Applications and the Grid

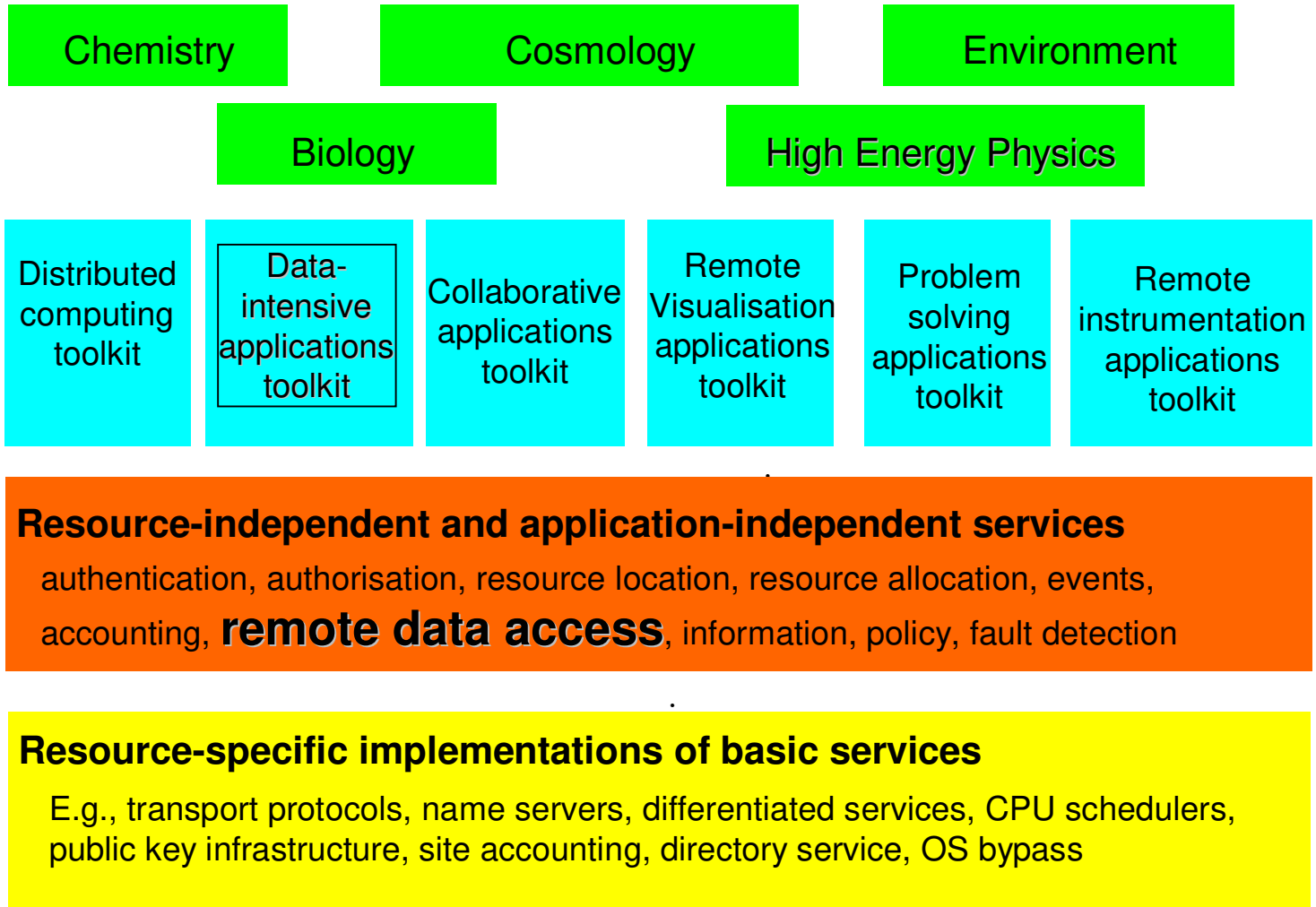
The European DataGrid Project Team

http://www.eu-datagrid.org

Applications and the Grid

- **An applications view of the the Grid**
- **Current models for use of the Grid in**
  - **High Energy Physics (WP8)**
    - **Initially: Atlas, Alice, CMS, LHCb**
    - **Now also Babar, D0….**
  - **Biomedical Applications (WP10)**
  - **Earth Observation Applications (WP9)**
- **Acknowledgments and references**

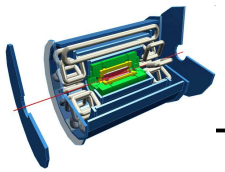
GRID Services: The Overview



What all applications want from the Grid (the basics)

- A homogeneous way of looking at a 'virtual computing lab' made up of heterogeneous resources as part of a VO(Virtual Organisation) which manages the allocation of resources to authenticated and authorised users
  - A uniform way of 'logging on' to the Grid
  - Basic functions for job submission, data management and monitoring
    - Ability to obtain resources (services) satisfying user requirements for data, CPU, software, turnaround……

LHC Computing (a hierachical view of grid…this has evolved to a 'cloud' view)



~PBytes/sec

Online System

~100 MBytes/sec

1 TIPS = 25,000 SpecInt95

PC (1999) = ~15 SpecInt95

- One bunch crossing per 25 ns
- 100 triggers per second
- Each event is ~1 Mbyte

Offline Farm
~20 TIPS

~100 MBytes/sec

~ Gbits/sec
or Air Freight

Tier 0

CERN Computer
Centre >20 TIPS

HPSS

Tier 1

US Regional Centre

Italian Regional Centre

French Regional Centre

RAL Regional Centre

• • •

Tier 2

ScotGRID++
~1 TIPS

Tier2 Centre
~1 TIPS

Centre
TIPS

Centre
TIPS

~Gbits/sec

Tier 3

Institute
~0.25TIPS

tute

stitute

Institute

Physics data cache

100 - 1000
Mbits/sec

Tier 4

Workstations

Physicists work on analysis "channels"

Each institute has ~10 physicists working on one or more channels

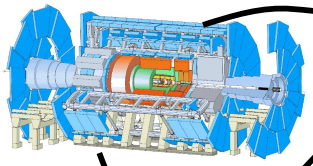
Data for these channels should be cached by the institute server

LHC Computing Requirements

Needed during a year of LHC operations		
Tape	Disk	CPU
29'400 TB	9'600 TB	$6.2 * 10^6$ SI95
In today's units:		
60 STK Silos	160'000 60GB disks	150'000 800 MHz CPUs

LHC Computing Review, CERN/LHCC/2001-004

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

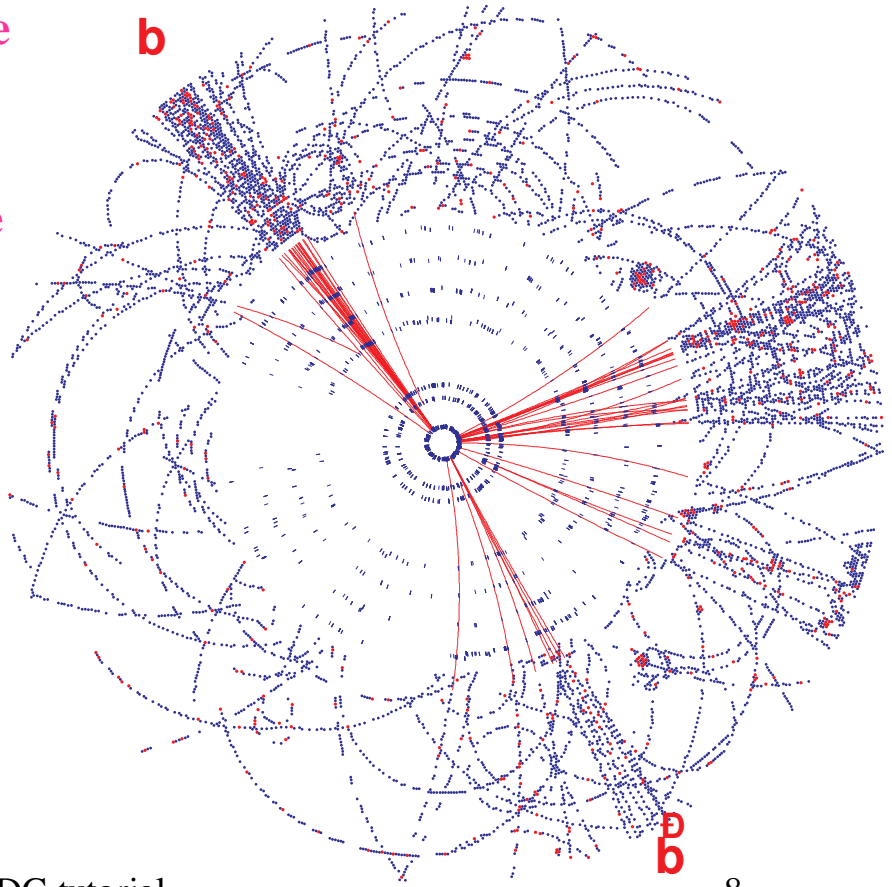
interactive physics analysis

les.robertson@cern.ch

CERN

HEP Data Analysis and Datasets

- **Raw data (RAW)** **~ 1 MByte**
  - **hits, pulse heights**
- **Reconstructed data (ESD) ~ 100 kByte**
  - **tracks, clusters…**
- **Analysis Objects (AOD)** **~ 10 kByte**
  - **Physics Objects**
  - **Summarized**
  - **Organized by physics topic**
- **Reduced AODs(TAGs)** **~1 kByte**
- **histograms, statistical data on collections of events**



HEP Data Analysis –processing patterns

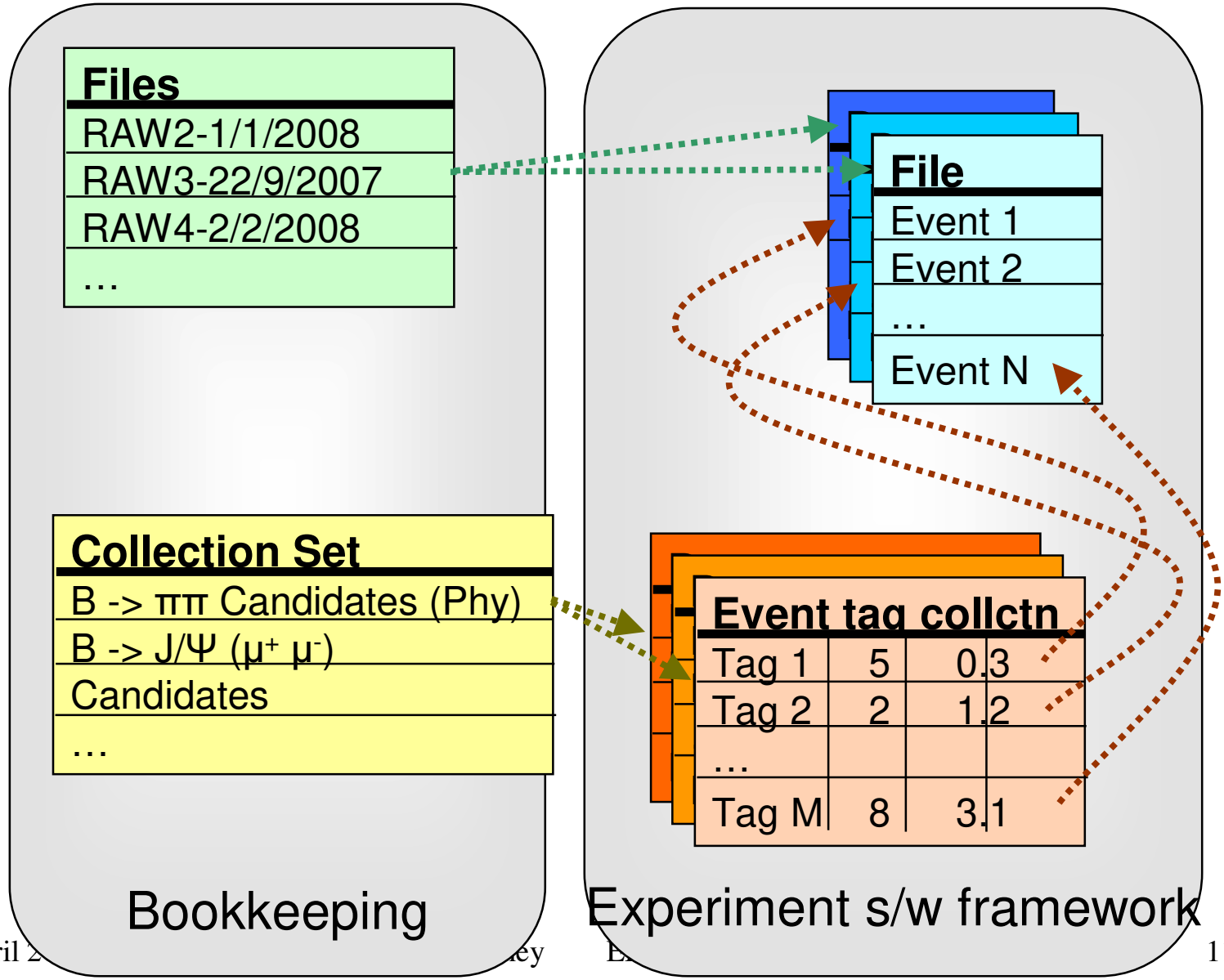
- **Processing fundamentally parallel due to independent nature of 'events'**
  - **So have concepts of splitting and merging**
  - **Processing organised into 'jobs' which process N events**
    - **(e.g. simulation job organised in groups of ~500 events which takes ~ day to complete on one node)**
      - **A processing for 10**6 events would then involve 2,000 jobs merging into total set of 2 Tbyte**
- **Production processing is planned by experiment and physics group data managers(this will vary from expt to expt)**
  - **Reconstruction processing (1-3 times a year of 10**9 events)**
  - **Physics group processing (? 1/month). Produce ~10**7 AOD+TAG**
  - **This may be distributed in several centres**

Processing Patterns(2)

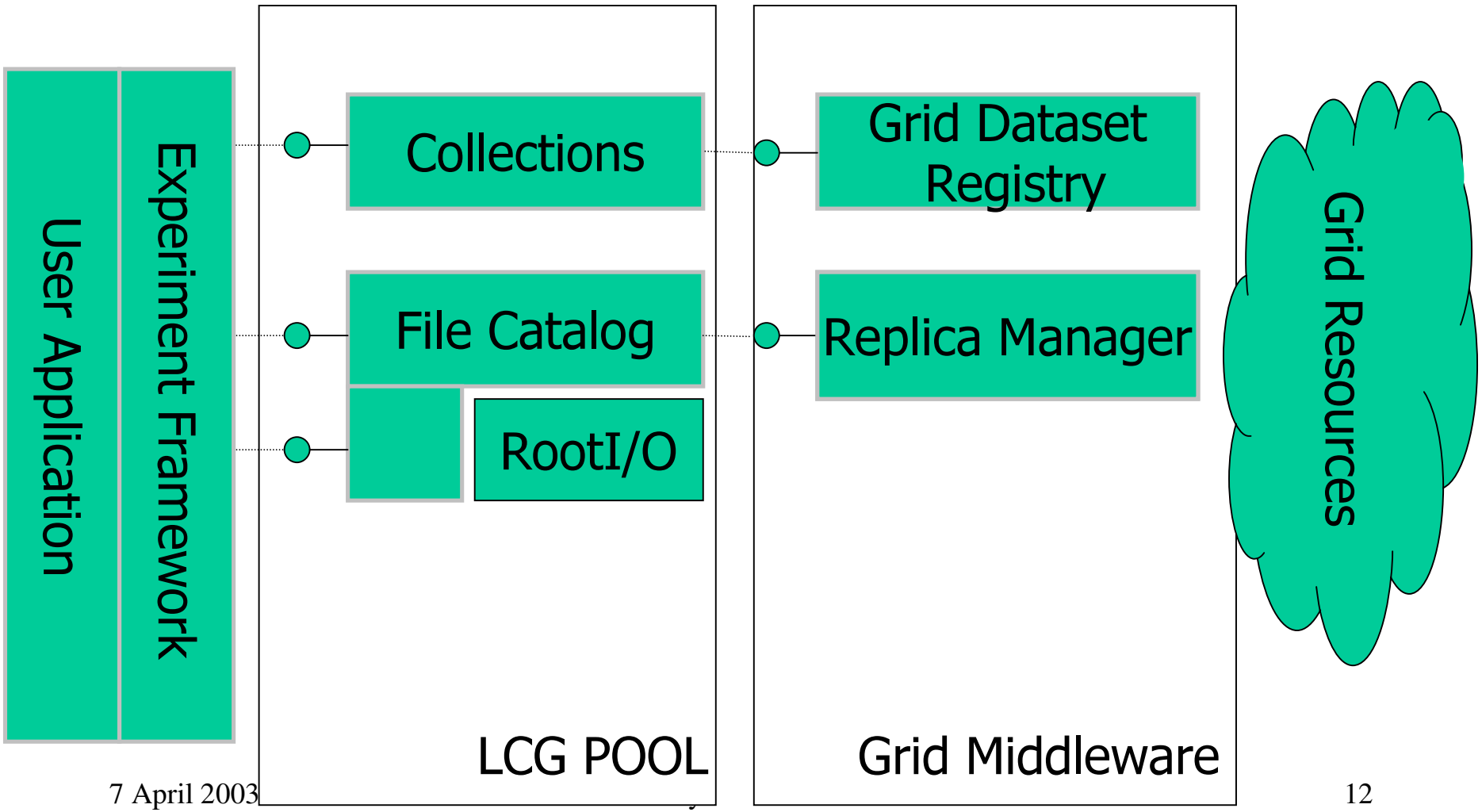
- **Individual physics analysis - by definition 'chaotic' (according to work patterns of individuals)**
  - **Hundreds of physicists distributed in expt may each want to access central AOD+TAG and run their own selections . Will need very selective access to ESD+RAW data (for tuning algorithms, checking occasional events)**

- **Will need replication of AOD+TAG in experiment, and selective replication of RAW+ESD**
  - **This will be a function of processing and physics group organisation in the experiment**

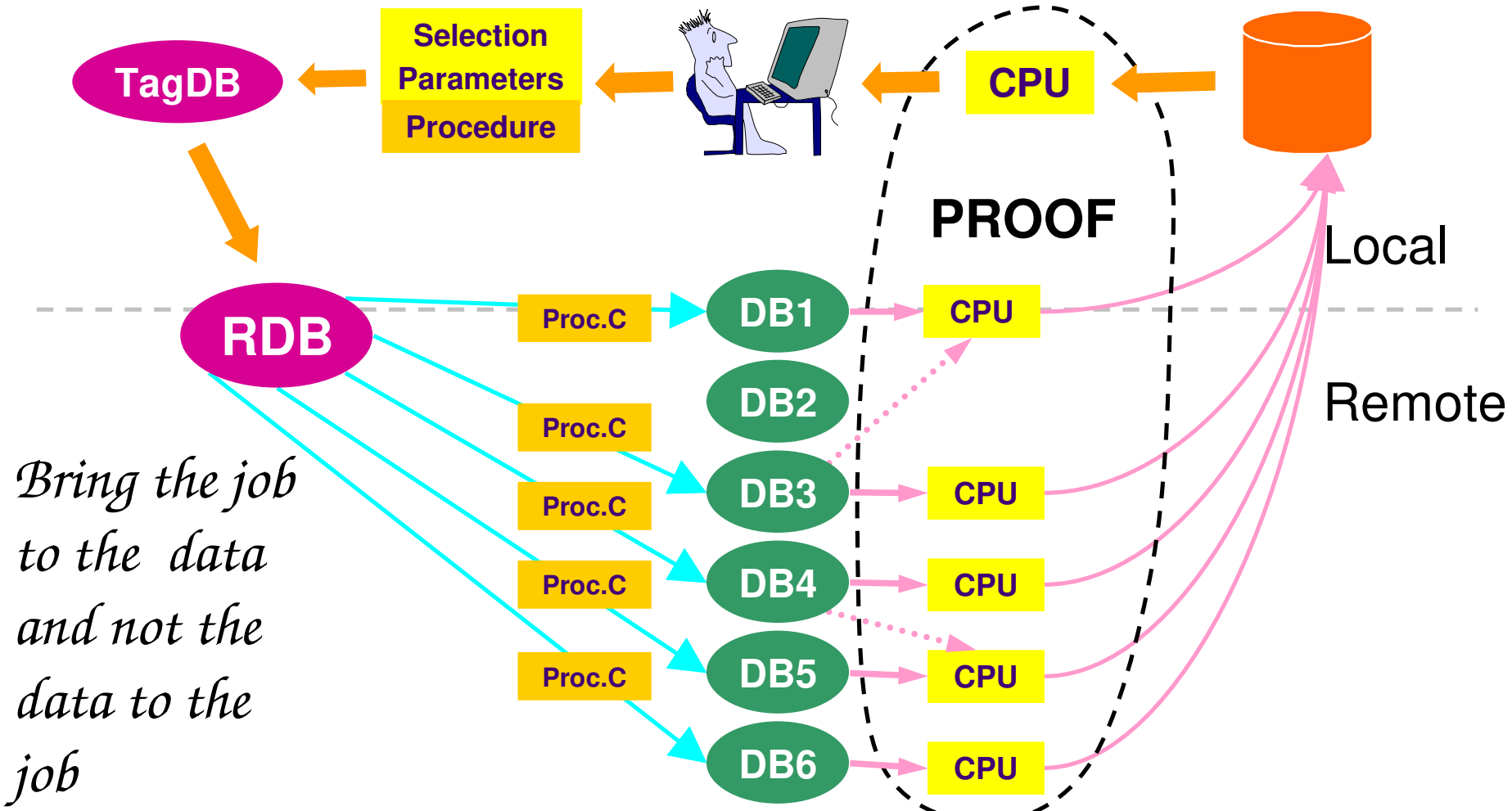
A Logical View of Event Data for physics analysis



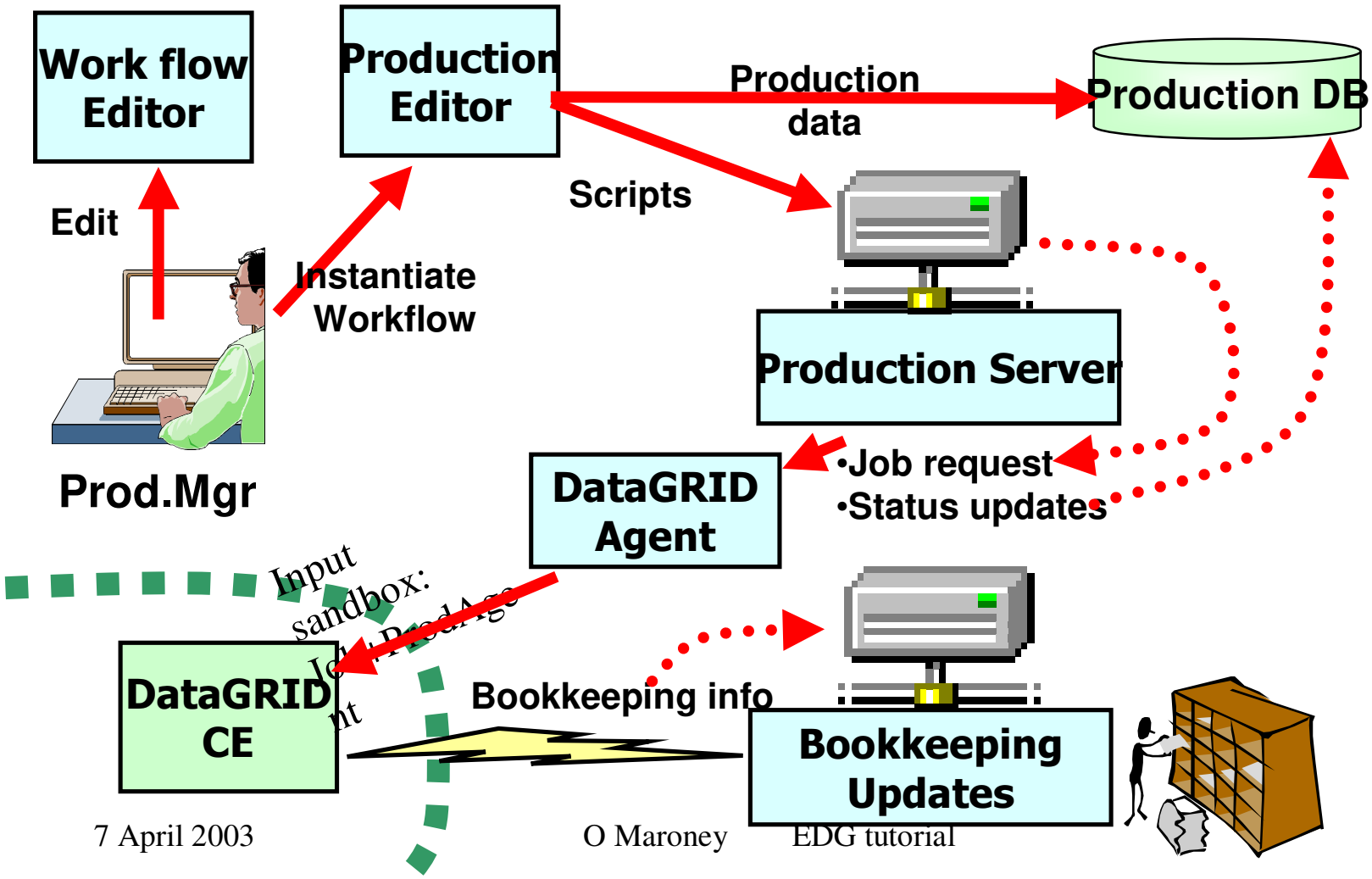
LCG/Pool on the Grid



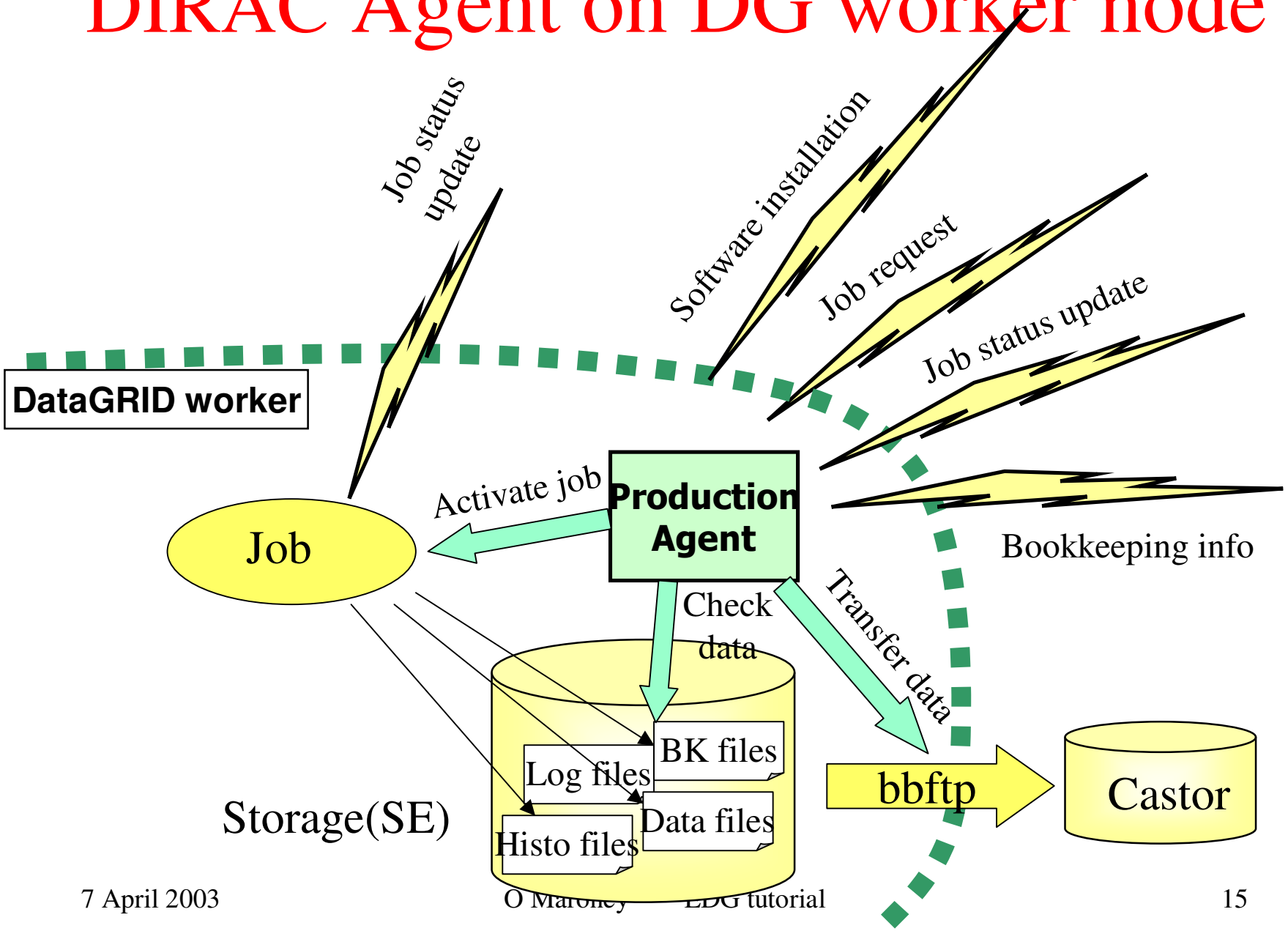
An implementation of distributed analysis in ALICE using natural parallelism of processing



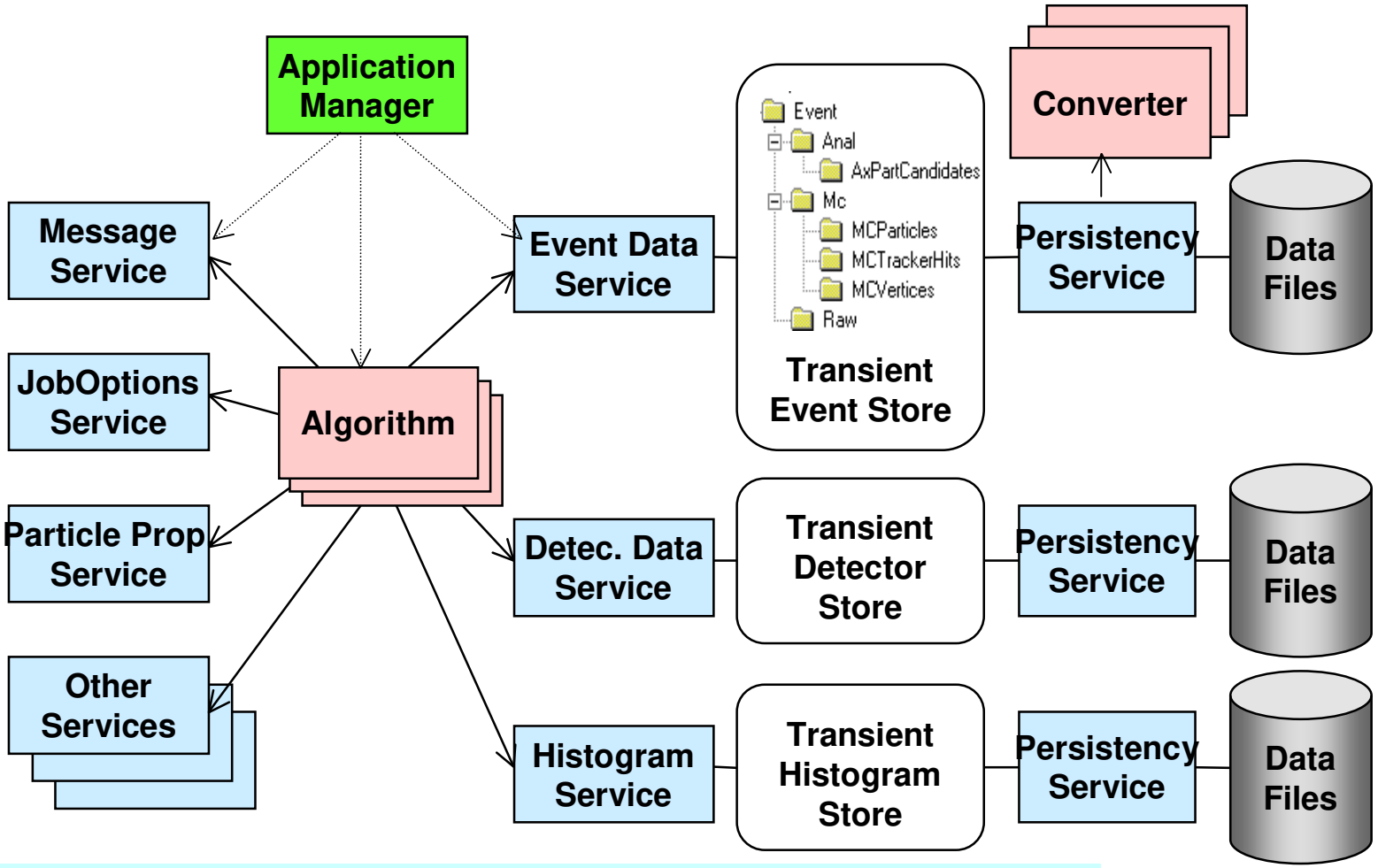
LHCb DIRAC: Production with DataGrid



DIRAC Agent on DG worker node



ATLAS/LHCb Software Framework (Based on Services)



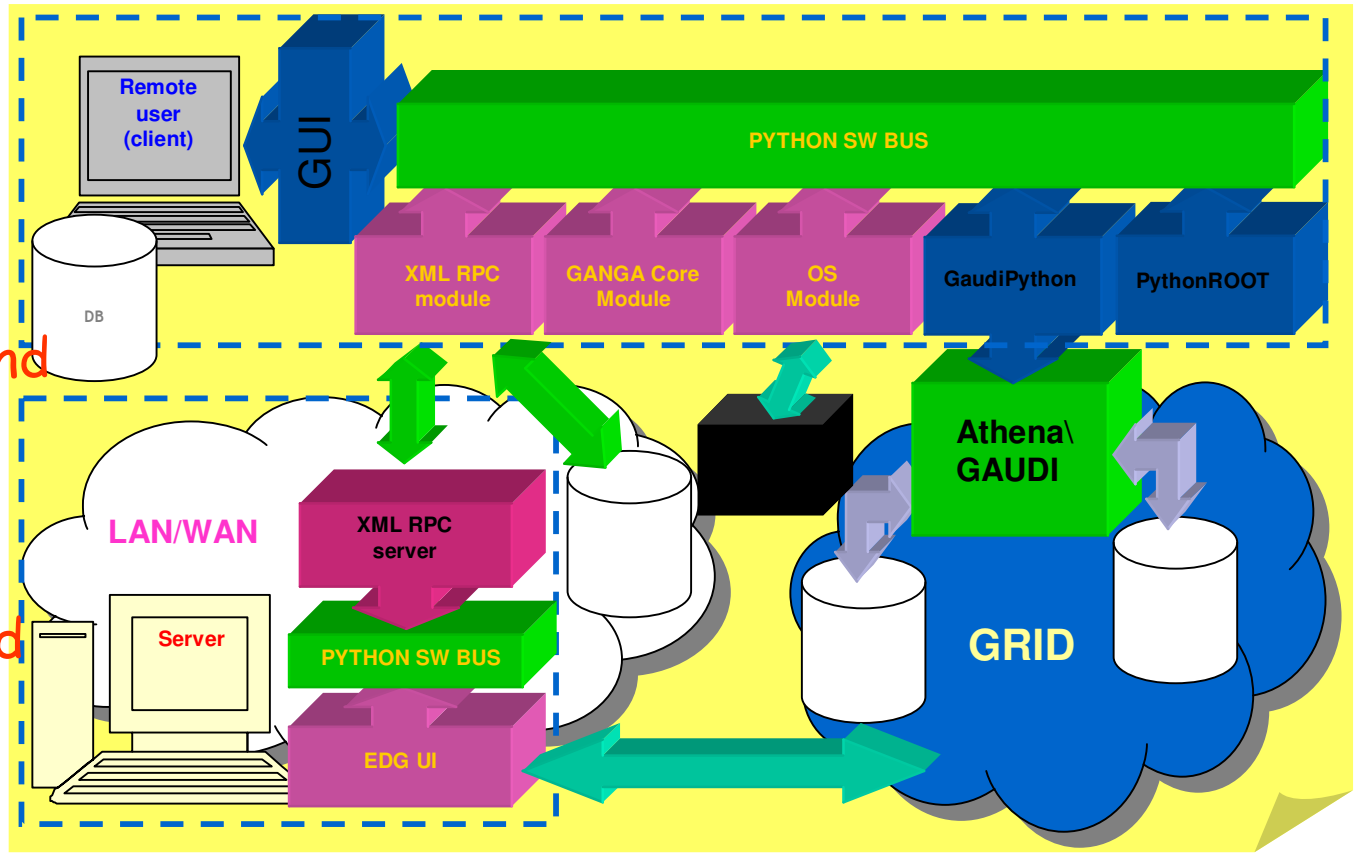
The Gaudi/Athena Framework – Services will interface to Grid (e.g. Persistency)

GANGA: Gaudi ANd Grid Alliance

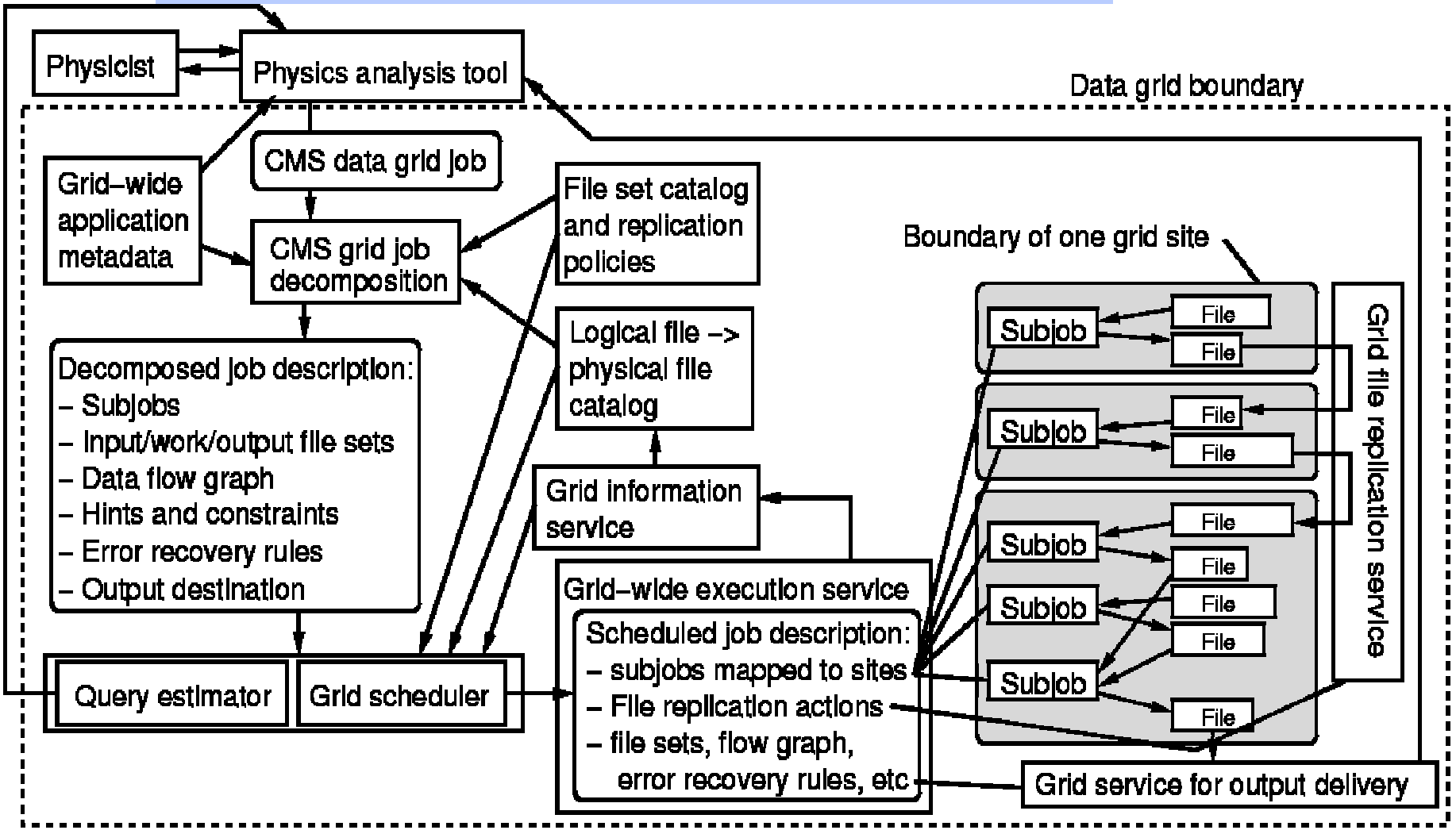
Joint Atlas/LHCb project

- Application facilitating end-user physicists and production managers the use of Grid services for running Gaudi/Athena jobs.

- a GUI based application that should help for the complete job life-time:
  - job preparation and configuration
  - resource booking
  - job submission
  - job monitoring and control



A CMS Data Grid Job



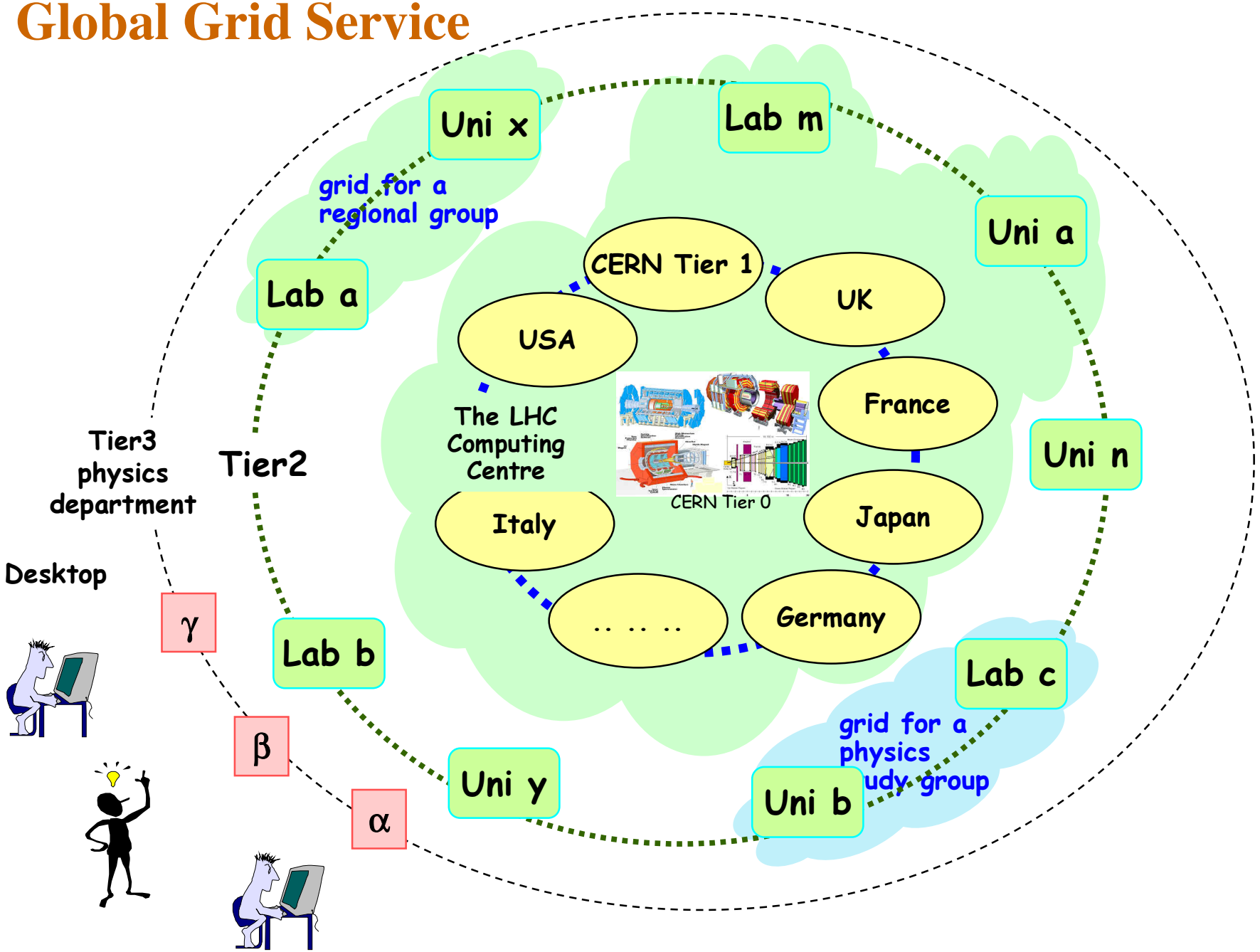
The CMS Stress Test

- CMS MonteCarlo production using BOSS and Impala tools.
  - Originally designed for submitting and monitoring jobs on a 'local' farm (eg. PBS)
  - Modified to treat Grid as 'local farm'
- December 2002 to January 2003
  - 250,000 events generated by job submission at 4 separate UI's
  - 2,147 event files produced
  - 500Gb data transferred using automated grid tools during production, including transfer to and from mass storage systems at CERN and Lyon
  - Efficiency of 83% for (small) CMKIN jobs, 70% for (large) CMSIM jobs

The CMS Stress Test

Site	CE	Number of CPUs	SE	Disk Space (GB)
CERN	lxshare0227	122	lxshare0393 lxshare0384	100 1000(=100*10)*
CNAF	testbed008	40+	grid007g	1000*
RAL	gppce05	16	gppse05	360
NIKHEF	tbn09	22	tbn03	35
LYON	ccgridli03 ccgridli08	120 400	ccgridli07	200
Legnaro	cmsgrid001	50	cmsgrid002	513(+513)
Padova	grid001	12	grid005	680
Ecole Polytechnique	polgrid1	4	polgrid2	220
Imperial College	gw39	16	fb00	450

Deploying the LHC Global Grid Service



les.robertson@cern.ch

LCG Grid Service

Interoperable grid using US and Europe LHC resources

- Taking services from US VDT 1.1.6, and EDG 1.4
- Adding services from EDG 1.5/2.0 as they become available

	July 2003				November 2003			
	Alice	Atlas	CMS	LHCb	Alice	Atlas	CMS	LHCb
Total stored data		150G	10T	0.5T	220T	1T	140T	6T
Total number of files		2K	100K	2K	200K	10K	850K	24K
Average data read/job		300M	20M	0.75G	2G	50M	2G	0.75G
Average data written/job		500M	200M	1.5G	2G	500M	1.5G	1.5G
Job submission rate (per day)		40	2,500	10		200	10K	130

From: Report of the Middleware Selection GDB Working Group 1

DataGrid Biomedical work package 10

- Grid technology opens the perspective of large computational power and easy access to heterogeneous data sources.
- A grid for health would provide a framework for sharing disk and computing resources, for promoting standards and fostering synergy between bio-informatics and medical informatics
- A first biomedical grid is being deployed by the DataGrid IST project

http://dbs.cordis.lu/fep-cgi/srchidadb?ACTION=D&SESSION=221592002-10-18&DOC=27&TBL=EN_PROJ&RCN=EP_RCN_A:63345&CALLER=PROJ_IST

Challenges for a biomedical grid

- **The biomedical community has NO strong center of gravity in Europe**
  - **No equivalent of CERN (High-Energy Physics) or ESA (Earth Observation)**
  - **Many high-level laboratories of comparable size and influence without a practical activity backbone (EMB-net, national centers,…) leading to:**
    - **Little awareness of common needs**
    - **Few common standards**
    - **Small common long-term investment**
- **The biomedical community is very large (tens of thousands of potential users)**
- **The biomedical community is often distant from computer science issues**

Biomedical requirements

- Large user community(thousands of users)
  - anonymous/group login
- Data management
  - data updates and data versioning
  - Large volume management (a hospital can accumulate TBs of images in a year)
- Security
  - disk / network encryption
- Limited response time
  - fast queues
- High priority jobs
  - privileged users
- Interactivity
  - communication between user interface and computation
- Parallelization
  - MPI site-wide / grid-wide
  - Thousands of images
  - Operated on by 10's of algorithms
- Pipeline processing
  - pipeline description language / scheduling

Diverse Users…

- Patient
  - has free access to own medical data
- Physician
  - has complete read access to patients data. Few persons have read/write access.
- Researchers
  - may obtain read access to anonymous medical data for research purposes. Nominative data should be blanked before transmission to these users
- Biologist
  - has free access to public databases. Use web portal to access biology server services.
- Chemical/Pharmacological manufacturer
  - owns private data. Need to control the possible targets for data storage.

…and data

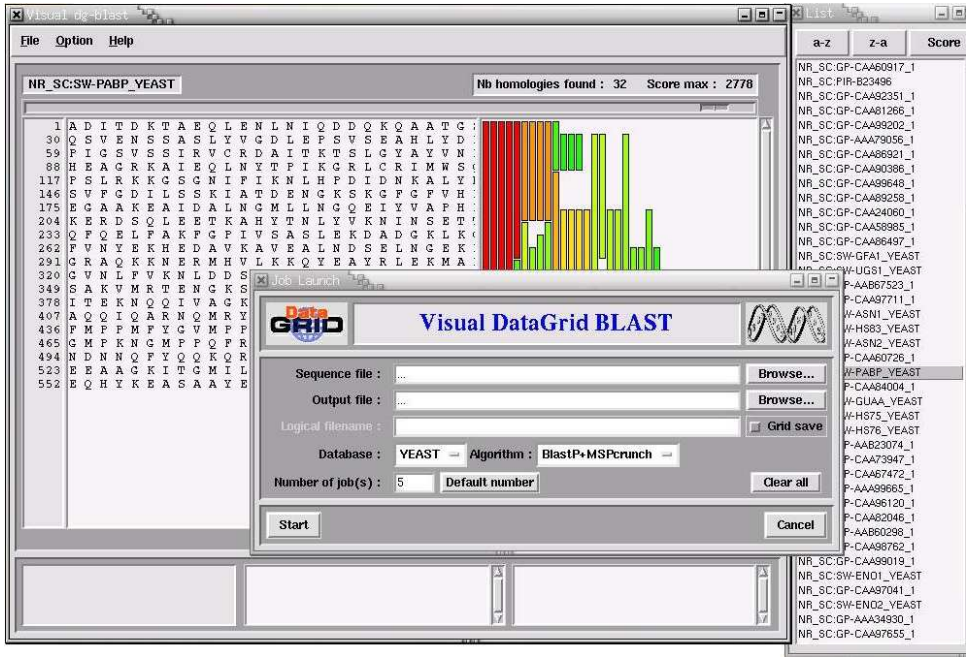
- Biological Data
  - Public and private databases
  - Very fast growth (doubles every 8-12 months)
  - Frequent updates (versionning)
  - Heterogenous formats
- Medical data
  - Strong semantic
  - Distributed over imaging sites
  - Images and metadata

Web portals for biologists

- Biologist enters sequences through web interface
- Pipelined execution of bio-informatics algorithms
  - Genomics comparative analysis (thousands of files of ~Gbyte)
    - Genome comparison takes days of CPU (~n**2)
  - Phylogenetics
  - 2D, 3D molecular structure of proteins…
- The algorithms are currently executed on a local cluster
  - Big labs have big clusters …
  - But growing pressure on resources – Grid will help
    - **More and more biologists**
    - **compare larger and larger sequences (whole genomes)…**
    - **to more and more genomes…**
    - **with fancier and fancier algorithms !!**

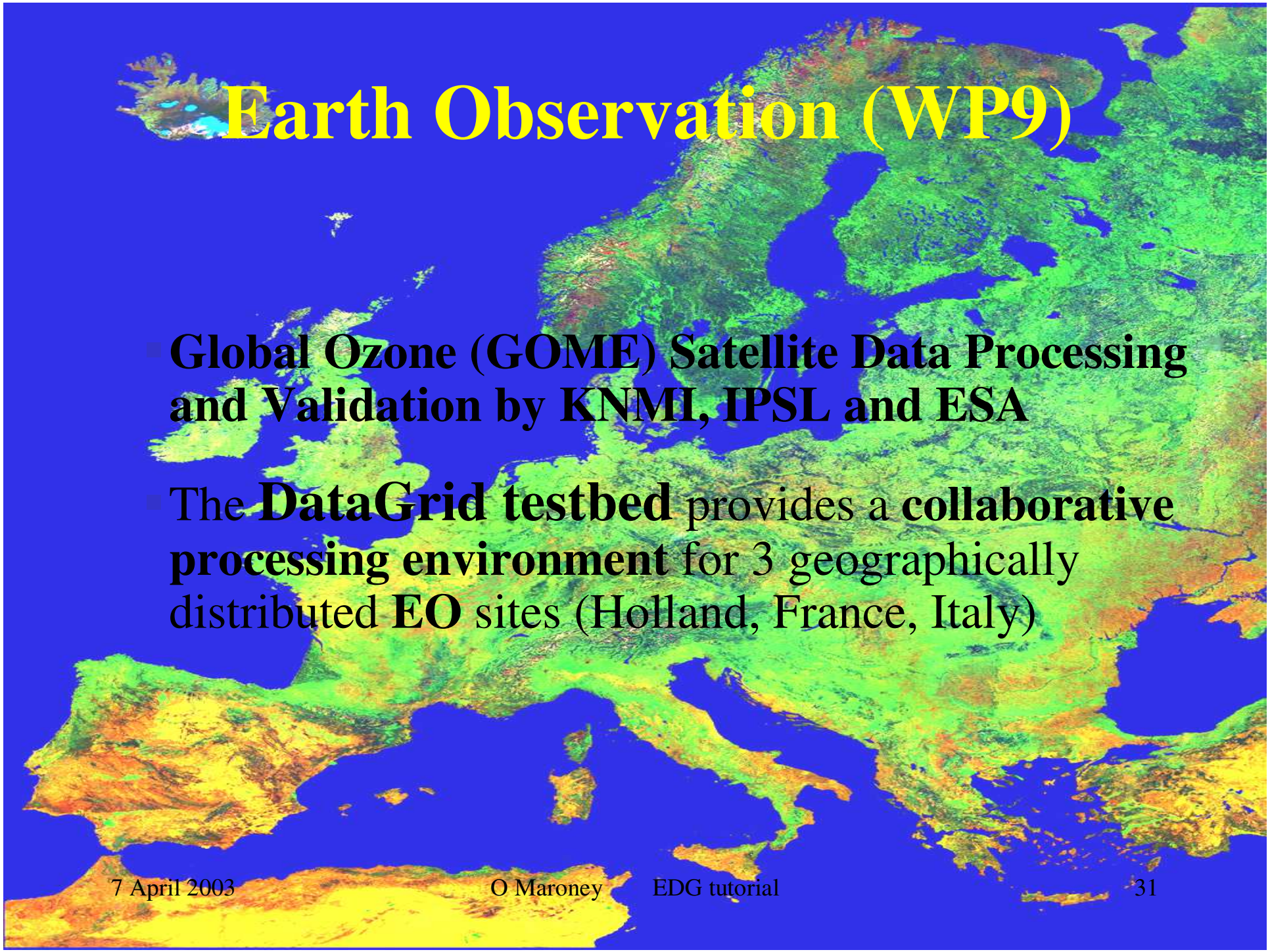
The Visual DataGrid Blast, a first genomics application on DataGrid

- A graphical interface to enter query sequences and select the reference database
- A script to execute the BLAST algorithm on the grid
- A graphical interface to analyze result
- Accessible from the web portal genius.ct.infn.it



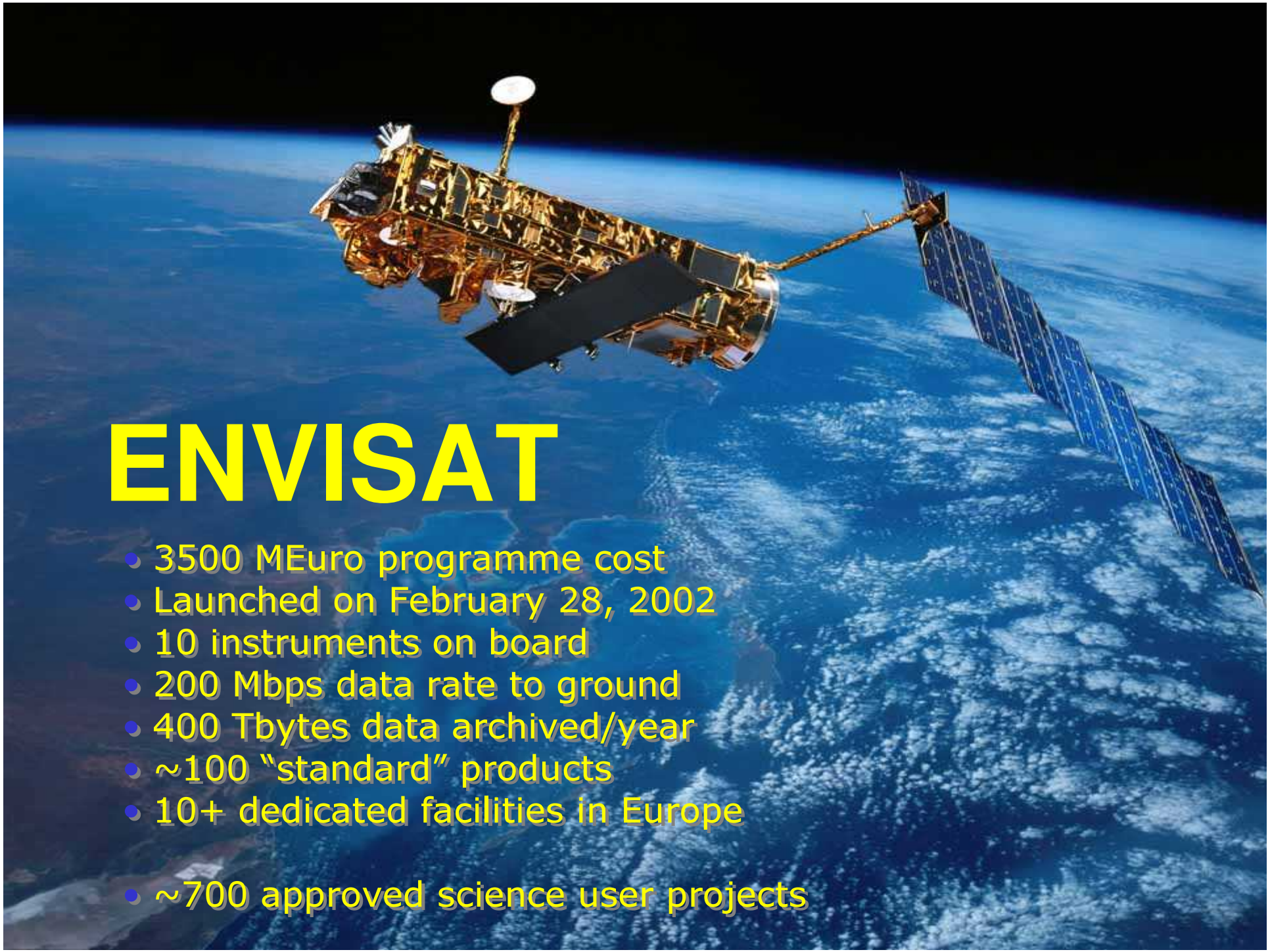
Other Medical Applications

- Complex modelling of anatomical structures
  - Anatomical and functional models, parallelizatoin
- Surgery simulation
  - Realistic models, real-time constraints
- Simulation of MRIs
  - MRI modelling, artifacts modeling, parallel simulation
- Mammographies analysis
  - Automatic pathologies detection
- Shared and distributed data management
  - Data hierarchy, dynamic indices, optimization, caching



Earth Observation (WP9)

- **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)


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

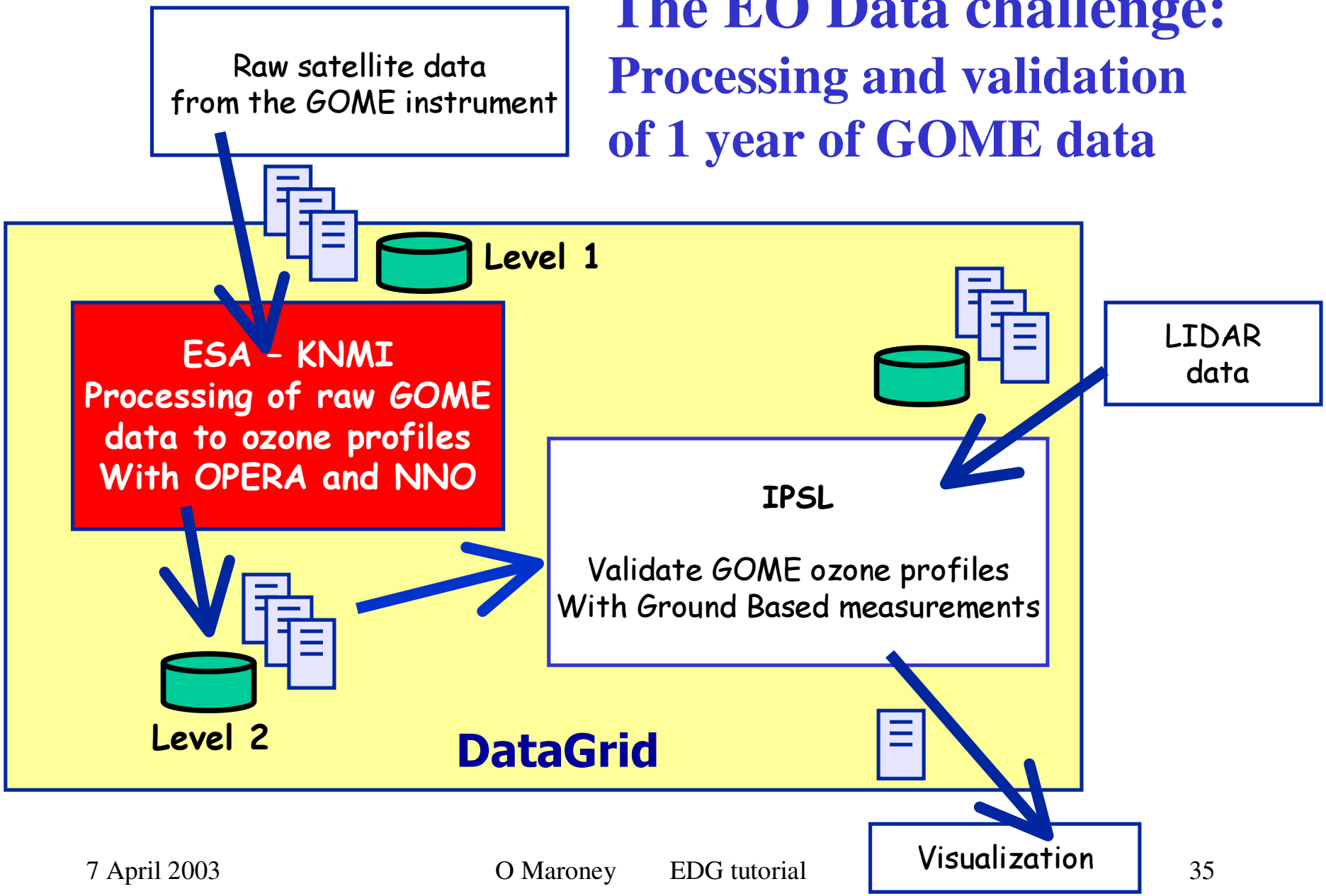
Earth Observation

- Two different **GOME** processing techniques will be investigated
  - **OPERA** (Holland) - Tightly coupled - using MPI
  - **NOPREGO** (Italy) - Loosely coupled - using Neural Networks
- The results are checked by **VALIDATION** (France). Satellite Observations are compared against ground-based LIDAR measurements coincident in area and time.

GOME OZONE Data Processing Model

- **Level-1** data (raw satellite measurements) are analysed to retrieve actual physical quantities : Level-2 data
- **Level-2** data provides measurements of OZONE within a vertical column of atmosphere at a given lat/lon location above the Earth's surface
- **Coincident** data consists of Level-2 data co-registered with LIDAR data (ground-based observations) and compared using statistical methods

The EO Data challenge:
Processing and validation of 1 year of GOME data



EO Use-Case File Numbers

1 Year of GOME data

Data	Number of files to be processed and replicated	Size
Level 1 (Satellite data)	4,724	15 Mb
Level 2 (NNO)	9,448,000	10 kb
Level 2 (Opera)	9,448,000	12 kb
Coincident (Validation)	12	2.5 Mb
Total:	**18,900,736 files**	**267 Gbyte**

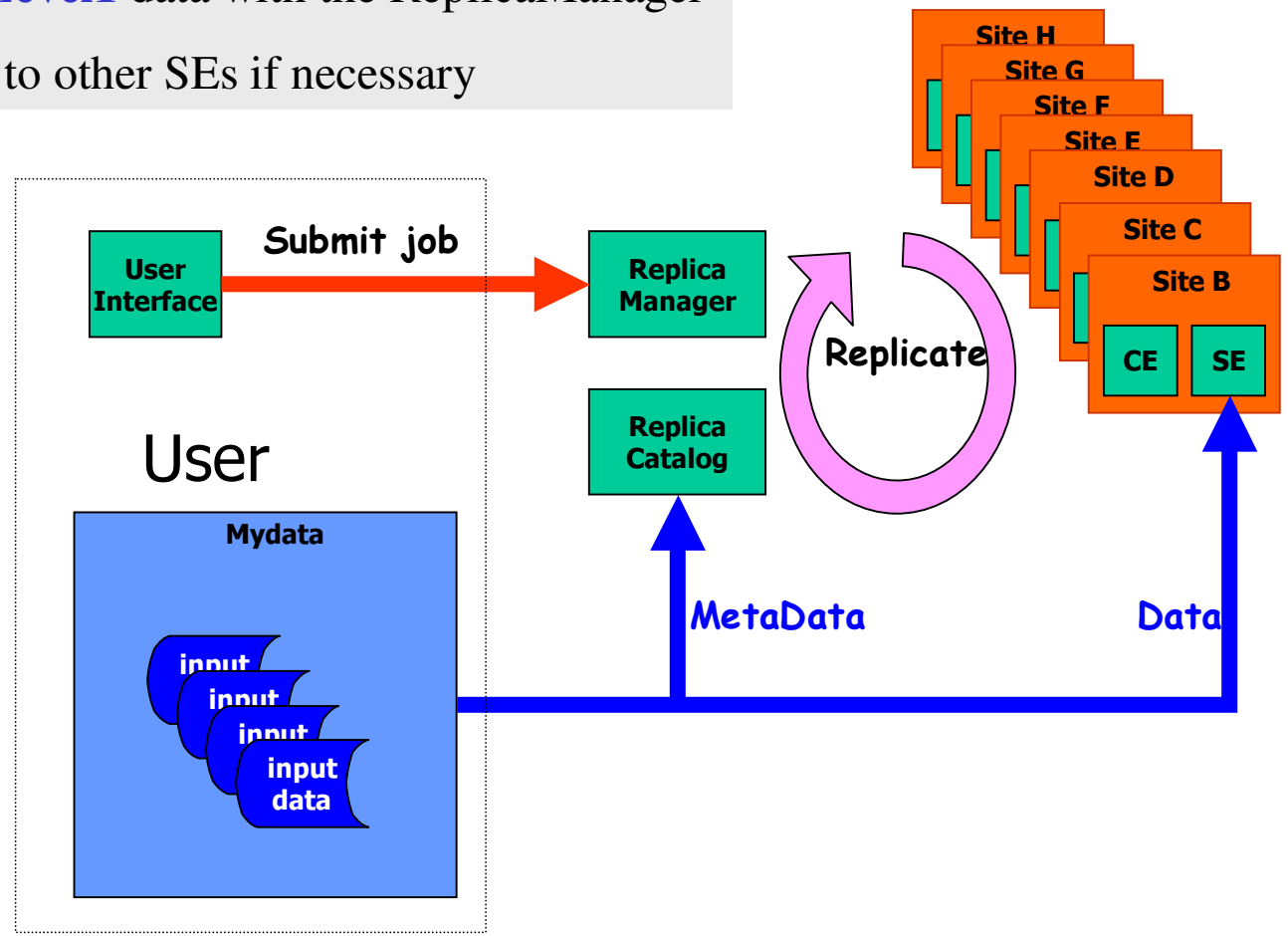
Part of a 5-year, global dataset

GOME Processing Steps (1-2)

Step 1: Transfer **Level1** data to the Grid Storage Element

Step 2: Register **Level1** data with the ReplicaManager

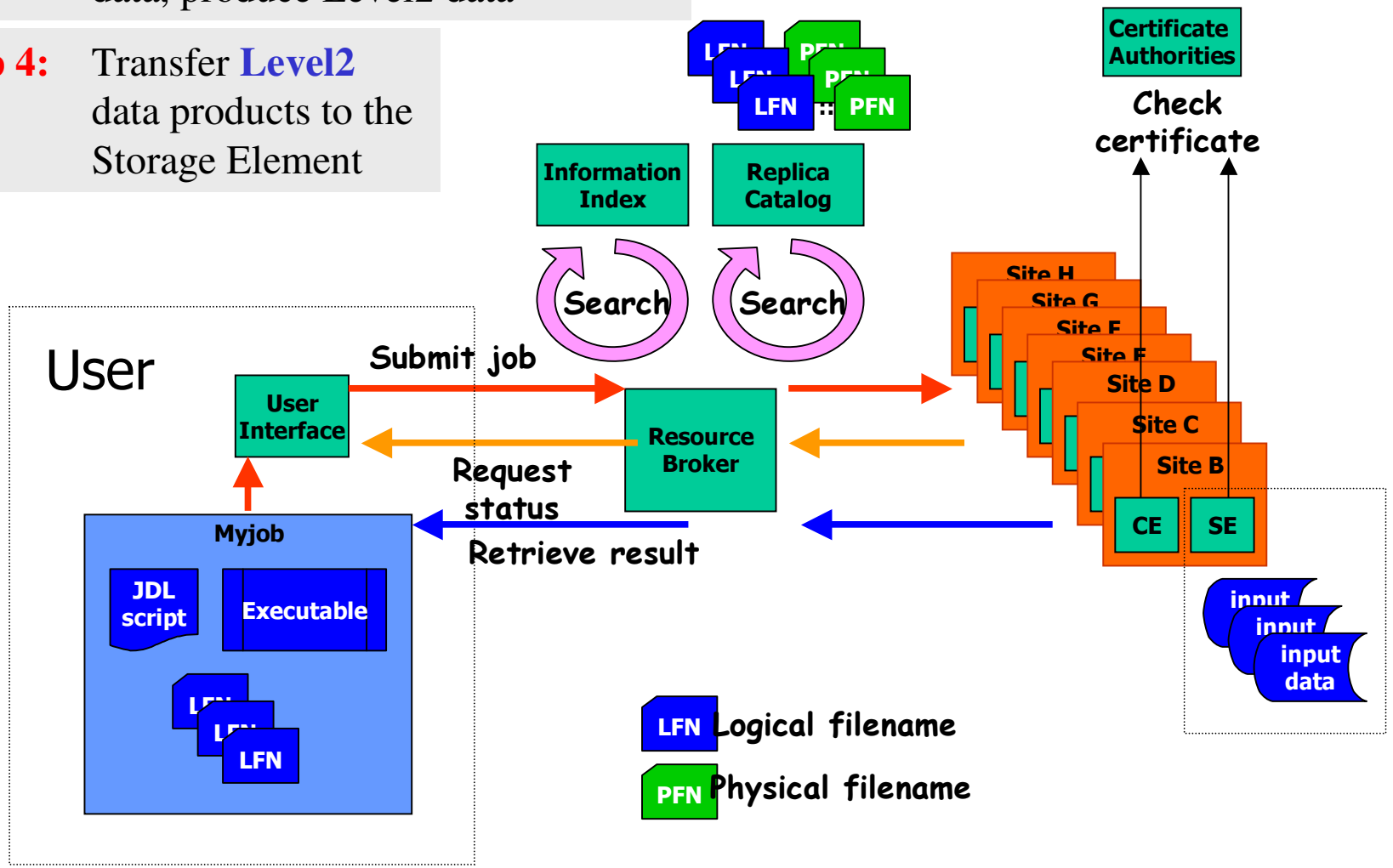
Replicate to other SEs if necessary



GOME Processing Steps (3-4)

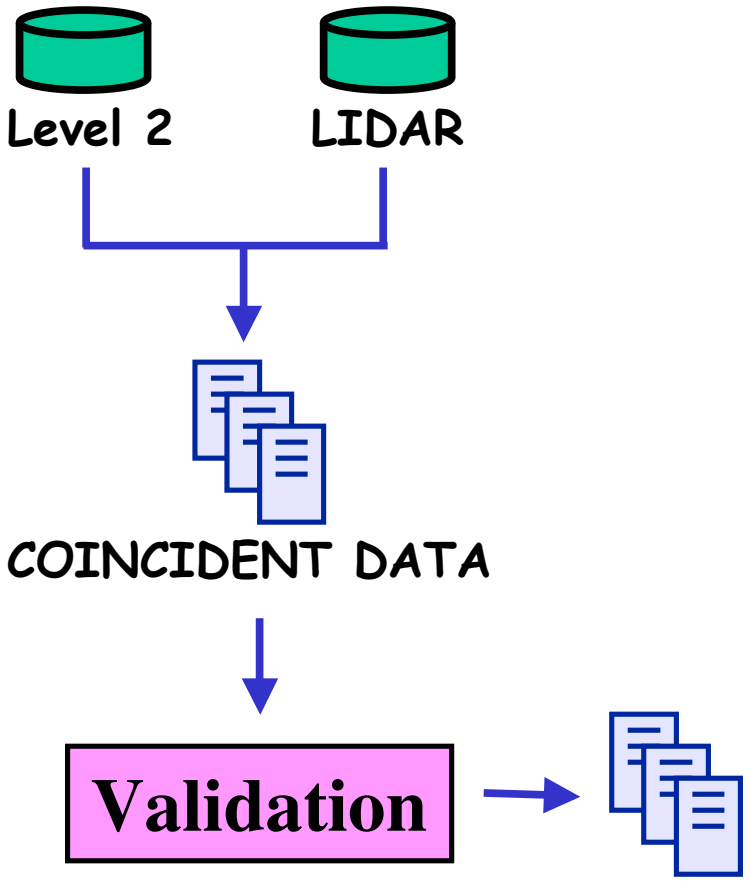
Step 3: Submit jobs to process **Level1** data, produce Level2 data

Step 4: Transfer **Level2** data products to the Storage Element

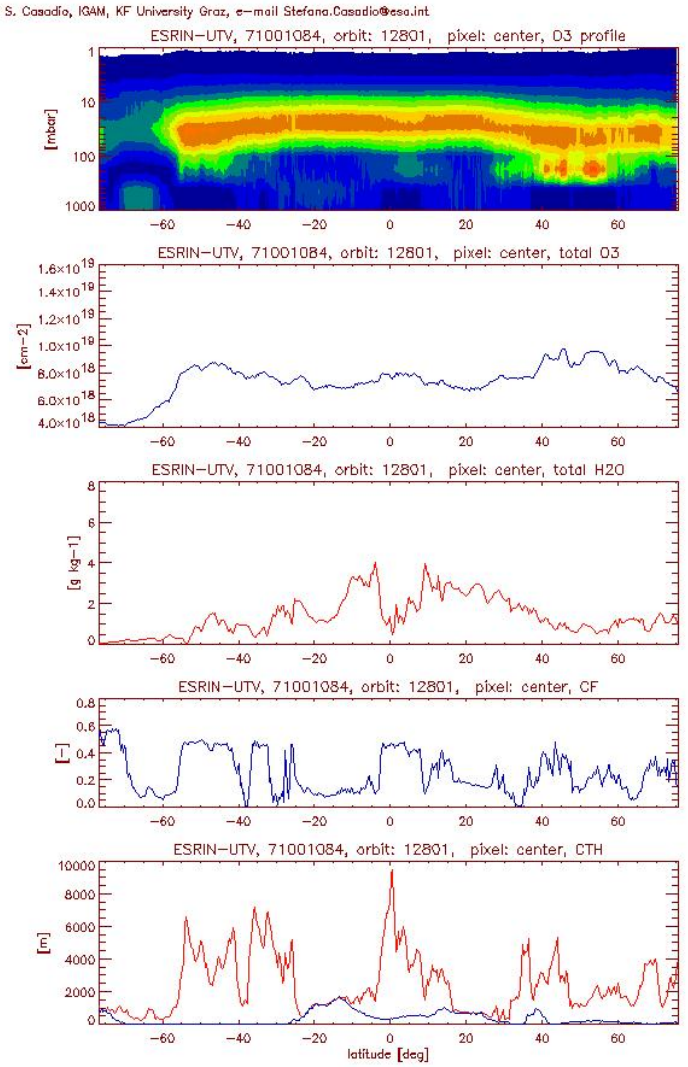


GOME Processing Steps (5-6)

Step 5: Produce Level-2 / LIDAR **Coincident** data perform **VALIDATION**



Step 6: Visualize Results



Summary and a forward look for applications work within EDG

- **Currently evaluating the basic functionality of the tools and their integration into data processing schemes. Will move onto areas of interactive analysis, and more detailed interfacing via APIs**
  - **Hopefully experiments will do common work in interfacing applications to GRID under the umbrella of LCG**
  - **HEPCAL (Common Use Cases for a HEP Common Application Layer) work will be used as a basis for the integration of Grid tools into the LHC prototype**

    http://lcg.web.cern.ch/LCG/SC2/RTAG4

- **There are many grid projects in the world and we must work together with them**
  - **e.g. in HEP we have DataTag,Crossgrid,Nordugrid + US Projects(GryPhyn,PPDG,iVDGL)**

- **Perhaps we can define shared project between HEP,Bio-med and ESA for applications layer interfacing to basic Grid functions.**

Acknowlegements and references

- **Thanks to the following who provided material and advice**
  - **J Linford(WP9),V Breton(WP10),J Montagnat(WP10),F Carminati(Alice),JJ Blaising(Atlas),C Grandi(CMS),M Frank(LHCb),L Robertson(LCG),D Duellmann(LCG/POOL) ,T Doyle(UK GridPP),M Reale(WP8)**
  - **F Harris(WP8), I Augustin(WP8) N Brook(LHCb), P Hobson (CMS), J Montagnat (WP10)**

- **Some interesting WEB sites and documents**

  - LHC Review http://lhc-computing-review-public.web.cern.ch/lhc-computing-review-public/Public/Report_final.PDF **(LHC Computing Review)**

  - **LCG** http://lcg.web.cern.ch/LCG
    http://lcg.web.cern.ch/LCG/SC2/RTAG6 **(model for regional centres)**
    http://lcg.web.cern.ch/LCG/SC2/RTAG4 **(HEPCAL Grid use cases)**
  - **GEANT** http://www.dante.net/geant/ **(European Research Networks)**
  - **POOL** http://lcgapp.cern.ch/project/persist/
  - **WP8** http://datagrid-wp8.web.cern.ch/DataGrid-WP8/
    http://edmsoraweb.cern.ch:8001/cedar/doc.info?document_id=332409 **(Requirements)**
  - **WP9** http://styx.srin.esa.it/grid
    http://edmsoraweb.cern.ch:8001/cedar/doc.info?document_id=332411 **(Reqts)**
  - **WP10** http://marianne.in2p3.fr/datagrid/wp10/
    http://www.healthgrid.org
    http://www.creatis.insa-lyon.fr/MEDIGRID/
    http://edmsoraweb.cern.ch:8001/cedar/doc.info?document_id=332412 **(Reqts)**