



# CYTED-Grid



## **Tecnología GRID como motor del desarrollo regional.**

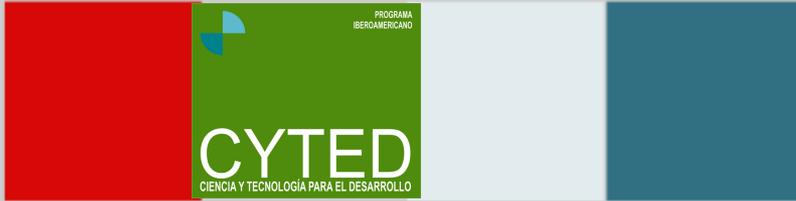
Acción de Coordinación de Proyecto de Investigación CYTED Grid 2006.



# Índice

- Presentación, objetivos, socios
- Paquetes de trabajo
- Estado del proyecto





# Presentación

## Acción de Coordinación de Proyecto de Investigación CYTEDGRID 2006.

- Creación de una infraestructura tecnológica y humana entre diversos grupos iberoamericanos dedicados al campo de la informática y con experiencia en el área de aplicaciones científicas potencialmente usuarias de la tecnología GRID.
- El **objetivo principal** de este proyecto es construir un GRID como soporte de diferentes tipos de aplicaciones, analizando hasta qué punto las soluciones disponibles actualmente son válidas, y proponiendo desarrollos nuevos que proporcionen la funcionalidad necesaria.





# Presentación

- Desarrollo de una infraestructura GRID
- Distribuida entre los participantes de la comunidad iberoamericana
- A disposición de investigadores no especialistas
- Para aplicaciones del área de salud y de gestión medio-ambiental
- Con énfasis en la generación de capacidades técnicas y **formación de recursos humanos**



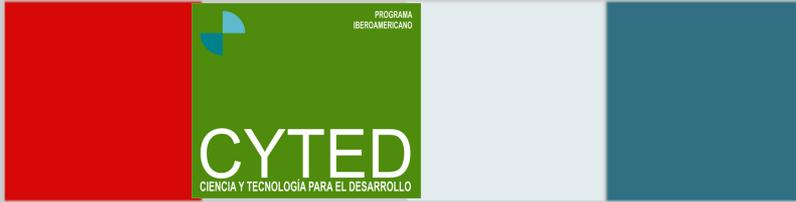


# Participantes

**19 instituciones en 13 países**

- UCM Universidad Complutense de Madrid **España**
- UAB Universidad Autónoma de Barcelona **España**
- UDELAR Universidad de la República **Uruguay**
- UH Universidad de la Habana **Cuba**
- UNLP Universidad Nacional de La Plata **Argentina**
- UNA Universidad Nacional de Asunción **Paraguay**
- UDEA Universidad de Antioquia **Colombia**
- UNAM Universidad Nacional Autónoma de México **México**
- UNAS Universidad Nacional de San Agustín **Perú**
- UMAG Universidad de Magallanes **Chile**
- ESPOL Escuela Superior Politécnica del Litoral **Ecuador**
- UPV Universidad Politécnica de Valencia **España**
- UdC Universidad de A Coruña **España**
- ULA Universidad de Los Andes **Venezuela**
- UNEB Universidade do Estado da Bahia **Brasil**
- UdP Universidade do Porto **Portugal**
- USB Universidad Simón Bolívar **Venezuela**
- UNSL Universidad Nacional de San Luis **Argentina**
- Universidade Federal de Campina Grande **Brasil**





- Conexión con otros proyectos GRID

- EGEE
- Embrace
- GRID-IT
- Metacentro Valenciano
- CrossGrid
- CLARA
- GRAMA (UNAM)
- PRAGMA (UNAM)
- **EELA**





# Objetivos

- Despliegue de una infraestructura de Grid basada en los recursos disponibles actualmente en los grupos participantes.
- Desarrollo de la capa de gestión de servicios de la plataforma Grid.
- Desarrollo de un portal Web como punto de acceso centralizado y transparente a los recursos del Grid.
- Puesta a punto de un conjunto de aplicaciones de impacto social y económico para la región que facilite su ejecución en la plataforma Grid.
- Contribución a la formación de formadores en las áreas temáticas del proyecto: cursos y seminarios de Postgrado. Desarrollo de tesis de Postgrado para estudiantes y docentes de la región en la temática del Proyecto.
- Generación de acciones de transferencia de tecnología a partir de los resultados del proyecto.
- Difusión de resultados y publicaciones conjuntas.





# Aplicaciones

- **Bioinformática:**

El conocimiento de la estructura tridimensional es fundamental para la comprensión de muchos procesos biológicos. Sin embargo, esta tarea requiere una ingente capacidad de cálculo y almacenamiento, al mismo tiempo que requiere una cuidada modelización del proceso de reconstrucción, durante el que se accede a diversos tipos de información. Es, pues, un caso que presenta retos importantes tanto desde el punto de vista del Grid computacional como del Grid de datos.



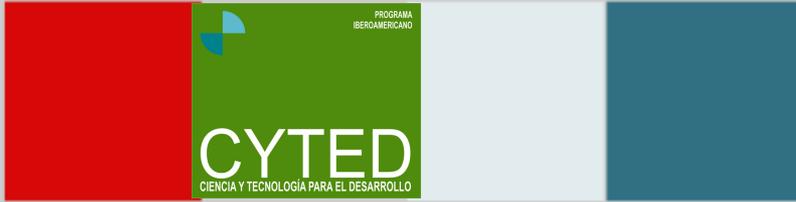
# Aplicaciones

- **Ciencias de la Salud:**

Estudio del uso de tecnologías Grid para la simulación de procesos biomédicos, procesamiento de datos médicos de gran dimensión (imágenes, signos vitales, genoma, etc), diagnóstico asistido por computador y minería de datos en bases de datos epidemiológicas.

- **Medio Ambiente:**

Simulación de incendios forestales e inundaciones, cuya importancia social hoy día es evidente, y que, computacionalmente, son ejemplos de aplicaciones paralelas con grandes exigencias de cálculo, claras candidatas a mejorar su rendimiento en entornos distribuidos como el propuesto.



# Coordinación





# WP1 Despliegue

## Coordinador: UPV

- Objetivo: Despliegue de la infraestructura Grid.
- Con soporte para:
  - Certificación (OpenCA)
  - IPv6 (mediante túneles)
- Actividades:
  - Autoridad de Certificación en Madrid
  - Distribución Knobus (UPV)





# WP2 Planificación y Tolerancia de fallos

- Planificación de Tareas a "bajo nivel" y Tolerancia a Fallos en sistemas Grid
  - Tarea 1: Planificación de tareas a "bajo nivel" en un Grid o un multicluster
  - Tarea 2. Desarrollo de una librería y un precompilador para checkpointing (UDC)
    - CPPC (Controller/Precompiler for Portable Checkpointing)
  - Tarea 3. Tolerancia a fallos transparente al usuario basada en protocolos rollback-recovery (UAB)
    - Se propone una arquitectura RADIC (Redundant Array of Independent Checkpoints)
  - Tarea 4. Tolerancia a Fallos basada en replicación de datos. (UAB)





# WP3 Herramientas

- Herramientas para el Grid: desarrollo de aplicaciones, monitorización y análisis
- Monitorización dinámica de aplicaciones Grid. (UAB)



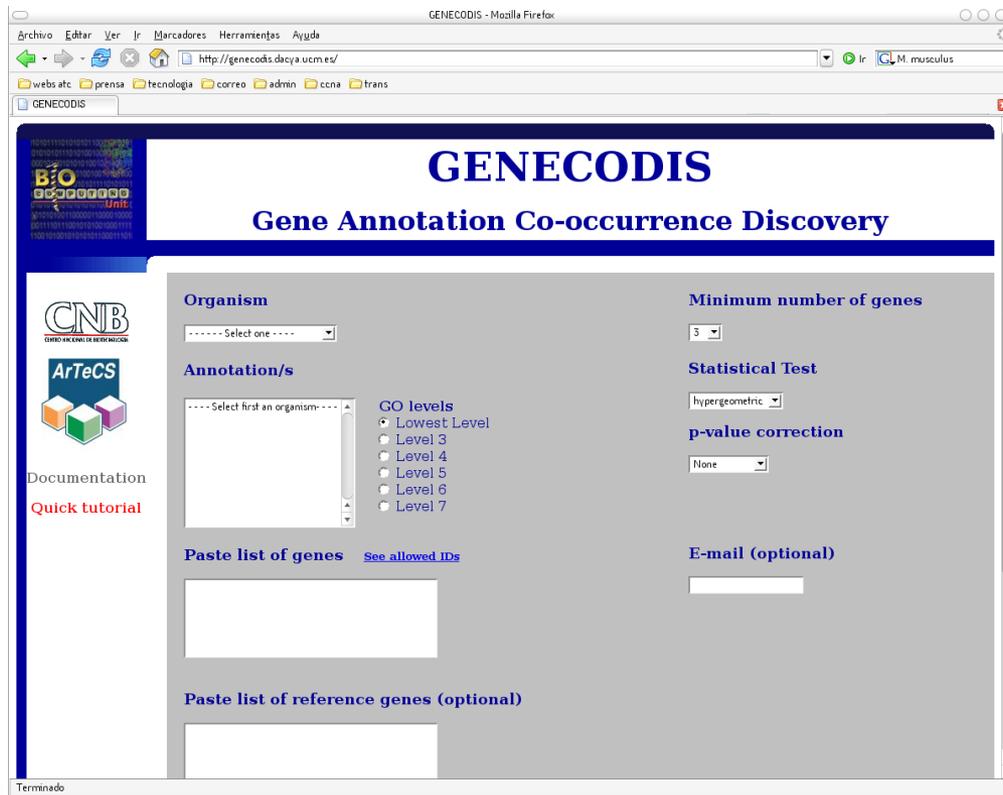


# WP4 Aplicaciones

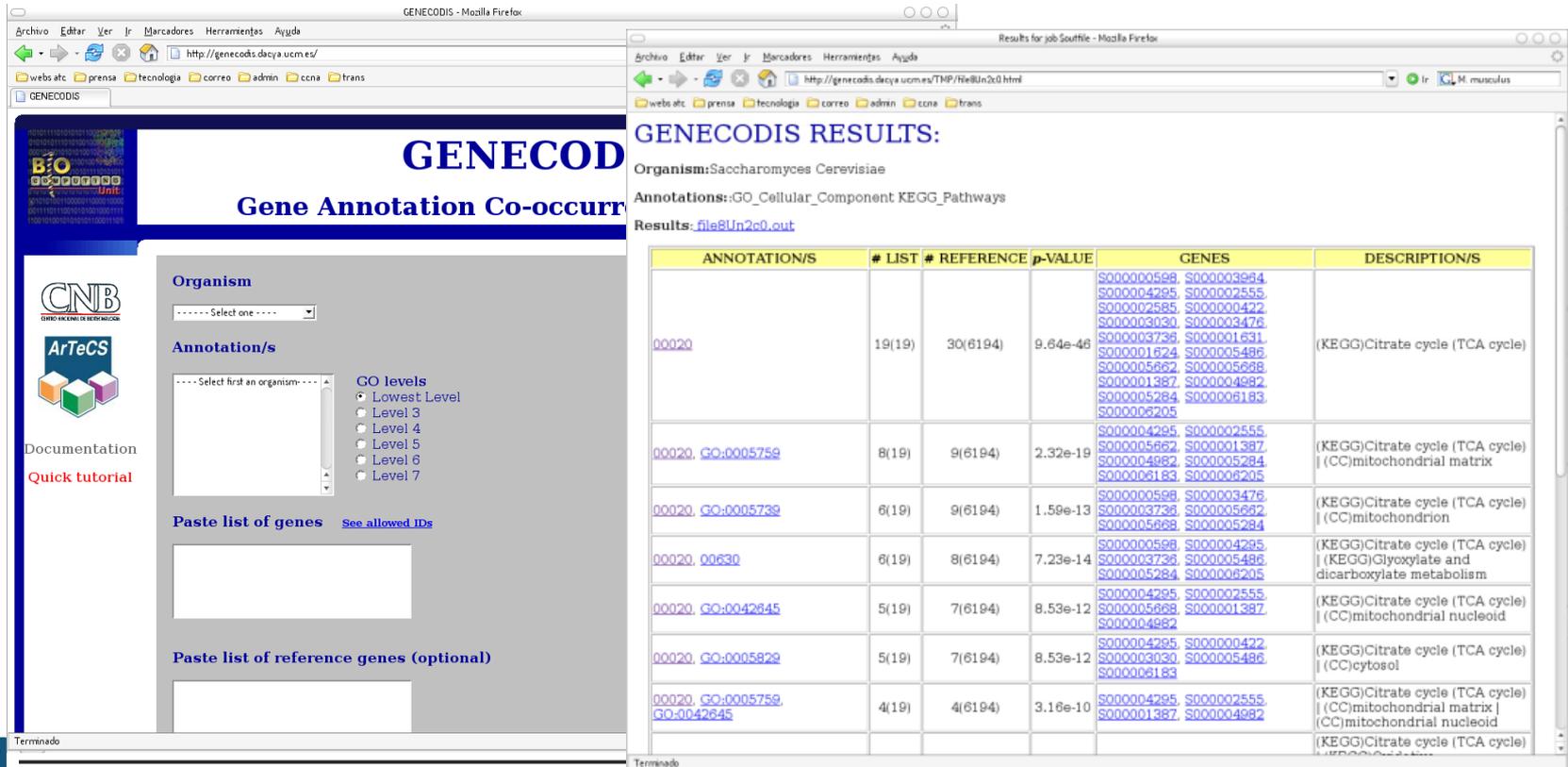
- Disponible ya
  - Gene Annotation Co-occurrence Discovery, Genecodis (UCM-CNB)
  - **<http://genecodis.dacya.ucm.es>**
- Disponibles durante 2007
  - Simulación de propagación de incendios (UAB)
  - Manejos de Flujos (UAB-UNLP)
  - BIOMNF (UCM-CNB)



## Genecodis (<http://genecodis.dacya.ucm.es>)



## Genecodis (<http://genecodis.dacya.ucm.es>)



**GENECODIS**  
Gene Annotation Co-occurrence

Organism:

Annotation/s:

GO levels:  
 Lowest Level  
 Level 3  
 Level 4  
 Level 5  
 Level 6  
 Level 7

Paste list of genes [See allowed IDs](#)

Paste list of reference genes (optional)

Terminado

---

**GENECODIS RESULTS:**  
 Organism: *Saccharomyces Cerevisiae*  
 Annotations: GO\_Cellular\_Component KEGG\_Pathways  
 Results: [file8Un2c0.out](#)

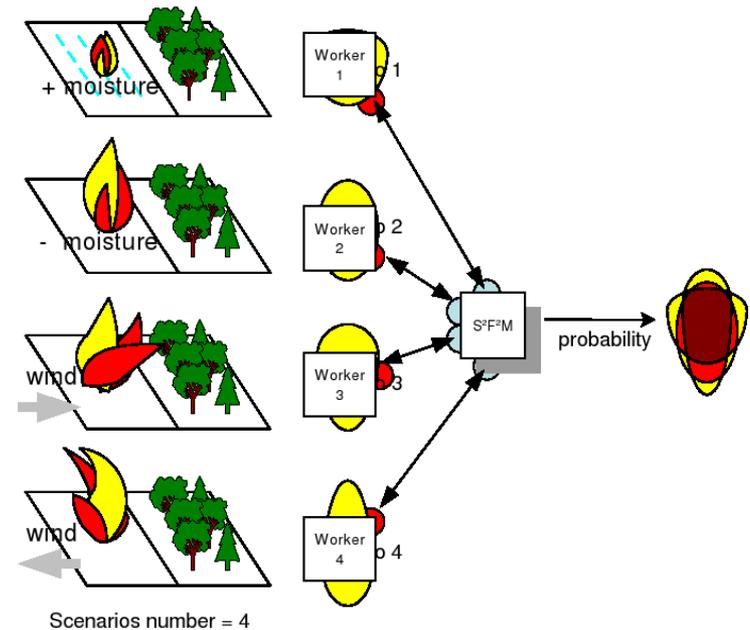
ANNOTATIONS	# LIST	# REFERENCE	p-VALUE	GENES	DESCRIPTIONS
<a href="#">00020</a>	19(19)	30(6194)	9.64e-46	<a href="#">S000000588</a> , <a href="#">S000003984</a> , <a href="#">S000004285</a> , <a href="#">S000002555</a> , <a href="#">S000002585</a> , <a href="#">S000000422</a> , <a href="#">S000003030</a> , <a href="#">S000003476</a> , <a href="#">S000003736</a> , <a href="#">S000001631</a> , <a href="#">S000001624</a> , <a href="#">S000005486</a> , <a href="#">S000005682</a> , <a href="#">S000005688</a> , <a href="#">S000001387</a> , <a href="#">S000004982</a> , <a href="#">S000005284</a> , <a href="#">S000006183</a> , <a href="#">S000006205</a>	(KEGG)Citrate cycle (TCA cycle)
<a href="#">00020</a> , <a href="#">GO:0005759</a>	8(19)	9(6194)	2.32e-19	<a href="#">S000004285</a> , <a href="#">S000002555</a> , <a href="#">S000005682</a> , <a href="#">S000001387</a> , <a href="#">S000004982</a> , <a href="#">S000005284</a> , <a href="#">S000006183</a> , <a href="#">S000006205</a>	(KEGG)Citrate cycle (TCA cycle)   (CC)mitochondrial matrix
<a href="#">00020</a> , <a href="#">GO:0005739</a>	6(19)	9(6194)	1.58e-13	<a href="#">S000000588</a> , <a href="#">S000003476</a> , <a href="#">S000003736</a> , <a href="#">S000005682</a> , <a href="#">S000005688</a> , <a href="#">S000005284</a>	(KEGG)Citrate cycle (TCA cycle)   (CC)mitochondrion
<a href="#">00020</a> , <a href="#">00630</a>	6(19)	8(6194)	7.23e-14	<a href="#">S000000588</a> , <a href="#">S000004285</a> , <a href="#">S000003736</a> , <a href="#">S000005486</a> , <a href="#">S000005284</a> , <a href="#">S000006205</a>	(KEGG)Citrate cycle (TCA cycle)   (KEGG)Glyoxylate and dicarboxylate metabolism
<a href="#">00020</a> , <a href="#">GO:0042845</a>	5(19)	7(6194)	8.53e-12	<a href="#">S000004285</a> , <a href="#">S000002555</a> , <a href="#">S000005688</a> , <a href="#">S000001387</a> , <a href="#">S000004982</a>	(KEGG)Citrate cycle (TCA cycle)   (CC)mitochondrial nucleoid
<a href="#">00020</a> , <a href="#">GO:0005829</a>	5(19)	7(6194)	8.53e-12	<a href="#">S000004285</a> , <a href="#">S000000422</a> , <a href="#">S000003030</a> , <a href="#">S000005486</a> , <a href="#">S000006183</a>	(KEGG)Citrate cycle (TCA cycle)   (CC)cytosol
<a href="#">00020</a> , <a href="#">GO:0005759</a> , <a href="#">GO:0042845</a>	4(19)	4(6194)	3.16e-10	<a href="#">S000004285</a> , <a href="#">S000002555</a> , <a href="#">S000001387</a> , <a href="#">S000004982</a>	(KEGG)Citrate cycle (TCA cycle)   (CC)mitochondrial matrix   (CC)mitochondrial nucleoid

Terminado

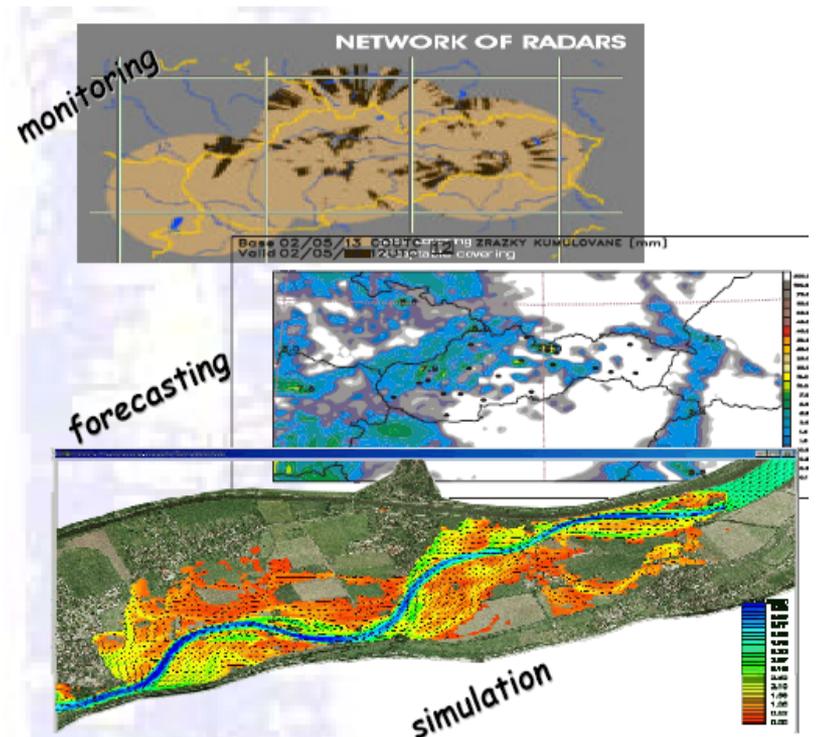


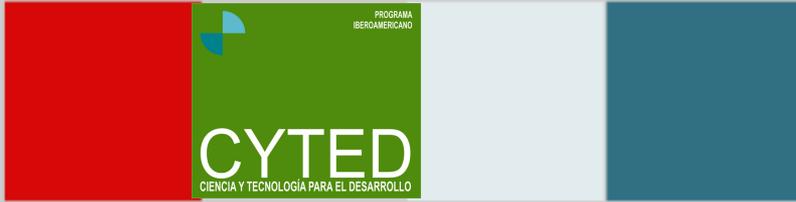
## • Simulation of Forest Fire Propagation

- S<sup>2</sup>F<sup>2</sup>M es una alternativa a la resolución de un problema mediante el uso de múltiples casos.



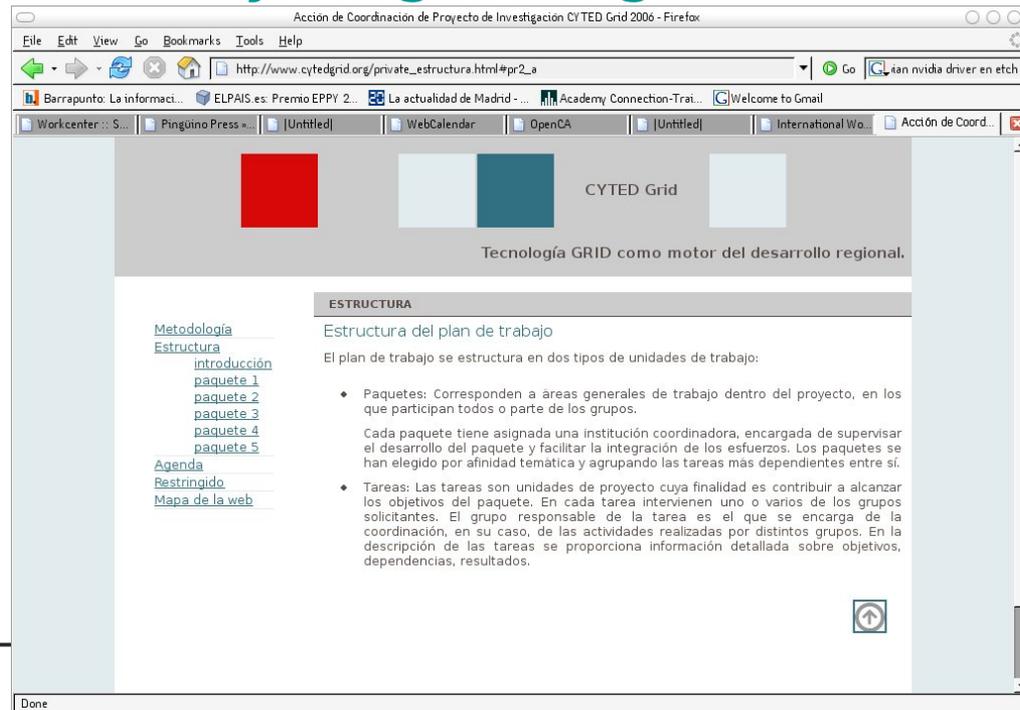
- **Simulation of Forest Fire Propagation**
  - S<sup>2</sup>F<sup>2</sup>M es una alternativa a la resolución de un problema mediante el uso de múltiples casos.
- **Flood management**
  - Los modelos simulan el flujo del agua a través de diversas estructuras fluviales para predecir el impacto de la inundación.





# Otros recursos

- Mailing List: [cytedgrid@cytedgrid.org](mailto:cytedgrid@cytedgrid.org)
- Forum
- Web site: <http://www.cytedgrid.org>





# CYTED-Grid



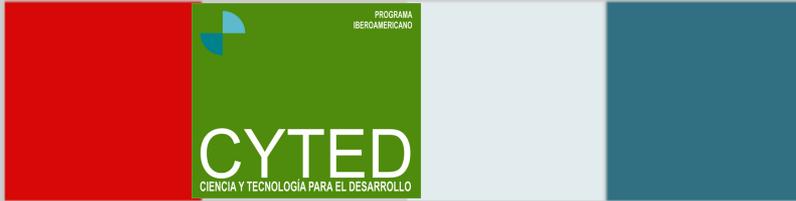
Gracias



# Software Grid

- **Middleware Actuales** : Principalmente se pueden clasificar en dos grupos:
  - **Basados en Globus Toolkit 2 (GT2)**
    - Orientado a la creación de una infraestructura de cálculo.
    - Proporciona una capa de bajo nivel con servicios básicos:
      - GRAM (Grid Resource Allocation and Management): Lanzamiento de trabajos de forma remota.
      - GridFTP: Acceso a ficheros.
      - MDS2 (Monitoring and Discovery System): Sistema de información.
      - RLS (Replica Locator Service): Gestión de replicas de ficheros.
      - ...
  - **Basados en Globus Toolkit 4 (GT4)**
    - Orientado a servicios grid.
    - Proporciona funcionalidad similar a GT2 pero implementada como servicios grid:
      - WSGRAM (Web Services Grid Resource Allocation and Management): Lanzamiento de trabajos de forma remota.
      - MDS4 (Monitoring and Discovery System): Sistema de información.
      - RFT(Reliable File Transfer): Transmisión de ficheros (usa GridFTP).
      - ...

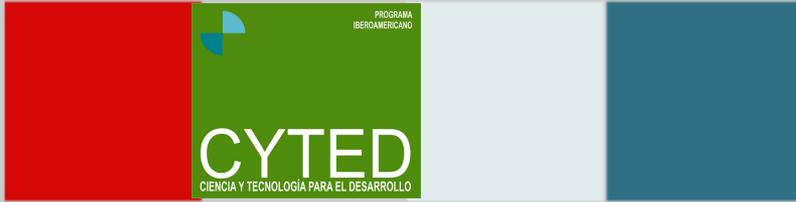




# Software Grid

- Elección del modelo de middleware:
- GT2:
  - Orientado a la creación de infraestructuras grid.
  - Mas asentado y estable.
  - Mejores prestaciones.
  - Mayor numero de desarrollos basados en él.
  - Desarrollo finalizado.
- GT4:
  - Misma funcionalidad que GT2 pero actualmente en desarrollo.
  - La mayoría de los proyectos en desarrollo actuales se basan en él.
  - Más orientado a la creación de arquitecturas de servicios grid, que a la creación de infraestructuras.
  - Menos tiempo en “producción” y por tanto menor estabilidad.





# Software Grid

- Middleware del proyecto EGEE
  - Creación de una infraestructura Grid de alta disponibilidad para la e-Ciencia en Europa.
  - Uno de los middleware basados en GT2 más destacados.
  - Gran implantación
  - Mucha funcionalidad
- LCG/gLITE tiene la ventaja de estar soportado por el CERN a través del principal proyecto internacional de e-Infraestructura (EGEE, EGEE2) y sus proyectos satélite (**EELA**, EUMEDGRID, EUCHINAGRID, SEEGRID) así como por la transición que la CE está planificando, EGEE/EGEE-2/EGI, para asegurar la sostenibilidad de esta e-Infraestructura.

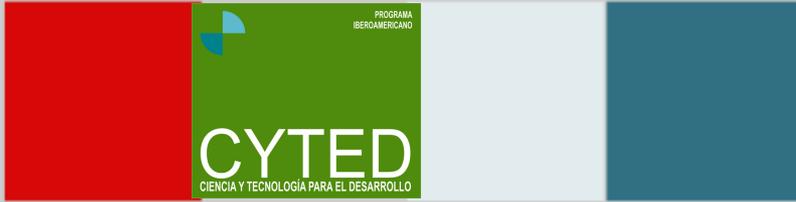




# Software Grid

- Glite 3.0
  - Proyecto EGEE
  - Basado globus 2.4
  - Sistema operativo Scientific Linux (Red Hat)
  - Versión 3.0.0 (4/5/06)
- Globus Toolkit 4.0.2
  - Basado en Web Services
  - Permite múltiples plataformas
  - Versión 4 estable
- Dada la orientación computacional del proyecto, GT2 podría ser una solución apropiada, !!pero es un desarrollo “finalizado”!!
- **GT4 es el middleware principal en el que se basan la mayoría de los desarrollos actuales.**

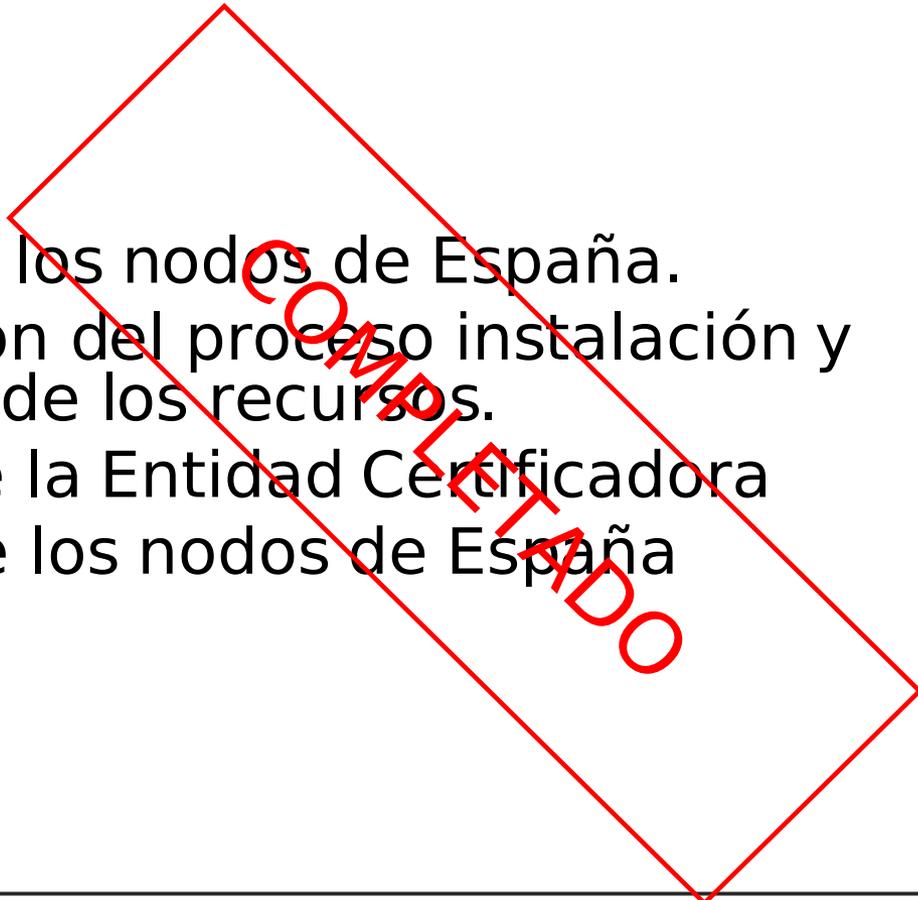




# Estado Actual

## Primera fase:

- Instalación de los nodos de España.
- Documentación del proceso instalación y configuración de los recursos.
- Despliegue de la Entidad Certificadora
- Integración de los nodos de España





# Estado Actual

## Despliegue de la e-infraestructura:

- Test de infraestructura Grid
- Instalación de los nodos de España (UPV, UCM, UAB, UDC).
- Documentación del proceso de instalación y configuración de los recursos.
  - LiveCD
  - Instalación estandar.
- Despliegue de la Entidad Certificadora <http://aker.dsic.upv.es:10443/pub>
- Integración de recursos.

COMPLETADO



## Despliegue de la e-infraestructura: Recursos disponible noviembre 2006

- UPV (2 cluster, 23 procesadores)
  - 6 dual AMD64 Opteron@2.4GHz, 1GB RAM.
  - 5 dual Xeon@2GHz, 1GB RAM.
  - 1 Pentium IV@3Ghz, 256MB RAM.
- – UAB (1 cluster , 8 procesadores )
  - 8 Pentium IV@2.8GHz, 512MB/1GB RAM.
- – UDC (1cluster, 26 procesadores)
  - 1 dual Xeon@3.2GHz, 4GB RAM.
  - 12 dual Xeon@1.8GHz, 2.8GHz or 3.2GHz, 1GB/2GB RAM.
  - 1 AMD Athlon 64 3200+, 1GB RAM.
  - 1 AMD Athlon XP 2600+, 512MB RAM.
- – UCM (2 cluster, 38 procesadores )
  - 2 Pentium IV@2GHz or 2.4GHz, 1GB RAM.
  - 8 AMD@2.0GHz dual core, 4GB RAM.
  - 8 Itanium II@900MHz dual, 2GB RAM (nodes) and 3GB RAM (front-end).

COMPLETADO



# Estado Actual

**Despliegue de la e-infraestructura:** Recursos a integrar antes de primer trimestre 2007

- Universidad de OPORTO
- Universidad de San Luis ( Argentina)
- Universidad de la Plata ( Argentina )
- Universidad de San Salvador de Bahía ( Brasil )

