

# The EELA Project

Inês Dutra - COPPE/UFRJ

Second Grid School XXVII Congresso da SBC IME/RJ, 05.07.2007

Realização Promoção







Rio de Janeiro SBC 2007

30 DE JUNHO A 06 DE JULHO DE 2007

www.eu-eela.org





## Outline

- What is EELA?
- Objectives
- Partners
- Organization
- WP2: Infrastructure
- WP3: Applications
- WP4: Dissemination



## What is EELA?

- E-infrastructure shared between Europe and Latin America
- 2-year project financed by the European Union
- Start date: 1/01/2006
- Total budget: 2,568,320 €
- Funding from the EC: 1,700,000 €
- Total effort in person-month: 1109
- Web site: www.eu-eela.org
- Contact person:
   Ramón Gavela
   email: ramon.gavela@ciemat.es



## **Objectives**

- Establish a collaboration network between European institutions where Grid expertise exists (e.g. EGEE project), and Latin American institutions where Grid activities are emerging.
- Set up a pilot e-Infrastructure in LA, interoperable with the EGEE one in Europe, allowing to run enhanced applications, thus enabling dissemination of knowledge and experience on Grid technology.
- Ultimately set up a steady framework for e-Science collaboration between Europe and Latin America.

**Human Network!!!** 



#### **EELA PARTNERS**

E-infrastructure shared between Europe and Latin America



10 Countries

21 Partners

Argentina: UNLP

Brazil: CEDERJ, RNP, UFF,

**UFRJ** 

Chile: REUNA, UDEC,

**UTFSM** 

Cuba: CUBAENERGIA

Mexico: UNAM

Peru: SENAMHI

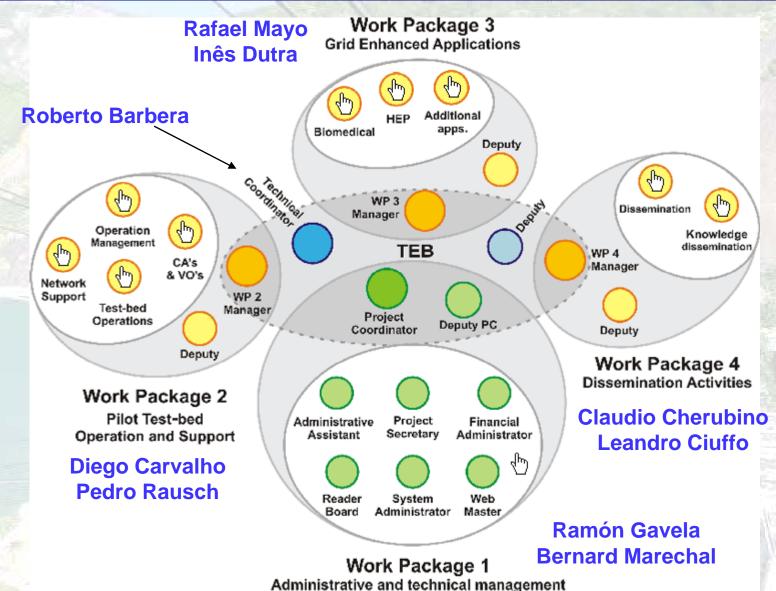
Venezuela: ULA

**CLARA** 



E-infrastructure shared between Europe and Latin America

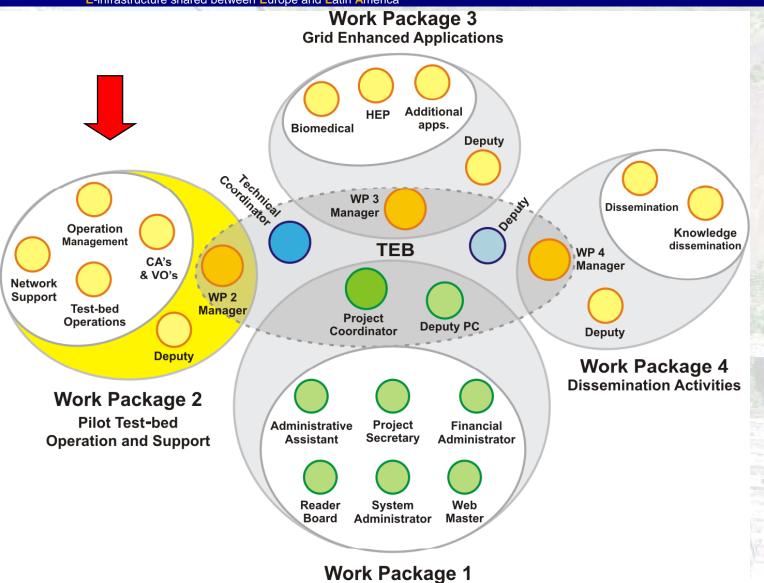
## **Organization**





## **WP2: Infrastructure**

E-infrastructure shared between Europe and Latin America



Administrative and technical management



## **WP2: The EELA Pilot Test-bed**



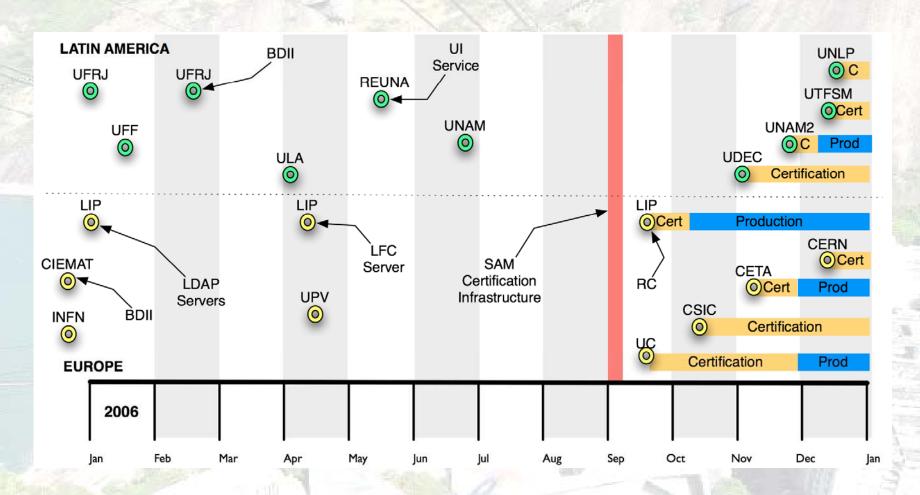




## **WP2:** number of sites

E-infrastructure shared between Europe and Latin America

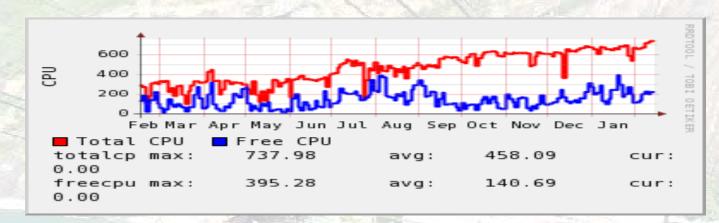
#### Evolution of the number of sites integrating the e-Infrastructure



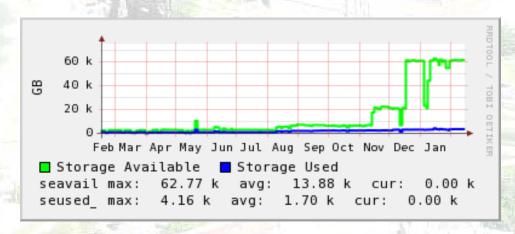
## **WP2: Resources**

E-infrastructure shared between Europe and Latin America

#### Resources integrated into the Test-bed



CPU Slots evolution



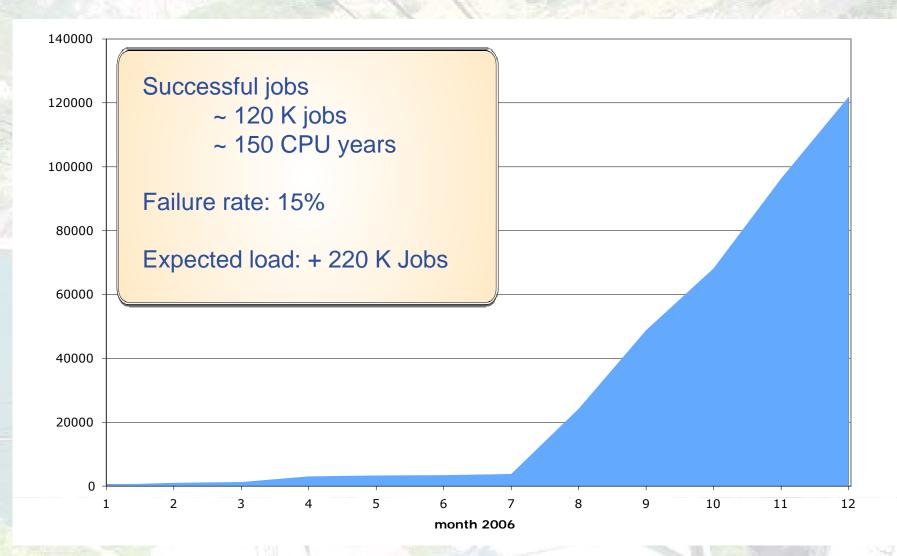
Storage Capacity evolution

Production sites (source GStat)



## WP2: Use of the infrastructure

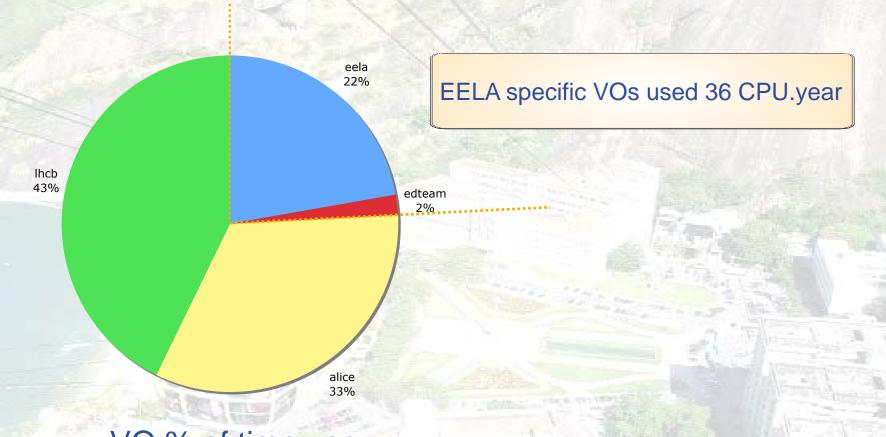
#### Statistics of use of the Pilot Test-bed





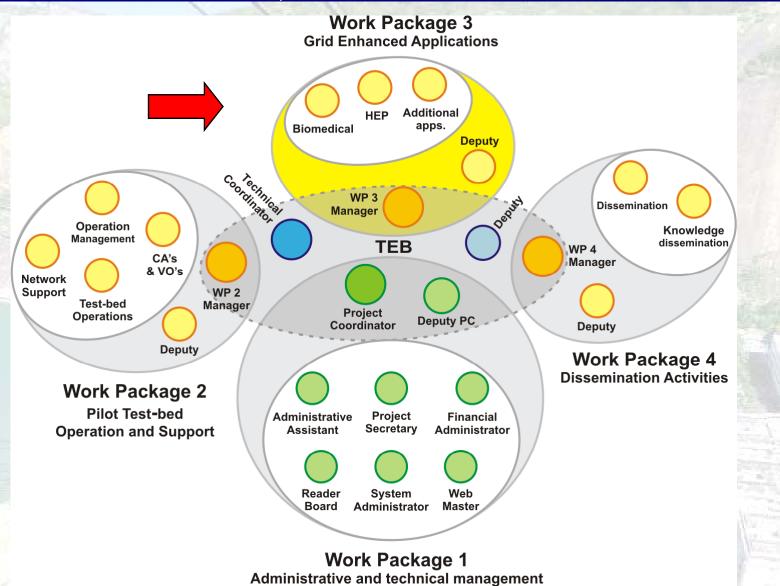
# WP2: Usage of the infrastructure E-infrastructure shared between Europe and Language of the infrastructure

Statistics of use of the Pilot Test-bed





## **WP3: Applications**



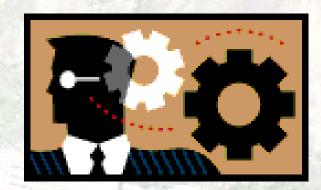


# **WP3: Applications**

E-infrastructure shared between Europe and Latin America

# OVERVIEW APPLICATIONS

- 3.1 Biomed
- 3.2 HEP
- 3.3 e-Learning & Climate





- WP3: Identification and support of Grid enhanced applications
  - Coordinated by CIEMAT
  - Identifies, selects and customizes relevant applications and tools suitable for the Grid dissemination process in:
    - T3.1 Biomedicine (CUBAENERGIA, ULA, UPV)
    - T3.2 High Energy Physics (CERN, CIEMAT, UFRJ, <u>UNAM</u>, UNLP, UTFSM)
    - T3.3 Additional Applications
      - e-Education (CECIERJ/CEDERJ, CIEMAT, CUBAENERGIA, <u>UFRJ</u>, UNAM)
      - Climate (UC, UDEC, SENAMHI)
  - Aims at being the place of information exchange between already gridified applications and future ones.

- GATE.
- Malaria Docking: WISDOM.
- Blast in Grids: BiG.
- Phylogeny.
- Participants: UPV (Coordinator: Vicente Hernández),
   CUBAENERGIA, ULA.



## **WP3: Task 3.1: GATE**

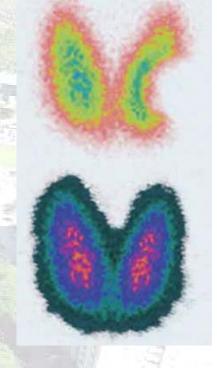
E-infrastructure shared between Europe and Latin America

#### Problem Addressed

- The computation of the <u>doses in radiotherapy treatments</u> is performed through analytical simplified models
- Environment for the <u>Monte-Carlo simulation of particle</u> <u>physics emission in the medical field</u> that is more accurate specially when dealing with heterogeneous tissues
- Very time-consuming

#### User Community

- The interest of the LA community is led by CUBAENERGÍA
- It is focused towards two main oncological problems:
  - Thyroid cancer.
  - Treatment of metastasis with P<sup>32</sup>.
- 9 centers in Cuba are interested (5 hospitals and 4 oncological centers and institutions)





#### **WP3: Task 3.1: GATE**

E-infrastructure shared between Europe and Latin America

- Outcome
  - >60 Computing resources have GATE Available
  - Gate is Installed on
    - tochtli.nucleares.unam.mx
    - ramses.dsic.upv.es
    - grid012.ct.infn.it
    - ce-eela.ic.uff.br
    - ce01.eela.if.ufrj.br
    - ce-eela.ciemat.es
  - A demonstration was performed in the frame of the EU-LAC Summit, Lisbon 28-29 April 2006



 Due to the lack of network connectivity from/to Cuba a local configuration has been set-up using the EELA stand-alone grid distribution



## WP3: Task 3.1: WISDOM

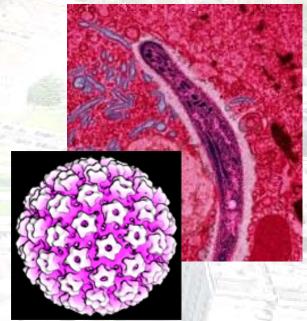
E-infrastructure shared between Europe and Latin America

#### Problem Addressed

- WISDOM (Wide In-Silico Docking of Malaria) is a deployment of a <u>high-throughput virtual screening platform</u> in the perspective of <u>in-silico drug discovery</u> for <u>neglected diseases</u>
- The in-silico docking is faster and much cheaper than the experimental docking, which is restricted to the most successful ligands obtained after the simulation process

#### LA Interest

- ULA is leading the interest in the LA
- Collaboration is started in the analysis of new targets for malaria. ULA has selected new targets of Plasmodium Vivax which have been included in large-scale docking experiments



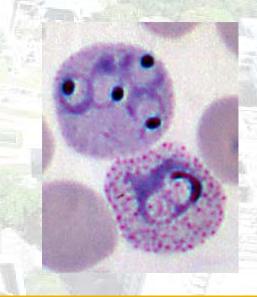


# First EELA WISDOM Data Challenge

E-infrastructure shared between Europe and Latin America

#### Outcome:

- 2 Targets proposed by ULA
- DC operated and coordinated by UPV
- Starting date: October the 23<sup>rd</sup>, 2006
- Ending date: January the 1st, 2007
- Number of original jobs of the <u>first target</u>. 2422. First data challenge experiment on completing all jobs
- Average computing time per job: 34,4 hours
- Total effective running time: 228 CPU.days
- Results obtained: 53 GBytes





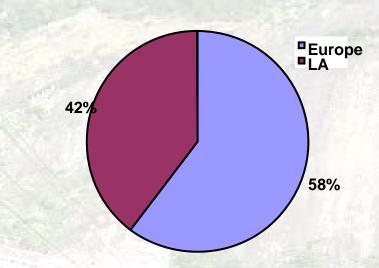
# First EELA WISDOM Data Challenge

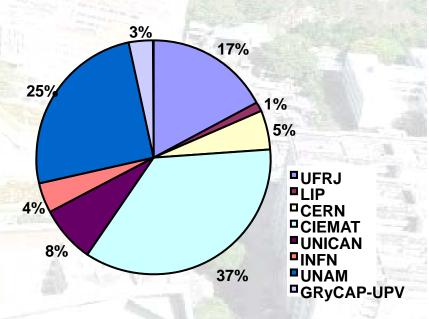
E-infrastructure shared between Europe and Latin America

#### Sites used:

FP6-2004-Infrastructures-6-SSA-026409

- CIEMAT, Since beginning
- GRYCAP-UPV, Since beginning
- INFN , Since beginning
- UNAM, Since beginning
- UFRJ, Since beginning
- UNICAN, LIP and CERN during the DC execution
- More than 40% of the jobs were effectively executed in LA sites







## WP3: Task 3.1: Blast in Grids

E-infrastructure shared between Europe and Latin America

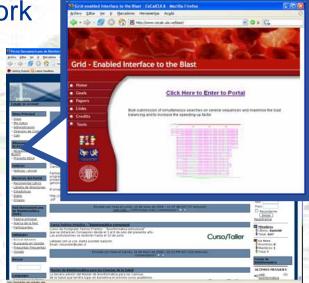
#### Problem addressed

- BiG: BLAST in Grids is a grid-enabled BLAST interface
- BLAST (Basic Local Alignment Search Tool) is a <u>computationally</u> <u>intensive bioinformatics procedure</u> applied to identify <u>compatible</u> <u>protein and nucleotids sequences</u> in <u>protein and DNA databases</u>
- Applications in drug development, phylogeny, etc.

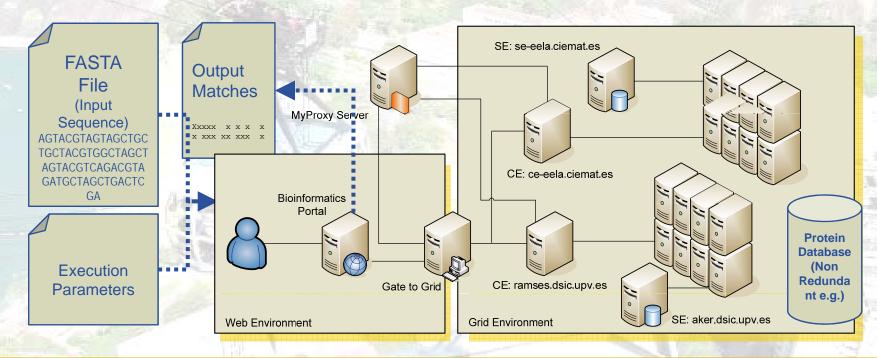
#### User community

 The Bioinformatics Ibero-American Network and Portal (http://portal-bio.ula.ve)

- This portal also provides several on-line applications for registered users
- It currently has almost 600 registered users from 70 countries (although 90% come from 10 countries)



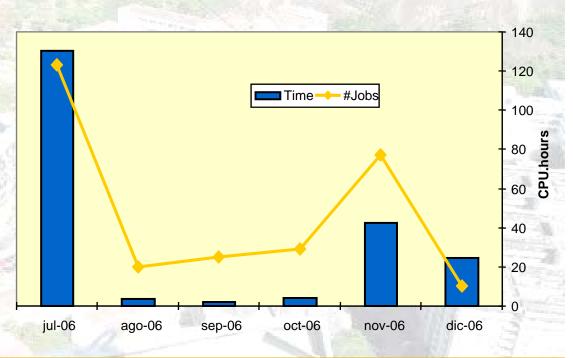
- Use of MPI-Blast kernel
- Enhanced security through a MyProxy server
- Fault tolerant on the client and server side
- Embeddable on a stand-alone application or web portal
- Splitting of input sequences and reference databases into multiple jobs. Deals with multiple databases simultaneously





# WP3: Task 3.1: Blast in Grids: Usage Report

- Period: Jul'06-Dec'06.
- **Usage statistics:** 
  - Number of jobs: 284
  - CPU consumed: 173 CPU.days
  - Resources used: ramses.dsic.upv.es 20-CPU queue
  - BiG is being used at the University of Los Andes to work on the complete genome of the Plasmodium Falciparum for the identification of DHFR antigenic proteins





# WP3: Task 3.1: Phylogeny

E-infrastructure shared between Europe and Latin America

#### Problem addressed

- A phylogeny is a <u>reconstruction of the evolutionary history of a</u> <u>group of organisms</u>. Very important for the analysis of the <u>resistance to treatments</u>
- Phylogeny tools consume much memory and computing time

#### User community

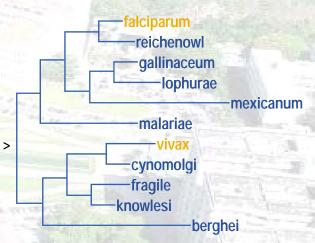
 The Bioinformatics Ibero-American Network and Portal (<a href="http://portal-bio.ula.ve">http://portal-bio.ula.ve</a>)

#### Solution proposed

 Use an mpi-enabled version of a widely used Bayesian inference application for phylogeny (MrBayes)

#### Achievements

 A grid-service to run parallel MrBayes executions is currently available at UPV





## **WP3: Task 3.2: HEP**

- HEP and GRID computing
- First HEP applications
  - ALICE experiment
  - LHCb experiment



## **EELA - HEP synergy**

E-infrastructure shared between Europe and Latin America

#### Mature applications

: guaranteed applications

#### Well supported

- Developed by LHC collaborations
- Support available to collaboration members

#### •HEP Experiments need GRID computing

- All experiments use grid based computing models
- HEP communities in Latin America require GRID
- HEP communities in LA are interested in collaboration with EELA

## Several EELA partners have HEP groups



## **HEP applications in EELA**

-infrastructure shared between Europe and Latin America

#### Applications of interest to EELA partners and other communities in Latin America

- Initial applications
  - ALICE experiment: Heavy ion physics at LHC
    - INFN-Catania, CIEMAT, and UNAM
  - LHCb experiment: B physics at LHC
    - UFRJ
- Other LHC applications
  - ATLAS experiment (General purpose)
    - UFRJ, UNLP and UTFSM
  - CMS experiment (General purpose)
    - CIEMAT; CMS has participants in EELA Brazil and Mexico, but no EELA partner from LA
- New projects
  - Pierre Auger Observatory
    - INFN-Catania, LIP, UFRJ, UNAM, and UNLP, others in EGEE



# E-infrastructure shared between Europe and Latin America Experiment in EELA

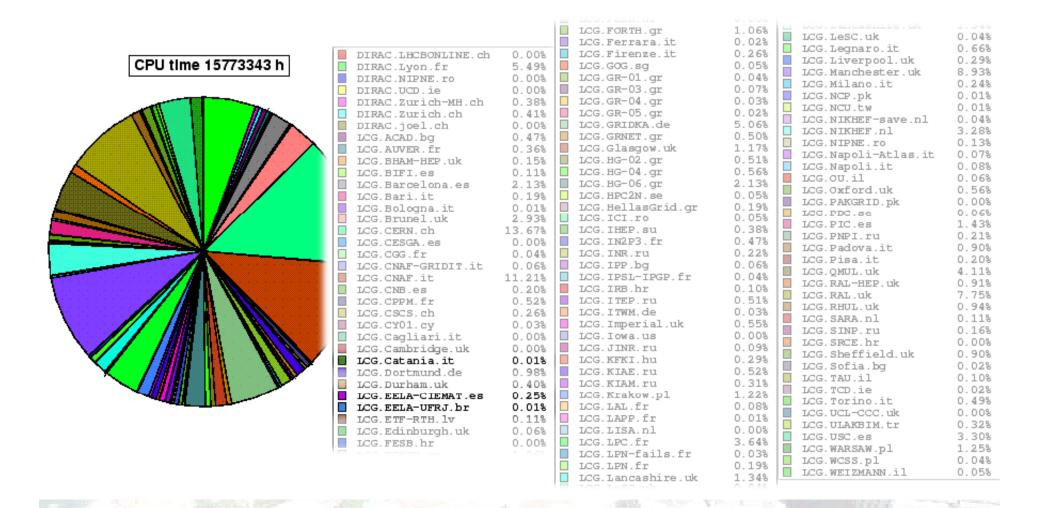


#### Usage info

	🦃 Jol	b parameters		Application software			Details		
Run# +	PID	Owner	Events	ROOT	ALIROOT	GEANT	Date	Output dir	
		podesta (15) 🔽					last year <u></u> ✓		
7	1711466	podesta					19.01.2007 03:10	/alice/cern.ch/user/p/podesta/demo/flow/output_nc/007/	test pro
6	1711465	podesta					19.01.2007 03:10	/all_v/c=n.ci_(vc_i_desta/demo/flow/output_nc/006/	test pr
5	1711443	podesta		3		ISS	(9) (2) (3)	/alloy/com.ch/w/com/sizedesta/demo/flow/output_nc/006/ /alloy/com.ch/wer/sizedesta/demo/flow/output_nc/005/	test pro
4	1711433	podesta					19.01.2007 03:10	/alice/cern.ch/user/p/podesta/demo/flow/output_nc/004/	test pro
3	1711432	podesta			rid id	hh	from	/a ce/cat/ spc/podesta/demo/flow/output_nc/003/ /a	test pro
3	1797427	podesta			VI-VIew B			/aG.U.str/_/podesta/pp_900GeV/03/	MC sim
2	1711421	podesta			000		19.01.2007 03:10	/ailce/cern.ch/user/p/podesta/demo/flow/output_nc/002/	test pro
2	1794911	podesta		v5/A\0	neri	ca	25.01.2007 20:06	/allice/cern.ch/user/p/podesta/pp_900GeV/02/	MC sim
1	673395	podesta			000 000 000 000 000 000 000 000 000 00		06.09.2006.21:08	/allce/cern.ch/user/p/podesta/demo/flow/output_nc/001/	test pro
1	1711410	podesta					19.01.2007 03:10	/alice/cern.ch/user/p/podesta/demo/flow/output_nc/001/	test pro
1	1794908	podesta		v5-13-02	v4-04-Rev-08	v1-6-1	25.01.2007 20:06	/alice/cern.ch/user/p/podesta/pp_900GeV/1/	MC sim
	1769912	podesta					23.01.2007 20:37	/alice/cern.ch/user/p/podesta/demo/flow/output_nc/test/	test ev
	1771607	podesta					23.01.2007 21:36	/alice/cern.ch/user/p/podesta/demo/flow/output_nc/test/	test ev
	1771617	podesta					23.01.2007 21:36	/alice/cern.ch/user/p/podesta/demo/	test ev
	1771691	podesta					23.01.2007 23:39	/alice/cern.ch/user/p/podesta/demo/	test ev
TOTAL: 15 jobs 0							Export folders		



## LHCb experiment in EELA





# ATLAS experiment in EELA

- UNLP has recently joined the ATLAS collaboration
- UTFSM has started to collaborate with ATLAS
- UFRJ ATLAS group expressed interest in using the EELA infrastructure
- CIEMAT already resource centre
- UNLP and UTFSM are setting up resource
  - Support ATLAS
  - To be shared with EELA

http://atlas.web.cern.ch/Atlas/index.html

- CMS is the only LHC experiment without participation from a Latin American EELA partner
- CMS groups in EELA member countries
  - Brazil
  - Mexico
  - Spain (CIEMAT)
- First contacts established
  - Presentation at the 1st EELA conference
  - Brazilian and Mexican CMS groups participated in the network survey

http://cms.cern.ch/



# **Pierre Auger Observatory**

E-infrastructure shared between Europe and Latin America

- Four EELA partners (more collaborators in EGEE)
  - INFN-Catania
  - LIP
  - UFRJ
  - UNAM
  - UNLP
- Start GRID use in 2007

http://www.auger.org/



## **Task 3.3: Additional Applications**

E-infrastructure shared between Europe and Latin America

# Task responsible for new applications (dynamic!)

- e-Learning
- Climate



## WP3: Task 3.3: Additional Apps

- -8 Partners/ 6 Countries (1 EU and 5 LA):
  - CEDERJ (Brazil)
  - CIEMAT (Spain) active
  - CUBAENERGIA (Cuba)
  - SENAMHI (Peru)
  - UniCan (Spain)
  - UDEC (Chile)
  - UFRJ (Brazil)
  - UNAM (Mexico)





#### What is it?

- Use of multimedia technology to develop and improve new learning strategies
- Characteristics
  - Data stored under the control of a Learning Management System (LMS)
  - Contents can be updated and exchanged with other systems

E-infrastructure shared between Europe and Latin America

#### • Why e-learning on a grid?

- Limitations of current systems
  - Scalability
  - Storage capacity
  - Availability
  - Computing power
  - Lack of some kinds of interactive capabilities
- Good side-effects:
  - Automatic data location (video, text, equipments, etc.)
  - Confidentiality
  - Extraction of relevant information, not only location
  - Data standardization

#### WP3: Task 3.3: e-Learning Applications





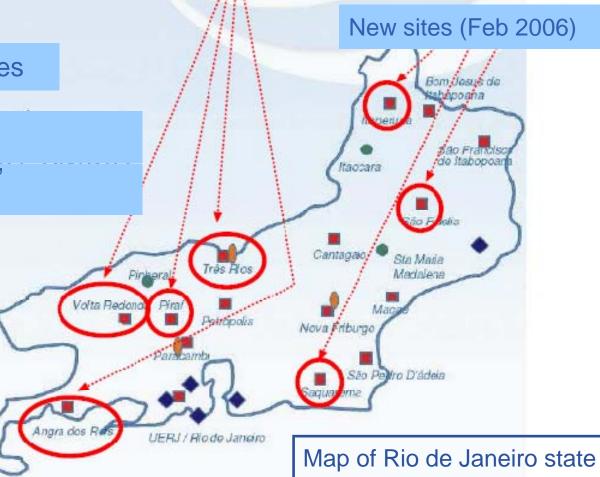
#### **WP3: Task 3.3: VoD**

E-infrastructure shared between Europe and Latin America

#### Sites of the CS course

- 16 Regional sites
- 3 Regional satellite sites
- 5 call centers at UERJ, UFRJ, UNIRIO, UFF and UENF
- 3 places dedicated to Science

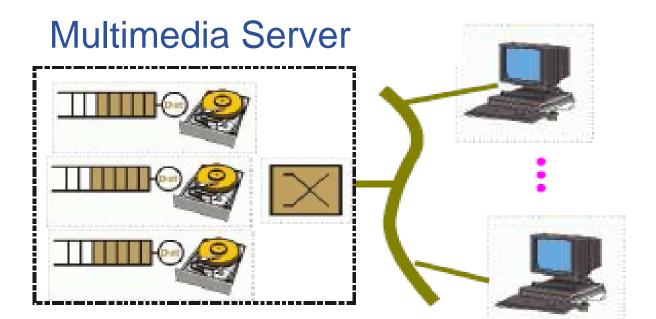






#### **WP3: Task 3.3: VoD**

E-infrastructure shared between Europe and Latin America

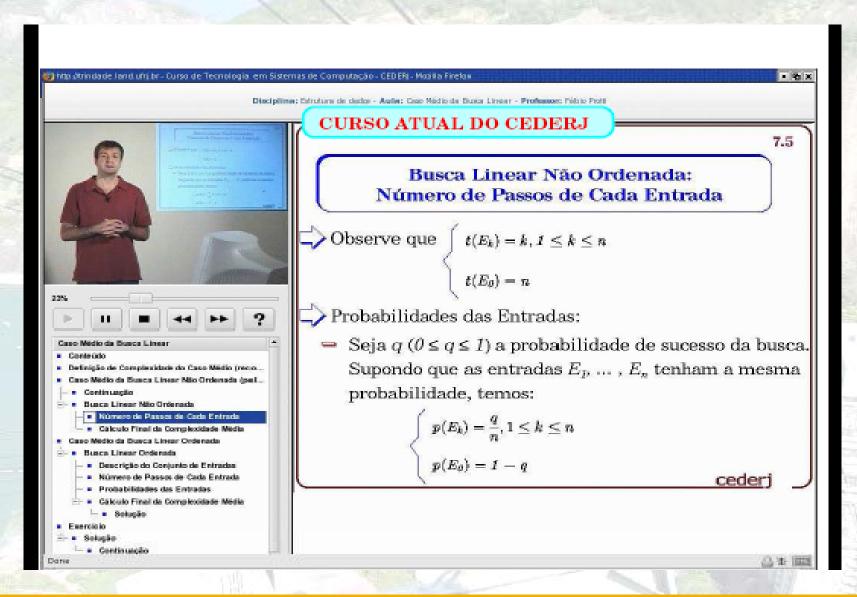


Computer Lab



FP6-2004-Infrastructures-6-SSA-026409

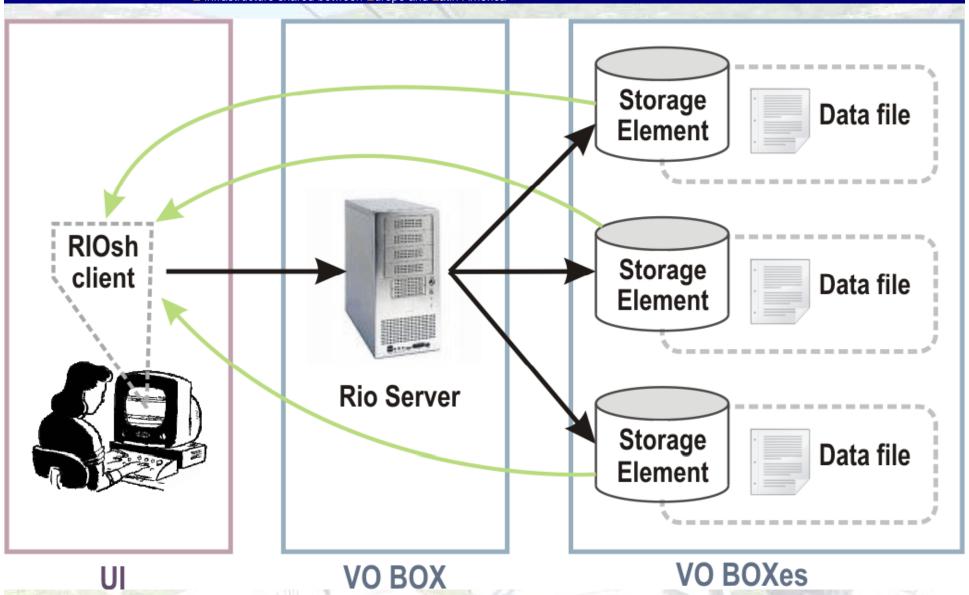
#### **WP3: Task 3.3: VoD**



- First prototype version to the EELA infrastructure during the 1<sup>st</sup> EELA Grid School (EGRIS-1)
- Client requests come from User Interfaces (UIs)
- Videos and data are stored on Storage Elements
- RIO server installed on entry points of the EELA infrastructure



#### **VoD Architecture on EELA**

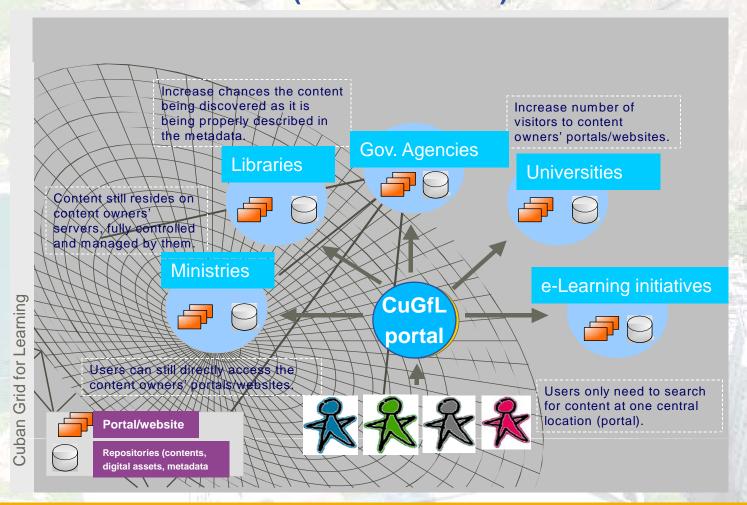




#### WP3: Task 3.3: CuGfl

E-infrastructure shared between Europe and Latin America

### Cuban Grid for Learning (CuGfL) (CUBAENERGIA)



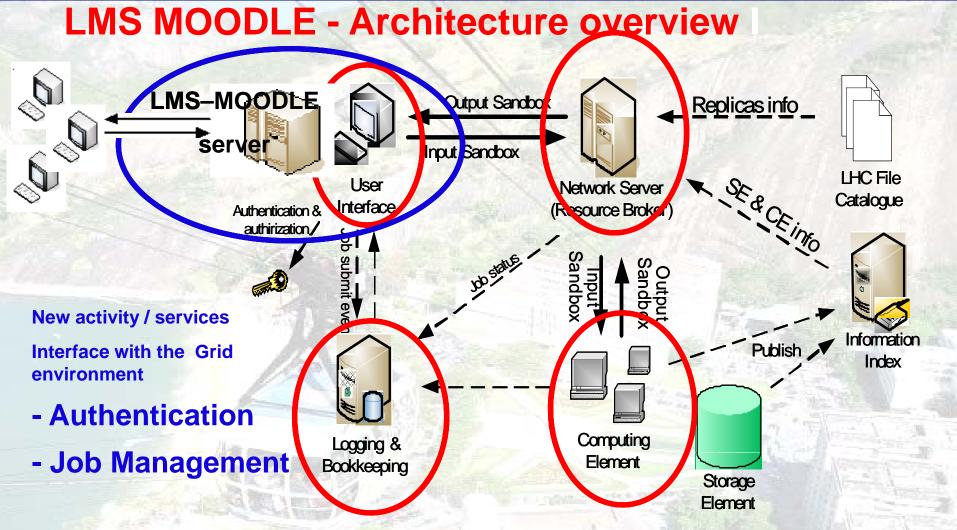
#### WP3: Task 3.3: CuGfl

- -Based on Moodle
- -Course management system to produce webbased courses that support a social constructionist framework of education
- -Implemented:
  - -Job Management Module
  - -Authentication Module



#### WP3: Task 3.3: CuGfl

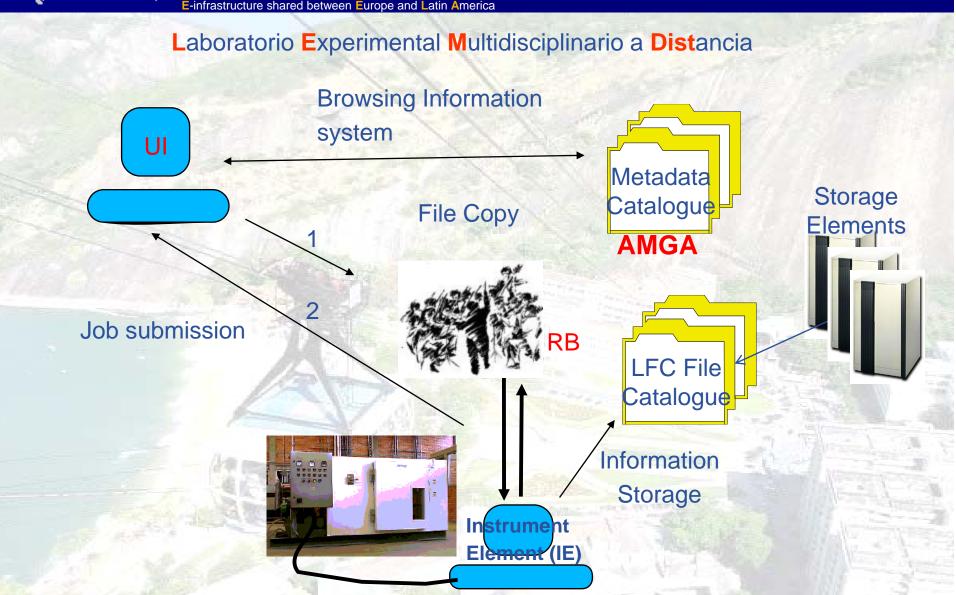
E-infrastructure shared between Europe and Latin America



Workload Management System components



#### WP3: Task 3.3: LEMDist





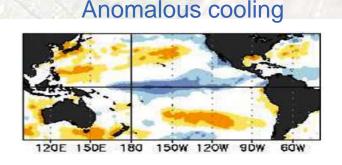
#### WP 3: Task 3.3. Climate

-infrastructure shared between Europe and Latin America

Goal: Predict Local Impacts of "El Niño" in Latin America
 A challenging problem for the climate community, with huge socio-economical impact in Latin America.

Anomalous heating

40N20NEQ20S40S120E 15DE 180 150W 120W 9DW 60W



- GRID helps to share computing resources, heterogeneous data, as well as know-how in a user-friendly form.
- A new integrated climate application developed in EELA from the scratch, with no similar counterpart in any other Earth Science/Climate EU Project.



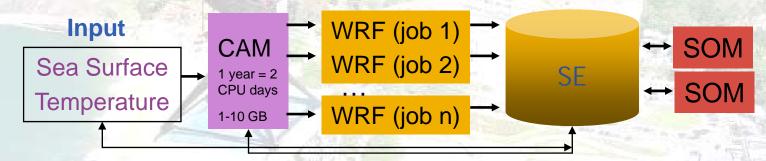
#### WP 3: Task 3.3: Climate

infrastructure shared between Europe and Latin America

#### Three applications have been identified (climate sequence):

- Global atmospheric circulation model (CAM) ...... Deployed!

This sequence poses several computational challenges Nontrivial dependent relationships among the applications.

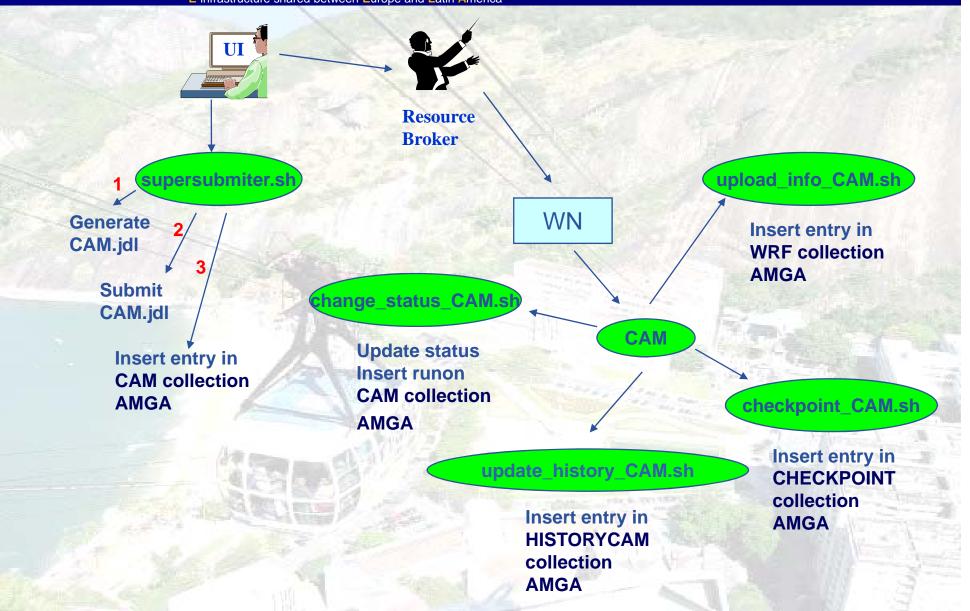


#### This sequence of jobs demands middleware solutions for:

- Preparing and submitting dependent jobs / data sharing (workflow).
- Restarting interrupted experiments.
- Manage metadata (for datasets and application status).



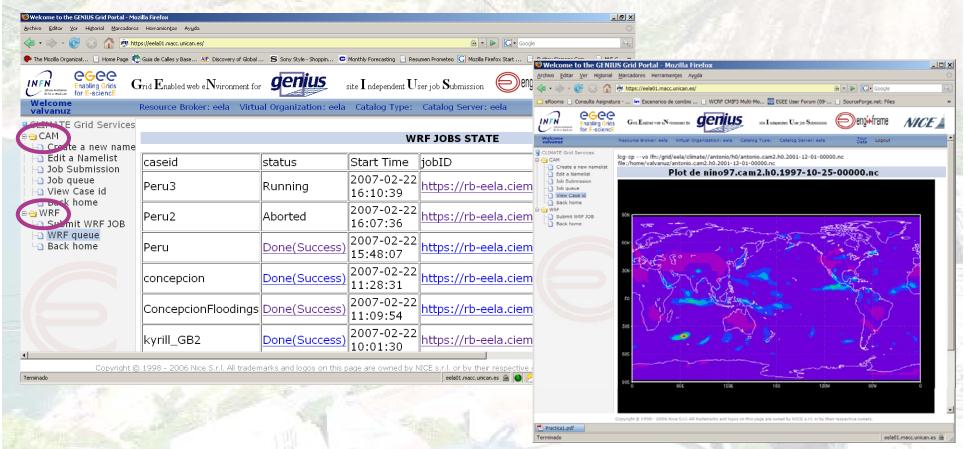
#### **Application Workflow**





#### WP 3: Task 3.3: Climate

- Using GENIUS to interact with the applications (CAM+WRF)
- In the future a climate specific portal will be developed (JSR168) to run and track scientific experiments.





#### For the near future, but established through 2006

- Increase and support the user communities
- Survey of new communities
  - MoU with more Projects
- 12 New applications interested in joining EELA



#### **New communities**

E-infrastructure shared between Europe and Latin America

## EELA has setup a procedure to accept new associated partners and their applications

http://www.eu-eela.org/eela\_mou.php



#### **New communities**

E-infrastructure shared between Europe and Latin America

#### **EGRIS-1**

EMBOSS (UNAM → MoU)

SegHidro (UFCG → MoU)

Distributed Simulation of Multiple Failure Events on Optical Networks (UNESP, UNICAMP, USP)



VOLCANO SONIFICATIONS (EGEE-INFN)

SATyrus (UFRJ)

PILP (UFRJ)



#### Conclusions

- Great achievements after the 1<sup>st</sup> EELA Grid School in Itacuruçá with EELA and non-EELA applications successfully ported to gLite
- More synergy and interaction among all partners
- People are very much involved in making the EELA infrastructure a success
- All of these apps will have a tremendous impact on LA







#### **Useful information**

E-infrastructure shared between Europe and Latin America

#### WP3 web page:

http://www.eu-eela.org/eela\_wp3.php

#### **WP3 documents:**

http://documents.eu-eela.org

#### **Forthcoming Events:**

- 3rd EELA Conference: December 3-5, 2007
- Second EELA Grid School, EGRIS-2



E-infrastructure shared between Europe and Latin America

# Thank you!!! Any questions?