

## **Enabling Grids for E-science, el grid al servicio de la e-ciencia**

*Carlos Fernández Sánchez.*

*Coordinador de sistemas HPC*

*Fundación Centro Tecnológico de Supercomputación de Galicia. CESGA*

### **Construyendo el grid global**

Hoy en día la web e internet se han hecho muy populares y todo el mundo reconoce su importancia. La web nos permite compartir información entre sitios dispersos a lo largo de todo el mundo, el grid pretende ir mas allá y permitir no sólo compartir información sino también recursos, sean estos capacidad de cálculo, almacenamiento, bases de datos, resultados de mediciones incluso en tiempo real, etc.

En la actualidad Europa es el continente que más está apostando por la tecnología grid, y la Unión Europea dentro de sus programas de desarrollo de infraestructuras de Tecnologías de la Información (FP6 IST), está financiando numerosos proyectos grid. Entre ellos el más importante es el proyecto Enabling Grids for E-Science (EGEE), a través del cual se pretende integrar las distintas iniciativas Grid nacionales, regionales y temáticas existentes, con el fin de desarrollar una infraestructura Grid para el soporte de la investigación científica. EGEE proporciona a la comunidad investigadora pública y privada acceso a recursos de computación, independientemente de su localización geográfica. Se trata por tanto de un proyecto de ámbito global, financiado en su mayor parte por la Unión Europea, y que recibe contribuciones de otros países como Estados Unidos, Rusia y Taiwán.

Para el desarrollo del proyecto, se utiliza las Redes de Investigación Regionales, Nacionales e Internacionales para la interconexión por internet de los diferentes centros. Además, se aprovecha la experiencia que ha sido generada en proyectos anteriores como Datagrid y Crossgrid de la Unión Europea y otras iniciativas como el programa de e-Ciencia del Reino Unido, el grid italiano INFN, Nordugrid y el Trillium de los Estados Unidos. Uno de los objetivos fundamentales es que esta infraestructura proporcione la interoperatividad necesaria con otros grids que ya se encuentran en funcionamiento, incluyendo a los Estados Unidos y Asia, con el fin de construir así una infraestructura grid que abarque todo el globo.

Así mismo, la infraestructura del proyecto también proporciona soporte a aquellas comunidades distribuidas de investigadores que trabajan conjuntamente y que quieran compartir sus recursos computacionales dentro del grid, actuando de este modo como proveedores y usuarios de los recursos del grid, siempre y cuando se acojan a las políticas de utilización comunes e imperativas para participar en la infraestructura del proyecto.

### **Estado del proyecto**

El proyecto EGEE comenzó oficialmente en Abril del 2004 y tiene una duración de dos años, aunque está concebido como un programa a cuatro años. Actualmente ya se han desarrollado los servicios básicos y existe una gran cantidad de recursos computacionales funcionando en modo de producción en la mayoría de las instituciones participantes en el proyecto. Además, los grupos de usuarios vienen incrementándose desde el inicio del proyecto. Está previsto que el proyecto tenga continuidad en una segunda fase tras los 2 primeros años y en la cual las instituciones vayan asumiendo progresivamente las funciones de operar y mantener la infraestructura Grid de forma estable y con soporte las 24 horas del día.

*Apenas un año después del inicio del proyecto (Marzo 2005), ya forman parte de la infraestructura del proyecto más de 90 centros de recursos (RC), que ofrecen en su conjunto un grid formado por*

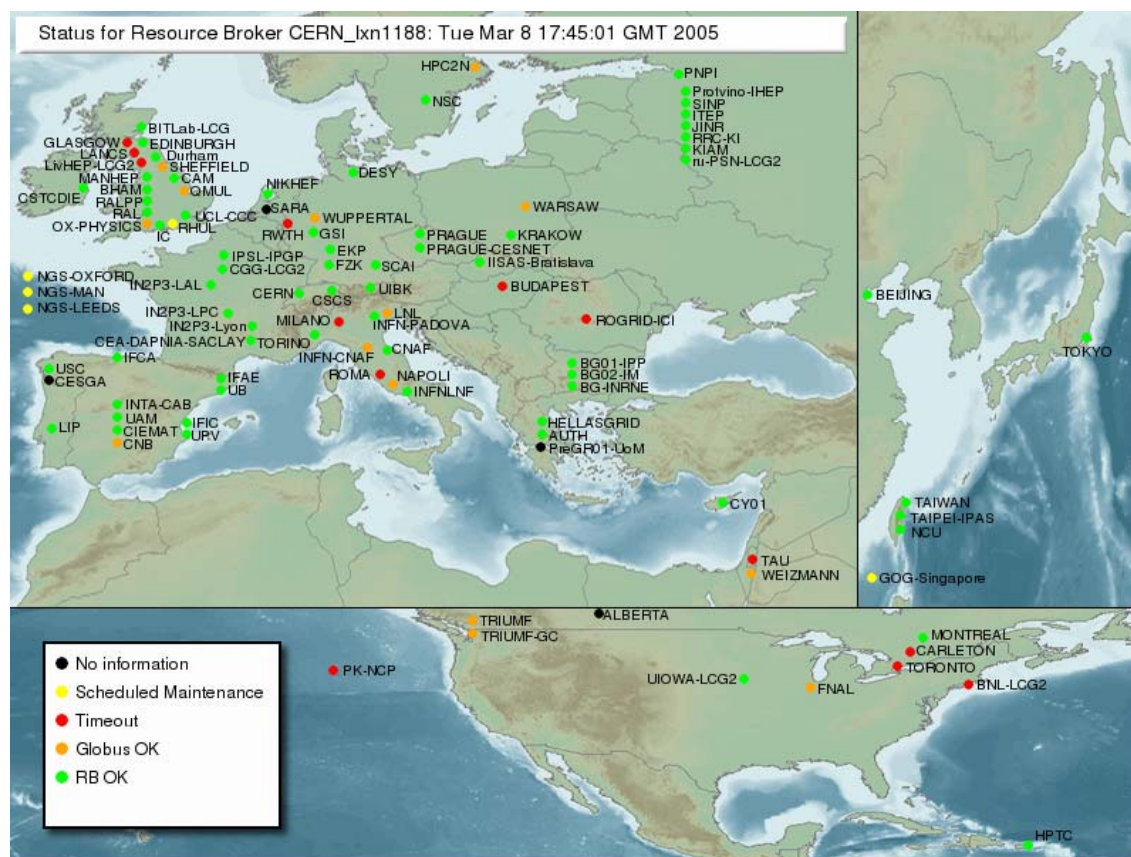
más de 9000 procesadores funcionando en modo de producción. Más de 2000 procesadores se encuentran localizados en centros que no forman parte directa del proyecto EGEE, en lugares como Taiwan, China, Estados Unidos, Japón, Canadá, Pakistán y la India. Ya hay más de 700 usuarios registrados utilizando este grid y formando parte de alguna de las más de 60 organizaciones virtuales existentes.

Los recursos estarán accesibles a las comunidades de usuarios y las organizaciones virtuales de acuerdo con las políticas de acceso acordadas y los niveles de servicio contratados.

## Participantes

Dentro del proyecto participan activamente más de 70 organizaciones punteras de 27 países, agrupados en Grids regionales, con el objetivo de aglutinar más de 20,000 procesadores al fin del proyecto, en la que será la mayor infraestructura Grid jamás desarrollada. La Unión Europea aporta 32 millones de euros al proyecto y los participantes financian una cantidad similar. La aportación humana dedicada al proyecto está compuesta aproximadamente por 600 personas-año a lo largo de 2 años.

Este gran número de investigadores participantes implica un reto en cuanto a la coordinación de las distintas actividades y a la comunicación entre estas personas con una distribución geográfica tan extensa, siendo este aspecto un proyecto dentro del propio proyecto. Para facilitar y coordinar el desarrollo de las actividades se han organizado estructuras Grid a nivel regional (federaciones), agrupando diferentes países dentro de una misma región geográfica próxima. Cada una de estas regiones tiene un miembro en el Comité de Gestión del Proyecto (Project Management Board), así



como el propio CERN, como líder del proyecto y las Redes de Investigación Europeas.

Las 12 federaciones existentes son:

1. CERN
2. Europa Central, incluyendo Austria, República Checa, Hungría, Polonia, Eslovaquia y Eslovenia
3. Francia
4. Alemania y Suiza
5. Irlanda y Reino Unido
6. Italia
7. Norte de Europa, incluyendo Bélgica, Dinamarca, Finlandia, Holanda, Bélgica, Dinamarca, Finlandia, Noruega y Suecia
8. Rusia
9. Sureste de Europa, incluyendo Bulgaria, Chipre, Grecia, Israel y Rumanía
10. Suroeste de Europa, en donde se encuentran España y Portugal
11. NRENS (National Research and Education Networks)
12. Estados Unidos

A su vez, las actividades del proyecto están divididas en paquetes de trabajo:

- NA1: Gestión global del proyecto
- NA2: Diseminación y divulgación
- NA3: Cursos para usuarios y tutoriales
- NA4: Identificación de aplicaciones para el grid y soporte
- NA5: Políticas de uso y cooperación internacional
- SA1: Soporte, Operación y gestión del Grid en Europa
- SA2: Gestión de los recursos de Red
- JRA1: Integración y re-ingeniería del middleware
- JRA2: Gestión de la calidad
- JRA3: Seguridad
- JRA4: Desarrollo de servicios de Red
- Industry Forum: Participación de las empresas en el Grid

### **El papel de España en el desarrollo del proyecto**

España y Portugal son los países integrantes de la federación del Suroeste de Europa (SWE) y las instituciones participantes, coordinadas por el PIC desde Barcelona, aparecen en la tabla 1:

<b>Siglas</b>	<b>Institución</b>	<b>Localización</b>	<b>Actividades</b>
CESGA	Fundación Centro Tecnológico de Supercomputación de Galicia	Santiago de Compostela	SA1
CNB	Centro Nacional de Bioinformática	Madrid	NA4
CSIC - IFCA	Consejo Superior de Investigaciones Científicas. Instituto de Física de Cantabria	Santander	SA1
CSIC-IFIC	Consejo Superior de Investigaciones Científicas. Instituto de Física Corpuscular	Valencia	SA1
INTA	Instituto Nacional de Técnica Aeroespacial	Madrid	SA1
LIP	Laboratorio de Instrumentação e Física Experimental de Partículas	Lisboa	SA1
PIC	Port d'Informació Científica	Barcelona	SA1

Red.es	Red.es	Madrid	NA5
UPV	Universidad Politécnica de Valencia	Valencia	NA4

Como participantes en el grupo de trabajo SA1, la federación SWE tiene como misión la de crear, operar, soportar y mantener la infraestructura de Grid a nivel de producción con el fin de proporcionar los recursos de computación, almacenamiento e instrumentación, así como el soporte técnico a los distintos centros de recursos (RC) existentes y potenciales dentro de la región.

Uno de los puntos fundamentales dentro de las actividades que deberán llevarse a cabo, reside en incorporar los recursos grid computacionales y de conocimiento ya existentes dentro de la región, a la infraestructura del proyecto EGEE. Entre las actividades que desarrollaremos se encuentran:

- **Servicios fundamentales de la infraestructura:** operar servicios de información, brokers de recursos, gestión de la información y administración de las organizaciones virtuales.
- **Monitorización y control del Grid:** monitorizar el estado operativo del grid, así como su rendimiento, aplicando las acciones correctivas necesarias para resolver los problemas que puedan aparecer, tanto en la infraestructura como en los servicios prestados.
- **Implantación del middleware<sup>1</sup> y captación de recursos:** validar las nuevas versiones del software e implantarlas en los centros de recursos del grid. Además, se proveerá el soporte necesario a cualquier centro que pueda aportar recursos al grid, ayudando en todo lo relacionado con la instalación del middleware y los procedimientos operativos.
- **Soporte a los usuarios y a los recursos:** responder y coordinar la resolución de problemas que puedan surgir entre los centros de recursos y los usuarios.
- **Gestión del Grid:** coordinar el cumplimiento de los anteriores objetivos por parte de los Centros de Operaciones Regionales (ROC) y los Centros de Infraestructura Fundamentales (CIC), a través de la negociación de acuerdos de nivel de servicio (SLA).
- **Colaboración Internacional:** llevar a cabo acuerdos con organismos en América, Asia y el Pacífico y asegurar la interoperatividad de las infraestructuras y servicios grid entre ellos.

Para poder realizar las actividades anteriores, existe una estructura de gestión del grid jerarquizada en los siguientes niveles:

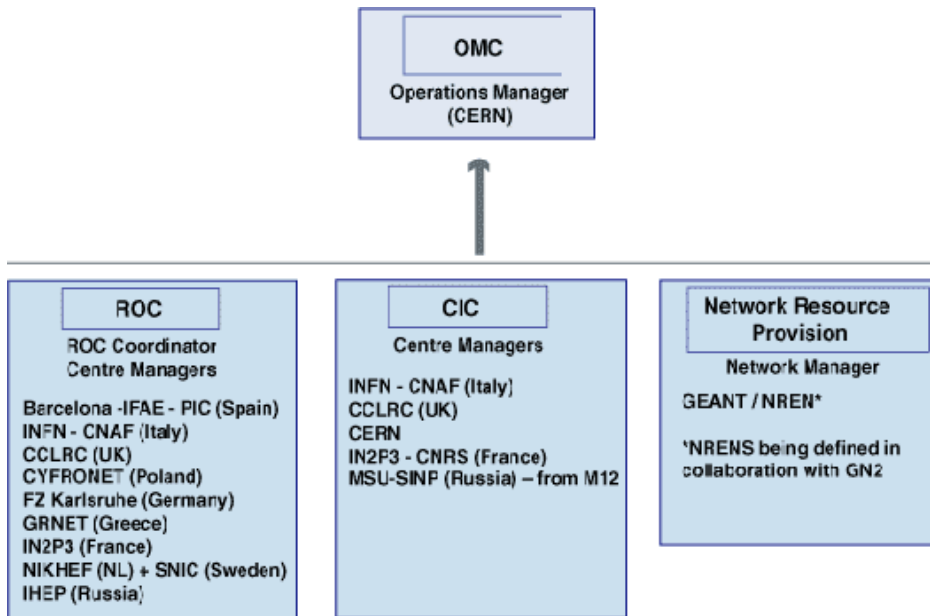
- **Nivel 1:** Centros de operaciones regionales (ROC), que se encargan de implementar el middleware Grid en los Centros de Recursos para conectar recursos al grid de EGEE y proporcionar el soporte de primer nivel a los usuarios y los Centros de Recursos de la región. Existe un ROC en cada una de las federaciones que participan en el proyecto, con la peculiaridad de que en la federación SWE las funciones del ROC se encuentran distribuidas entre varias instituciones.
- **Nivel 2:** Centros de Infraestructuras Fundamentales (CIC) que proporcionan la infraestructura de servicios del Grid más crítica, operando los servicios esenciales para la conexión de los usuarios a los recursos. Los CICs también soportarán a los ROCs, garantizando el funcionamiento del Grid las 24 horas del día. Actualmente existen 4 CICs en toda Europa.

---

<sup>1</sup> Middleware: conjunto de componentes software, intermedios entre el sistema operativo y las aplicaciones, y en el que se sustenta la arquitectura Grid.

- **Nivel 3:** Centro de gestión de operaciones (OMC), el cual es único y se encarga de administrar las operaciones del conjunto del Grid de forma centralizada.

Además de esta estructura de gestión, los centros de recursos (RC) se encargan de proporcionar y administrar la capacidad de computación, el almacenamiento y los recursos de red. Estos RCs trabajan coordinados por los ROCs, de donde también reciben el soporte técnico u supervisión necesarios para el despliegue de la infraestructura.



### El papel de las empresas en el proyecto EGEE

Las empresas pueden participar en el proyecto EGEE de tres modos diferentes:

**Actuando como participantes dentro del proyecto.** A través de la colaboración con los distintos participantes del proyecto, las empresas tienen la oportunidad de contribuir en el desarrollo de actividades específicas, incrementando así sus conocimientos en la tecnología grid.

**Actuando como usuario final.** Determinados sectores empresariales e industriales, se pueden beneficiar como usuarios potenciales de la infraestructura Grid a través de proyectos de I+D que demanden de grandes recursos computacionales y de almacenamiento.

**Actuando como proveedor de servicios,** ya que la construcción de un Grid eficiente y de producción requiere que las propias empresas se involucren en el largo plazo, y de esta forma, participen en el mantenimiento de los servicios Grid más críticos, tales como los centros de soporte y atención telefónica, por ejemplo.

### Qué proyectos científicos utilizan la infraestructura EGEE.

Cualquier disciplina científica con altas demandas de computación se puede beneficiar de la infraestructura del proyecto EGEE. En la práctica, existen dos áreas o aplicaciones piloto que han sido seleccionadas para guiar la implementación inicial y para certificar el rendimiento y funcionalidad de los desarrollos. Una es el Grid para la Computación del Colisionador de Hadrones (LHC), el cual se basa en una infraestructura grid para almacenar y analizar petabytes ( $10^{15}$  bytes) de datos, tanto reales como procedentes de simulaciones de experimentos de física de altas energías

desarrollados en el CERN. La otra área o reto de aplicación se localiza en el campo de la biomedicina, en donde un gran número de comunidades investigadoras se enfrentan a retos igualmente complejos, como el data-mining aplicado a la genómica, y la indexación de las bases de datos médicas procedentes de centros hospitalarios, lo que en la práctica representa decenas de terabytes de información por hospital y año. Este tipo de información añade al proyecto la dificultad de asegurar la confidencialidad y seguridad de los datos almacenados y transmitidos a través de la red.

**Referencias y más información en:**

Web del proyecto: <http://public.eu-egee.org/>

Web de la federación SWE: <http://grid.ifca.unican.es/egee-sa1-swe/>

Centro de operaciones del grid: <http://goc.grid-support.ac.uk/>