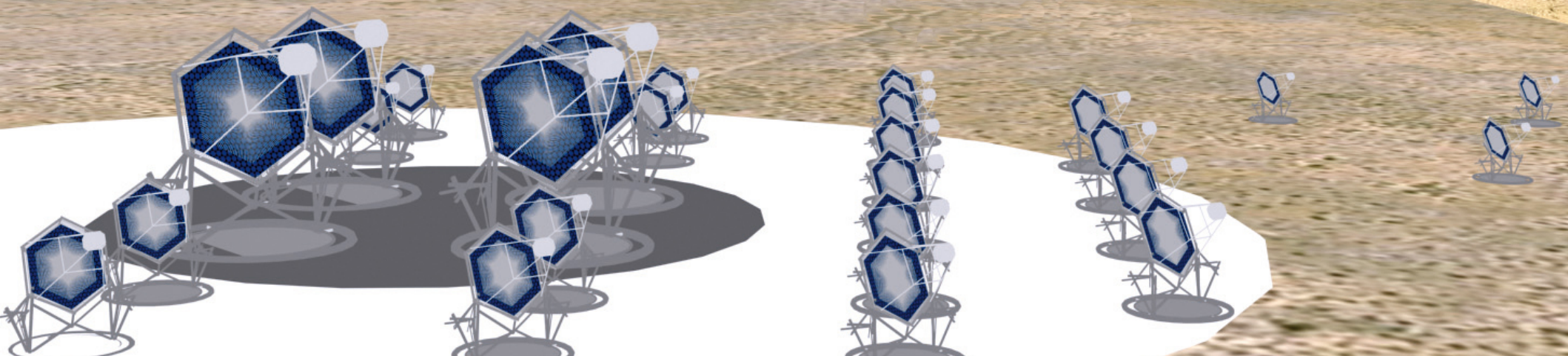
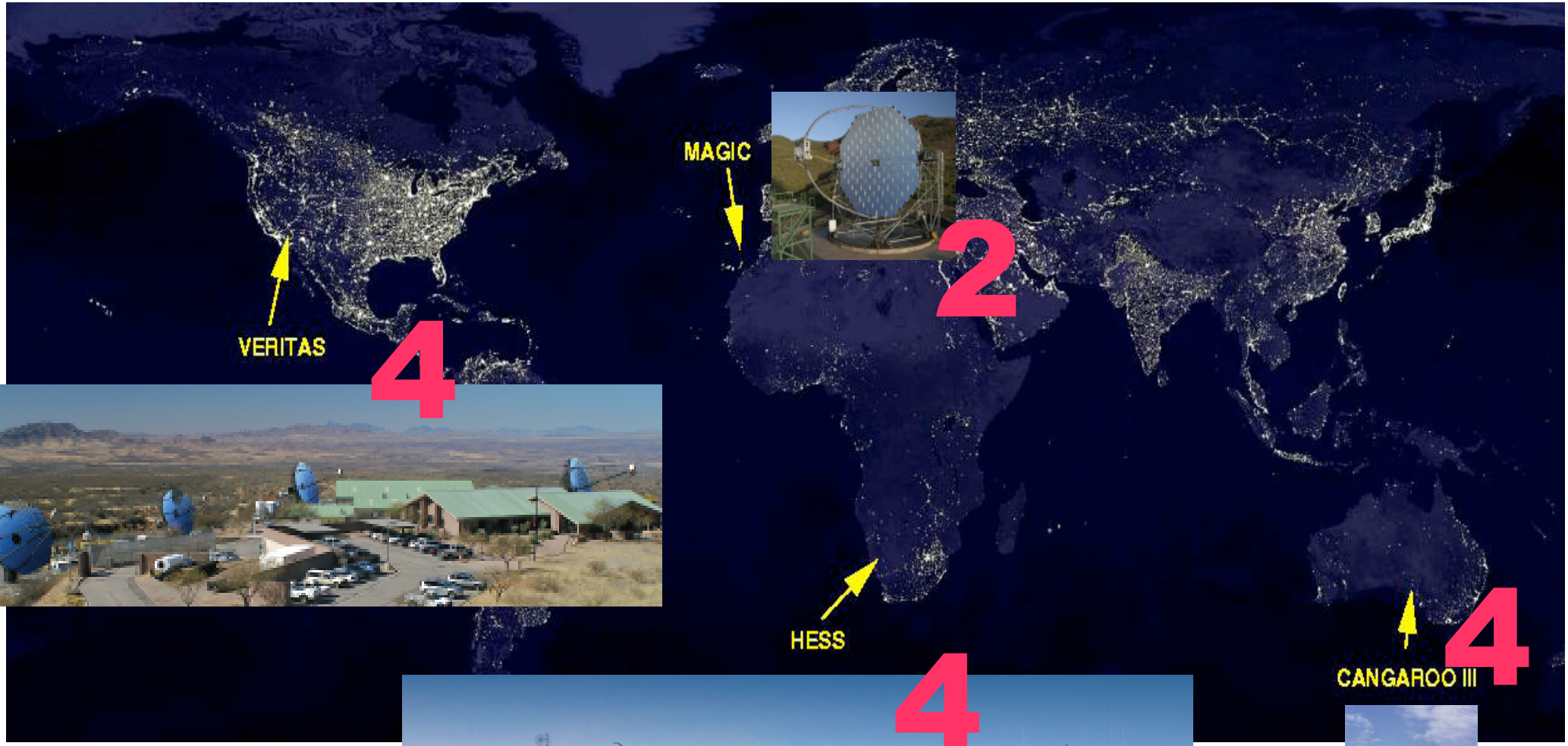
The CTA Computing Grid Project

Nukri Komin

LAPP, CNRS/IN2P3 Annecy-le-Vieux

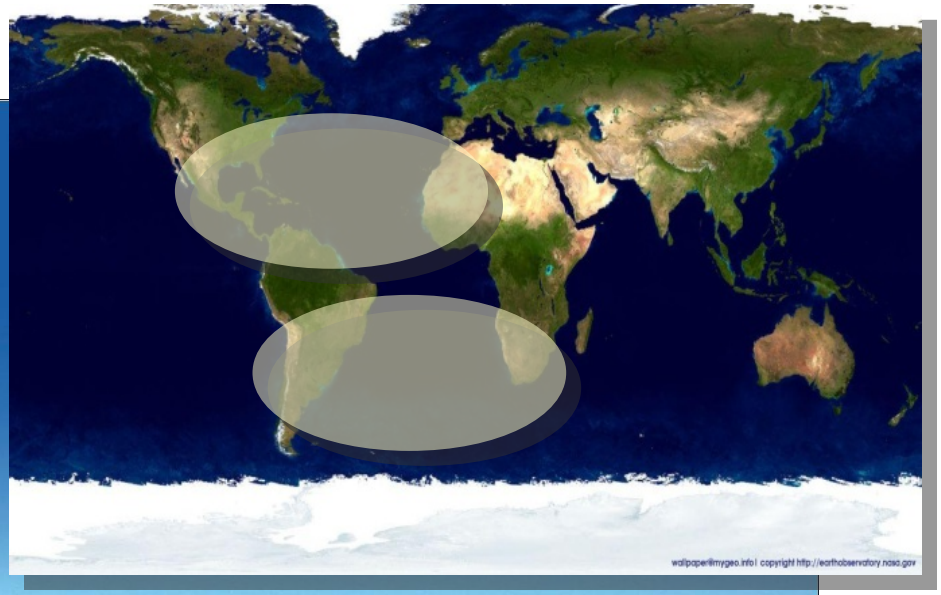


Current Cherenkov Telescopes



Cherenkov Telescope Array

- large array, 30 – 100 telescopes in 3 sizes, 2 sites
  - remote site
- Preparatory Phase 2010-2013
- 100 institutes
  - 22 countries



Current Activities

- Preparatory Phase
  - investigation of the science possible with CTA
  - administrative and technical aspects of building and operating a telescope array
- technical aspect: computation
- we (LAPP Annecy, LUPM Montpellier) propose Grid Computing for
  - Monte Carlo simulation
  - Data Management and Processing
  - CTA Science Gateway

CTA Virtual Grid Organisation

- Why do we use the Grid?
  - institutes can provide easily computing power
    - minimal man power needed, usually sites already supporting LHB
  - can be managed centrally (e.g. for massive simulations)
  - distributed but transparent for all users (compare HESS)
- CTA virtual organisation: vo.cta.in2p3.fr
  - French name, but open to everyone (renaming almost impossible)
  - VO manager G. Lamanna @ LAPP

CTA Virtual Grid Organisation

- 14 CEs in 5 countries
  - very heterogeneous (CPUs, RAM, scratch space)
  - 3 big sites: CC Lyon, DESY Zeuthen, Cyfronet Poland, PIC
  - many small sites (~100 CPUs)
  - 30k logical CPUs, shared with other VOs
  - ~1000 – 2000 CPUs for CTA at any time (based on experience)

			logical CPUs	for CTA
France	Annecy	IN2P3-LAPP	725	200
	Lyon	IN2P3-CC	9068	200...600
	Paris	GRIF	7495	several 100
		OBSPM	112	
	Bordeaux	M3PEC	384	
	Montpellier	LUPM	152	100
Germany	Zeuthen	DESY-ZN	786	400
	Dortmund	UNI-DORTMUND	1832	
	Munich	MPPMU	856	
Spain	Barcelona	PIC	2368	200
		IFAE	360	
	Madrid	CIEMAT-LCG2	836	
Poland	Cracow	CYFRONET-LCG2	5104	several 100
Greece	Thessaloniki	GR-01-AUTH	230	

CTA Virtual Grid Organisation

- LFC catalogue
- each site with several 100GB up to 10 TB local disk space
- massive storage (several 100 TB):
  - CC Lyon (including tapes), DESY Zeuthen, Cyfronet
- massive storage for large temporary files
  - simulations: corsika, will be kept for reprocessing
  - corsika file size 20-30 GB for 100000 proton showers

Grid Monte Carlo Production

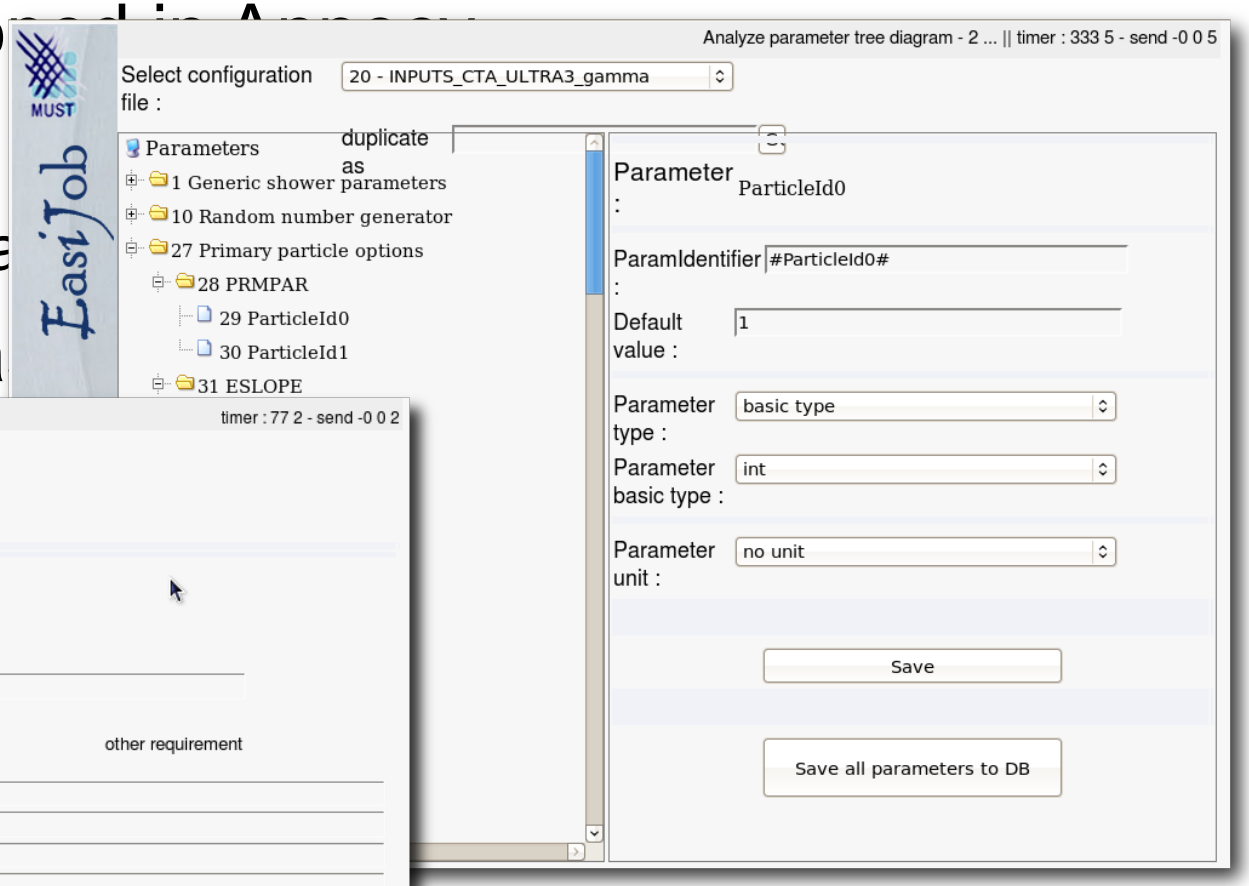
- first massive use of Grid: simulations
  - corsika (http://www-ik.fzk.de/corsika/)
  - sim_telarray (K. Bernlöhr [arXiv:0810.5722])
- protons, hard spectrum (index 1.5), 275 telescopes
  - high requirements
    - up to 4 GB RAM
    - 10 GB local scratch disk space
  - only subset of CEs can handle these requirements
- statistics
  - 54800 good quality runs, ~ 5TB
  - 9.9 10^8 sim_telarray showers
  - efficiency: 10^8 sim_telarray showers per month
    - many problems solved, next round much more efficient

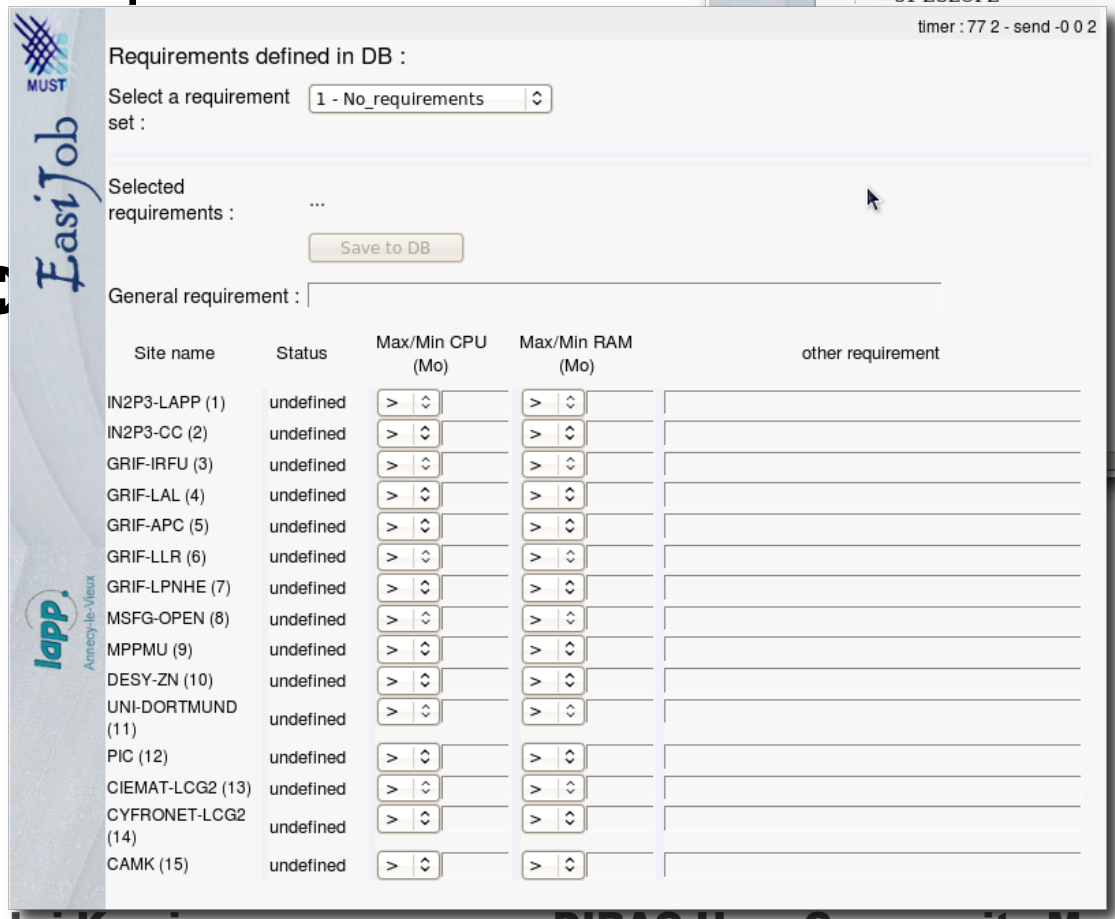
Job Configuration and Submission

- EasiJob, tool developed in Annecy
  - task configuration, job definition and submission
  - web interface and data base
- requirements are based on published parameters
  - e.g. `GLueHostMainMemoryRAMSize > 2000`
  - these requirements are interpreted differently at each site
- jdl requirements adapted for each site
  → allows fine tuning adapted to each site
  - close interaction with site admins

Job Configuration and Submission

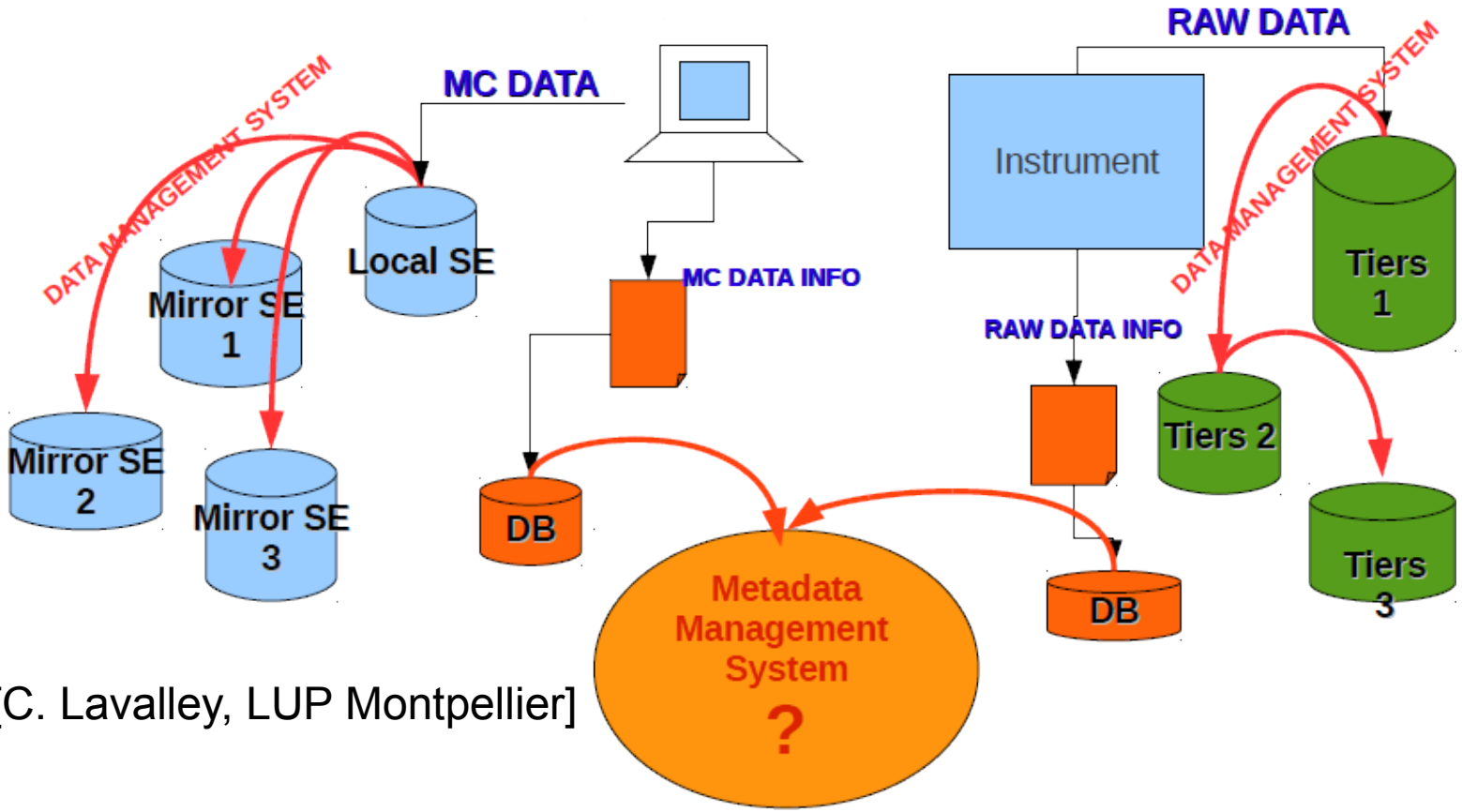
- EasiJob, tool develo
  - task configuration, job
  - web interface and data
- requirements are ba





Meta Data Management

- meta data: data about data
  - logical and physical file name, production parameters, etc.
- meta information can be in several data bases
- we will evaluate AMI (Atlas Meta Data Interface)



[C. Lavalley, LUP Montpellier]

Bring the User to the Data

- data on the Grid (simulations), bring the analysis as well
- Grid User Interface needed
  - certificate infrastructure
  - software to download files and submit jobs
  - easy installation on all platforms
- user tasks
  - analyse simulation files (no official software yet)
  - software development
- we evaluate two ways to make Grid UI available
  - Virtual Machine
  - Dirac

Virtual Machine

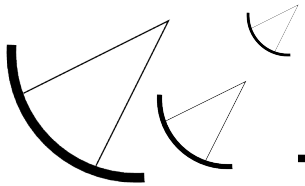
- CernVM (http://cernvm.cern.ch/portal/)
- contains all Grid infrastructure
- easy to install ✔
- will be distributed for CTA users
- will be developed for cloud computing platform

	Windows	Linux	Mac
VirtualBox	✔	✔	✔
vmware	✔	✔	✔
KVM	-	✔	-
Xen	-	✔	-

DIRAC as User Interface

- easy to install ✔
- user does not need to understand all grid details
- easy job creation, submission and monitoring
- where are we?
  - currently being tested (me with help of Ricardo)
  - DST production for simulated files
- want to open it soon to collaboration
  - analysis tasks of individual CTA users

Analysis Chain



telescopes

observation and calibration runs

raw data (level 0)

calibration

calibrated camera images (level 1)

reconstruction

photon list (level 2)

analysis

sky maps, lightcurves, spectra (levels 3 and 4)

e.g. *Fermi* Science Tools
software available for Linux, Mac, Windows

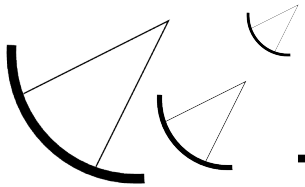
internal data

published data

Data Rates

- Raw Data: 1 GB/s, 1..3 PB/year (level 0)
  - production during night time, max. 8..10h per day
  - 29 day cycle with peak at new moon
- Reconstructed Data (level 2, available to public)
  - about 10% of raw data
  - computing requirements:
    - 1h of raw data needs ~200 CPUs*days (today)
    - based on HESS Model++, 28min of 4 telescopes needs 3x 1Ms [M. de Naurois]
- Results (level 3 and 4, available to public)
  - requirements: to be evaluated

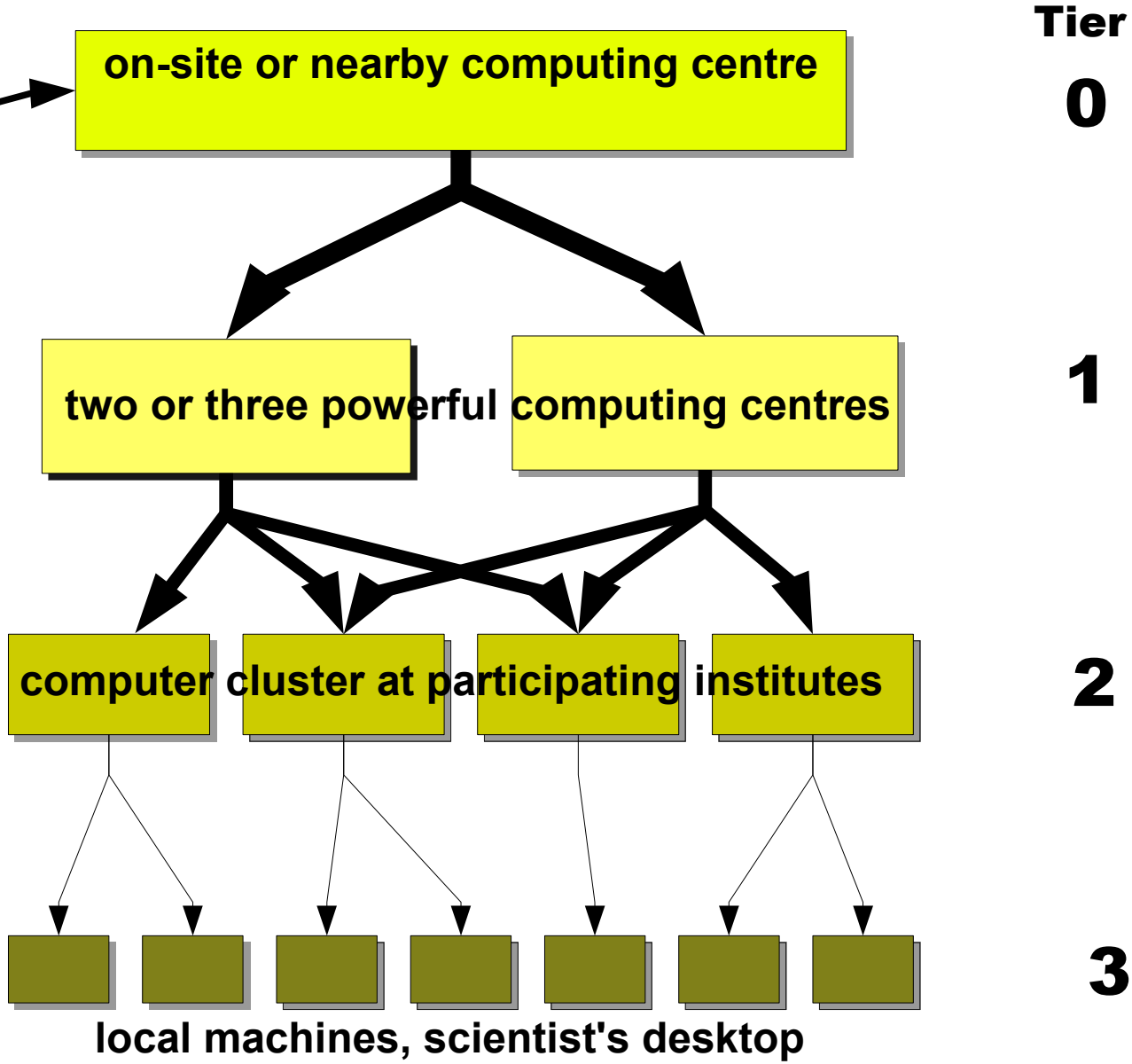
Data Flow



telescopes

raw data

remote site:
- local computing centre or
- fast internet link



Tier 0...3

- Tier 0
  - data source, at the telescopes
- Tier 1
  - backup of raw data
  - calibration, reconstruction
  - 2 or 3 sites
- Tier 2: participating institutes
- Tier 3: scientist's machine
- will be set up soon for MC data chain

Summary

- Cherenkov Telescope Array
  - future, large array for gamma-ray astronomy
  - preparatory phase 2010 – 2013
- Grid Computing
  - 14 CEs, *O*(1000) CPUs, very heterogeneous
  - massive Monte Carlo simulations
  - data pipeline to be set up
- current needs (Dirac?)
  - meta data base
  - user interface
- nothing is decided, may be no Grid at all!!