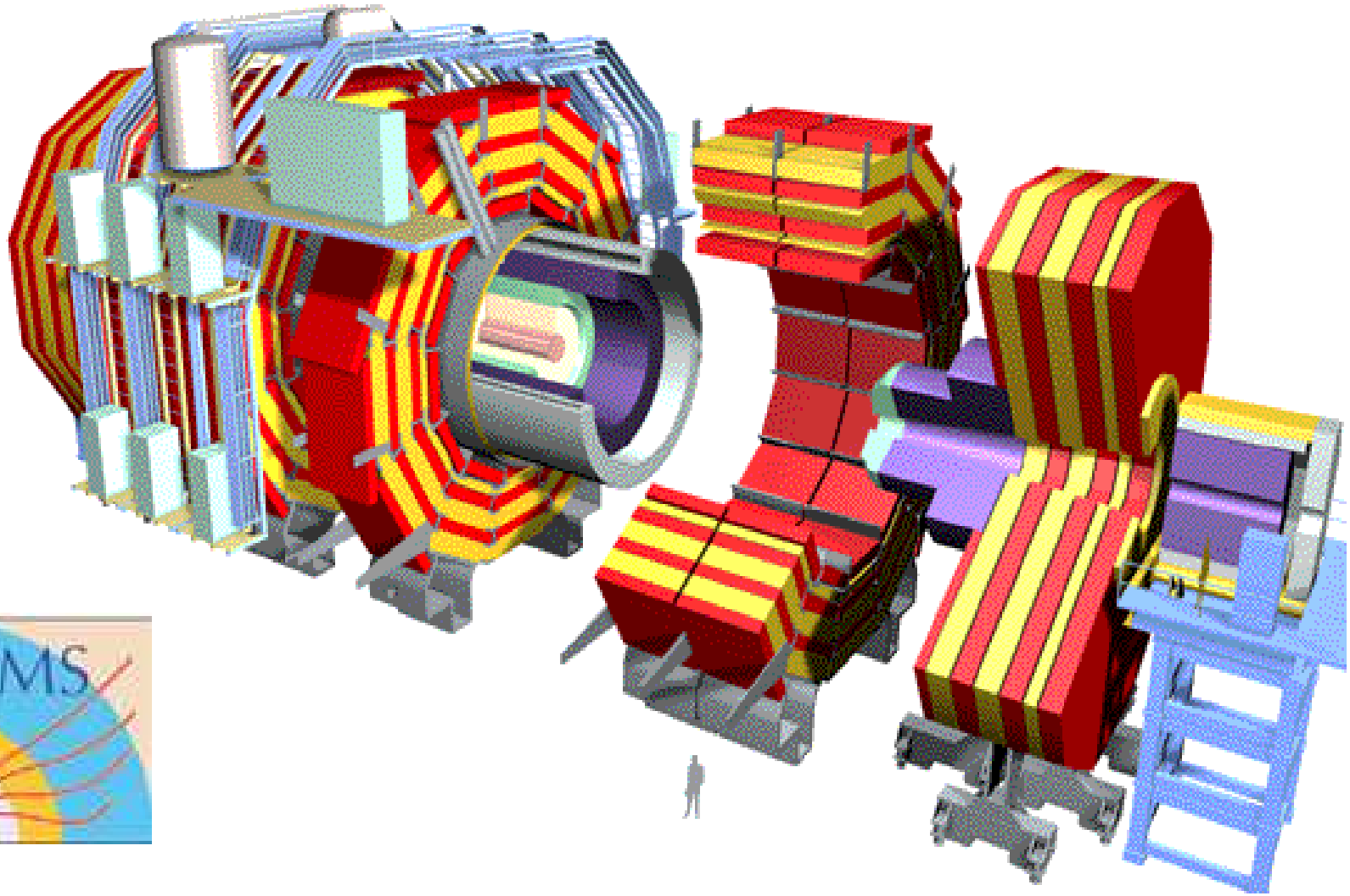
Europe: 267 institutes, 4603 users
Elsewhere: 208 institutes, 1632 users

Data Flow in LHC

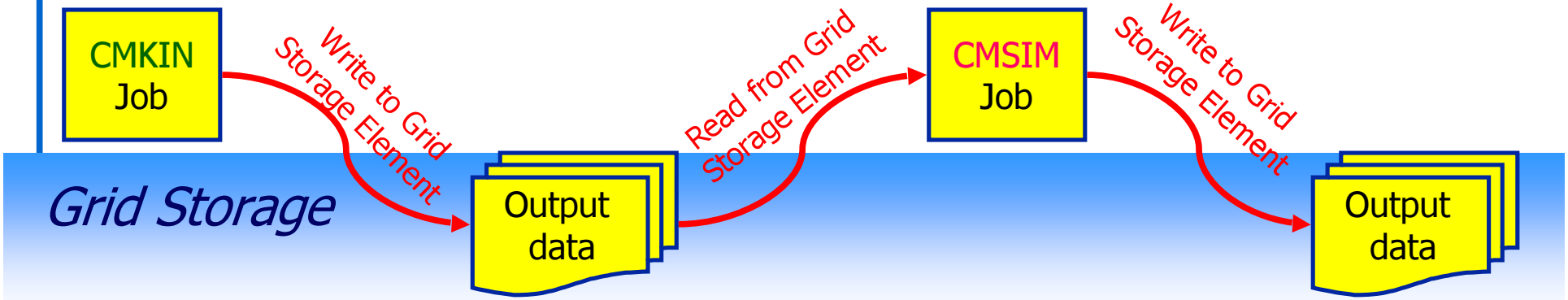




Example: CMS Monte Carlo Production



CMS jobs description



CMKIN : MC Generation of the proton-proton interaction for a physics channel (dataset)

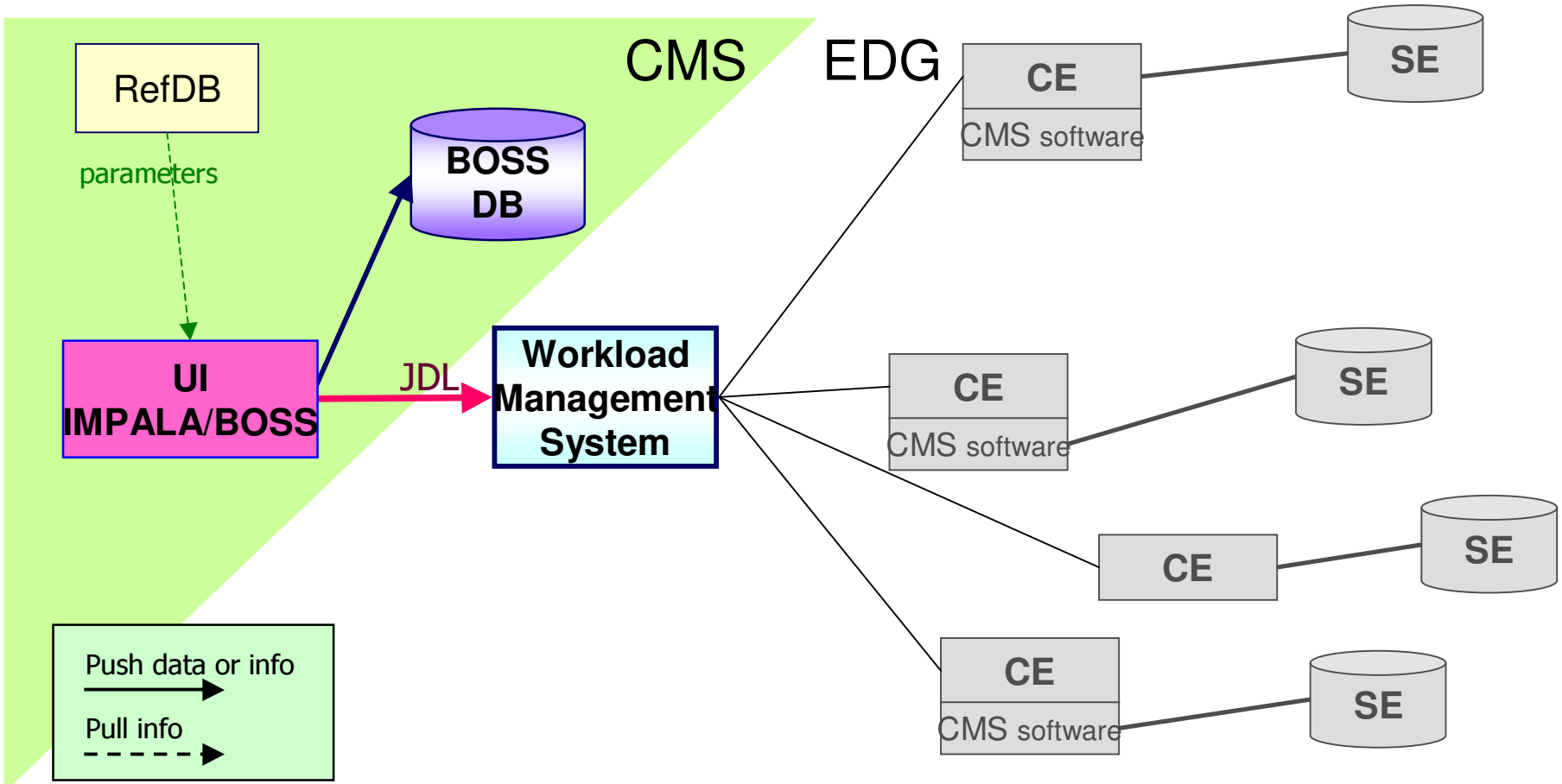
CMSIM: Detailed simulation of the CMS detector, processing the data produced during the CMKIN step

	size/event	time*/event
CMKIN	~ 0.05MB	~ 0.4-0.5 sec
CMSIM	~ 1.8 MB	~ 6 min

* PIII 1GHz 512MB ~ 46.8 SI95

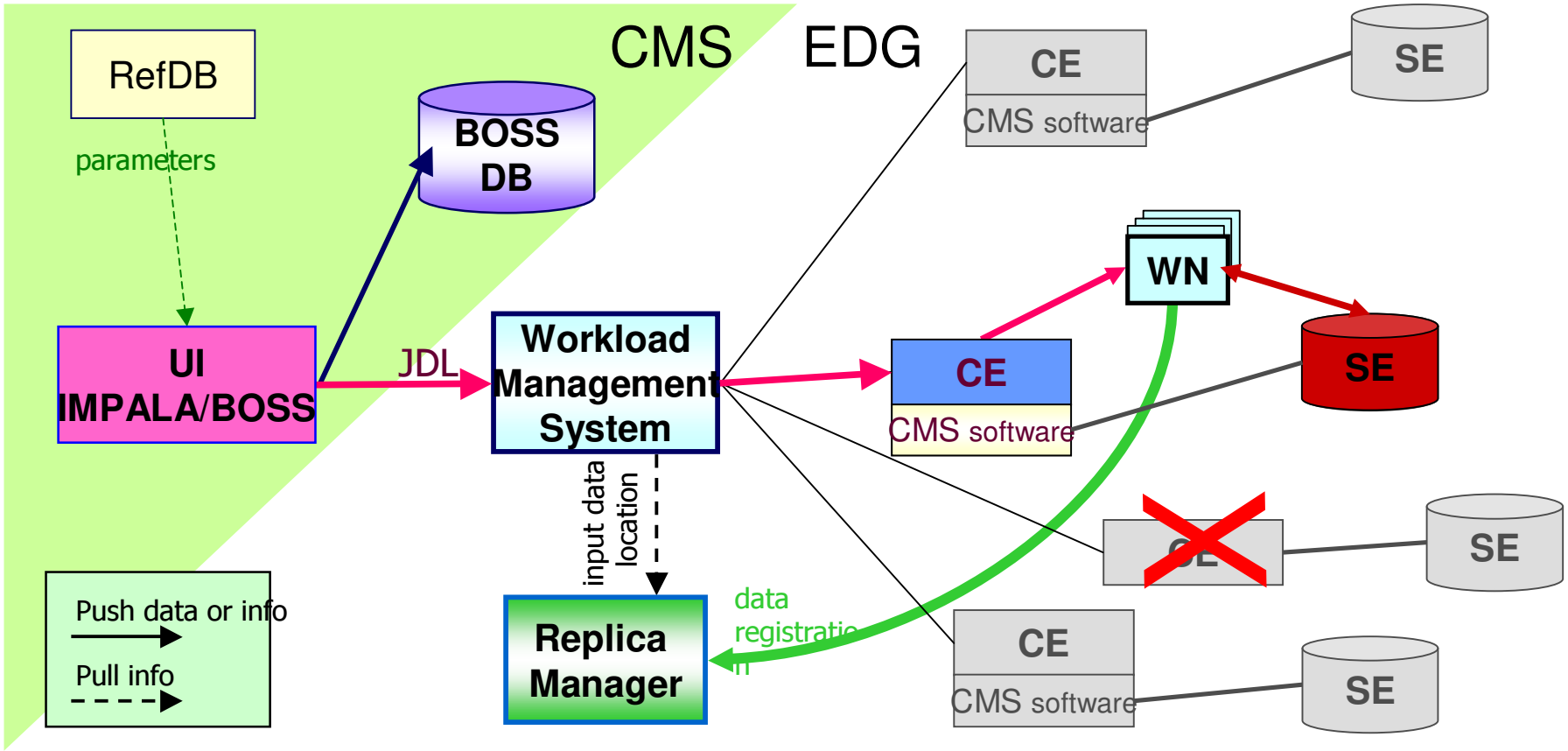
CMS production components interfaced to EDG middleware

- Production is managed from the EDG User Interface with IMPALA/BOSS
- CMS Virtual Organization server at NIKHEF (Amsterdam)



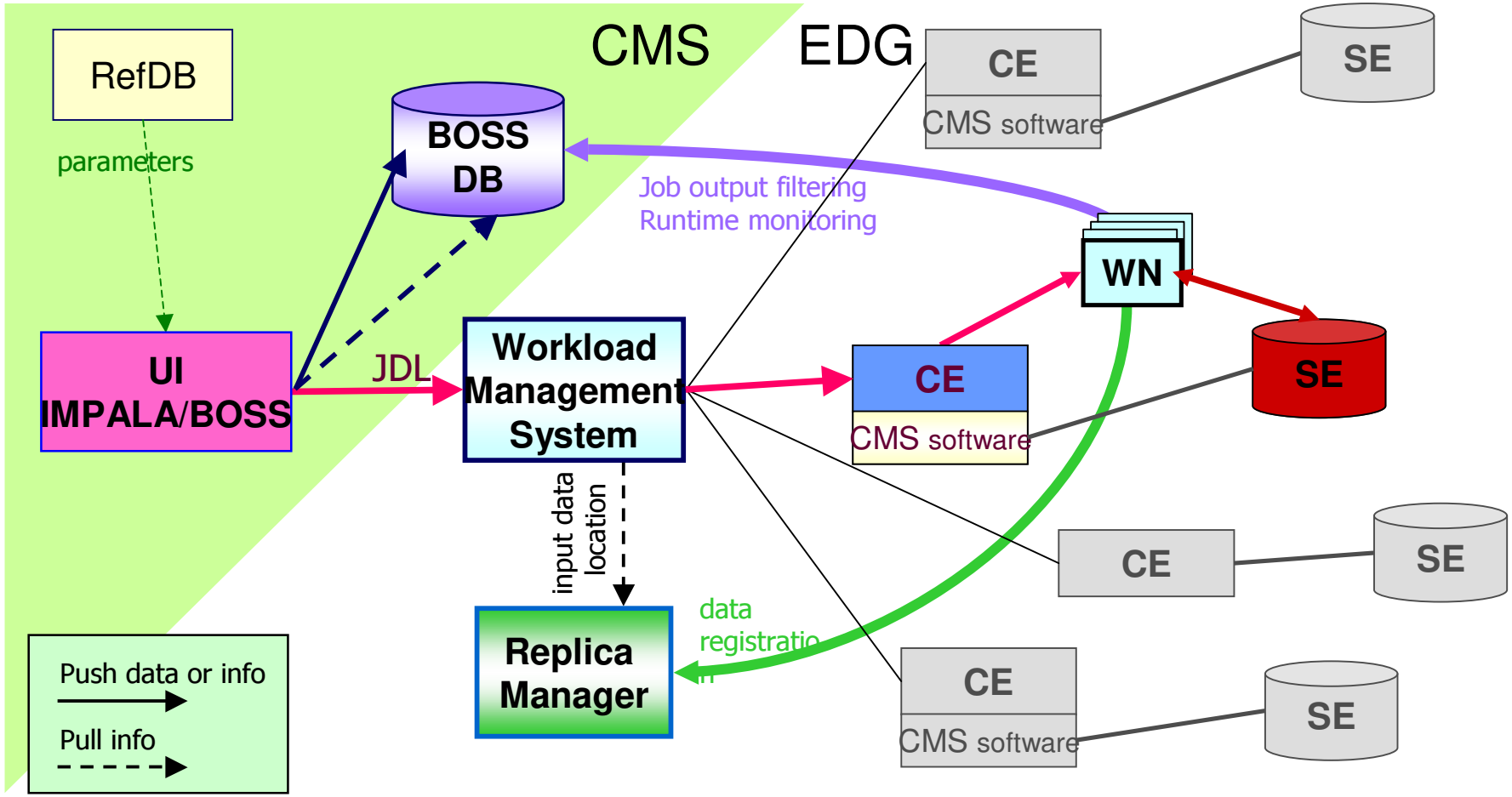
CMS production components interfaced to EDG middleware

- CMKIN jobs running on all EDG Testbed sites with CMS software installed
- CMSIM jobs running on CE close to the input data
- produced data: scripts for batch replication to a dedicated SE



CMS production components interfaced to EDG middleware

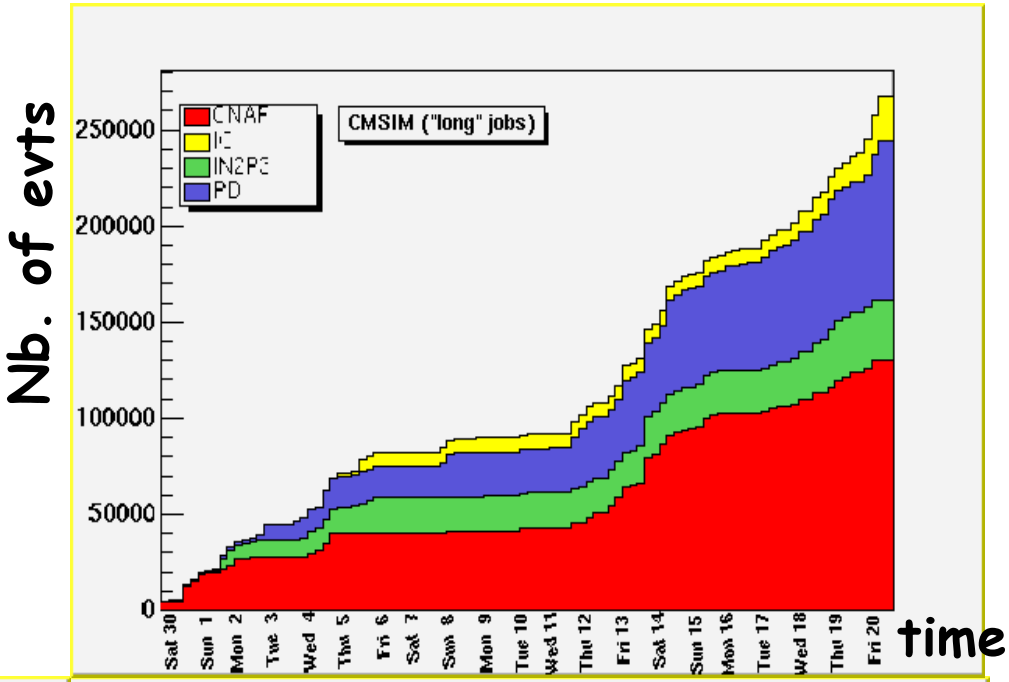
- Job monitoring and bookkeeping: BOSS DBs, EDG Logging & Bookkeeping service

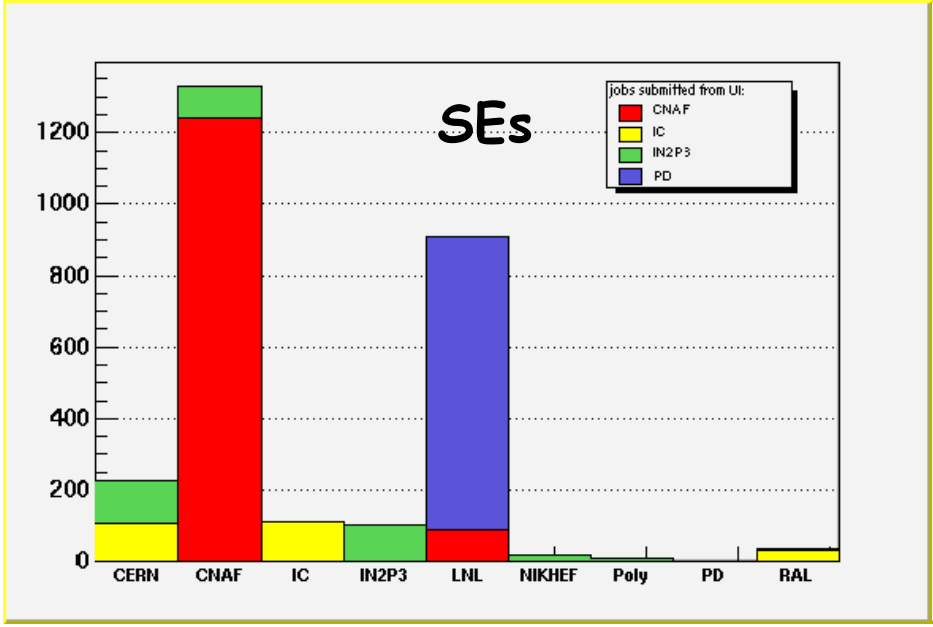


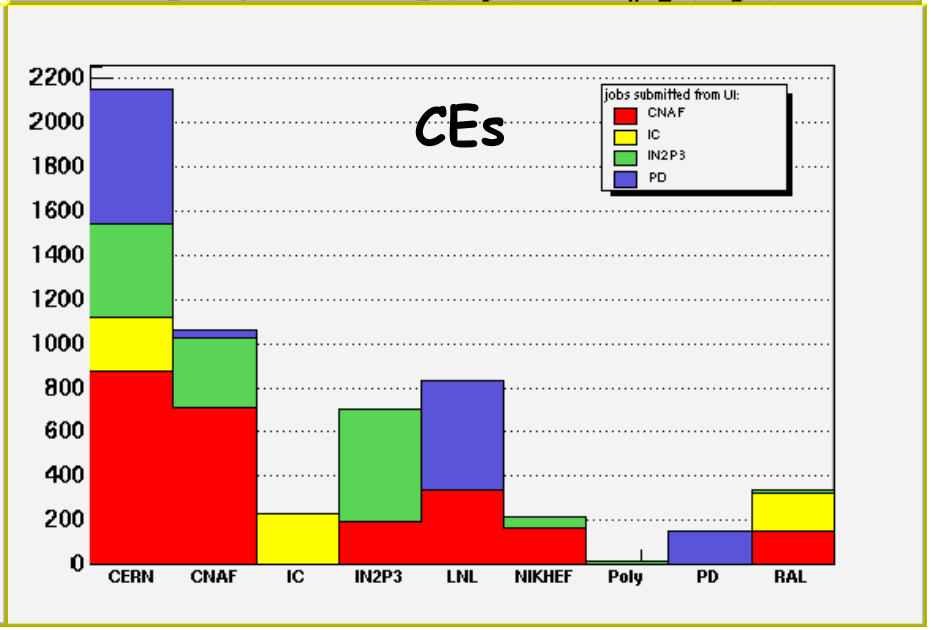
CMS use of the system (Statistics)

Events Production within EDG is part of the Official CMS production

http://cmsdoc.cern.ch/cms/production/www/html/general/







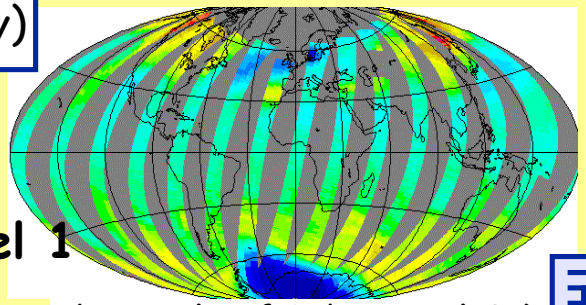
Summary of CMS work and plans for use of EDG middleware

- RESULTS
  - We can distribute and run CMS s/w in the EDG environment
  - We have generated ~250K events for physics with ~10000 jobs in 3 week period
- OBSERVATIONS and PLANNING for the future
  - We were able to quickly add new sites to provide extra resources
  - There was a fast turnaround in bug fixing and installing new software
  - The stress test was labor intensive (since software was developing)
  - Release EDG 2.0 should fix the major problems and allow for enhanced scalability,and we look forward to evaluating it and using it in our Data Challenge work

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)

LIDAR
data (7 stations,
2.5MB per month)

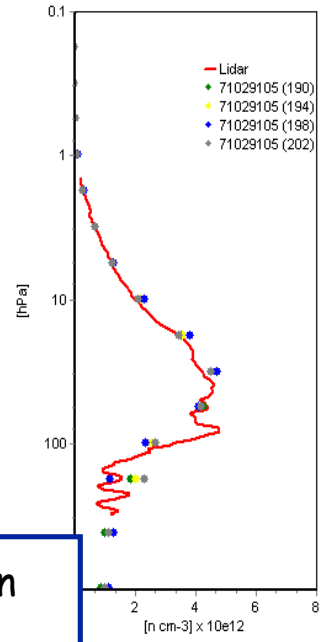
ESA(IT) KNMI(NL)
Processing of raw GOME
data to ozone profiles.
2 alternative algorithms
~28000 profiles/day

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

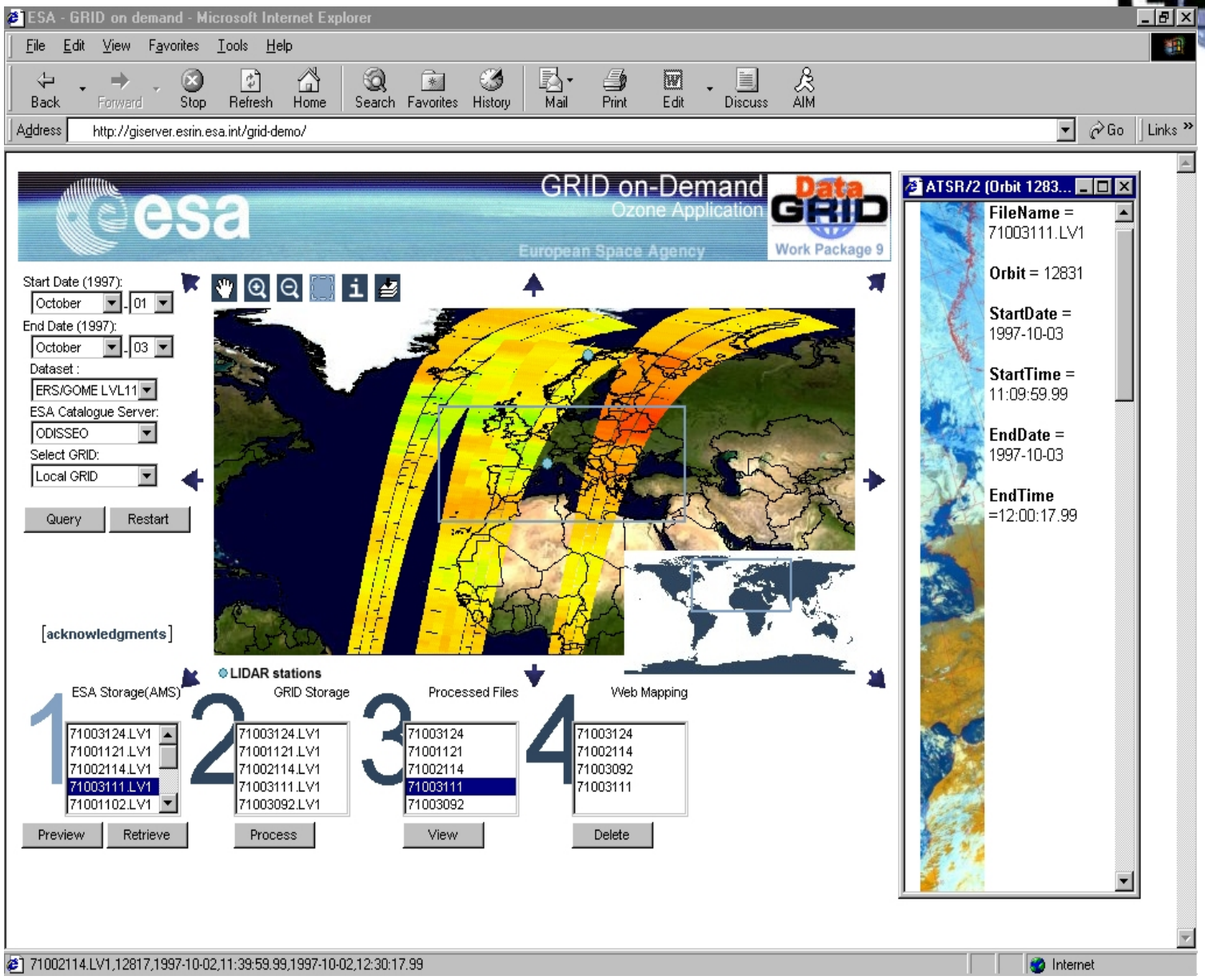
Level 2

DataGrid environment

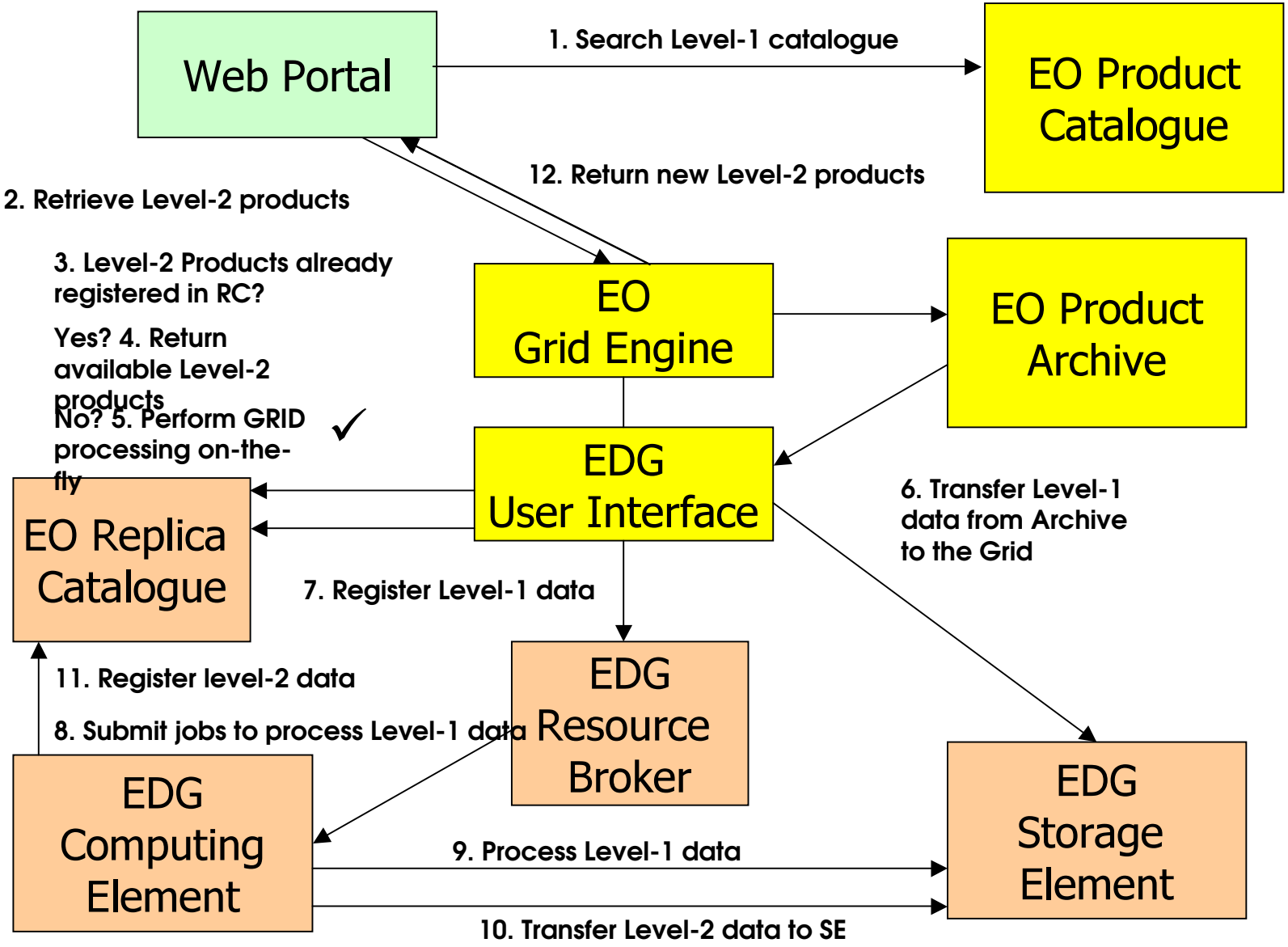
Visualization
& Analyze



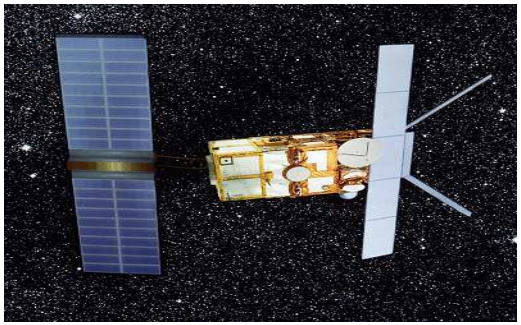
EO WebMap Portal



Processing Sequence



GOME Ozone Profile Validation



ERS/GOME satellite

50 km

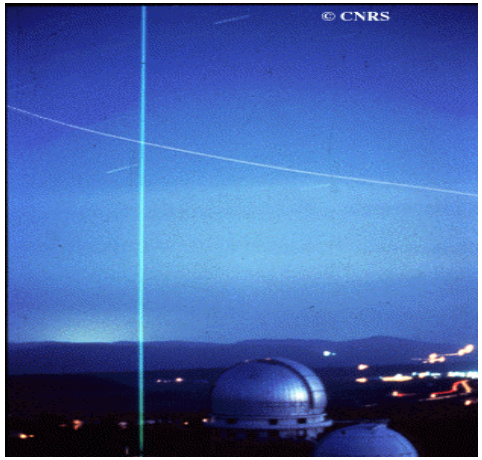
10 km

OZONE LAYER

✓ Goals of the DataGrid application

validate satellite data with all ground based data available in an easy way:

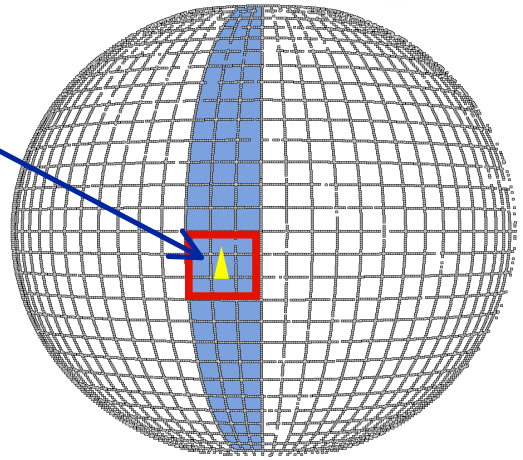
- Comparison of ozone profiles provided by satellite with lidar data in different locations and times (see the web portal)
- Statistical comparison and analysis in order to improve algorithms.



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

Queries and data information retrieval from the Lidar metadata catalogue

3

Submission of the Job in the GRID

GRID Portal

1

Lidar data catalogue

GRID

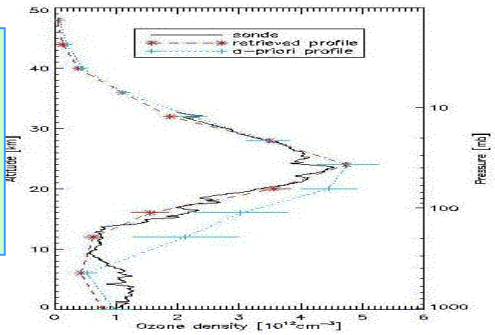
Storage Elements with Lidar data

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.

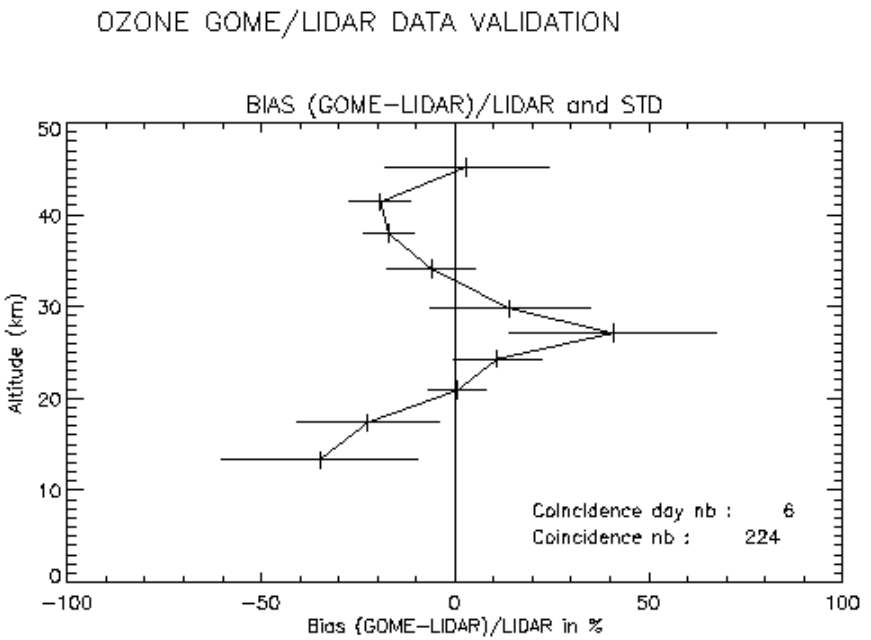
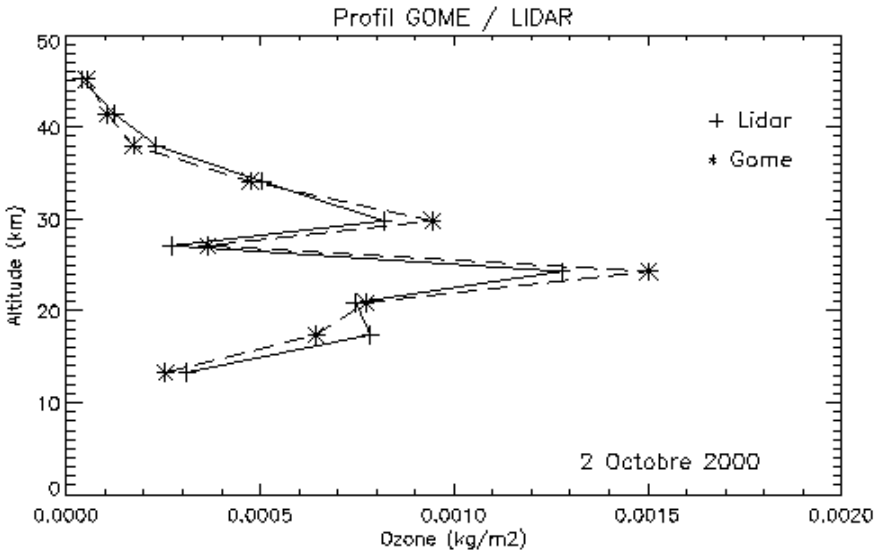


Figure 2 :

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



Perspectives for Biomedical Applications

- Grids open new perspectives in large scale genomics analysis
  - Complete genome annotation
  - Cross-genomes analysis
  - Data mining on distributed databases
  - Pipelining of huge automatic bio-informatics analysis
- Medical image processing
  - Large databases processing
  - Anatomy and physiology modeling
  - Epidemiological studies

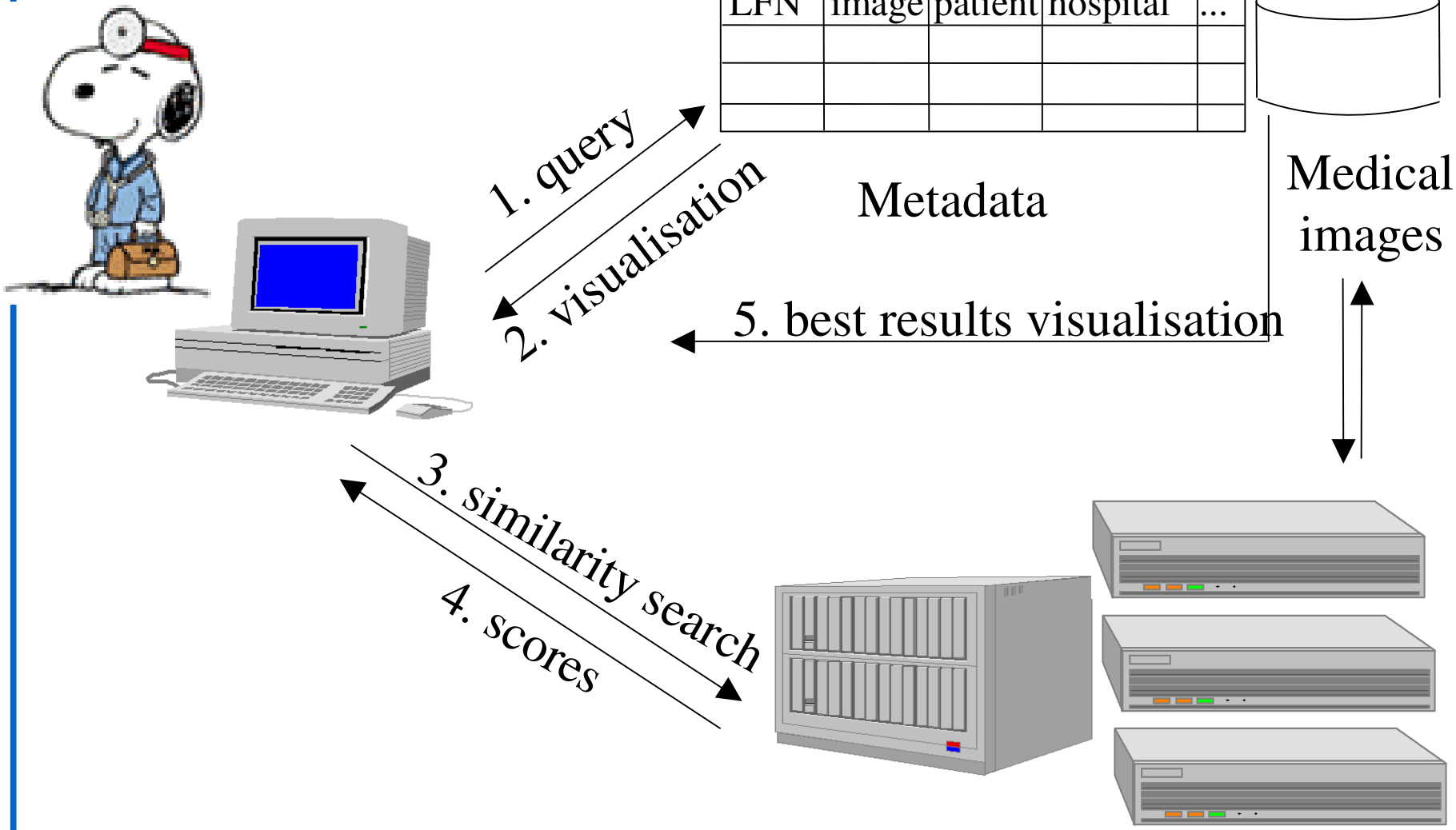
Biomedical Applications

Applications deployed
Applications tested on EDG
Applications under preparation

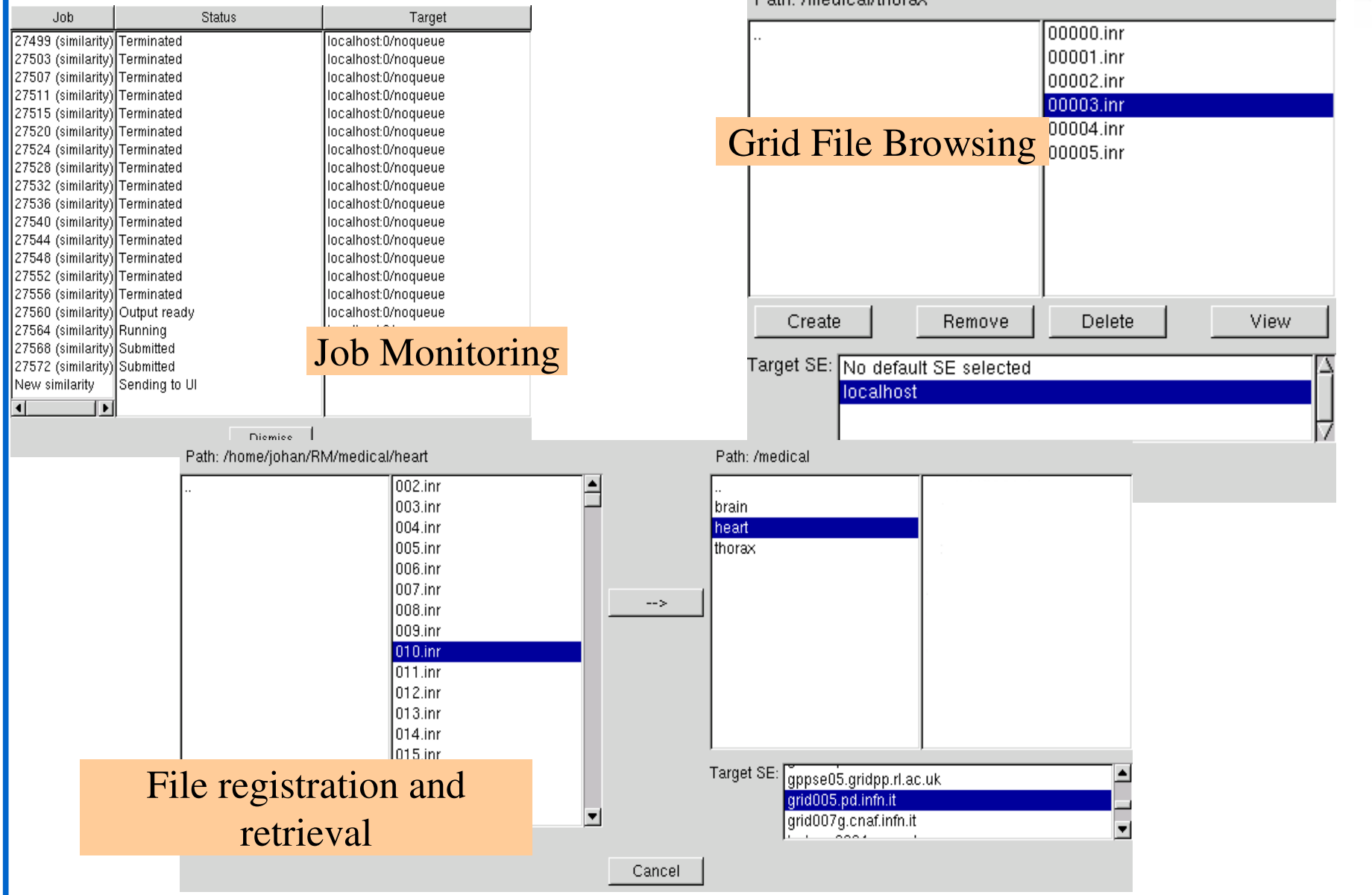
- **Bio-informatics**
  - Phylogenetics : BBE Lyon (T. Sylvestre)
  - Search for primers : Centrale Paris (K. Kurata)
  - Statistical genetics : CNG Evry (N. Margetic)
  - Bio-informatics web portal : IBCP (C. Blanchet)
  - Parasitology : LBP Clermont, Univ B. Pascal (N. Jacq)
  - Data-mining on DNA 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

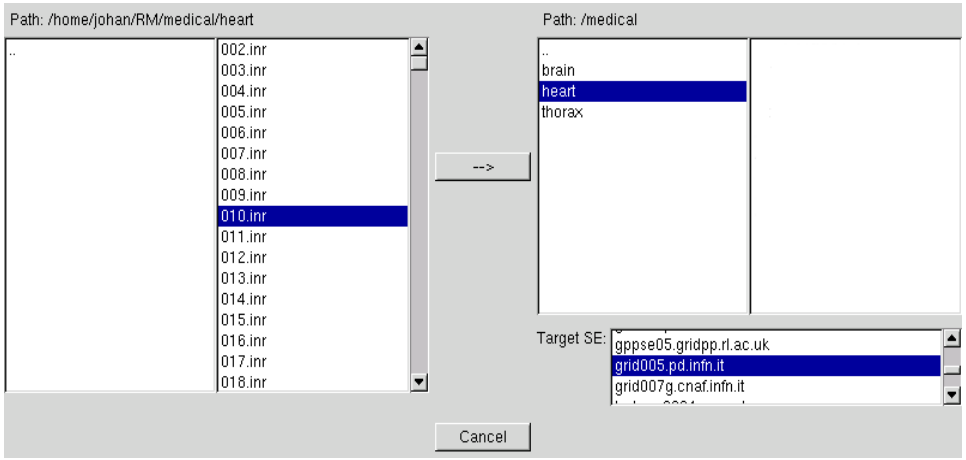


Graphic layer



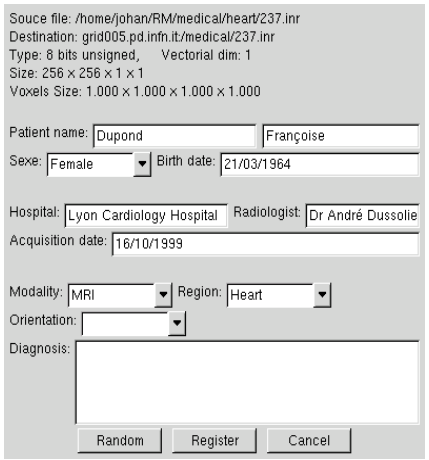
Graphical Interfaces

Image registration



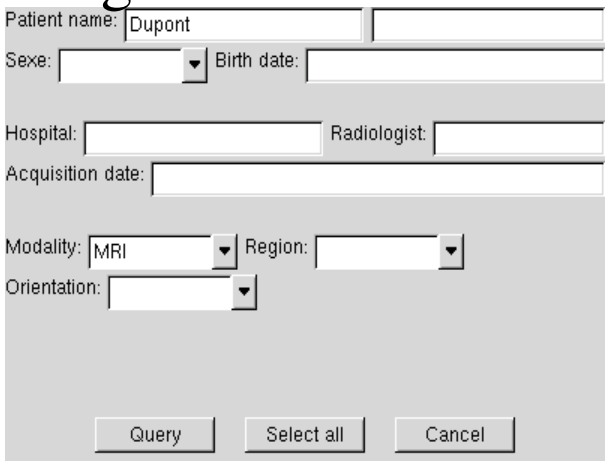
Local files

Grid files

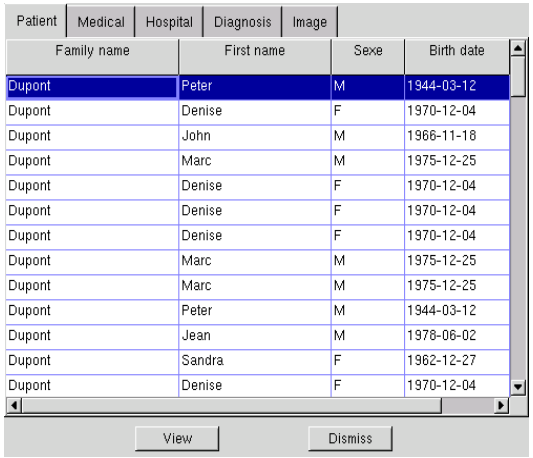


Metadata

Image retrieval

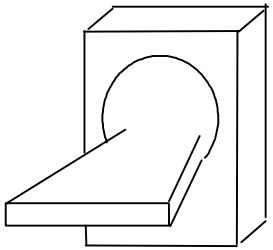


Query over metadata

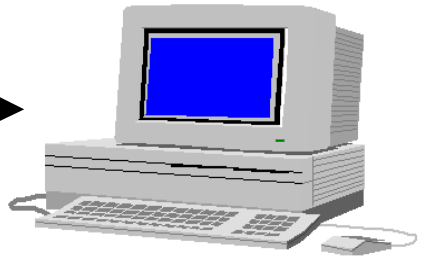


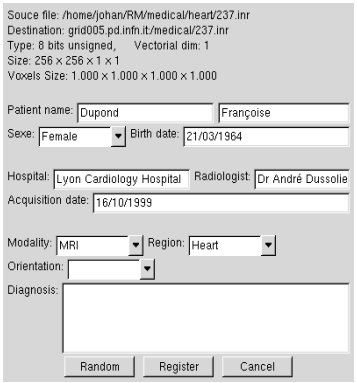
Query result

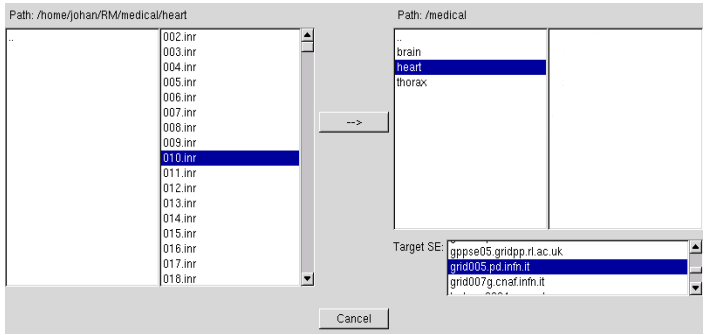
Image Registration



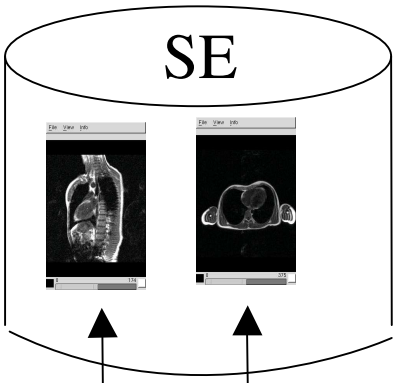
Imager





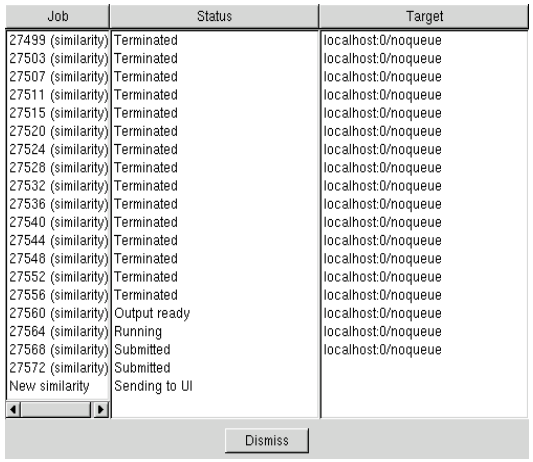


LFN	image	patient	hospital	...

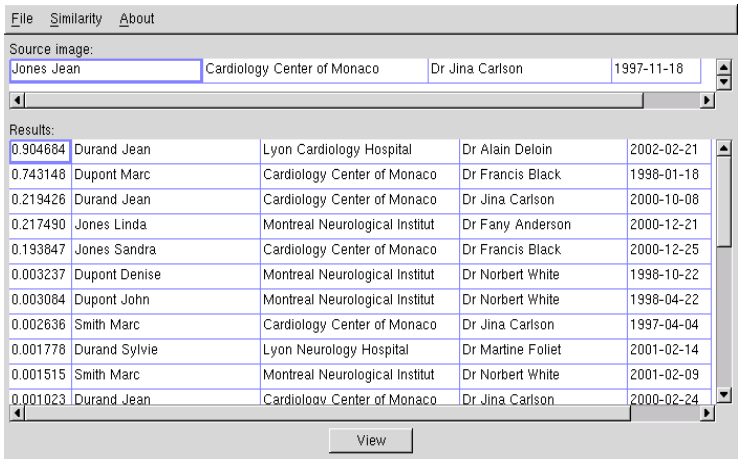


Similarity search

Similarity computation

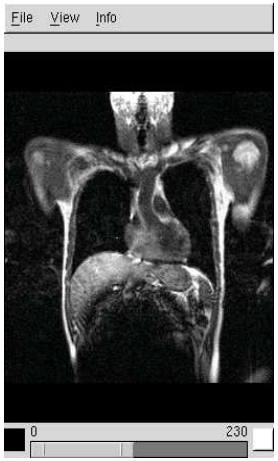


Job monitoring



Ranked list of images

Results visualization

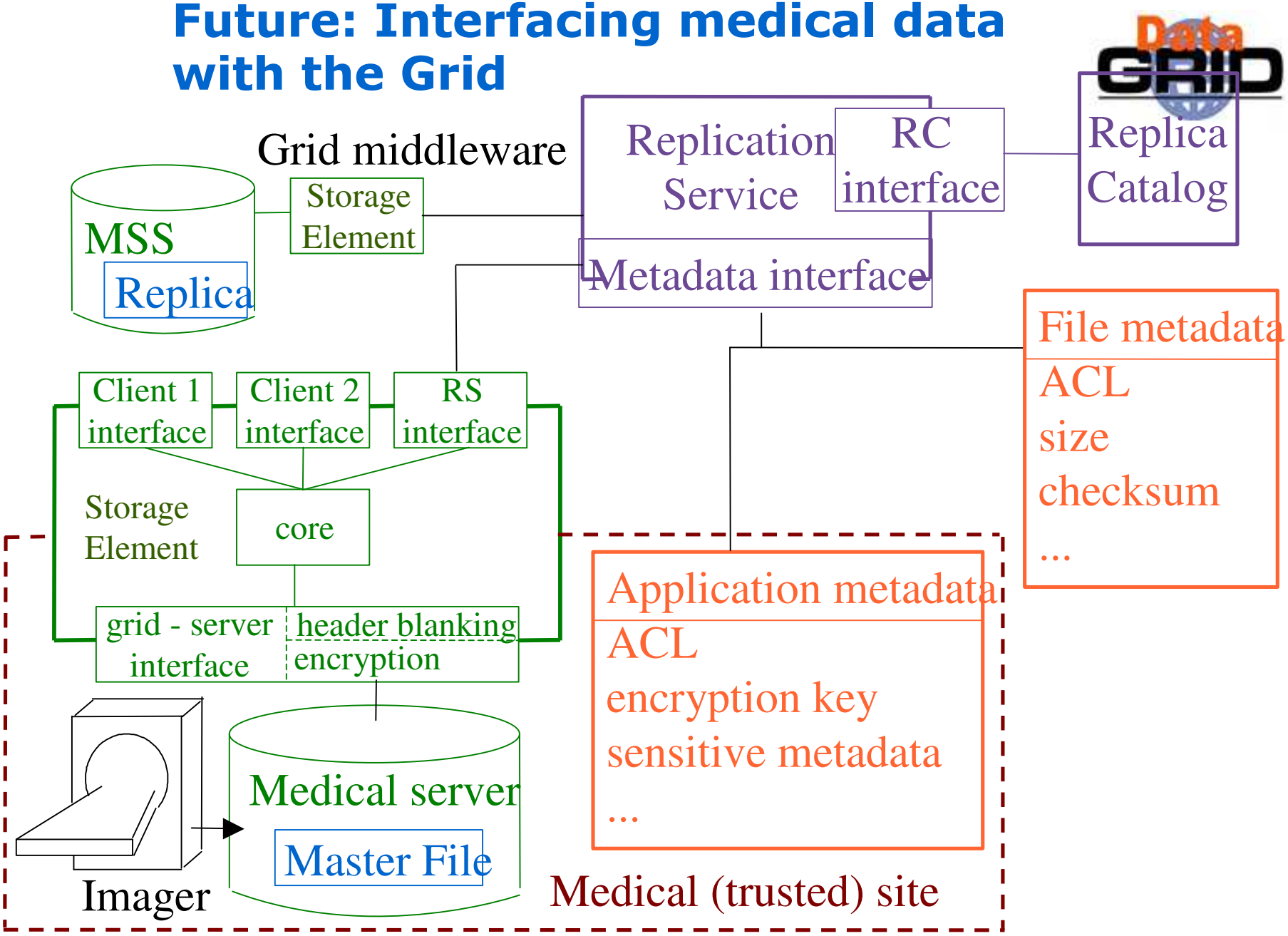


Source image

Most similar images

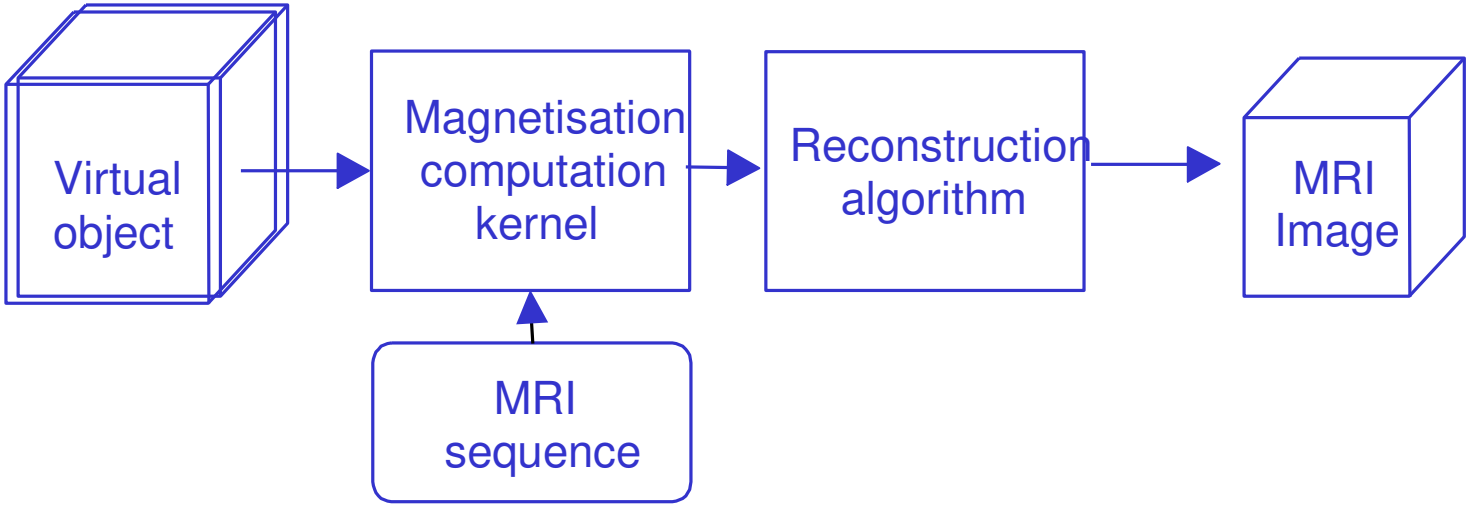
Low score images

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/

The End