CTA report

<u>Luisa Arrabito</u>, Ricardo Graciani, Claudia Lavalley

DIRAC User Workshop 29th – 31st October 2012, CPPM - Marseille

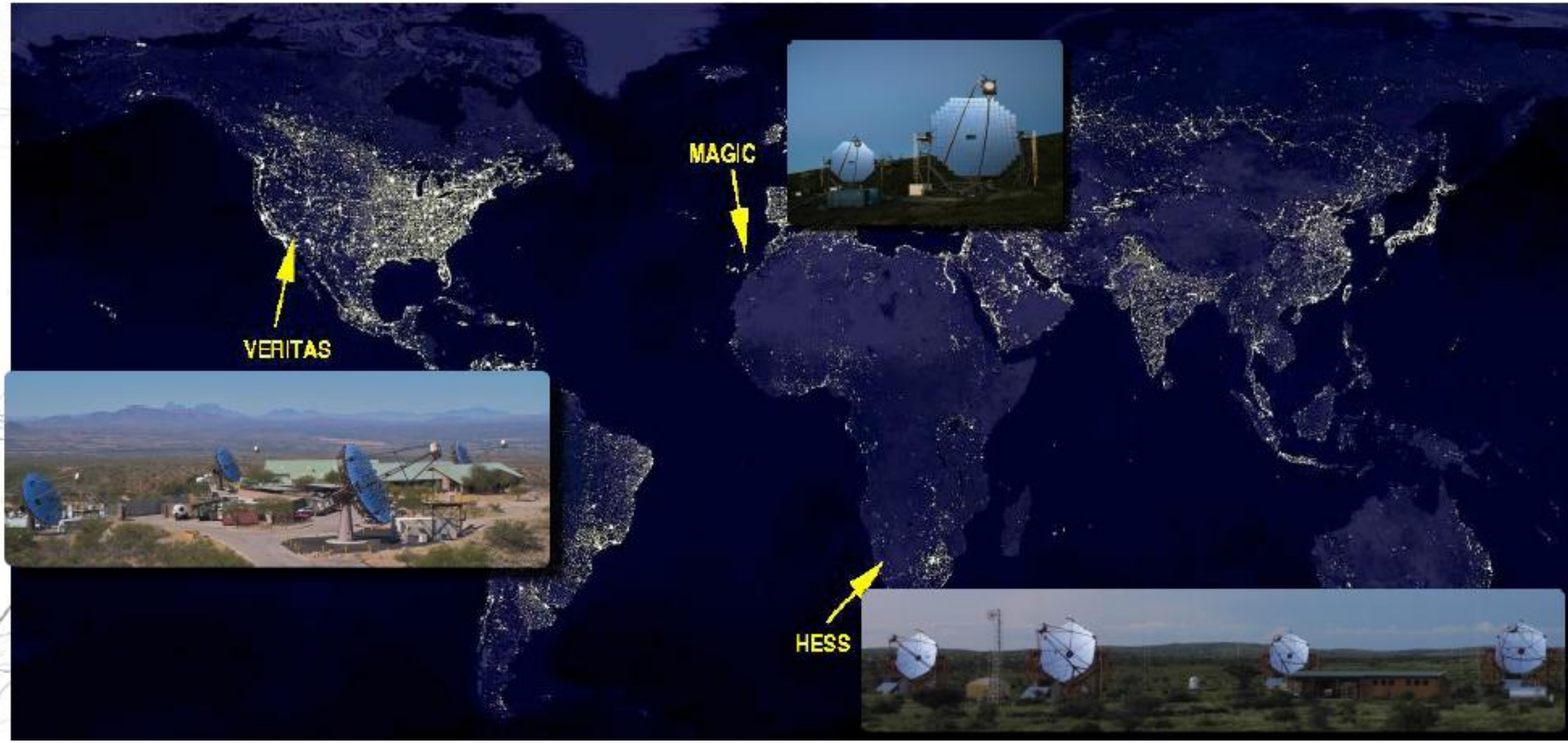
CTA report

Outlook:

- CTA project
- CTA computing
- MC simulations on the Grid
- Current use of DIRAC
- Conclusions and perspectives

Gamma-Ray Astronomy

- Astronomy at High Energy > 10s GeV
- Currently **3 experiments with 2-5 Cherenkov Telescopes**:
  operational since 2003 (H.E.S.S.)
- Discovery of hundred's of sources in the last 8 years
- Observation of non-thermal universe:
  supernova remnants, pulsar wind nebulae, galaxies

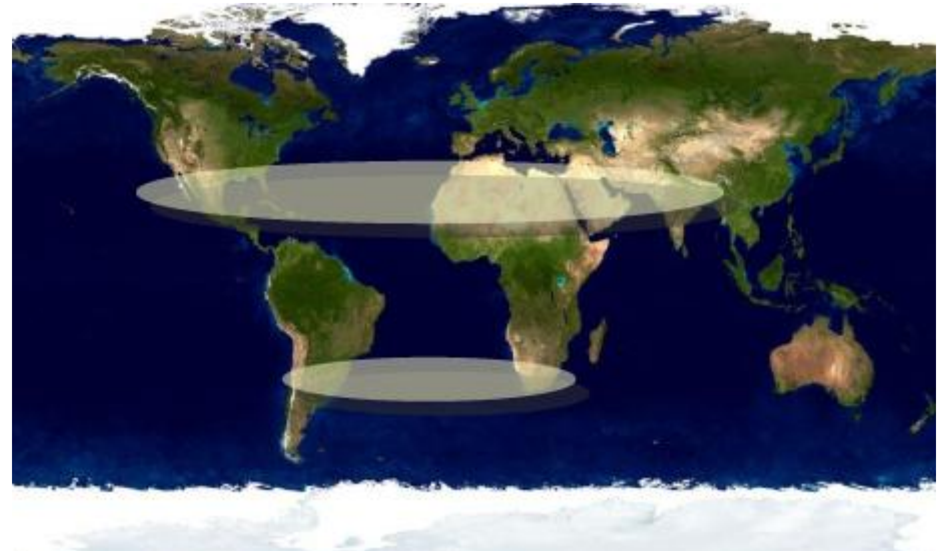


Next generation: Cherenkov Telescope Array (CTA)

Cherenkov Telescope Array

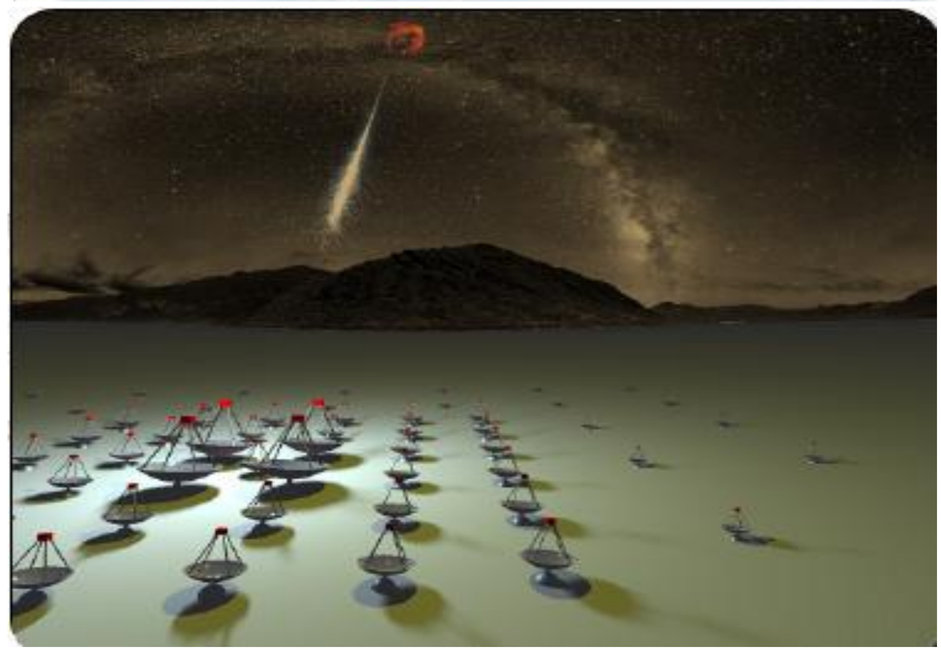
- **2 arrays of 50-100 Cherenkov Telescopes of 3 different sizes in each hemisphere**
- **Enhanced performances:**

- 10x sensitivity with respect to current experiments
- Energy domain HE/VHE (20 GeV-200 TeV)
- Improved angular resolution and collection area



Scientific goals

- Investigation of the **origin of the Galactic Cosmic Rays**
- Astrophysical phenomena (e.g. AGNs, GRBs)
- Fundamental physics and cosmology



Next generation: CTA

CTA Project

- Consortium of > 800 scientists in 25 countries
- Currently Preparatory Phase (2011-2013):

- Find sites
- Build prototypes of telescopes
- Define optimal array layout
- ***Prepare data pipeline and data distribution***

- The construction of the full array planned to be completed around 2018
- Operate as an observatory

CTA Computing tasks

Preparatory Phase:

MC simulation:

- Study of telescope and array design
- Algorithm optimization

Operation Phase:

- Raw data acquisition on remote site
- Data calibration and reconstruction
- Analysis and data publication
- MC simulation:
  - optimization of event selection cuts
  - generation of acceptance areas
  - energy reconstruction and resolution tables
  - point spread function

CTA Computing: Operation Phase

Computing needs

- **Data rate: 0.1 – 13 GB/s**
- Total amount of Raw Data: **2 – 25 PB/ year**
- **Processing time** for 1 hour of observation (see Table below)

	Raw data rate (GB/s)	Calibration (HS06h)	Reconstruction (HS06h)
CTA South Array	0.3- 4.0	3000-20000	6000-20000
CTA North Array	0.1-1.3	1500-6000	3000-6000

Additional Computing power will be needed for:

- Periodical data re-processing
- Real-time data pre-calibration and reconstruction
- MC simulations
- High-level data analysis

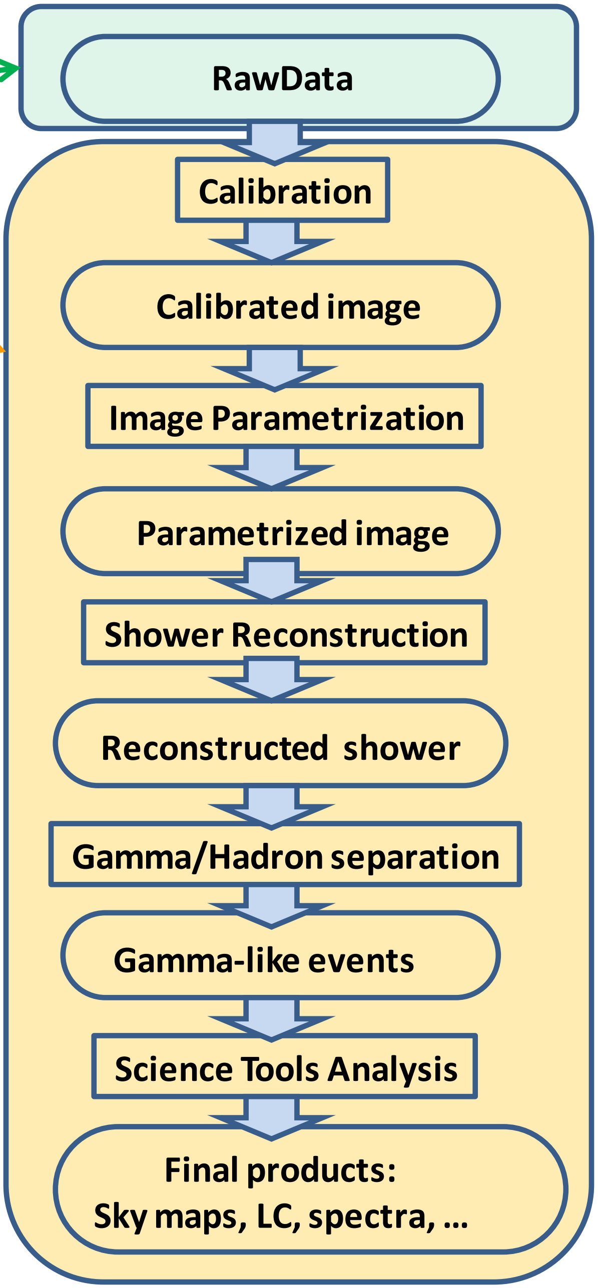
-> Overall computing time still not estimated

- ***No computing model yet:***

  -> Distributed Computing approach will be explored for CTA operation

CTA Computing: MC simulations

MC simulation:
- Production
- Analysis



Production

1) Air showers and Cherenkov light
- corsika sw [http://www-ik.fzk.de/corsika/]

2) Simulation of telescope response
- sim_telarray sw [Bernlohr 2008, AIP Conf. Proc., Vol. 1085, 874-877]

Analysis

- No official CTA analysis software is available
- Analysis sw coming from the current IACTs:

- more than one analysis chain for each experiment (HESS, MAGIC,VERITAS)
- some of them are proprietary software

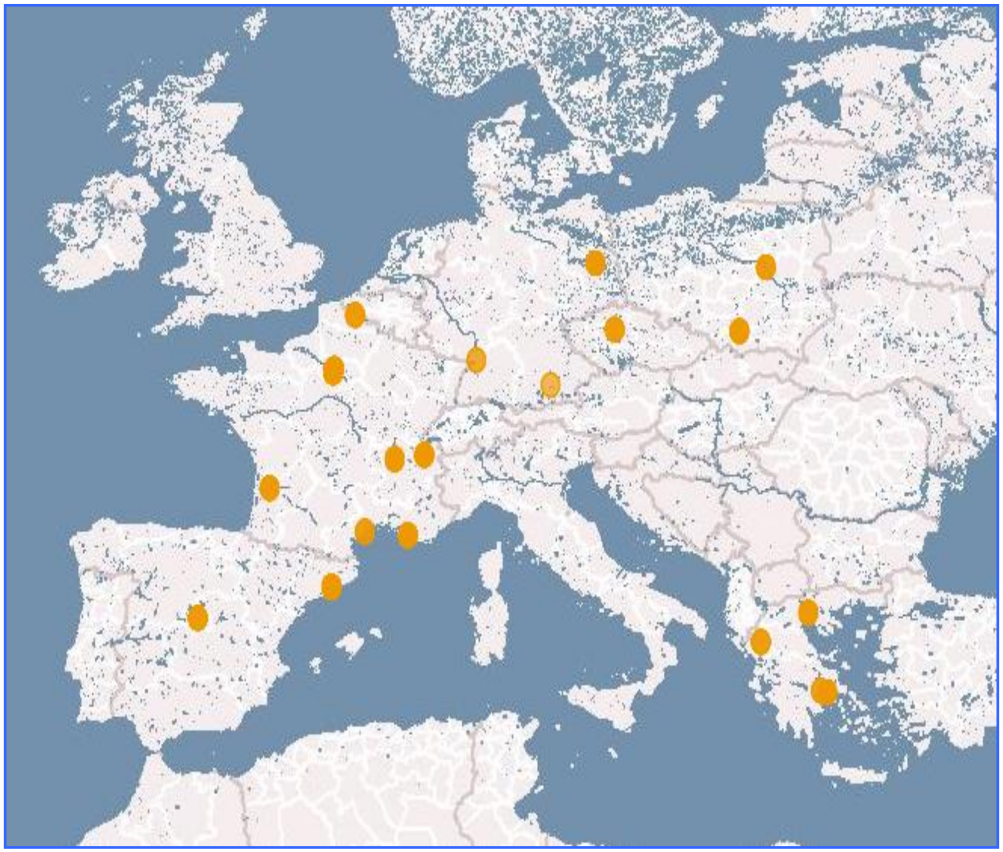
Computing needs for a typical MC production

- **~150k jobs**
- corsika + sim_telarray in a single job (about ~3h)
- producing **about 200 TB**

MC simulation on the EGI Grid: past experience

Grid used for MC productions since 2008
- Submission and configuration tool based on GANGA
- Parallel productions on local clusters

17 EGI Sites spread in 6 countries



Computing resources
1000 – 2000 CPUs simultaneously available based on past experience

Storage resources
- **390 TB** distributed in all Sites and shared with other VOs
- **430 TB dedicated to CTA:**

1) IN2P3-CC Disk and Tape
2) DESY-ZEUTHEN
3) CYFRONET

MC simulations on the Grid: DIRAC

• **First proposal of using DIRAC to handle MC simulations (production and analysis) one year ago by LUPM and UB teams**

Goals:

- Optimise MC production
- Provide tools for User Analysis

Motivations for using DIRAC:

- Integrate *heterogeneous resources*
- Implementation of the *Pilot Job mechanism -> high efficiency*
- Central management of *VO policies*
- *Flexible framework* allowing an easy customization
- Handle bulk submission
- Fine grained job monitoring/accounting (web based)
- *Save MetaData* informations in the DIRAC File Catalog allowing users' queries through Metadata
- *Fault-tolerance* design for core services
- User interface (python API, command-line, web portal)

Current use of DIRAC

What's done so far?

DIRAC instance for CTA

- 3 servers at PIC:

- Central services
- Agents
- Sandboxes

- 1 server at CC-IN2P3:

- Storage Element Service for log files and Failover
- Configuration System mirror

Current use of DIRAC

What's done so far?

MC Production

- Implementation of DIRAC modules to run massive simulations (corsika+sim_telarray)
- **Use of the DIRAC File Catalog** to save Metadata (identify specific functionalities to handle CTA Data Model)
- Not tested yet at large scale, since the next MC campaign is not yet ready

MC Analysis

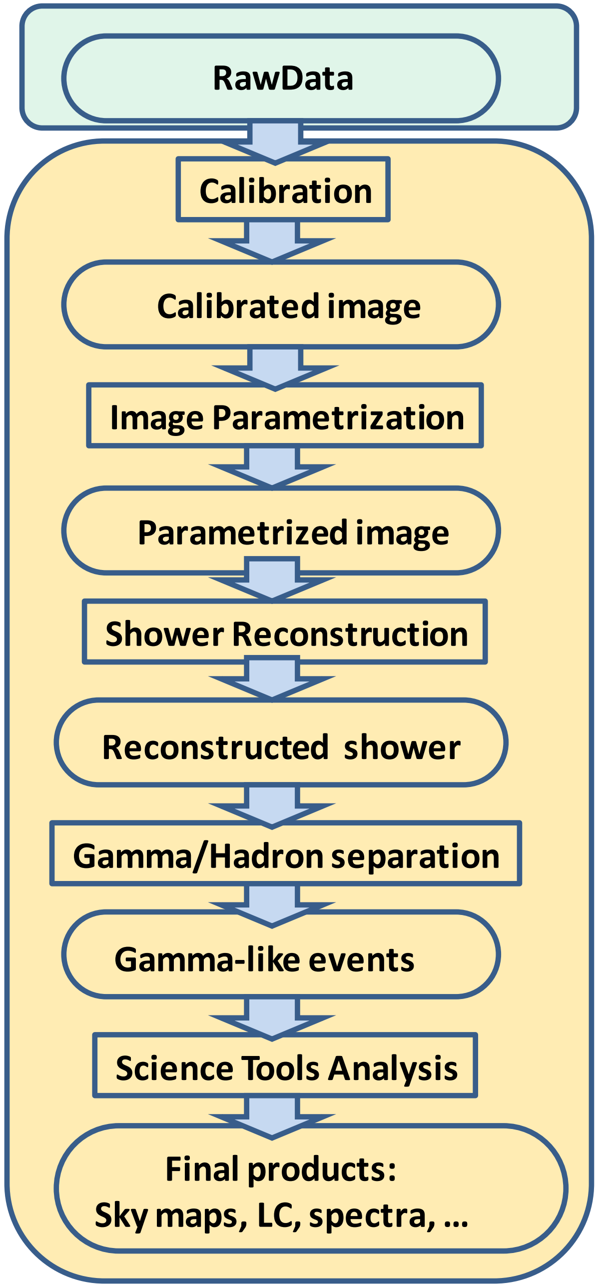
- Implementation of **DIRAC modules** to run some of the analysis chains used in CTA:

1) read_hess, HAP (from HESS)

2) Recently some work has started for the **MAGIC sw**

3) High-level analysis sw

- Work on **end-user documentation**
- Got first user feedback

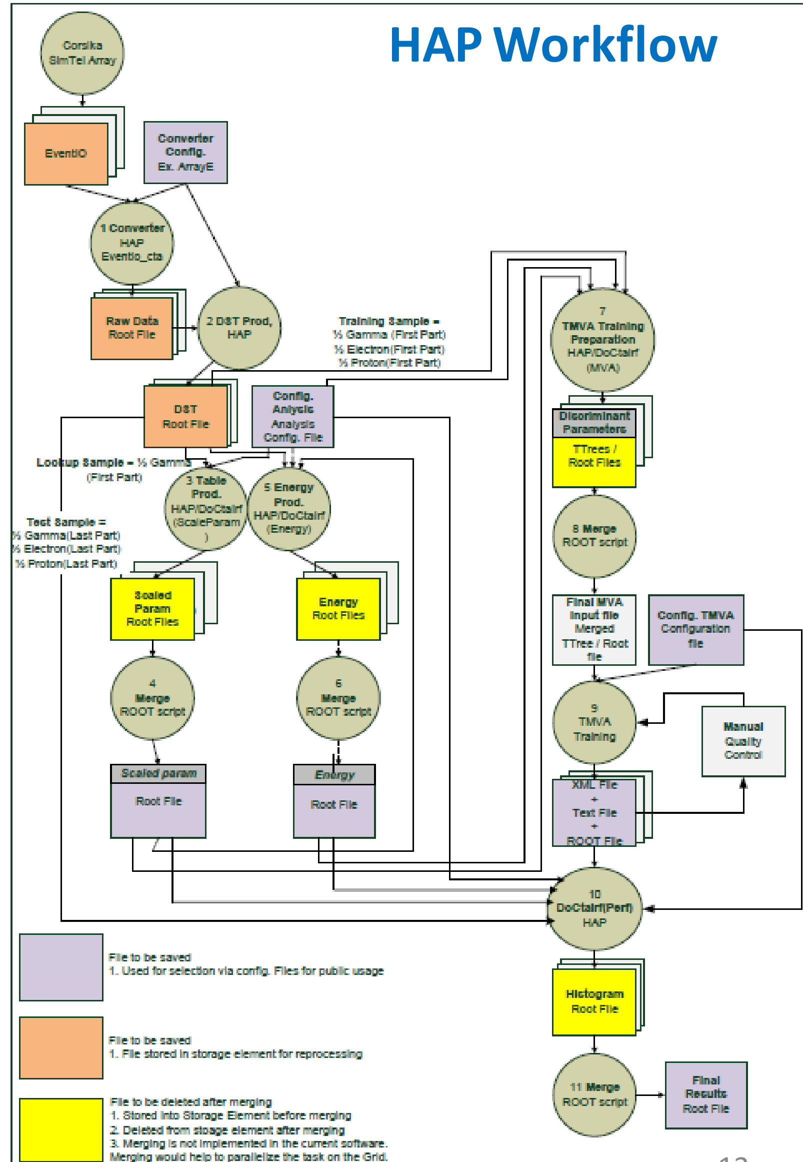


Current use of DIRAC: HAP analysis chain

Complex workflow

- Many interdependent steps linked by I/O
- Different CPU and I/O profiles
- Different data products (temporary and persistent files)
- Quality checks on data products
- Auxiliary jobs: Merging/ Format Conversion

- Currently: a DIRAC module for each step
- Next: combine them into workflows



Documentation

http://www.cta-observatory.org/ctawpcwiki/index.php/CEIN/DIRAC_evaluation_for_CTA

(CTA internal)

Documentation for users with no knowledge about the Grid

Section dedicated to examples for CTA applications

User experience (I)
(by P. Jean/IRAP)

Scientific goal:

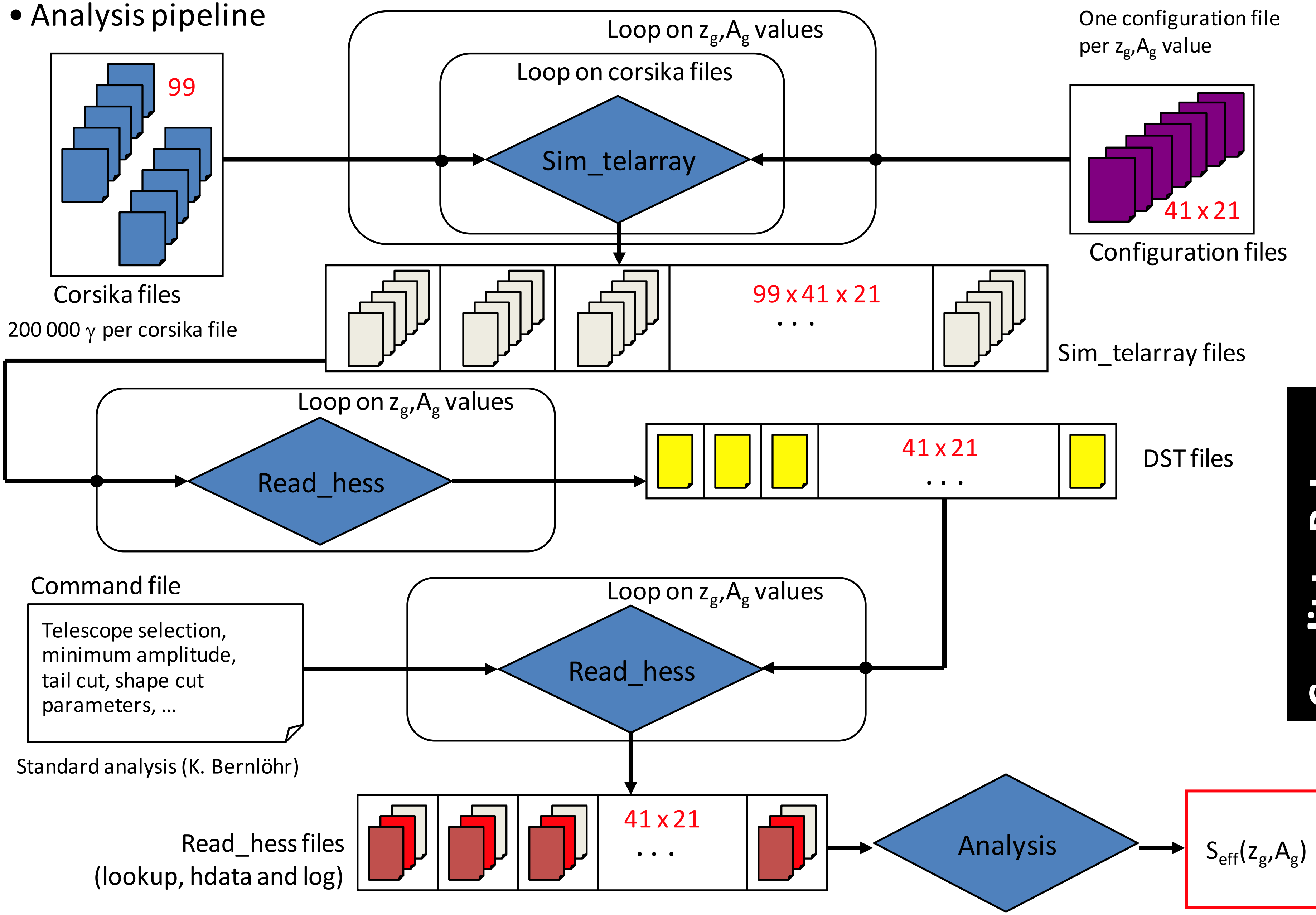
Calculate CTA effective area as a function of source angular position with respect to global CTA axis. The goal is to determine if we can gain effective area and global field of view if CTA is in a 'divergent mode' compared to the 'normal mode' for which all telescope point at the same direction

Program:

- Source scan over (21x21) bins in (zenith, azimuth) made twice for 'normal' and 'divergent' modes
- *High computing needs*
- *Use DIRAC to access Grid resources*

User experience (II)

- Analysis pipeline

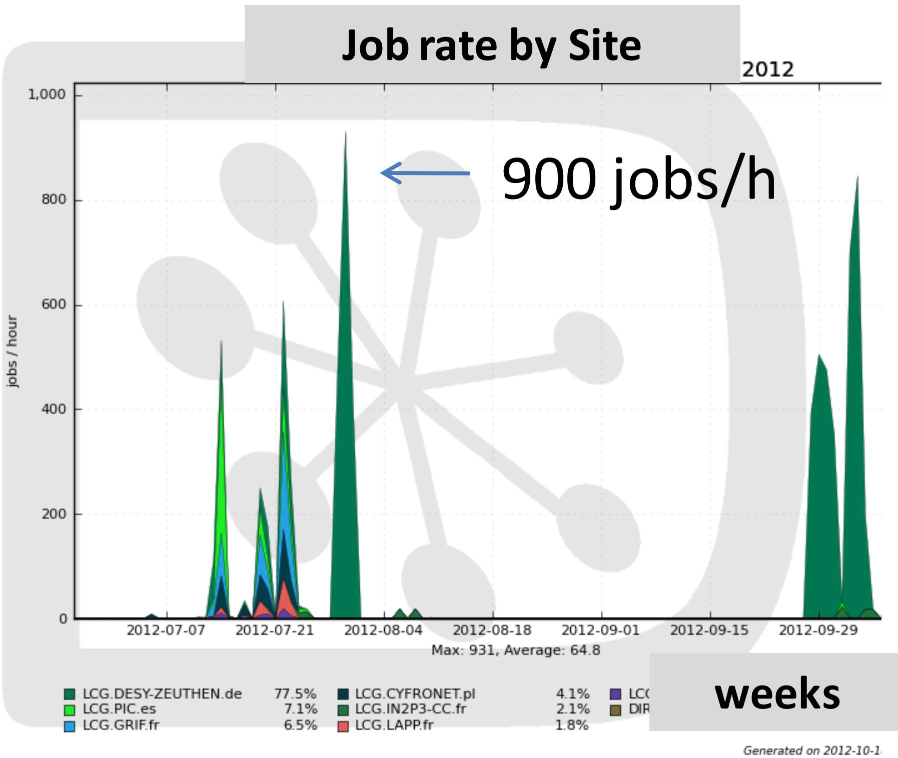


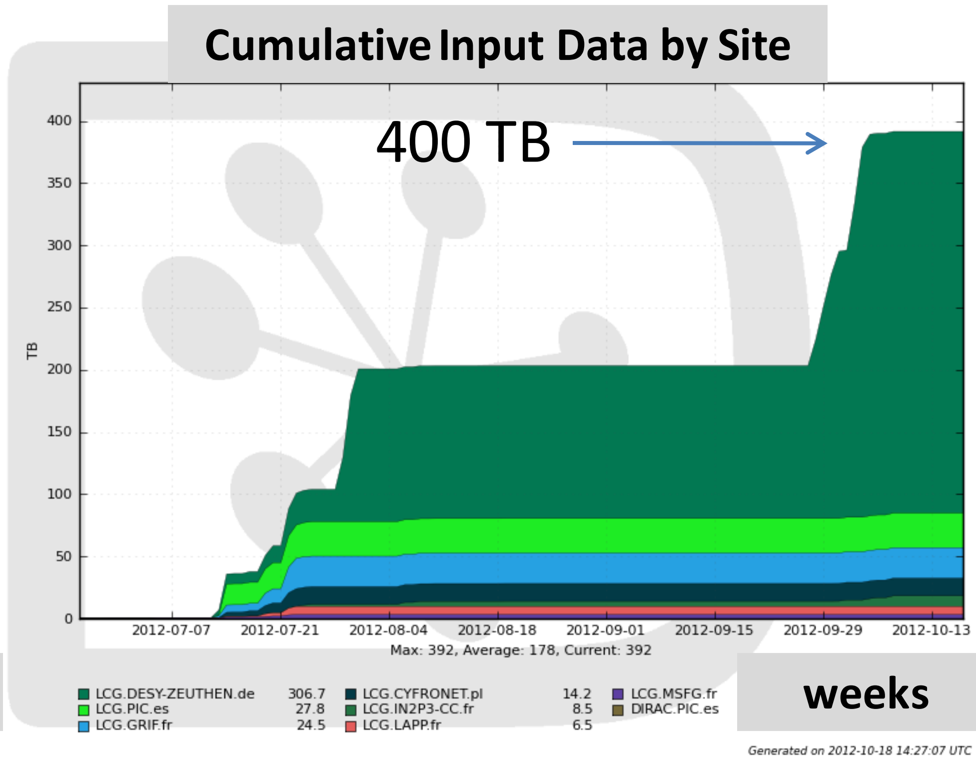
Credit to P. Jean

User experience (III)

About 4 weeks to run the program (July-October 2012):

- 180k jobs
- 400 TB processed
- 12 TB Output Data

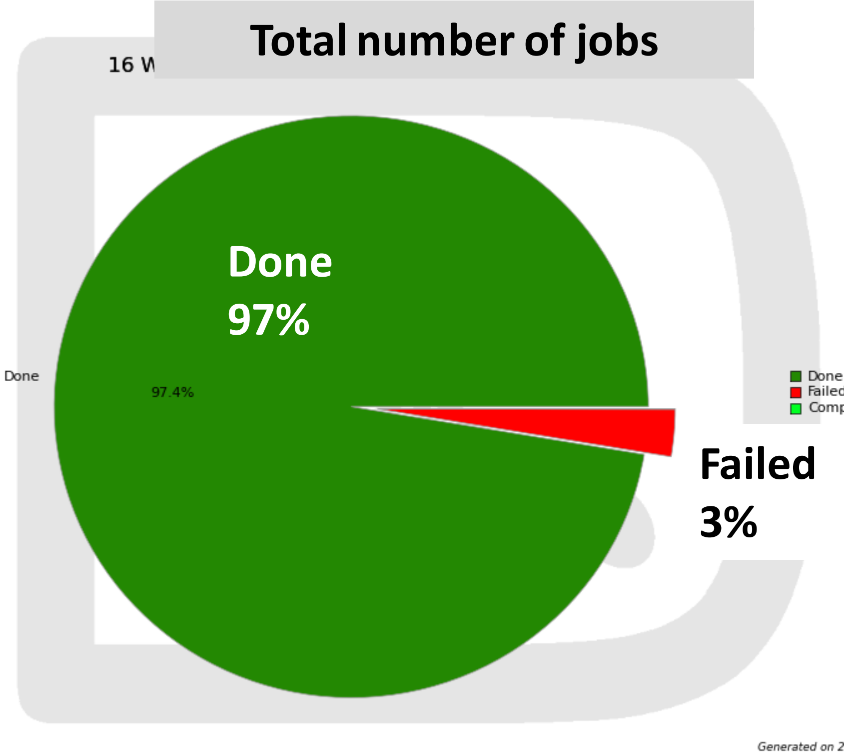


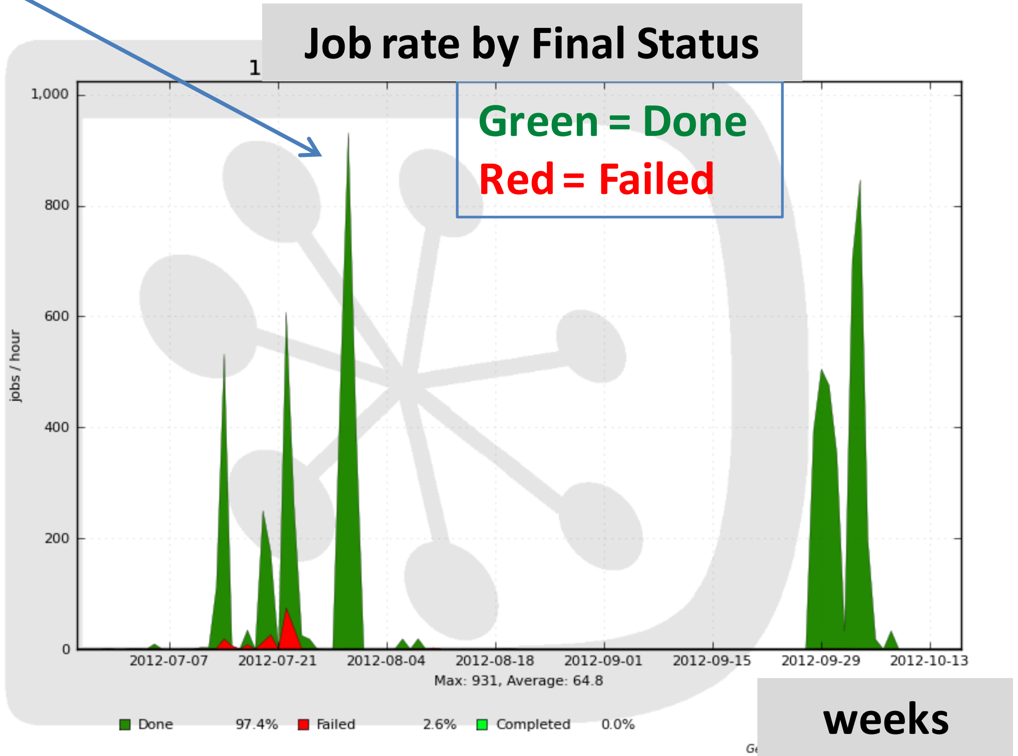


User experience (IV)

Job Success rate:

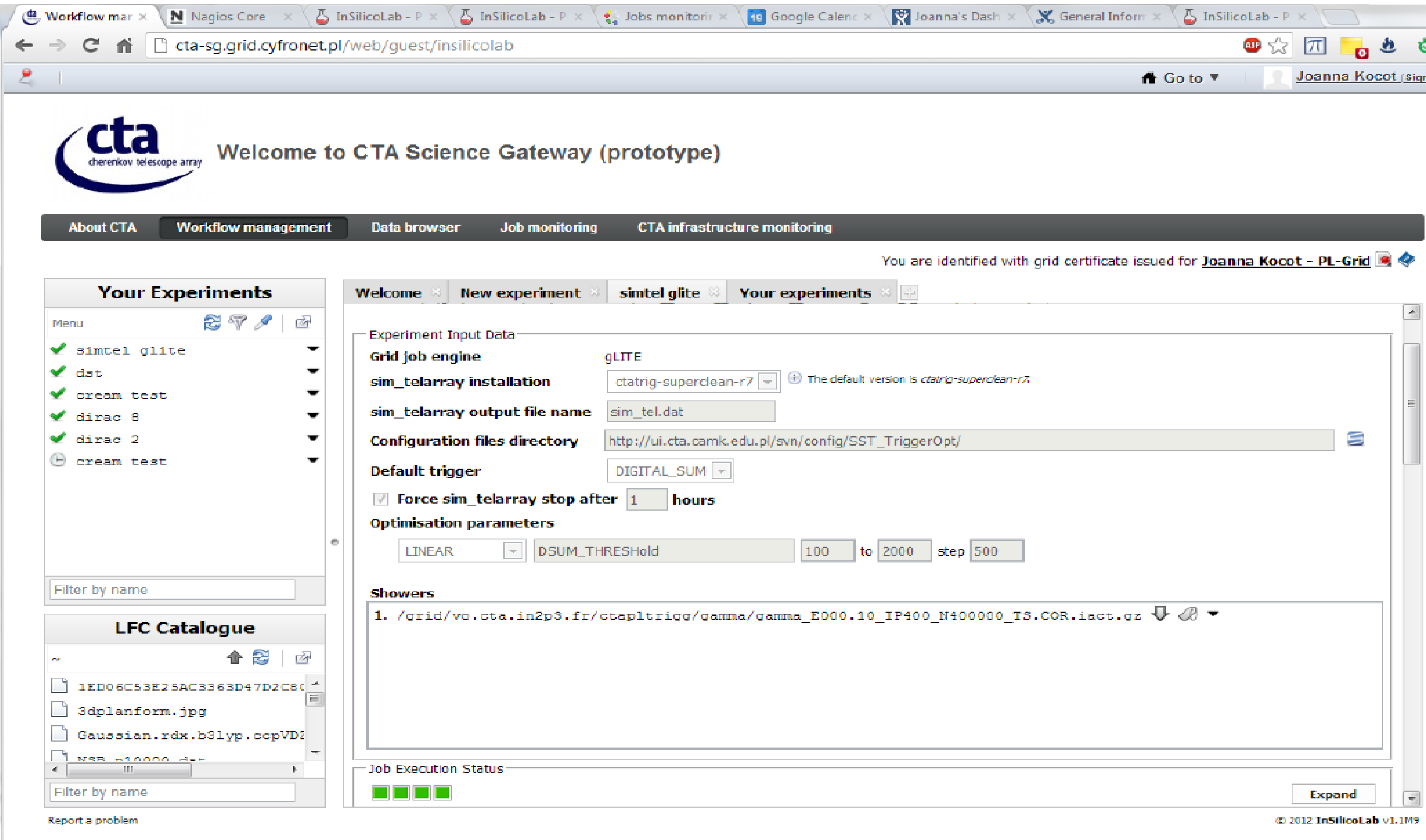
- overall **> 97%** (including first phase of test)
- 100% in last bunches





Toward a CTA Science Gateway

Work in progress to integrate DIRAC CTA modules in a web portal (prototype by CYFRONET)



Conclusions and perspectives

- CTA is the next generation instrument in the VHE gamma-ray astronomy
- It will be producing several PB/year
- DIRAC is currently used to handle MC simulations on the Grid
- We found out that all needed functionalities for MC production are covered
- User Analysis experience is also positive
- DIRAC File Catalog is currently being tested
- Work in progress to integrate DIRAC CTA modules in a CTA Science Gateway prototype
- Need to define a Computing Model for future CTA operation phase