CTA

Cherenkov Telescope Array Computing Grid and e-infrastructures

Giovanni Lamanna

LAPP - Laboratoire d'Annecy-le-Vieux de Physique des Particules,

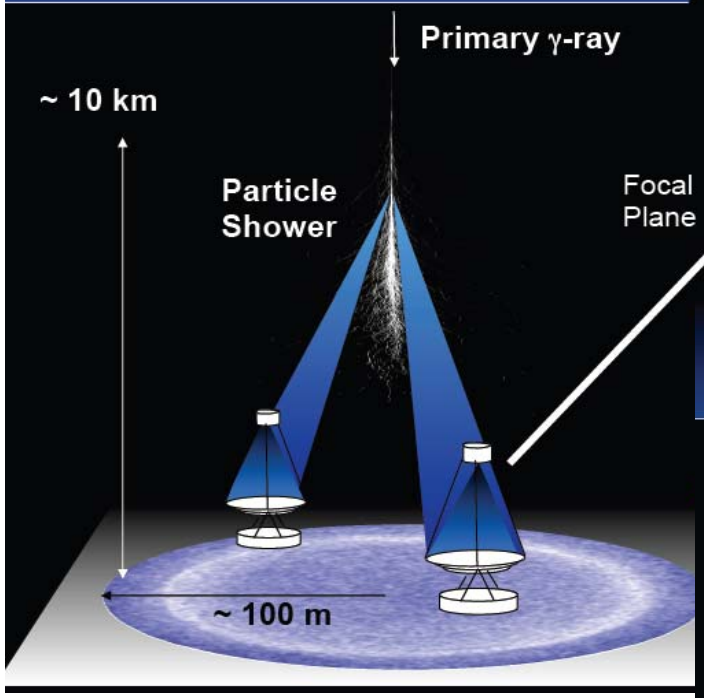
Université de Savoie, CNRS/IN2P3, Annecy-le-vieux, France

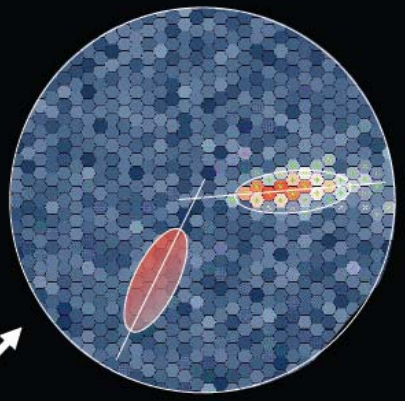
Lyon, 7-8 October 2010

Outline

- Introduction: What is CTA ?
- CTACG project: CTA design study and MC production
- CTA Observatory E-Infrastructures
- CTA data access
- Perspectives
- Conclusion

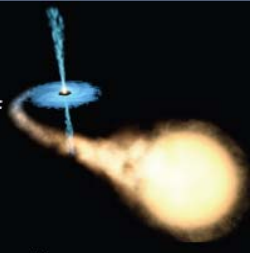
Cherenkov Technique

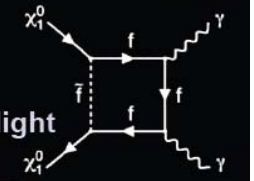


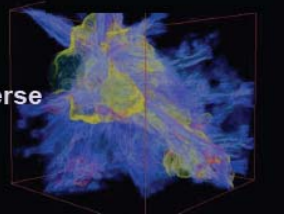


What can we learn using VHE γ-rays?

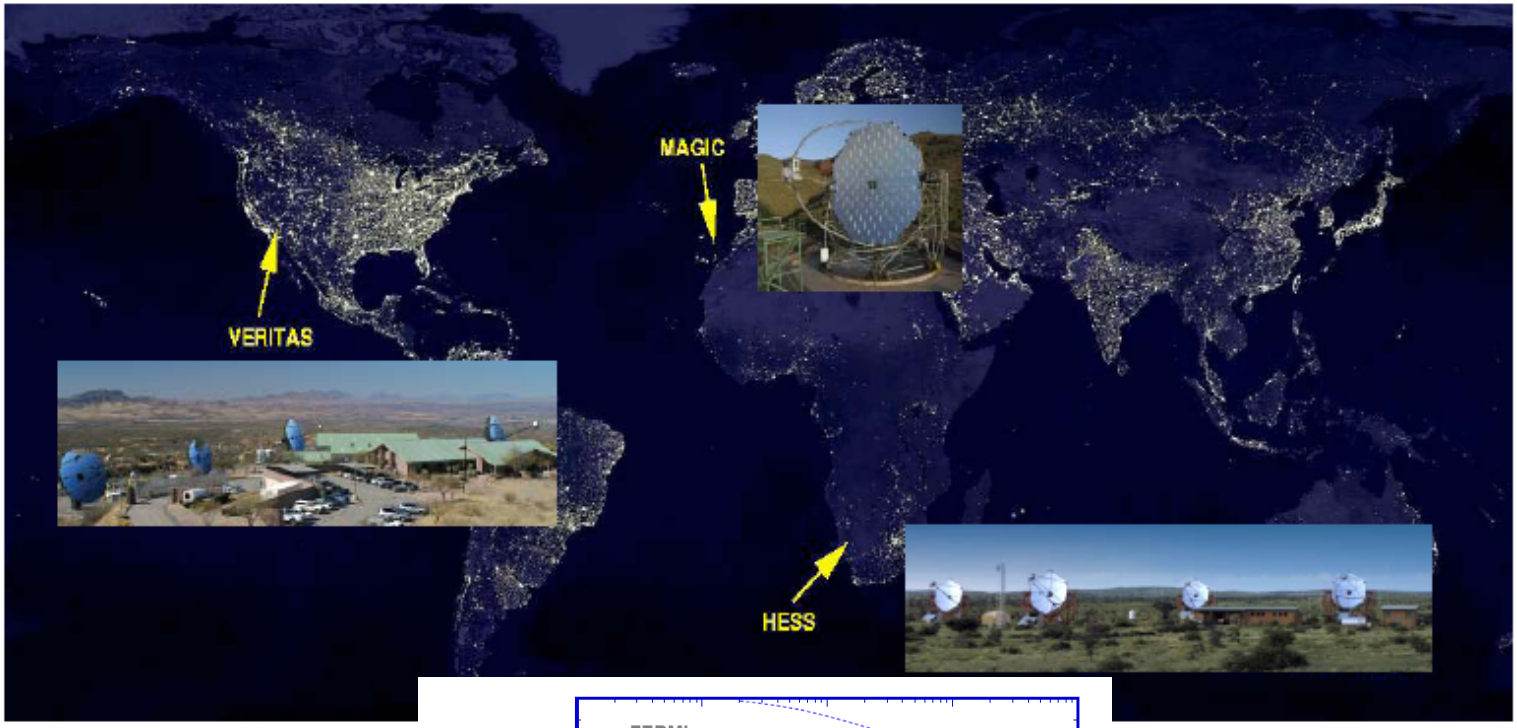
- **High energy astrophysics**
  - Acceleration, propagation and energy losses of ultra-relativistic particles in astrophysical environments
    - Specific question:
      - Where are the "cosmic rays" accelerated?
- **Astroparticle Physics**
  - Indirect search for Dark Matter
  - Search for energy dependence of the speed of light
- **Cosmology**
  - Indirect measurement of the Extragalactic Background light
    - Probing the star formation history of the universe
  - Non-thermal content of galaxy clusters

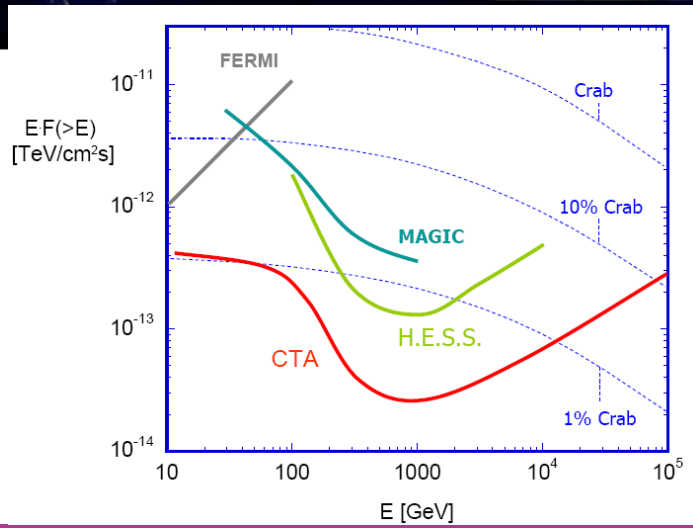






VHE Gamma-ray infrastructures





Cherenkov Telescope Array

- CTA is a new project for ground based gamma-ray astronomy planned to consist of several tens of Imaging Atmospheric Cherenkov Telescopes (IACTs)
- CTA is an Astroparticle research infrastructure aimed to work as an observatory providing services making gamma-ray astronomy accessible to the entire community
- **The CTA international consortium (after a Design Study phase) is currently committed in its (EC-FP7 funded) Preparatory Phase (2011-2013)**

The CTA consortium

The CTA international consortium counts > 100 institutes from > 22 countries:

1) Armenia,
2) Bulgaria,
3) Croatia,
4) Czech Republic,
5) Finland,
6) France,
7) Germany,
8) Greece,
9) Ireland,
10) Italy,
11) Netherlands,
12) Poland,
13) Spain,
14) Switzerland,
15) UK,
16) Denmark,
17) Sweden,
18) Argentina,
19) Japan,
20) Namibia,
21) South Africa,
22) USA.

Design Concepts for the
Cherenkov Telescope Array
CTA

An Advanced Facility for Ground-Based
High-Energy Gamma-Ray Astronomy

The CTA Consortium

May 2010

- **CTACG (CTA Computing Grid) aims at:**
  - Optimising the application of grid technologies for the CTA:
    - Simulation
    - Data processing
    - Storage
    - Offline analysis
    - Virtual Observatory interface

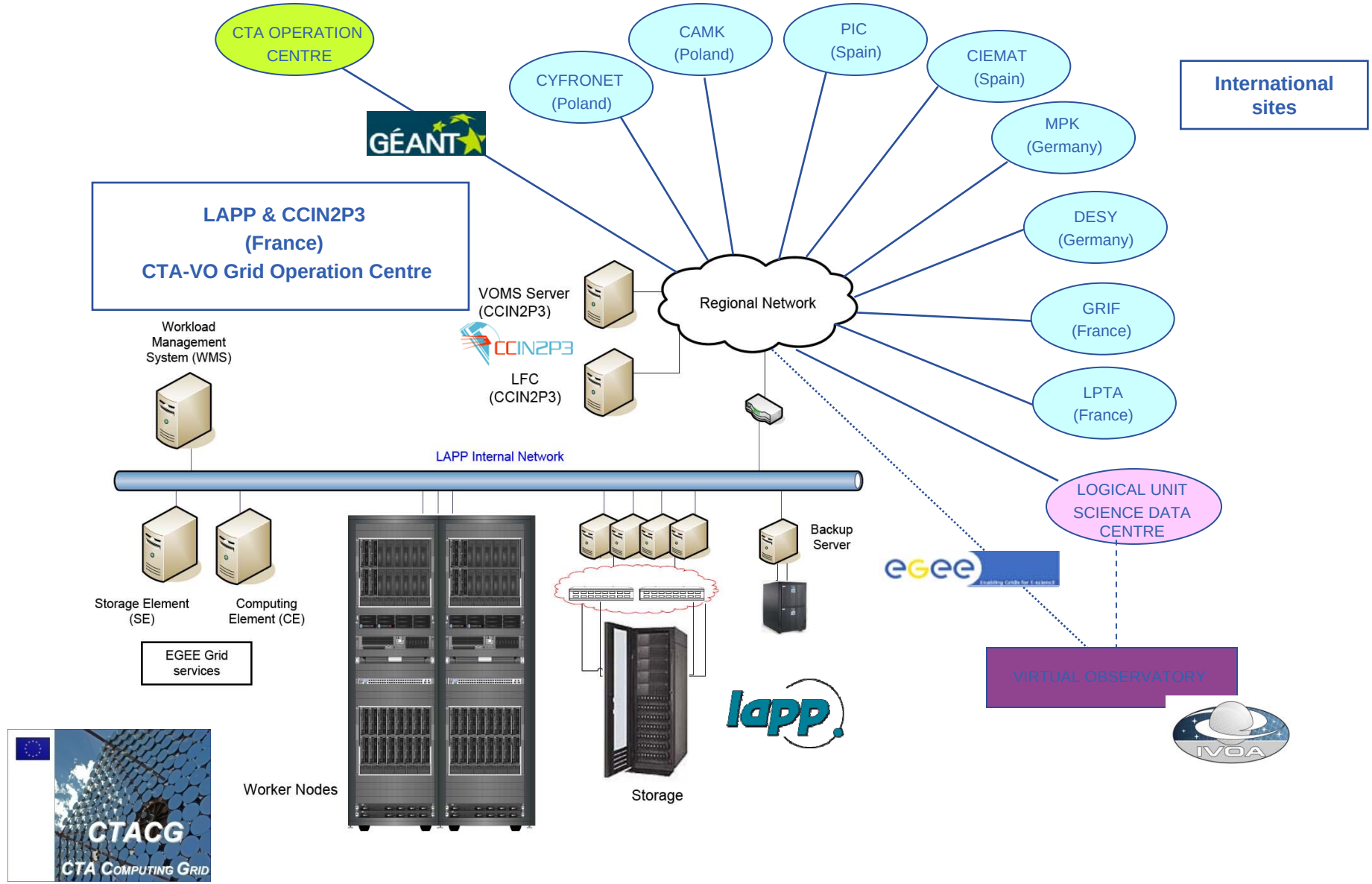
Contingent motivation in 2008:

- Computing resources for MC simulations in the Design Study phase

Why GRID?

- CE and SE to fulfill Monte Carlo production and data reduction requirements
- Central production and control
- Distributing data produced by “distributed” resources (Worker nodes)
- User access to computing resources and analysis tools
- A step toward a coherent CTA data management system
- (At a cost of some initial efforts) Benefit of LCG experience, technical support and mainframe (e.g. middleware, security, db, tools,…)
- Opportunity for LCG Tier2 to open up to a multi-disciplinary context for mutual exchange of experience (e.g. conjugating VObs and EGI-middleware, facing different requirements from the Astroparticle community) towards Astrop.- “Data Centers”.

From CTACG to CTA E-Infrastructures



CTACG Infrastructure

- **Grid infrastructure:** a dedicated EGEE VO named vo.cta.in2p3.fr created in 2008 and including 14 sites in 4 countries (France, Germany, Spain , Poland)
- **Resources overview**

Country	Site	Available Storage (TB)	Available CPUs
France	IN2P3-LAPP	15	760
	IN2P3-CC	280	4745
	GRIF-IRFU	10	4670
	MSFG-OPEN	10	104
Germany	MPPMU	10	856
	DESY-ZN	10	672
Spain	pic	7	1365
	CIEMAT-LCG2	1	824
Poland	CYFRONET-LCG2	100	1984
	CAMK	5	0
Total	**All CTA sites**	**448**	**15980**

Available software:

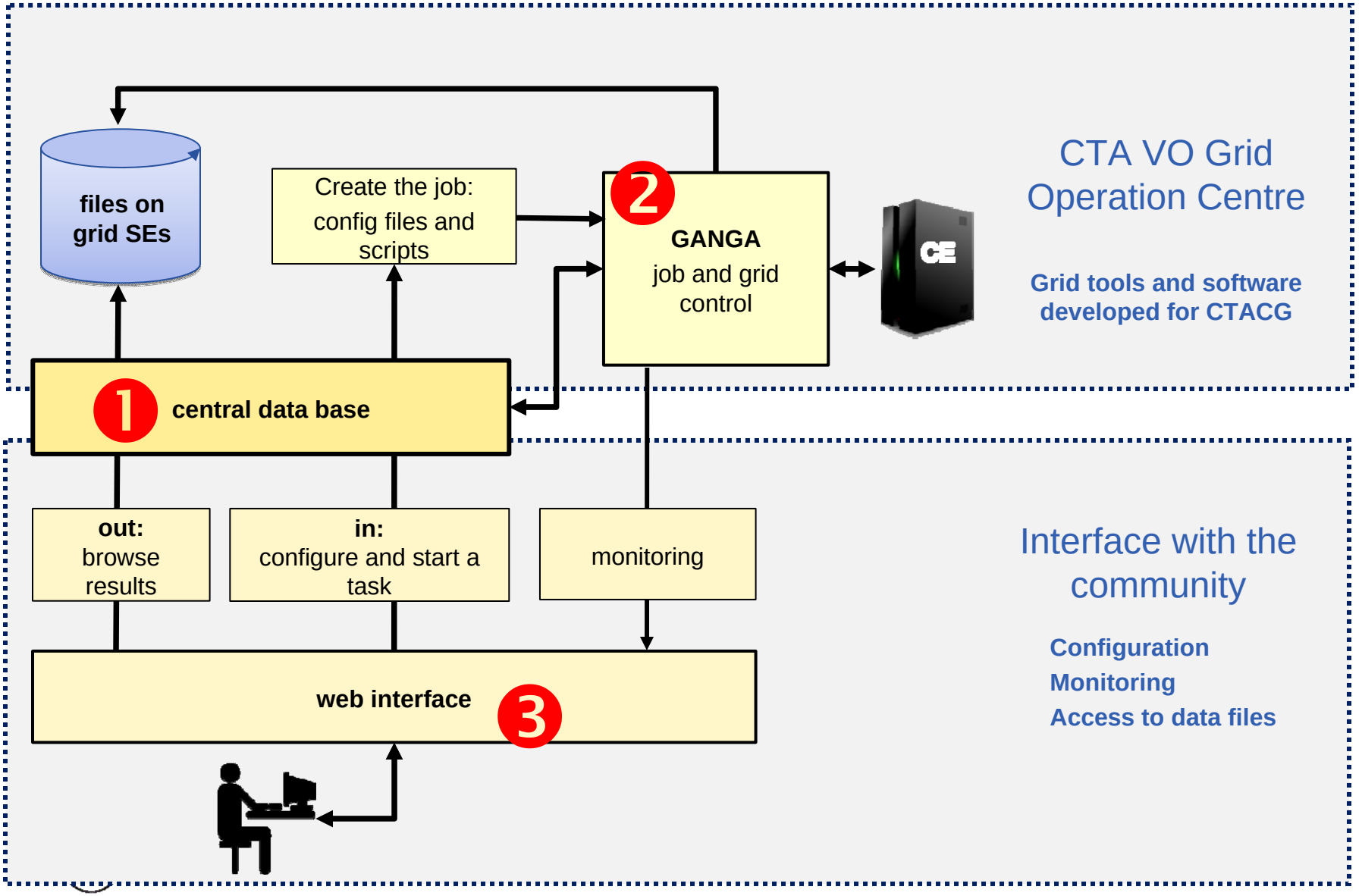
Corsika + sim_telarray

⇒ ~ 450 TB available (280 TB used)

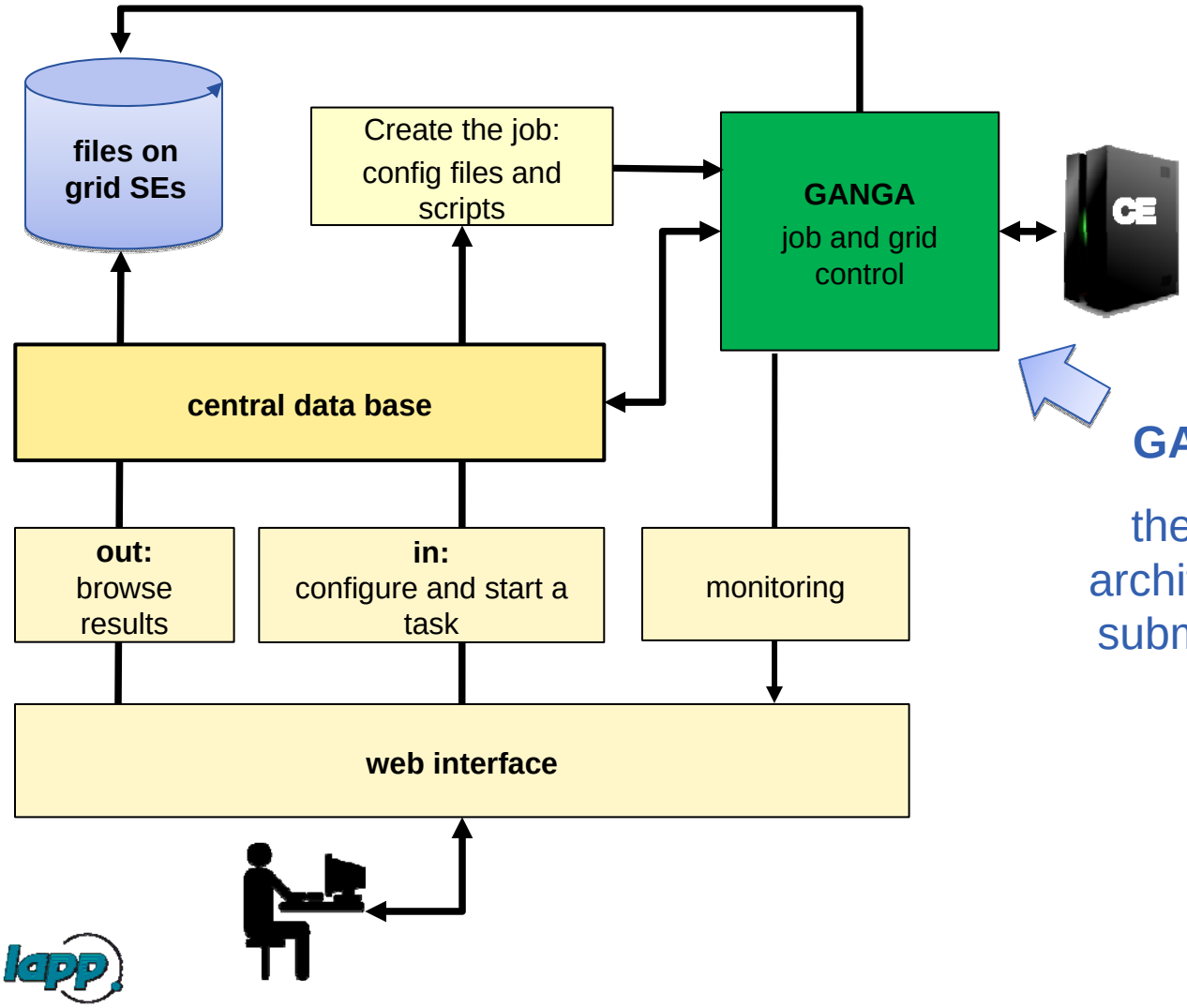
⇒ fulfills MC requirements:
> 400 TB SE, > 300 CPUs CE

- **Requirements :**
  - Automated/permanent tool
  - Generic and self consistent tool
  - Embedded monitoring, logging and bookkeeping systems
  - Tool based on already tried and tested EGEE Grid tools
  - Easily configurable tool to run simulations, analysis or other tasks
  - Smart management of Grid resources, according to available resources
  - Monitoring of sites, tasks, jobs, output files
- **From a user point of view:**
  - No specific Grid knowledge needed to run the jobs and get information and output files from the jobs

CTACG Architecture



CTACG - Ganga Grid interface



GANGA Grid interface

the core of CTACG Grid architecture that enables job submission and grid control

CTACG – GANGA Grid interface

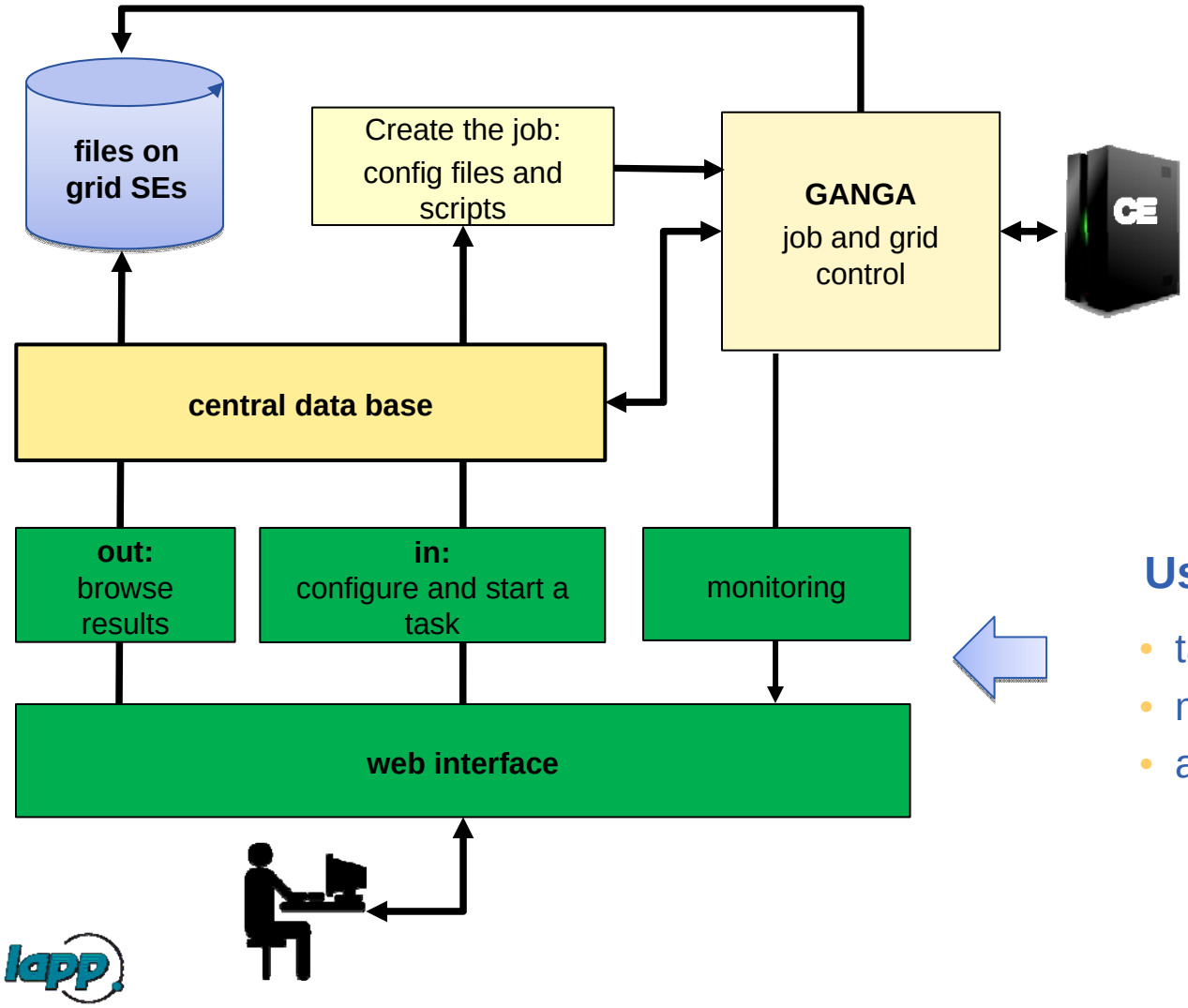
- **Based on GANGA**
  (Gaudi/Athena aNd Grid Alliance: http://ganga.web.cern.ch)
  - Grid front-end for job definition and management in Python: developed, used and maintained by the LCG (LHC Computing Grid) community
  - CTA Grid interface developed at LAPP
    (by S.Elles and O.Arnaez - Atlas LAPP group)

CTACG – GANGA Grid interface

- **Main features:**
  - Runs as a daemon for CTA
    ⇨ infinite loop that automatically launches all tasks defined in the central database
  - Automated start of new tasks once defined
  - Stop, restart, delete jobs/tasks possible at any moment on demand
  - Embedded Grid resources management tool
  - Logging, bookkeeping, monitoring systems

- **Current status: applied to MC production:**
  - Centrally done at LAPP
  - Will then be open to all users

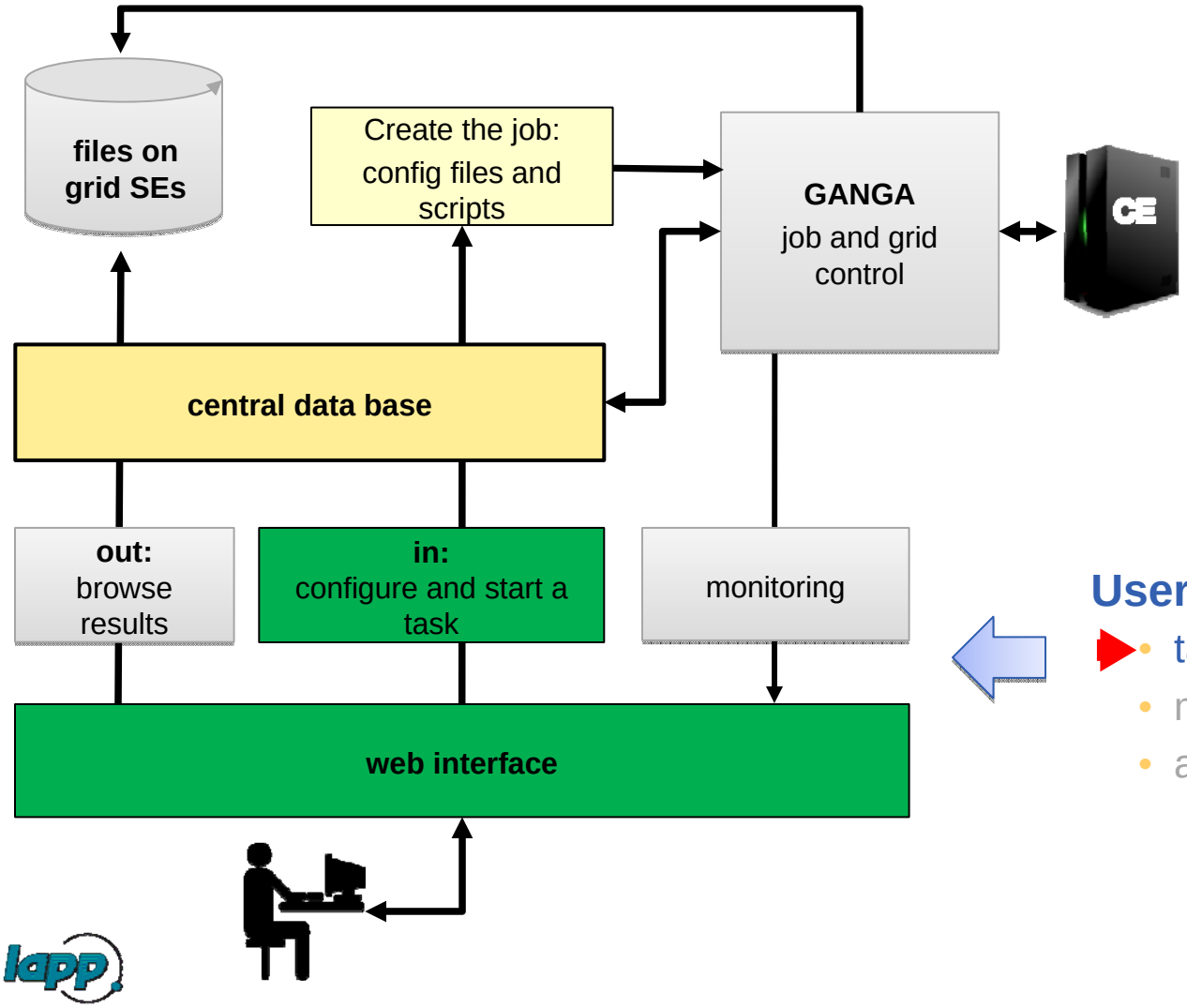
CTACG - User interface



User Interface

- task configuration
- monitoring
- access to result files

Task configuration



User Interface

- task configuration
- monitoring
- access to result files

Software configuration tool

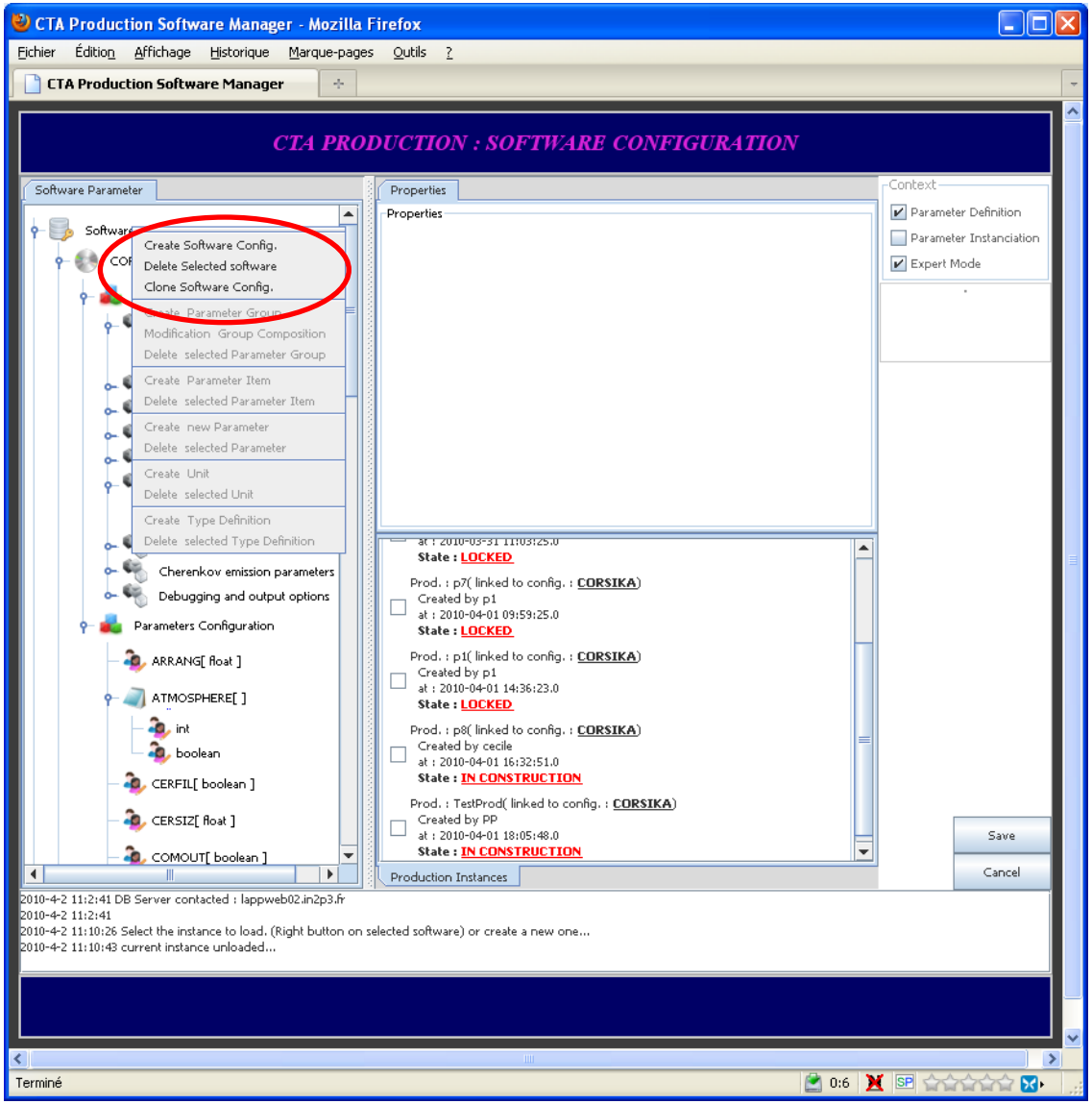
- **Web interface developed in Java at LAPP enabling for each software:**
  - Software parameters creation (lists, one by one)
  - Setting of values for these parameters (some of them mandatory)
  - Cloning of a software configuration (ex: need to add a few parameters when new software released)
  - Cloning of all parameters values (ex: need to run the same software with different input values)

- **Central database automatically updated upon changes**

  2 available levels:
  - Administrator level with full access to the tool
  - User level allowing to change only a few parameters (chosen by the administrator)

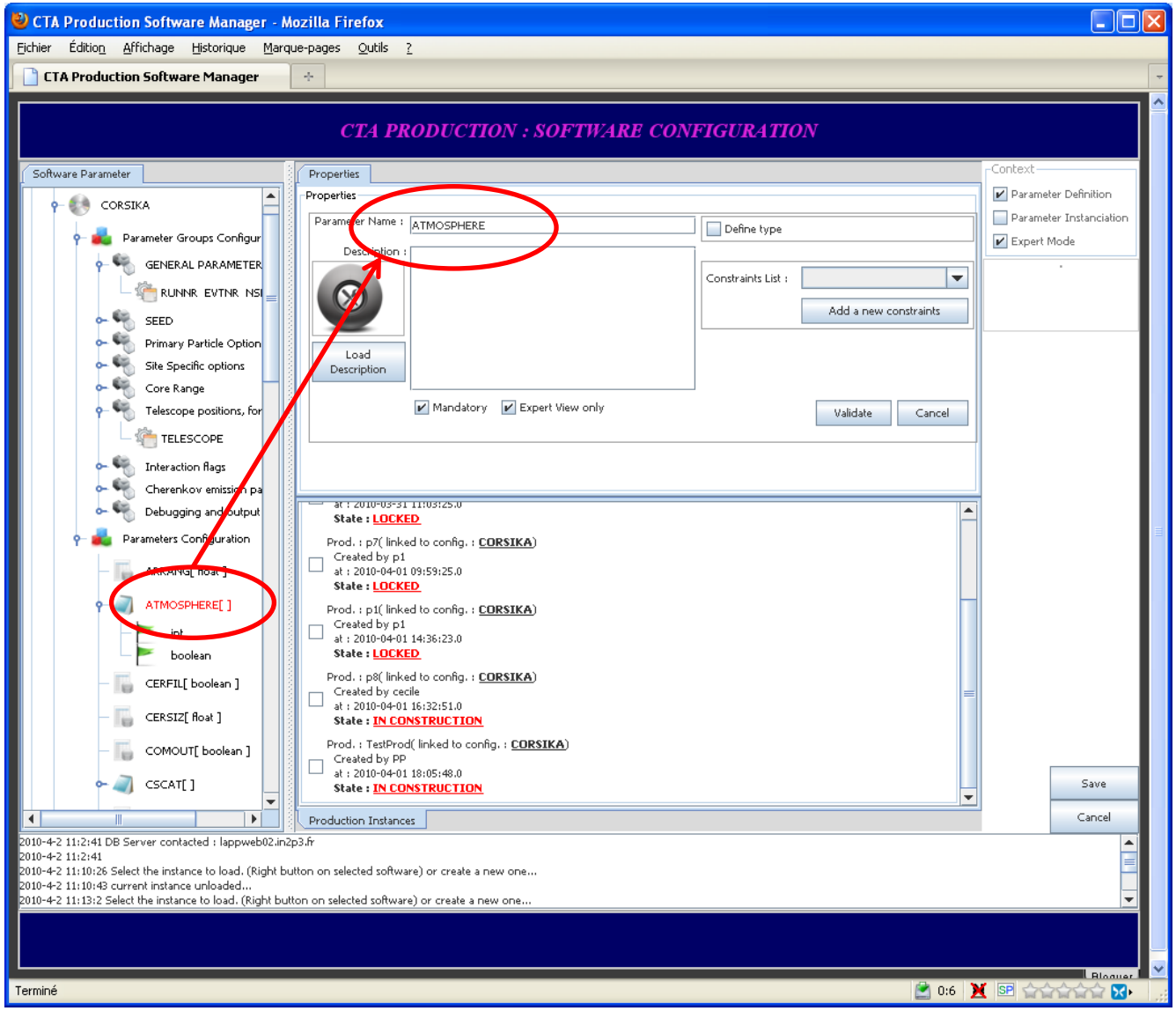
Software configuration tool

Create or clone a software configuration



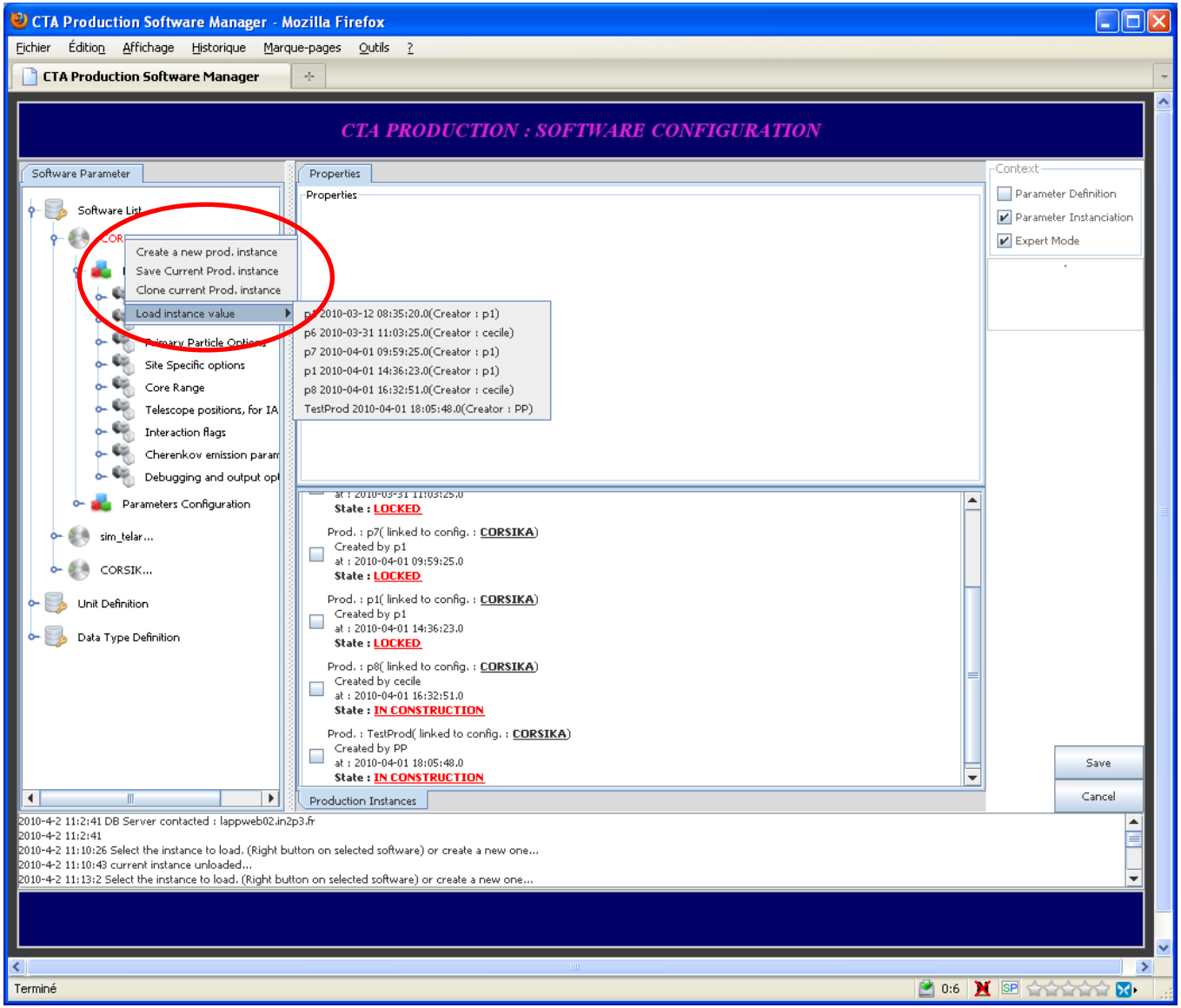
Software configuration tool

Create parameters and groups of parameters in the selected software configuration



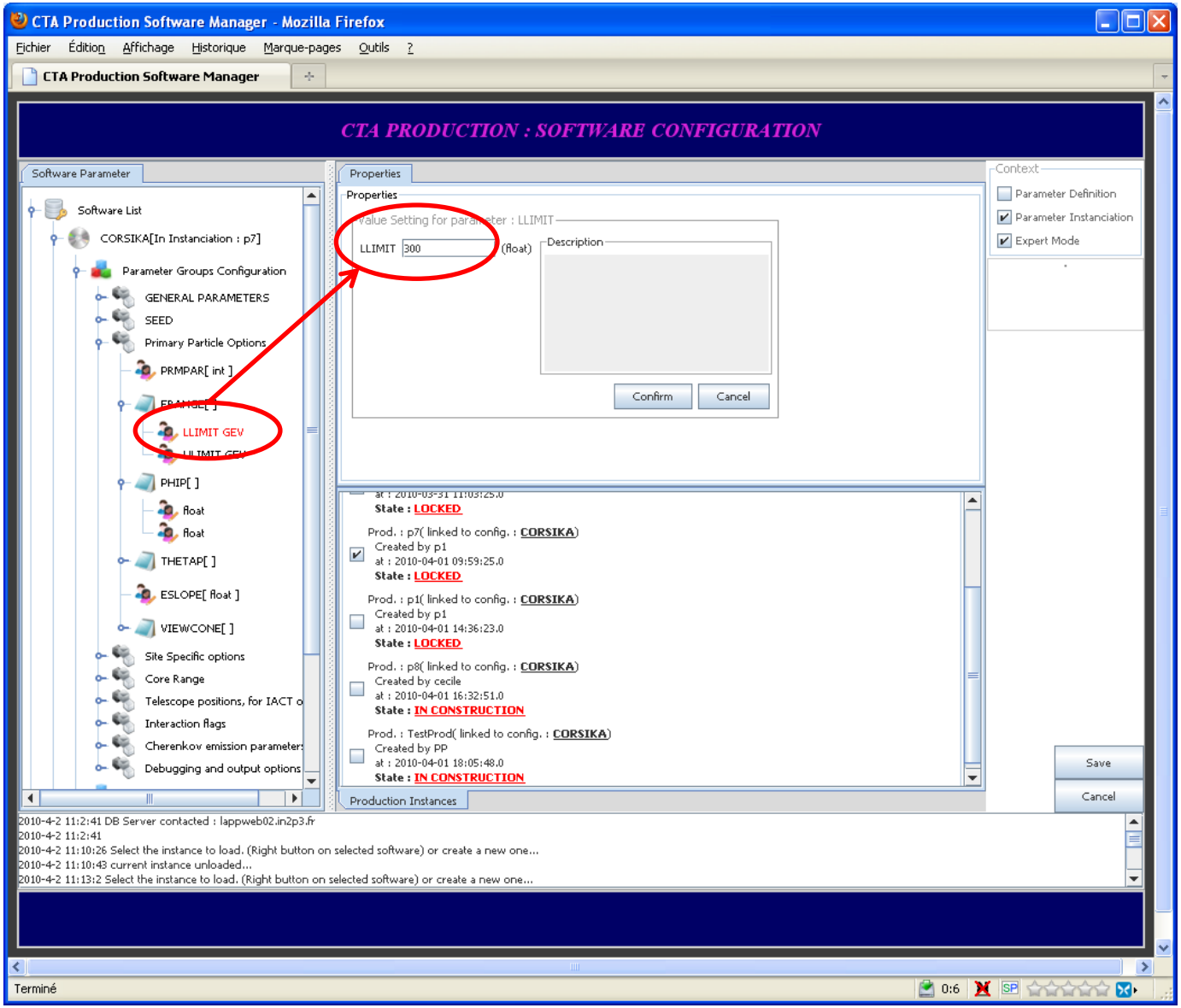
Software configuration tool

Load, create or clone
a production instance



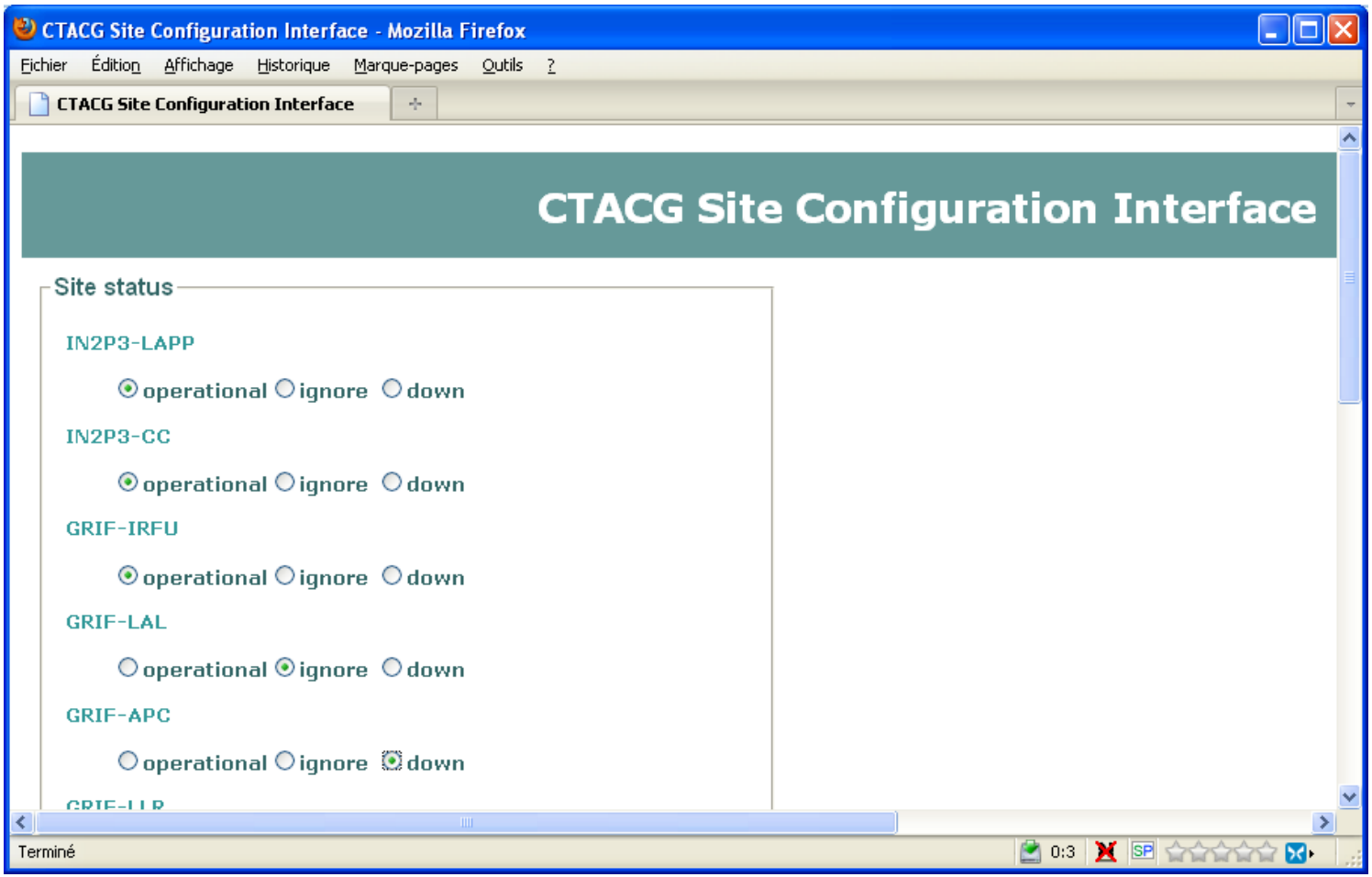
Software configuration tool

Set values to mandatory parameters of a production instance



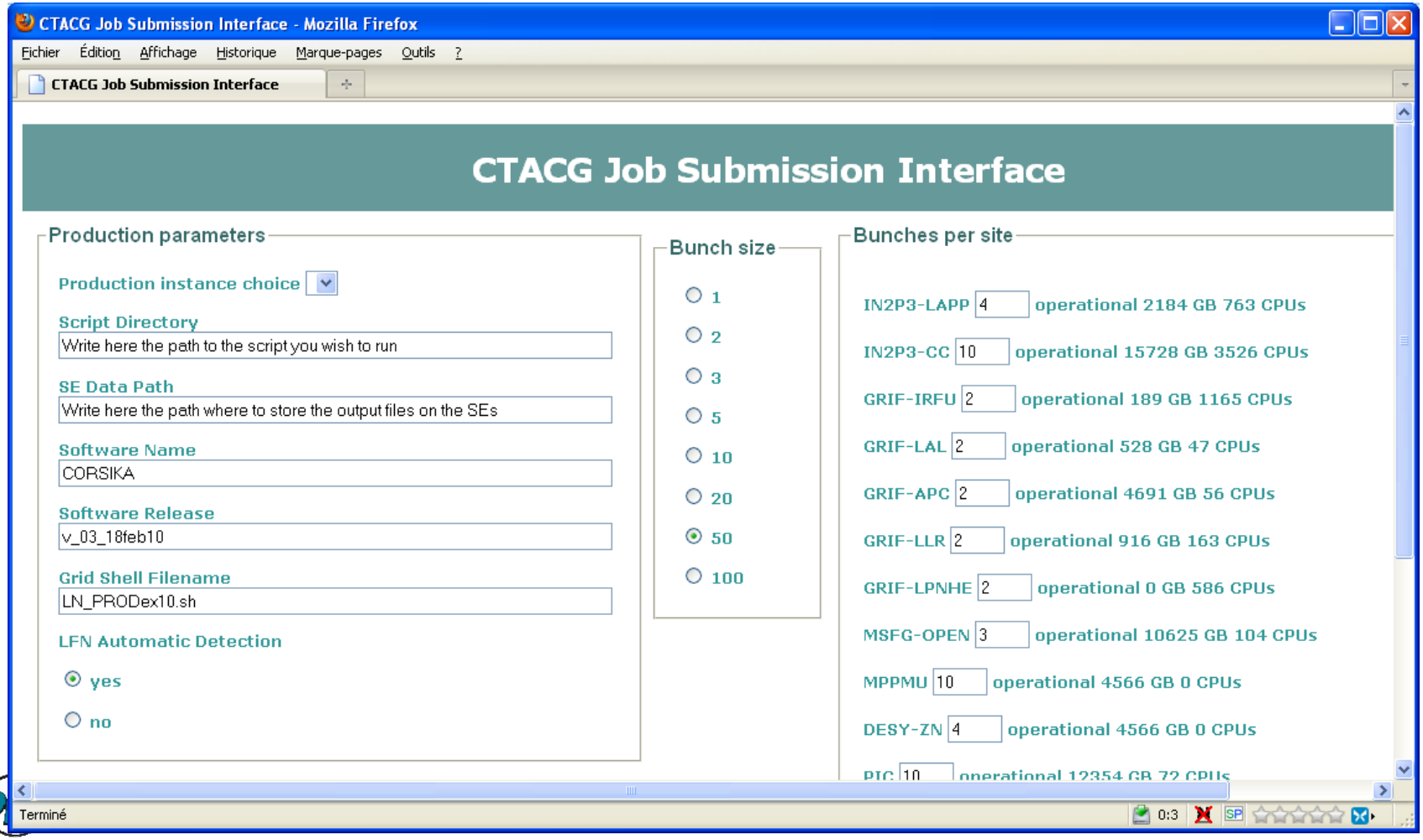
User interface: site configuration

- Enables the administrator to choose on which sites to run according to all information he has (EGEE broadcast, failed jobs on one site, ...)



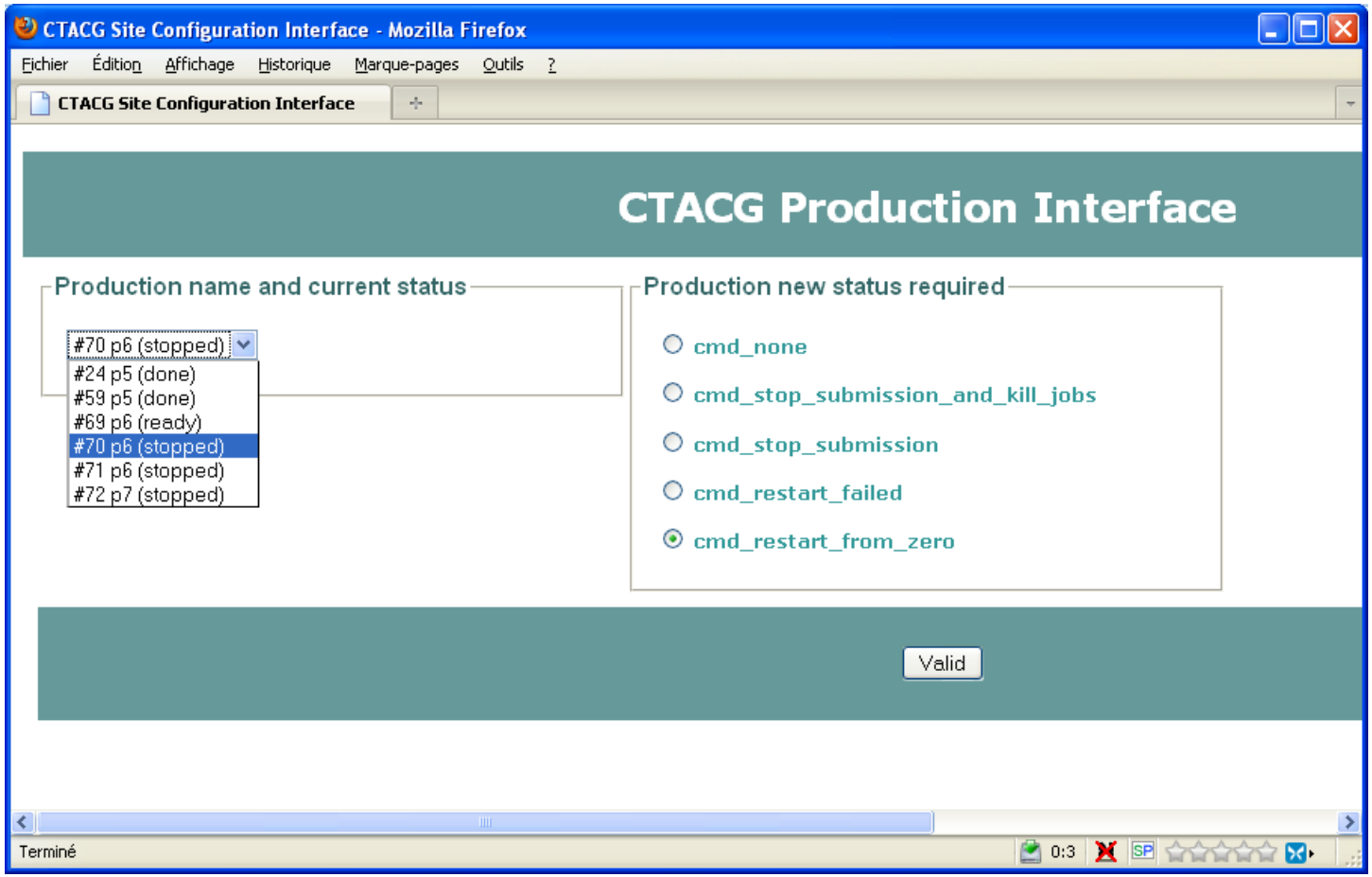
User interface: jobs submission

- Enables the administrator to choose scripts and executables to be submitted, number and subsets of jobs to run according to available resources; to start the jobs submission.



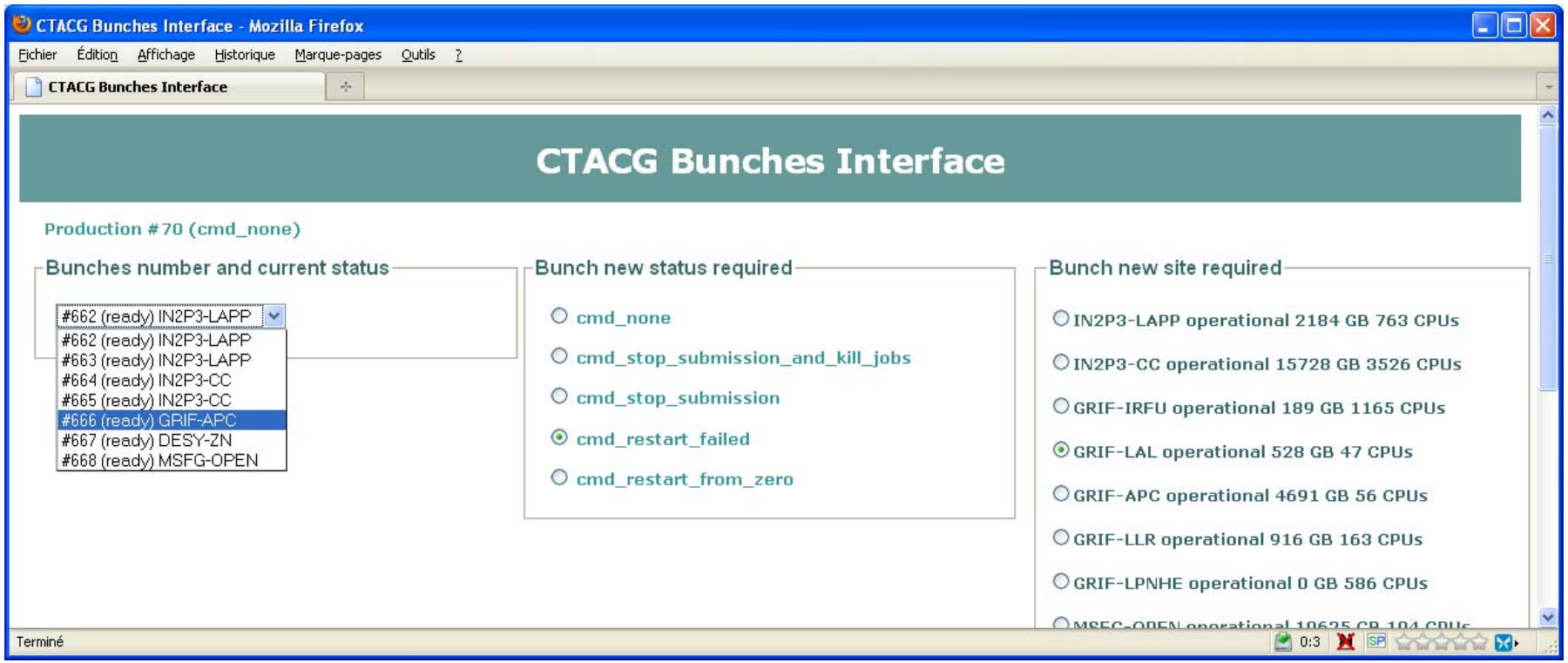
User interface: production

- Enables the administrator to select an existing production and change its status.

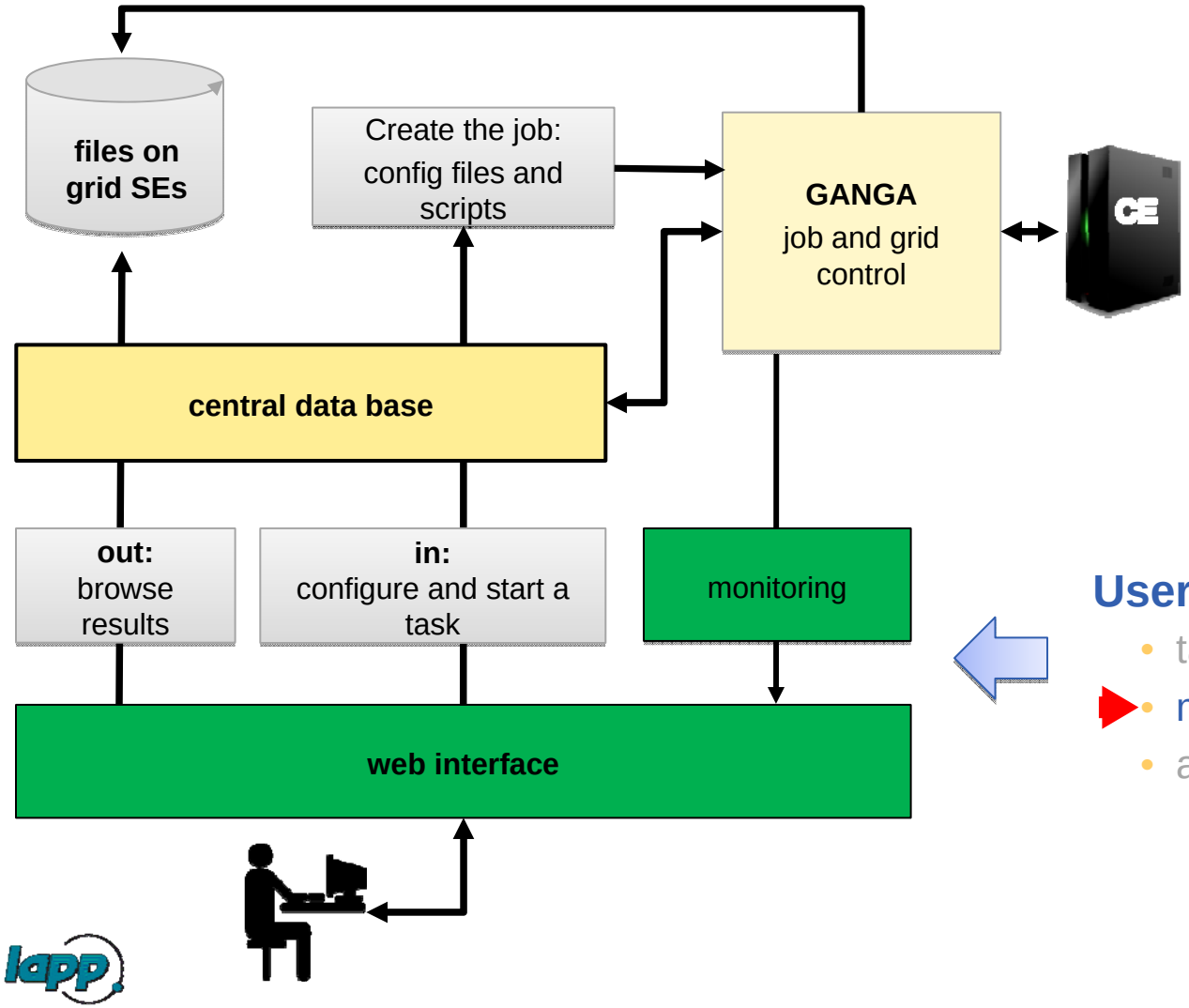


User interface: subsets

- Enables the administrator to select a subset of jobs among an existing production and change its status.



Monitoring

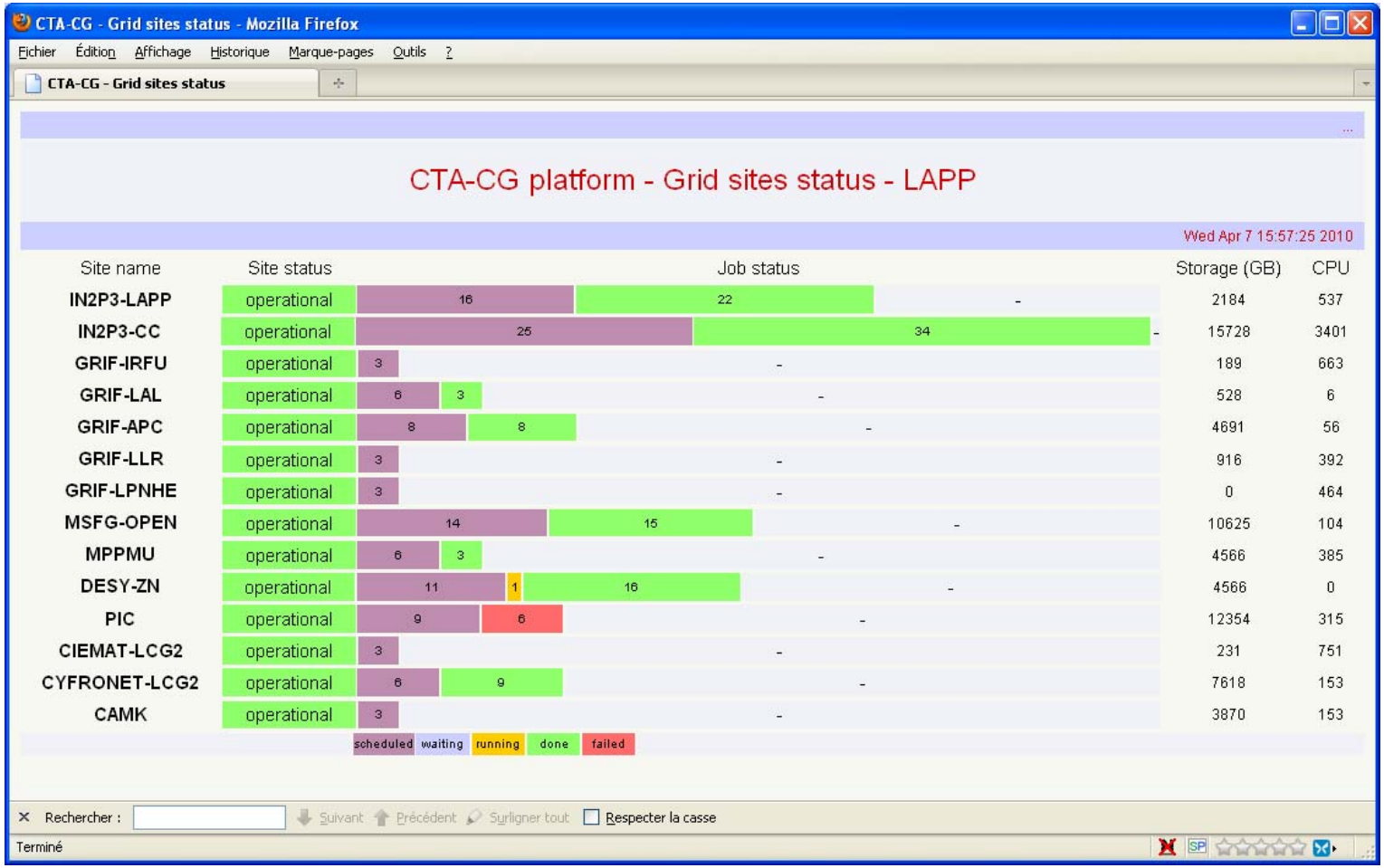


User Interface

- task configuration
- monitoring
- access to result files

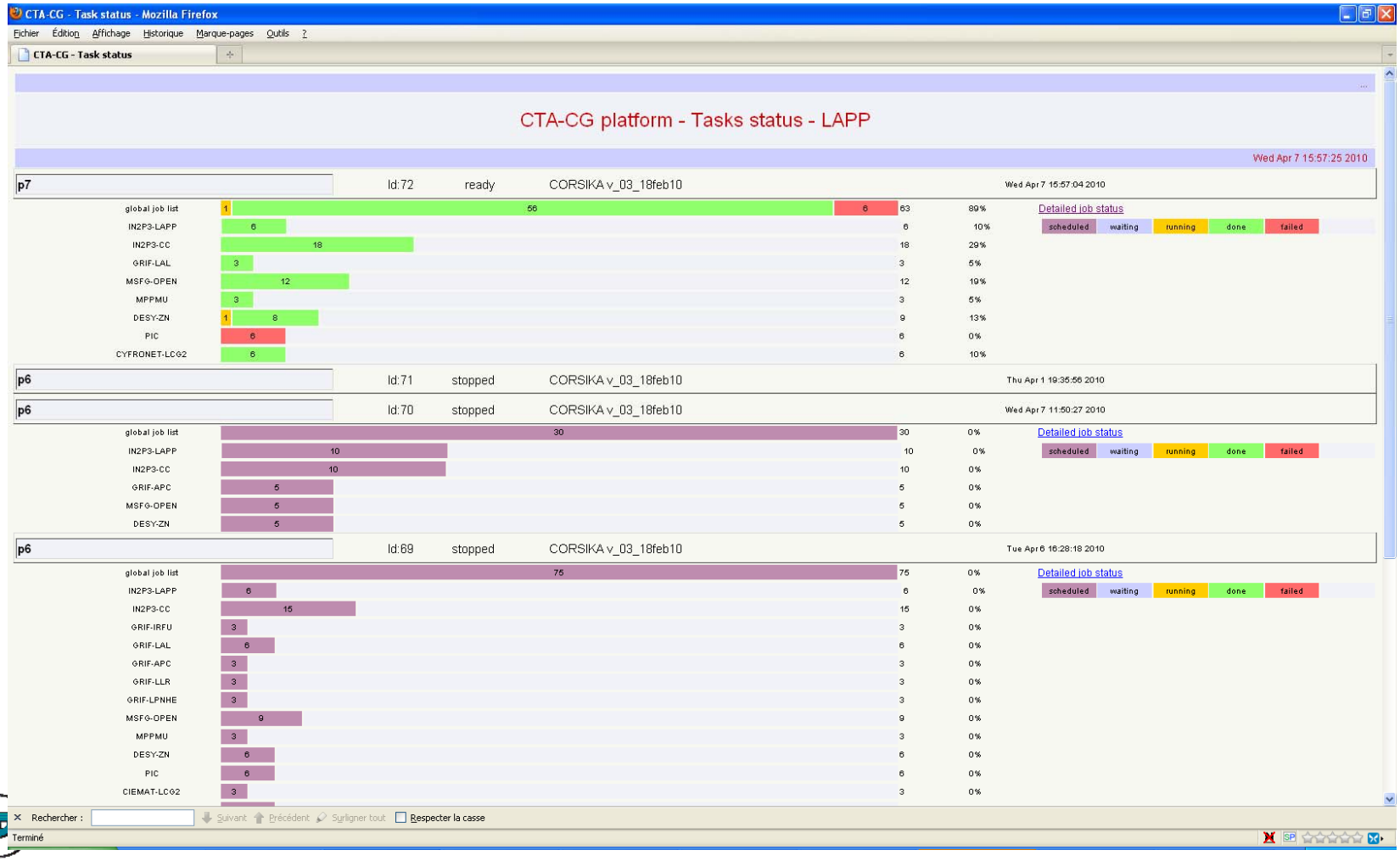
Sites Monitoring

- Display of site status, storage and CPUs available at each site
- Detailed job status per site



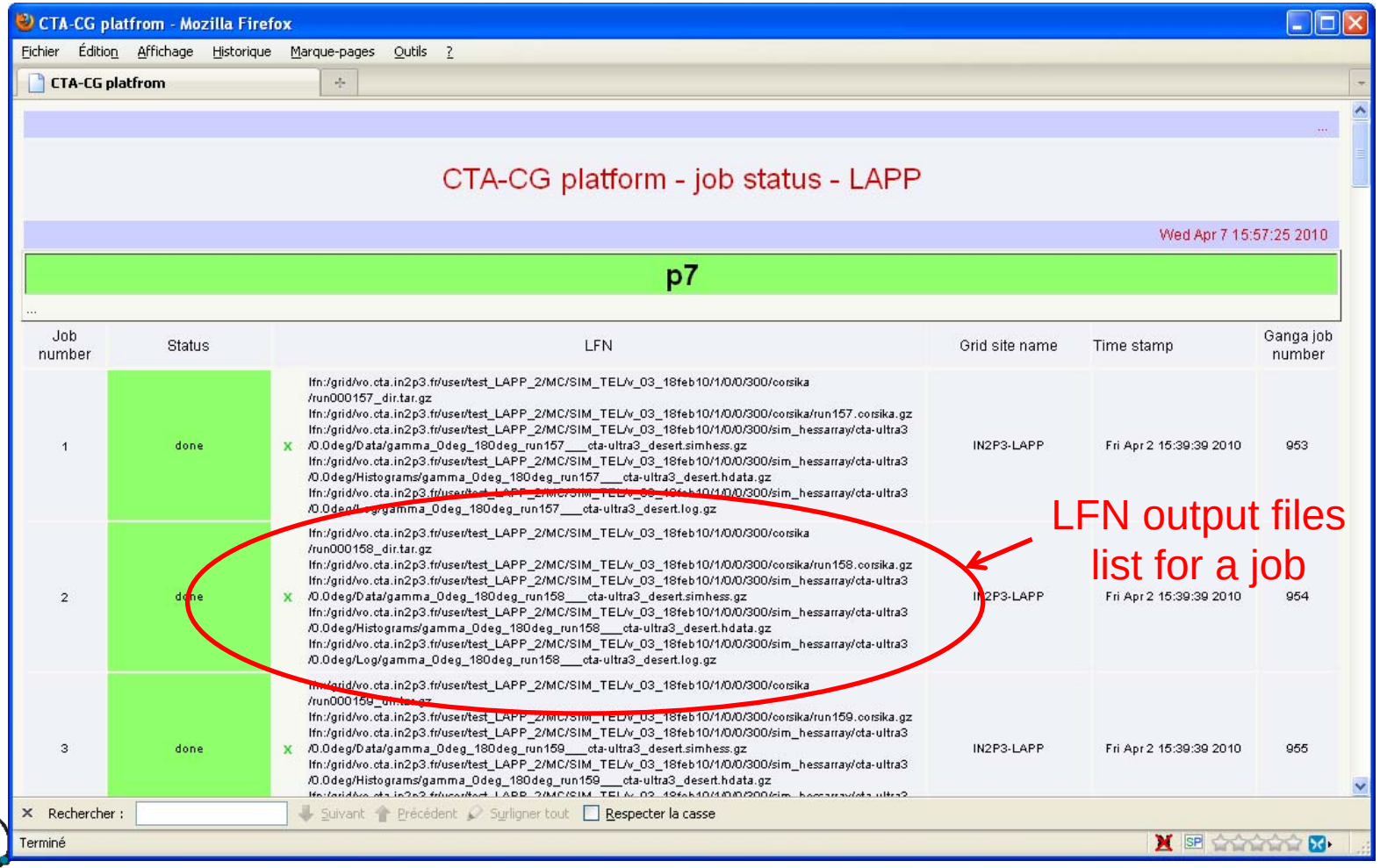
Tasks Monitoring

- Display of tasks status (global and per site)
- Detailed job status per site

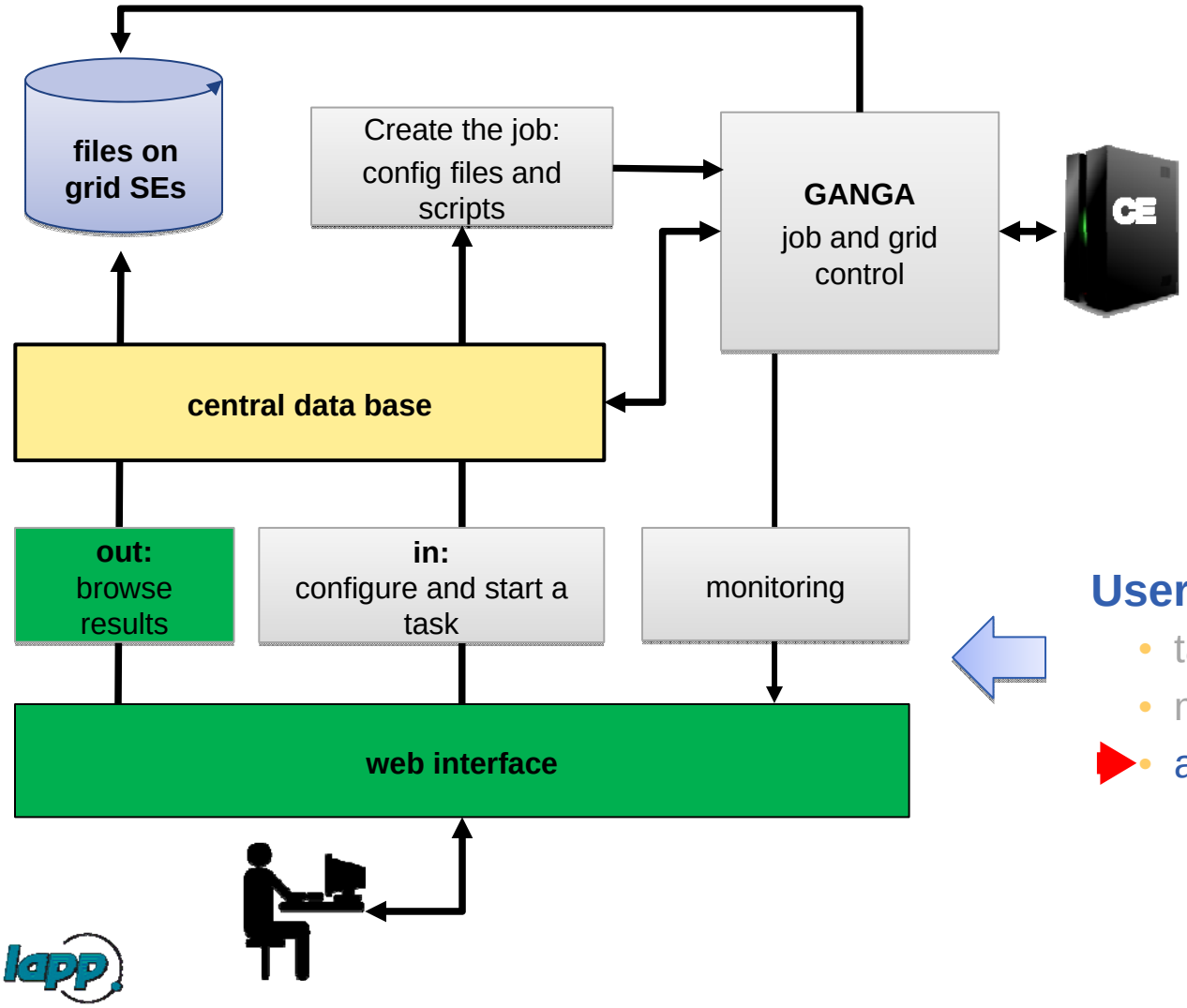


Detailed job Monitoring

- Display of jobs status within a task
- Detailed list and status of output files in SEs



Access to result files



User Interface

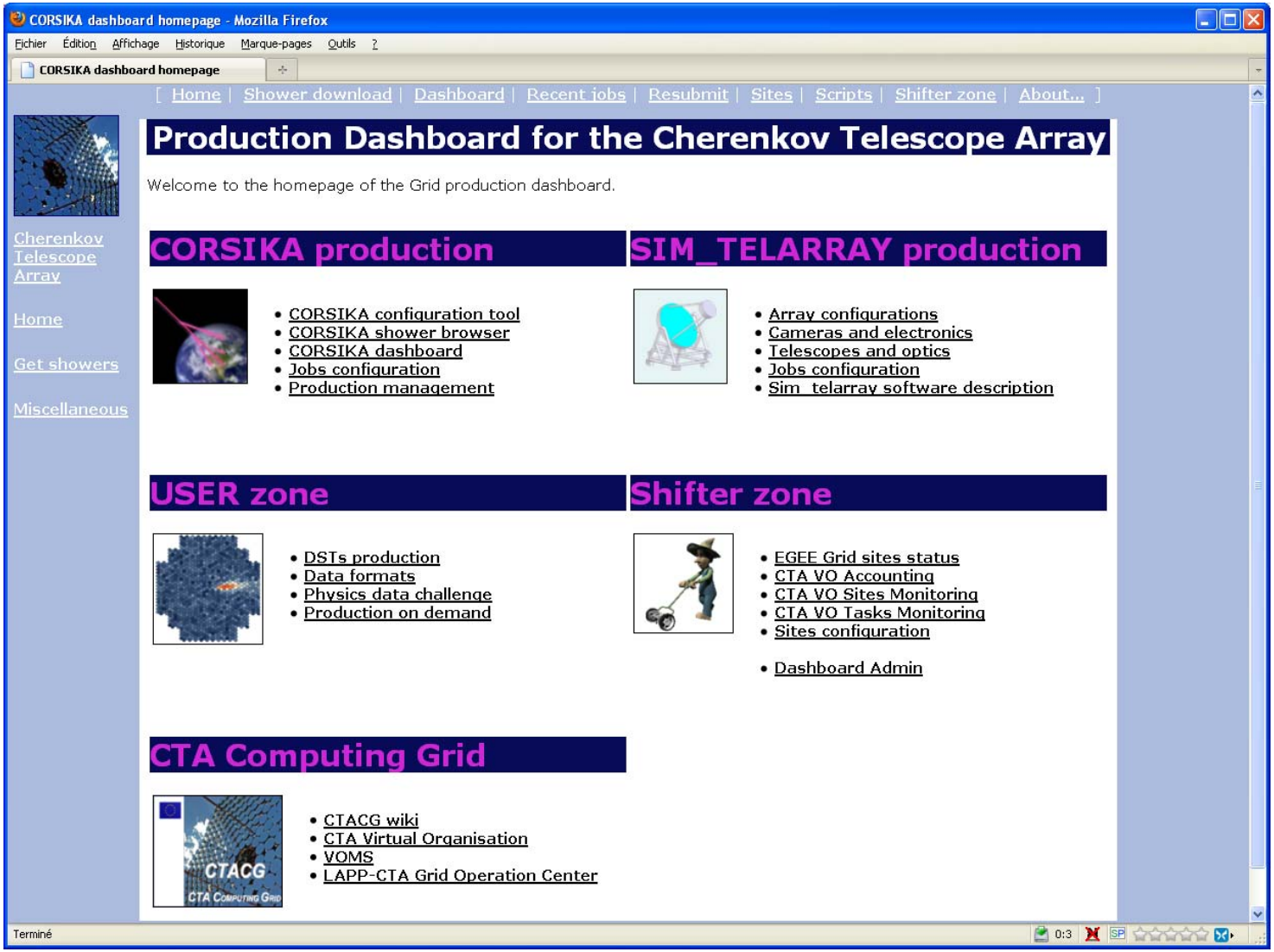
- task configuration
- monitoring
- access to result files

Files download

- Search engine through a web interface:
  - Users can list files matching the research criteria
  - In the future, they will be able to download these files
- **Status**: in progress (inspired by Auger experience)

CTACG dashboard

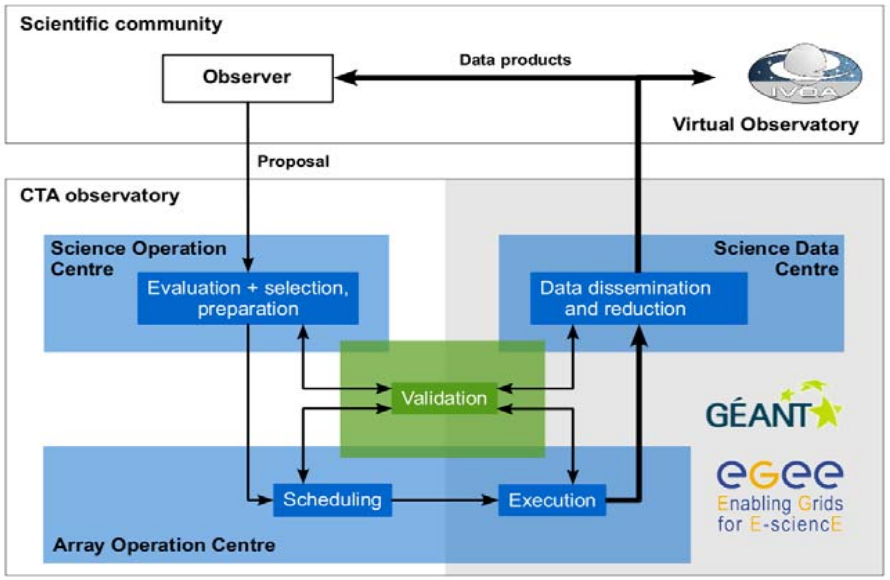
- Inspired by the Auger dashboard (4th EGEE User Forum Catania)
- Used to access and run all the previous tools from a unique web page
- **Status**: operational, not yet online



CTA E-Infrastructures

The CTA Observatory main logical units :

- **Science Operation Centre:** organisation of observations
- **Array Operation Centre:** the on-site service
- **Science Data Centre:**
  - Software development
  - Data analysis
  - Data reduction
  - Data archiving
  - Data dissemination to observers



Total expected data volume from CTA: 1 to 10 (?) PB per year

(main data stream for permanent storage is of the order of 1 (10 ?) GB/s)

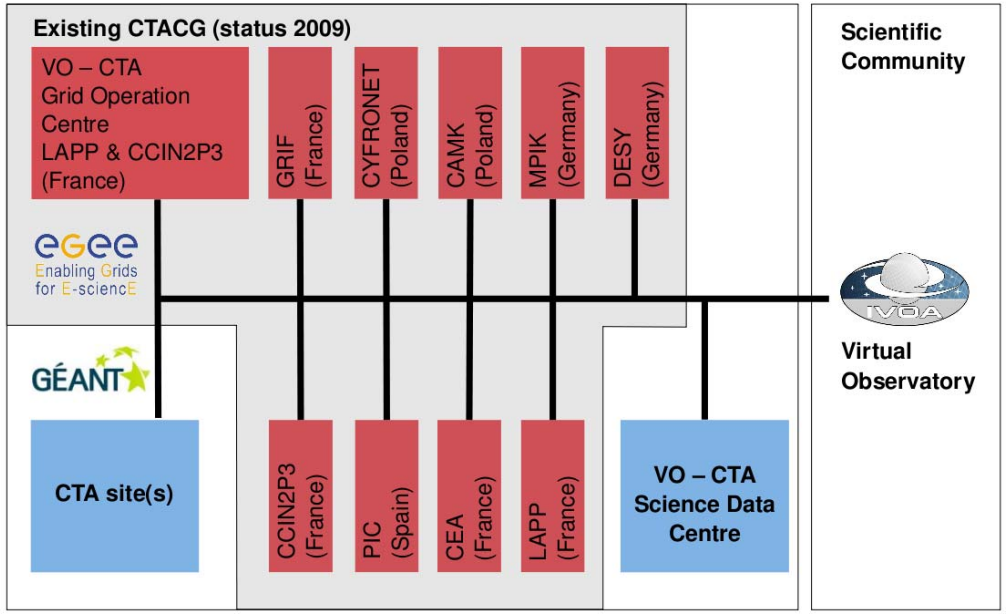
MC requirements: tens of CPU years, hundreds of TB

Existing ICT-based infrastructures, such as EGEE and GEANT, are potential solutions to provide the CTA observatory with best use of e-infrastructures.

Workflow and "logical sites"

Development of an integrated application of e-infrastructures like EGEE-Grid, GÉANT and IVOA for the architecture of the CTA observatory and…
Deployment of end-to-end services in support of a wide scientific community for open data access for CTA and VHE gamma ray astronomy.

EGEE-Grid infrastructures and EGEE-Grid middleware for distributed data storage, data analysis and data access are considered the most efficient solution for the CTA e-infrastructure (CTACG).



A dedicated CTA Work Package

1) Data Management System and Data Pipelines

(deliverables: framework, data applications and data products)

The data management system begins at the data acquisition interface and flows through to the data products accessed by end users

2) Data Management Facilities

(deliverables: network, computing model, global system and specific roles of facilities)

The data flow, on the way, moves through different facilities supporting data management according to different tasks, as well as end user sites that may conduct science using CTA data.

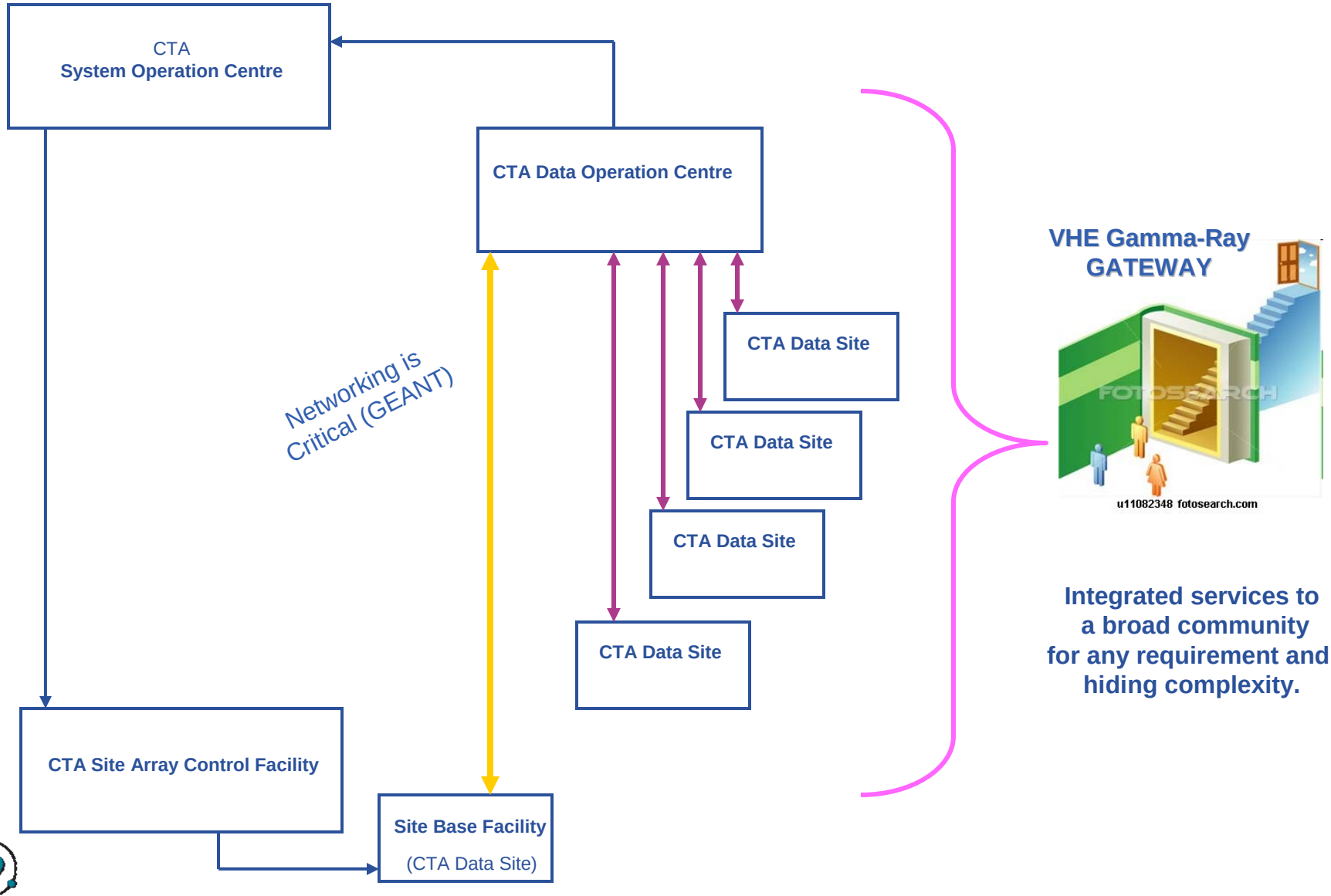
3) Data Access

(deliverables: services for data and computing sharing and public openness)

End users and end user sites that may conduct science using CTA data require different level of data access and different tools, resources and services to be part of the community (VHE gamma ray gateway). Further more users want to pipeline resources on their own computing infrastructure.

-> Development for USER REQUIREMENTS

A possible workflow



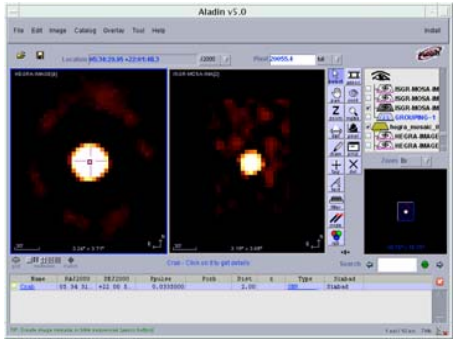
… toward a "VHE γ-ray Science GATEWAY"

VSG, through a Web-interface (… Virtual Machine) will provide:

- A complete, portable and easy to configure user environment for developing and running CTA data analysis locally and on the Grid (login, proxy on demand, selecting data level, data sets, software, and computing);

- Grid services for jobs and resources monitoring, archive access, telescope and observations monitoring, and support to scientist to port their software to the GRID;

- MC datasets and open production for deeper analysis of intermediate level data, On-the-fly reprocessing from the Archives Instruments calibration monitoring and trend analysis;



- Astronomical tools and services integrating VObs standards and gLite-EGEE middleware for analysis of high level data sets;

- Tools and IVOA interfaces for multi-wavelength A&A analysis;

- Forum and scientific documentation for an up-to-date scientific outreach and strengthening of the VHE gamma ray community.

VHE γ-ray Science GATEWAY

This project conjugates the long-term presence and experience in EGEE/EGI of the particle physics community and the activities of the astronomical community mainly devoted to the Grid solutions for public access and archive of observatory data.

Major challenges in CTA e-infra. applications shared between two communities :

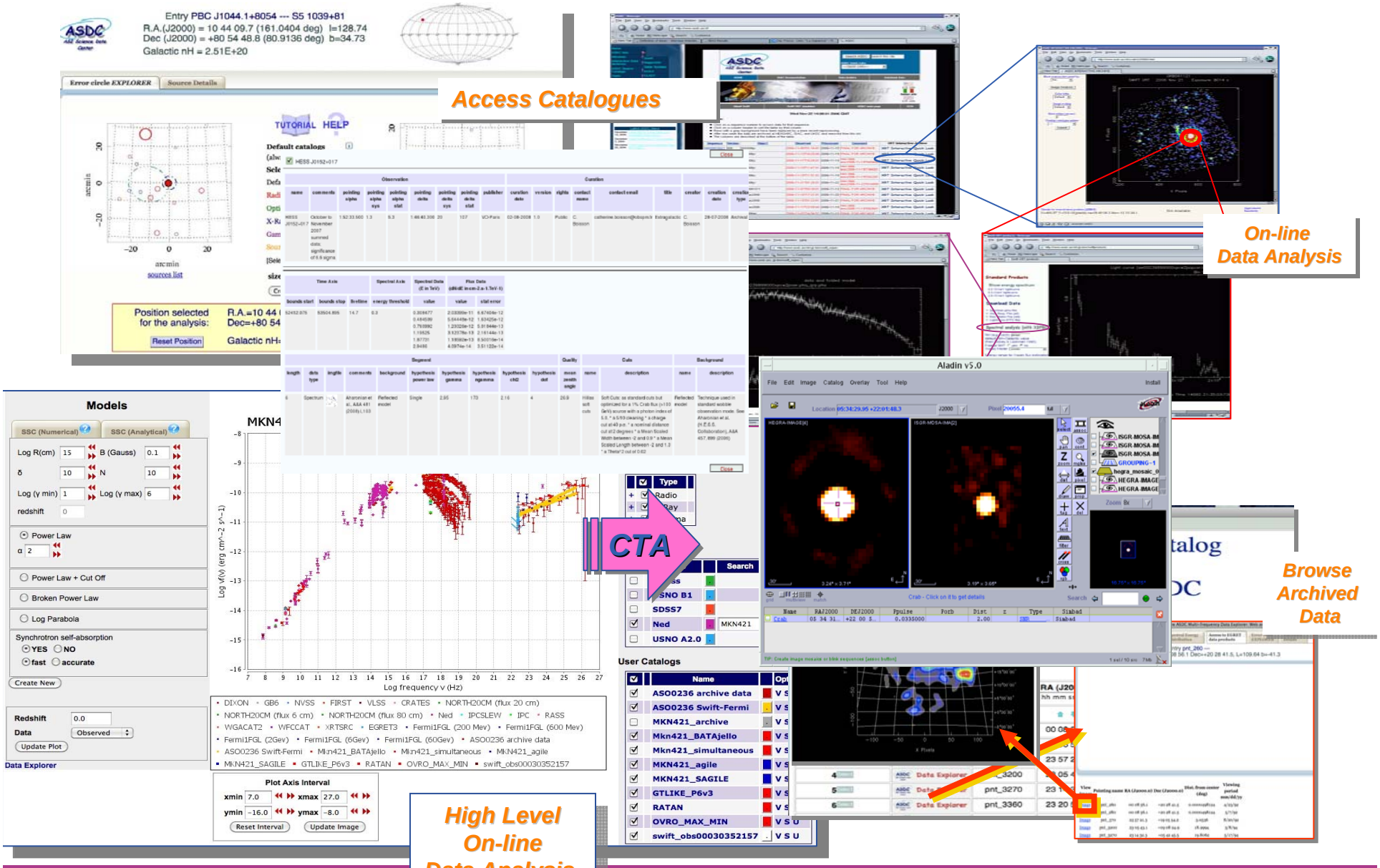
A&A

- Astrophysical data unified content descriptors
- Vocabulary for astronomical quantities extended to VHE gamma rays
- Astronomical catalogs and metadata
-Standard grammar for database queries

PP

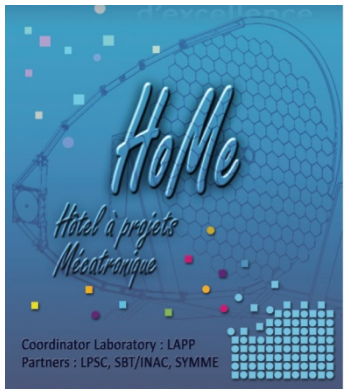
- Hierarchical data model definition and corresponding implementation of software and computing solutions (for processing and storage)
- Evolution towards open software and data access
- Development of Grid Analysis platform at all levels (Virtualization)
- Implementation of EGEE Grid services for the community.

From low level data …To high level data and (MWL) data archive



CLOUD

… but which one ?



HoMe project

Partners : **LAPP**, LPSC, SBT/INAC & SYMME

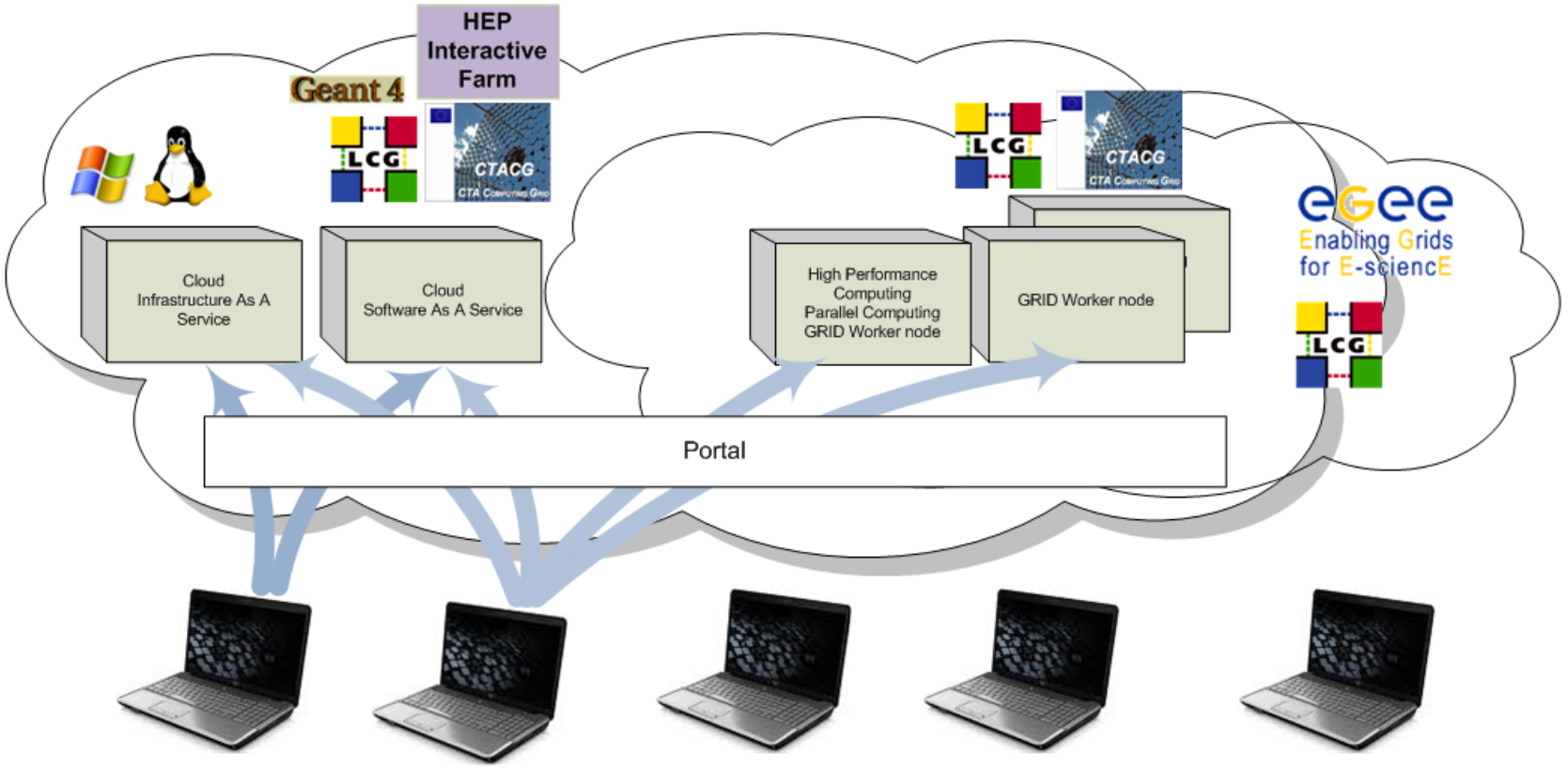
Computing work package : GRACIA (Grid And Cloud Infrastructure for the Alps)

evolve from current EGI grid and high performance computing infrastructure to cloud flexible services to better meet the research community users and socio-economic partners computing requirements

Computing is seen as services :

- Software as a service (SaaS): complete application offered as a service on demand.
- Infrastructure as a service (IaaS): delivers basic storage and computing capabilities as standardized services over the network (based on a on-demand virtual images deployment system)
- Platform as a service (PaaS): encapsulates a layer of software and provides it as a service that can be used to build higher level services

Computing infrastructure



Socio-économic partne

Scientists from University, Research laboratories, International Research Collaborations

One platform suitable for cloud, high performance and grid computing.

Conclusion

- **CTACG:**
  - Current developments fulfil requirements
  - CTA MC Production successfully tested and running
- **Perspectives:**
  - Further adaptation of these tools to other CTA software (e.g. data analysis)
  - Extension of Ganga tools
  - Provide public access to data and software

- **“CORE”: CTA Observatory E-infrastructures:**
  - Development of the CTA Data Management System
  - Data access from low (including computing resources access) to high level (interface with IVOA and running MWL VObs data on EGEE-middleware)
  - Towards a distinction between **Data Operation Centre** and **Science Data Centre** (more virtual than real, e.g. interactive services to be used by any “classical” data centre and … by a CTA SOC ?)

- **Further questions/comments:**
  - http://lappwiki01.in2p3.fr/CTA-FR/doku.php?id=cta-computing-grid-public
  - ctacg-support (at) lapp.in2p3.fr