

Libro blanco
e-CIENCIA en España
2004

Libro blanco

e-Ciencia en España

Edita:

FECYT (Fundación Española para la Ciencia y la Tecnología)©

Diseño y Maquetación:

Global Diseña

Impresión:

Depósito Legal:

M-9625-2005

Resumen ejecutivo

Hace más de 30 años Leonard Kleinrock, uno de los padres fundadores de Internet, apuntaba que probablemente veríamos la propagación de los servicios de computación que, como la luz o el teléfono, llegarían hasta casas y oficinas. Años más tarde, en 1998, esta idea empezaba a materializarse con “The Grid: Blueprint for a Future Computing Infrastructure”, de Ian Foster y Carl Kesselmann, y el inicio de la Globus Alliance, creada para organizar la investigación y el desarrollo para crear tecnologías básicas para el Grid.

La e-Ciencia se entiende como el conjunto de actividades científicas desarrolladas mediante el uso de recursos distribuidos accesibles a través de Internet. Hoy en día, cálculo, almacenamiento e información, entre otros, constituyen los principales recursos utilizados y compartidos mediante la red. La evolución de las redes de comunicaciones de alta velocidad dedicadas a la investigación y de las tecnologías y aplicaciones colaborativas están creando un escenario idóneo para la interacción entre investigadores.

Aunque existen diversas tecnologías para la compartición y el acceso de recursos distribuidos, el Grid parece haberse erigido como el estándar. La herramienta más utilizada en entornos científicos parte de Globus Toolkit, un *software* de código fuente abierto impulsado por la Globus Alliance que permite la construcción de *grids*. En poco más de cinco años, tanto en Estados Unidos como en Europa se han puesto en marcha diferentes proyectos Grid, algunos de los cuales tienen una significativa participación española.

Así, tras el desarrollo inicial de Internet durante los últimos treinta años y la revolución que ha supuesto el World Wide Web, se han desarrollado *grids* de cálculo, de datos y de servicios. Tras esta última fase, la siguiente meta es el llamado Grid de conocimiento para compartir además experiencia y conocimiento científico y técnico. Este concepto justifica el nombre de e-Ciencia.

Para el desarrollo de la e-Ciencia, es básico disponer de redes de comunicaciones. Desde hace más de una década, España cuenta con una red nacional académica y de investigación, RedIRIS, que actualmente conecta a más de 260 instituciones, principalmente universidades y centros de I+D, a los que también proporciona conectividad externa con otras redes de investigación y con la Internet global comercial. Dentro del Estado, cuatro comunidades autónomas (Andalucía, Catalunya, Galicia y Madrid) disponen de redes propias de alta velocidad que enlazan con la red estatal.

El conjunto de estas infraestructuras, redes y *middleware*, permitirá a la comunidad investigadora acceder a los recursos de manera sencilla, ágil, segura, permanente y económica. Los diferentes centros (autonómicos, universitarios...) existentes en el Estado disponen de infraestructura de cálculo y de almacenamiento que podrá estar a disposición de una comunidad más amplia a través del Grid. También la información generada y almacenada en repositorios podrá ser compartida eficientemente con personas autorizadas. Además, se podrán operar remotamente instrumentos como telescopios, microscopios, capturadores de imagen, entre otros.

Diversas áreas de investigación podrán beneficiarse del desarrollo de la e-Ciencia. En este libro blanco se han considerado ocho áreas (Astronomía y Espacio, Biomedicina y Ciencias de la Salud, Ciencia y Tecnología de Materiales, Ciencias de la Tierra, Física, Ingeniería, Química y Tecnologías para la Sociedad de la Información), aunque el programa de e-Ciencia aportaría sin duda beneficios a otras aplicaciones.

El desarrollo de la e-Ciencia, en general, tendrá un impacto científico con la explotación eficiente de centros o recursos de excelencia y la existencia de nuevas formas de compartir conocimiento; un impacto tecnológico, ya que permitirá abrir nuevos mercados y nuevas formas de colaboración y desarrollo de proyectos; y un impacto social, pues proveerá de acceso para vencer la brecha tecnológica. Demorar su desarrollo alejaría demasiado España de los países que ya disponen de un programa de e-Ciencia.

Para evitarlo se propone:

1. Al MEC, que lance una acción estratégica incluida en el PNID 2004-2007, que cree un Comité Asesor y/o de Gestión, que nombre representantes para la coordinación internacional y que asegure la disponibilidad de recursos.
2. A la FECYT, que nombre un coordinador interno y un Grupo de Expertos, que facilite los encuentros para realizar acciones concretas (estructuración de una plataforma nacional, formación y difusión, y organización de una jornada internacional).
3. A RedIRIS, que coordine el lanzamiento de una plataforma Grid nacional a corto plazo, que afiance tecnologías, las inventaríe y las ofrezca a los grupos, que proponga la creación de experiencias piloto en cada campo, y que fortalezca la conectividad hasta el usuario final.

Preámbulo

La Comisión de Ciencias e Ingenierías de la Fundación Española de Ciencia y Tecnología (FECYT) impulsó la constitución de un Grupo de Expertos para elaborar un libro verde que identificase los recursos existentes en e-Ciencia, reconociese su necesidad para distintas aplicaciones y determinase las acciones para su uso global. Este libro verde fue el punto de partida para el debate que tuvo lugar en Madrid el pasado 8 de octubre y que reunió un centenar de expertos. Este debate ha servido para editar el presente libro blanco, que recoge sus recomendaciones de acciones y políticas de apoyo a la e-Ciencia en España.

El Grupo de Expertos encargado de la elaboración de ambos libros está integrado por Mateo Valero, de la UPC, y José Manuel Báez, de la FECYT (coordinadores), Miquel Huguet, del CESCA (relator), Rafael García Tamarit, de la FECYT (secretario), y como vocales Luis Balairón, del INM, Víctor Castelo, del CSIC, Manuel Delfino, de la UAB, Javier García Tobío, del CESGA, Javier Jiménez Sendín, de la UPM, Jesús Marco, del CSIC, Fernando Martín Sánchez, del ISCIII, y Daniel Ponz, de la ESA. Este Grupo de Expertos ha contado también con la colaboración de Lluís Anglada, del CBUC, Juan José Moreno, de la UPM, y Santiago Olivella y Pablo Ordejón, ambos del CSIC.

El libro blanco se divide en seis partes. En la primera se describen los conceptos fundamentales de la e-Ciencia y en la segunda se presenta la infraestructura necesaria para su desarrollo. Ésta consiste en *middleware*, que permite utilizar por parte de las aplicaciones, de forma conjunta o coordinada, recursos disponibles en diversos centros; redes de comunicaciones, que posibilitan gracias a su gran capacidad de transmisión la puesta en práctica de actividades de e-Ciencia, y recursos a compartir: de cálculo, de almacenamiento, de información, etc.

En la tercera parte se muestran diversas áreas de aplicación en las cuales el desarrollo de la e-Ciencia, en general, tendrá un impacto científico, tecnológico y social. Así, se describen los usos y necesidades, retos a afrontar, retornos y beneficios para cada una de estas áreas que, evidentemente, deben considerarse una muestra y no una lista completa ni excluyente.

En la cuarta parte se muestran algunos proyectos Grid en desarrollo en Europa, como CrossGrid, European Data Grid y Enabling Grids for E-science in Europe, que tienen participación española.

Por último, se incluyen las conclusiones y las acciones y recomendaciones recogidas durante los debates del Grupo de Expertos y de la jornada del pasado 8 de octubre para el desarrollo de un programa de e-Ciencia en España.

1. ¿Qué es la e-Ciencia?9
2. Infraestructura necesaria12
2.1. <i>Middleware</i>12
2.2. Redes de comunicaciones16
2.3. Recursos a compartir21
3. Aplicaciones26
3.1. Astronomía y Espacio26
3.2. Biomedicina y Ciencias de la Salud28
3.3. Ciencia y Tecnología de Materiales30
3.4. Ciencias de la Tierra31
3.5. Física33
3.6. Ingeniería34
3.7. Química35
3.8. Tecnologías para la Sociedad de la Información36
3.9. Algunos ejemplos clave38
4. Algunos proyectos en desarrollo41
5. Conclusiones43
6. Acciones y recomendaciones45

1. ¿Qué es la e-Ciencia?

Se entiende la e-Ciencia como el conjunto de actividades científicas desarrolladas mediante el uso de recursos distribuidos accesibles a través de Internet. Hoy en día, cálculo, almacenamiento e información, entre otros, constituyen los principales recursos utilizados y compartidos mediante la red. La evolución de las redes de comunicaciones de alta velocidad dedicadas a la investigación y de las tecnologías y aplicaciones colaborativas están creando un escenario idóneo para la interacción entre investigadores. Por este motivo, si bien la e-Ciencia puede llevarse a cabo de forma individual, ésta es más efectiva cuando va unida a una colaboración global.

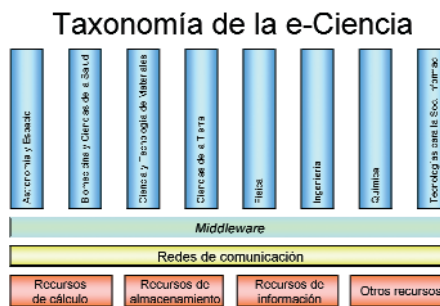


Figura 1. Taxonomía de la e-Ciencia

En la taxonomía de la e-Ciencia (figura 1) se han representado tres capas horizontales que corresponden a los recursos accesibles (de cálculo, de almacenamiento, de información y otros recursos), las redes de comunicación que permiten el acceso a éstos, y el *middleware* o *software* intermediario. Esta última capa permite a las aplicaciones utilizar de forma conjunta, o coordinada, recursos disponibles en localizaciones remotas. Estas tres capas son comunes para el desarrollo de la e-Ciencia en cualquiera de las ocho áreas de investigación representadas que serán comentadas en este documento y que, evidentemente, no son las únicas que pueden beneficiarse de la e-Ciencia.

La e-Ciencia utiliza la tecnología Grid para simplificar el acceso a los recursos y aprovechar su utilización, aunque no es el único método para conseguirlo. Otras tecnologías como el telnet, el ftp o el P2P permiten también la accesibilidad a recursos distantes. Según la definición de uno de los padres del Grid, Ian Foster¹, “un Grid es un sistema que coordina recursos, que no están sujetos a un control centralizado, usando interfaces y protocolos estándares, abiertos y de propósito general para proveer de servicios relevantes”.

¹ “What is the Grid? A three point checklist”, *Grid Today*, 2002.

Ya en 1969, Leonard Kleinrock, uno de los padres de Internet, decía: “Probablemente veremos la propagación de los servicios de computación (“computer utilities”) que, como actualmente la luz o el teléfono, llegarán hasta casas y oficinas de todo el país”.

Kleinrock ya asemejaba estos servicios computacionales a los servicios telefónicos o eléctricos, de ahí el término Grid, con el que se denomina en inglés la red eléctrica. Entre un Grid eléctrico y uno computacional, no obstante, existen algunas diferencias. Mientras que la electricidad es producida más económicamente centralizada en una gran central, se caracteriza por una métrica muy simple y se accede a través de una interfaz de conexión estandarizada, los recursos computacionales son producidos más económicamente de forma distribuida con servidores o *clusters*, se caracterizan por una métrica altamente compleja y se acceden mediante interfaces de *software* en proceso de estandarización.

Tras el desarrollo inicial de Internet durante los últimos treinta años y la revolución que ha supuesto el World Wide Web, se desarrolla el concepto de Grid de cálculo con el fin de compartir y utilizar recursos de computación. Este concepto se amplía con la utilización de recursos de almacenamiento masivo distribuido, dando lugar a Grid de datos.

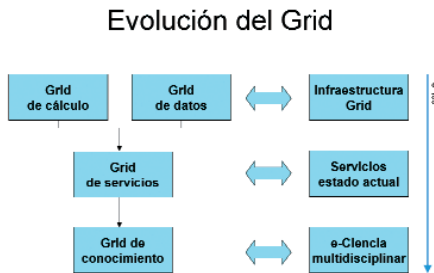


Figura 2. Evolución del Grid

Estos primeros pasos en el concepto de Grid están evolucionando para constituir el llamado Grid de servicios, que depende de cada área temática con servicios específicos. Estamos actualmente en esta fase. La meta prevista en esta evolución es el llamado Grid de conocimiento para compartir no sólo cálculo, datos o servicios sino también experiencia y conocimiento científico y técnico (figura 2). Este concepto justifica el nombre de e-Ciencia.

Desde el punto de vista de la ciencia, la infraestructura de e-Ciencia proporcionará al investigador los recursos necesarios (computacionales, informativos, etc.) para su trabajo desde una única interfaz que tenga las siguientes características:

- **Sencilla:** su uso ha de ser intuitivo como en el caso de los navegadores web.

- **Transparente:** el investigador no tiene porqué conocer cómo ni dónde se le provee del servicio requerido.
- **Ágil:** el tiempo de respuesta ha de estar dentro de los márgenes aceptables conocidos en las aplicaciones tradicionales de puesto de trabajo, es decir, desde unos pocos segundos o incluso menos de una décima de segundo para determinadas tareas, hasta algunas decenas de segundos para respuestas elaboradas.
- **Fiable:** debe garantizar los resultados obtenidos.
- **Segura:** debe controlar el acceso a recursos e información.
- **Permanente:** ha de estar disponible siempre que se necesite, al igual que la luz eléctrica, el agua, etc.
- **Económica:** el investigador ha de ser consciente de los costes incurridos para evitar despilfarro de la infraestructura -que perjudicaría a otros investigadores-, pero el coste ha de ser menor que tener su propia infraestructura, que normalmente le dará menos prestaciones pero más satisfacciones simplemente por el hecho de poseerla.

2. Infraestructura necesaria

Para el desarrollo de la e-Ciencia, se requiere de una infraestructura básica que será común para todas las aplicaciones. Esta infraestructura corresponde a las tres capas horizontales mostradas en la figura 1 y que serán desarrolladas en los apartados siguientes.

2.1. Middleware

El *middleware* permite utilizar por parte de las aplicaciones, de forma conjunta o coordinada, recursos disponibles en diversos centros. La tecnología más utilizada en entornos científicos parte de Globus Toolkit, complementada con otros paquetes. Globus Toolkit es un *software* de código fuente abierto impulsado por la Globus Alliance² que permite la construcción de *grids*.

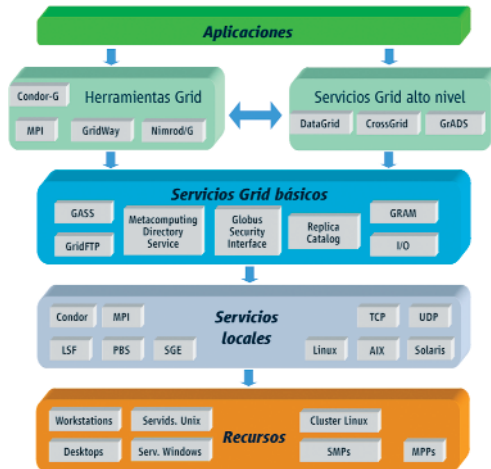


Figura 3. Tecnología Grid: Taxonomía de Globus

Globus se constituye de servicios locales para la explotación eficiente de recursos (gestores de colas *batch*, librerías de programación paralela, herramientas de depuración y monitorización...), servicios Grid básicos de autenticación y acceso a los datos, información y estructura de los recursos y ejecución de aplicaciones y seguimiento, servicios Grid de alto nivel, de acceso rápido y eficiente a los servicios básicos, y herramientas Grid para el desarrollo de aplicaciones (figura 3).

² <http://www.globus.org>

Servicios locales

Se incluye el desarrollo de herramientas para explotar los recursos en la Intranet del centro de investigación. En este campo existen en la actualidad muchos servicios disponibles para realizar un uso eficiente de los recursos locales, de los cuales podemos destacar las herramientas de monitorización de los diferentes recursos físicos del *cluster*, los gestores de colas *batch*, las librerías de programación paralela, y las herramientas de depuración y monitorización de aplicaciones.

Servicios Grid básicos

El Global Grid Forum es el organismo encargado de definir los estándares de servicios y protocolos necesarios para crear la infraestructura o tecnología Grid. Aunque existen otras tecnologías Grid como Legion³, Unicore⁴ o MOL⁵, la mayoría de los proyectos Grid actuales se están construyendo basados en los servicios y protocolos proporcionados por el Globus Toolkit. La tecnología Globus ha sido seleccionada como estándar de facto por las 12 compañías más importantes del sector de computación de altas prestaciones⁶ (Compaq, Cray, SGI, Sun Microsystems, Entropia, IBM, Microsoft, Platform Computing y Veridian en Estados Unidos; y Fujitsu, Hitachi y NEC en Japón). La versión actual de Globus basada en Open Grid Services Architecture (OGSA) muestra una clara convergencia hacia la tecnología de Web Services utilizada en el campo del e-Business. El Globus Toolkit (GT) apuesta por servicios Grid basados en Web Services. Esta evolución representa una gran oportunidad para lograr una amplia aceptación y difusión de la tecnología Grid, que puede extenderse, al igual que lo hizo el WWW, desde su ámbito original en el área de la computación científica, al de las aplicaciones comerciales.

Globus Toolkit es una colección de componentes *software open-source* y *open-architecture* diseñados para soportar el desarrollo de aplicaciones de alto rendimiento sobre entornos distribuidos tipo Grid. Realmente se trata de un conjunto de componentes autónomos que permiten al diseñador construir un Grid. Cada componente proporciona un servicio básico como autenticación, asignación de recursos, información, comunicación, detección de fallos y acceso remoto a datos. Los sistemas y aplicaciones Grid se pueden desarrollar empleando los siguientes servicios y protocolos como elemento básico:

- GT CORE: Implementa los estándares de OASIS WSRF (Web Services Resources Framework) y WSN (Web Service Notification)⁷.

³ <http://www.cs.virginia.edu/~legion>.

⁴ <http://www.unicore.org>.

⁵ <http://www.uni-paderborn.de/pc2/projects/mol>.

⁶ <http://www.globus.org/developer/news/20011112a.html>.

⁷ La versión anterior (GT3) se basaba en OGSi (Open Grid Service Infrastructure), que ha desaparecido en la versión actual.

- Seguridad, formada por dos servicios:
 - Grid Security Infrastructure (GSI), que permite la autenticación y comunicación segura sobre redes abiertas.
 - Community Authorization Service (CAS), que define a los proveedores de recursos políticas de control de acceso para comunidades.
- Gestión de datos (Data Management), incluye tres servicios:
 - GridFTP, protocolo para la transferencia segura y fiable optimizada y su API Global Access to Secondary Storage (GASS Transfer API).
 - Reliable File Transfer Service (RFT), un servicio OGSA que permite controlar y monitorizar transferencias que usen GridFTP.
 - Replica Location Service (RLS), que mantiene y provee de acceso a información de localización de datos.
- Gestión de recursos (Resource Management), formado por:
 - Grid Resource Allocation and Management (GRAM), que provee una interfaz para solicitar y usar recursos remotos para la ejecución de trabajos.
 - Servicios de información (Information Service).
- Monitoring and Discovery System (MDS3), que proporciona información sobre los recursos disponibles en el Grid y su estatus.
- XIO: Globus XIO es una utilidad de entrada/salida que proporciona una API para el acceso a sistemas de IO.

Servicios Grid de alto nivel

A pesar del tremendo esfuerzo realizado por la comunidad científica, la ejecución y la gestión de trabajos en un Grid resulta una tarea ardua y difícil, debido principalmente a la naturaleza dinámica y compleja que caracteriza los *grids*. Habitualmente el usuario ha de encargarse manualmente de todos los pasos involucrados en la ejecución de un trabajo, como descubrimiento y selección de recursos, inicialización, envío, monitorización, migración y finalización.

Para facilitar el uso del Grid, se están desarrollando servicios Grid de alto nivel y herramientas que realicen de forma automática y eficiente los pasos anteriores, además de

adaptar la ejecución de un trabajo a las condiciones dinámicas del Grid (disponibilidad, carga, coste, etc.) así como a las demandas dinámicas de la aplicación (tiempo máximo de ejecución, presupuesto, necesidad de *hardware/software* específico, etc.).

Herramientas Grid

En este nivel se incluyen herramientas de más alto nivel como librerías de programación, entornos especializados para la resolución de problemas y otras herramientas de ayuda al desarrollo de aplicaciones. Estas herramientas se basan en los componentes básicos y de alto nivel.

Aunque el Global Grid Forum⁸ está coordinando un gran esfuerzo colectivo de estandarización, el Grid representa todavía una tecnología en vías de consolidación y su desarrollo está en confluencia con el de los Web Services. Dentro del mundo industrial, también se está apoyando esta tecnología a través de la Enterprise Grid Alliance⁹.

Otras tecnologías

Aunque Globus es la tecnología dominante, existen proyectos como el Korea@home¹⁰ de computación distribuida, y aplicaciones como Nodzilla¹¹ para la compartición de archivos, que utilizan la tecnología P2P. Korea@home es un proyecto basado en la agregación de potencia de cálculo de PC's voluntarios aprovechando la infraestructura Internet de alta velocidad (conexiones Ethernet en domicilios particulares) existente en el país (figura 4). Nodzilla es una aplicación que permite crear un sistema distribuido y tolerante a fallos que, además de la compartición de archivos, ofrece servicios distribuidos como el *streaming* de vídeo y el almacenamiento seguro de ficheros. Esta aplicación incorpora también funcionalidades que otras aplicaciones P2P no ofrecen, como la "supervivencia" de los ficheros aunque el original desaparezca, la encriptación segura en la comunicación, etc.

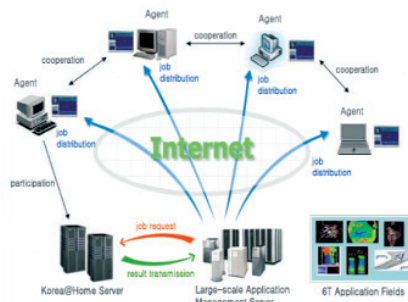


Figura 4. Computación distribuida con P2P: Korea@home

⁸ <http://www.ggf.org>.

⁹ <http://www.gridalliance.org>.

¹⁰ <http://www.koreaathome.org/eng>.

¹¹ <http://membres.lycos.fr/nodzilla>.

2.2. Redes de comunicaciones

La tecnología de red y el despliegue de infraestructuras, fundamentalmente fibras ópticas, por parte de operadores y otros agentes, están permitiendo una alta disponibilidad de medios de transmisión a larga distancia y accesos que posibilitan la utilización de altas velocidades, del orden de los gigabits por segundo (Gbps), y por tanto la puesta en práctica real de forma operativa de actividades de e-Ciencia.

La liberalización del mercado de las telecomunicaciones, y por tanto el fin de los monopolios y la instauración de la competencia, ha permitido disponer de velocidades de hasta 10 Gbps a precios asequibles. En los últimos 4 años los cambios fueron espectaculares, permitiendo que las redes académicas y de investigación construyeran redes como la red paneuropea Géant¹², con un núcleo de 10 Gbps ya desde finales de 2001. La Comisión Europea ha anunciado recientemente la renovación de esta red con una inversión de 93 M€ para cuatro años¹³. Esta red incorporará nuevos servicios de altas prestaciones como conectividad *end-to-end*, movilidad y acceso itinerante (*roaming*).

2.2.1. Red troncal nacional de RedIRIS

España comienza en 1988 su iniciativa de red académica y de investigación nacional, RedIRIS¹⁴ (figura 5). La red actual, después de una fase de topología en estrella, ha sido renovada en el 2003 de forma que se ha desplegado una nueva infraestructura de red mallada con un núcleo de 2,5 Gbps y al menos dos conexiones por nodo autonómico.

La red resultante dispone de 18 puntos de presencia, 10 enlaces de 2,5 Gbps, 13 de 622 Mbps y 6 de 155 Mbps. Su presupuesto en el 2003 ha sido de 16,8 M€ para operaciones corrientes y 2,1 M€ para inversión¹⁵.

Actualmente, RedIRIS conecta a más de 260 instituciones, principalmente universidades y centros de I+D, a los que también proporciona conectividad externa con otras redes de investigación y con la Internet global comercial.

¹² <http://www.dante.net/geant>. Aunque hay otras redes anteriores, como EARN o Hepnet, el primer antecedente de Géant de cierta entidad y con protocolos abiertos es IXI (abril de 1990-93), luego vienen otras redes intermedias como Europanet (1996-97), TEN-34 (1997-98) y TEN-155 (1998-2001).

¹³ <http://www.hoise.com/primeur/04/articles/weekly/AE-PR-10-04-15.html>.

¹⁴ <http://www.rediris.es>.

¹⁵ Para las distintas redes se indica el presupuesto de 2003 en millones de euros (M) a efectos orientativos, aunque pueda no ser el más representativo a largo plazo. Por ejemplo, para RedIRIS el presupuesto normalmente será algo menor ya que en el 2003 se produjo la renovación de su red.

Dentro del Estado, cuatro comunidades autónomas (Andalucía, Catalunya, Galicia y Madrid) disponen de redes propias de alta velocidad que enlazan con la red estatal. Éstas, sobre todo en las comunidades pluriprovinciales, facilitan la cohesión territorial y la conectividad con RedIRIS.

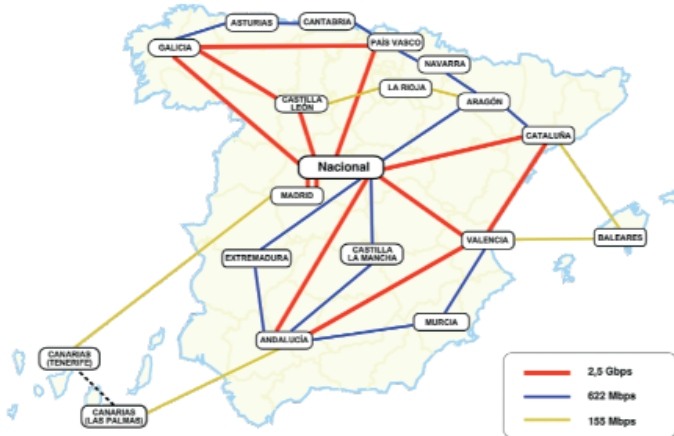


Figura 5. Topología de RedIRIS

2.2.2. Redes autonómicas

La Red Informática Científica de Andalucía (RICA)¹⁶, creada en 1991 y gestionada por el Centro Informático Científico de Andalucía (CICA), es de tecnología ATM y su tipología es de estrella (figura 6), donde el nodo central de Sevilla tiene conexiones con el resto de capitales andaluzas a 155 Mbps. Sin embargo, está previsto que finalice el próximo año su actualización a una red parcialmente mallada que mejorará su disponibilidad y fiabilidad. El presupuesto de RICA en el 2003 ha sido de 1,45 M€ (operaciones corrientes, sin personal) y la inversión, 114.000€.



Figura 6. La red en Andalucía

¹⁶ <http://www.cica.es/comu/rica.es.html>

En Catalunya, la Anella Científica¹⁷, creada en 1993 y gestionada por el Centre de Supercomputació de Catalunya (CESCA), utiliza tecnología Gigabit Ethernet desde su renovación en el 2003. La velocidad de su troncal es de 10 Gbps y las instituciones se pueden conectar hasta el gigabit, aunque la mayoría lo hacen a velocidades que van de los 10 Mbps a los 100 Mbps (figura 7). Su presupuesto en el 2003 ha sido de 1,20 M€ para operaciones corrientes y 0,15 M€ para inversión.

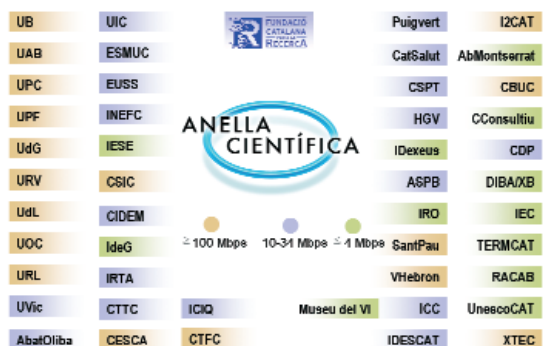


Figura 7. La red en Catalunya: la Anella Científica

En Galicia, el Centro de Supercomputación de Galicia (CESGA) se encarga de la gestión de la Red de Ciencia e Tecnoloxía de Galicia (Recetga), en funcionamiento desde 1993. Se trata de una red propia basada en tecnología Gigabit Ethernet y ATM con un ancho de banda en el troncal de hasta 1 Gbps (figura 8). La topología es mallada, con un nodo central situado en Santiago de Compostela. En el 2003, el presupuesto para operaciones corrientes ha sido de 0,43 M€ y el de inversión, 1,32 M€.



Figura 8. La red en Galicia: Recetga

¹⁷ <http://www.cesca.es/es/comunicacions/anella>

En la Comunidad de Madrid, REDIMadrid¹⁸, gestionada por la Fundación para el Conocimiento madri+d, tiene su centro de operación de red en el Centro Técnico de Informática del CSIC. Esta red utiliza tecnología Gigabit Ethernet y dispone de tres anillos (norte, sur y este) que conectan, en una primera fase, nueve instituciones (figura 9). Esta red, creada en el 2003, ha contado en el 2004 con un presupuesto de 1,46 M€.



Figura 9. La red en Madrid: REDIMadrid

En el País Vasco está en construcción la RED I2BASK, que estará previsiblemente en funcionamiento en enero de 2005. Esta red tendrá un troncal SDH a 2,5 Gbps que conectará Donostia, Arrasate, Vitoria, y la UPV en Lejona (figura 10). A estos puntos de acceso se conectarán diferentes parques tecnológicos. Esta red ha contado con un presupuesto para el 2004 de 1,10 M€ y de 1,45 M€ para el 2005.

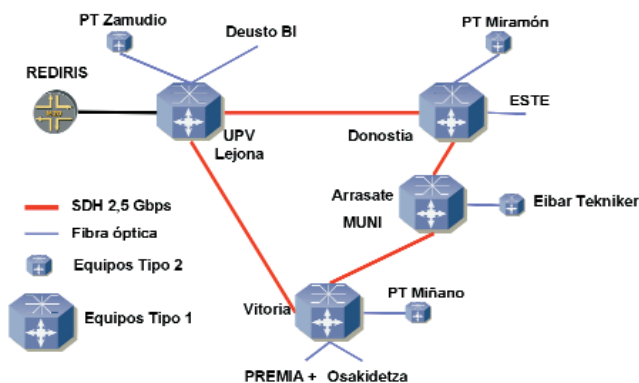


Figura 10. La red en el País Vasco: RED I2BASK

¹⁸ <http://www.madrimasd.org/RedTelematicaMadrid>.

2.2.3. Conexiones externas de RedIRIS

La conexión a Géant, que une RedIRIS con las redes de investigación europeas y con toda una red global de la investigación mundial, es un acceso a 10 Gbps que utiliza un enlace de 2,5 Gbps como *backup*. El nodo de Madrid conecta con el núcleo de Géant a través de dos enlaces de 10 Gbps hasta París y Milán.

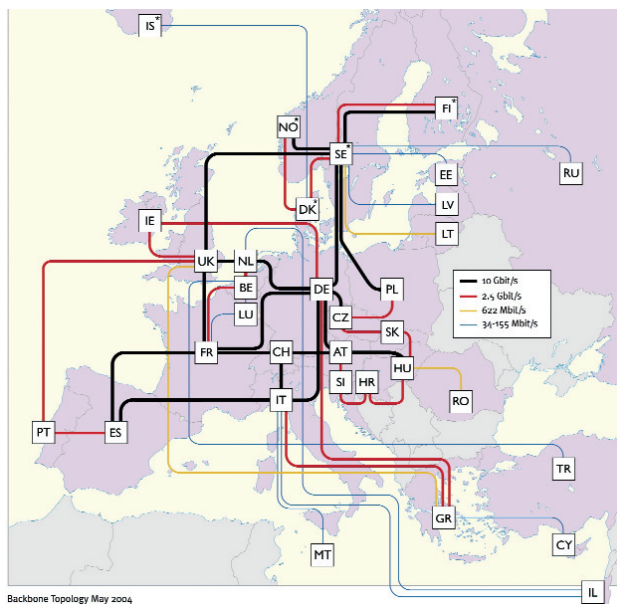


Figura 11. Topología de Géant

Además, el acceso de la red portuguesa al nodo de Madrid, que se realizaba a 622 Mbps, va a ser de 2,5 Gbps (aunque por el momento limitado a 1,2 Gbps), por lo que el nodo de Madrid tiene además un enlace de 2,5 Gbps hasta Lisboa y desde allí un enlace de 2,5 Gbps a Londres (figura 11). Así, dispone de una salida alternativa de alta velocidad desde la Península hacia el núcleo de Géant. Actualmente, el acceso de RedIRIS es de la máxima velocidad en este momento en Géant y el acceso de 2,5 Gbps adicional proporciona mayor flexibilidad al poderse usar para casos especiales donde haga falta una separación del tráfico que circula por el de 10 Gbps.

Es muy importante la actividad desarrollada en los últimos años, con una implicación especial de RedIRIS, para la creación de una red global de la investigación. Se participa en el proyecto Eumedconnect para el establecimiento de una red del Mediterráneo que conecte la ribera sur con Géant. El proyecto está entrando ya en fase operativa y se prolongará hasta junio de 2006. En el entorno latinoamericano, y después del estudio de viabilidad de la construcción de una red y su conexión a Europa, está en marcha el proyecto

ALICE, que ha colaborado en la creación de RedCLARA¹⁹ y su conexión a Géant, operativa desde septiembre de 2004, lo que repercutirá en la colaboración académica y científica entre Latinoamérica y Europa.

2.2.4. Conexiones de RedIRIS con la Internet global comercial

Las conexiones con redes fuera de la Intranet de la investigación en el mundo comercial usando la Internet global se realizan, por un lado, mediante puntos de intercambio de tráfico y, por el otro, a través de operadores internacionales. Para el intercambio de tráfico, RedIRIS dispone de una conexión en Espanix²⁰, el punto neutro español, de 2 Gbps. Además, RedIRIS también tiene una conexión en CAT-NIX²¹, el punto neutro de Intercambio de tráfico en Catalunya, albergado en las instalaciones del CESCA.

Para el tránsito, RedIRIS dispone de dos conexiones a 622 Mbps a través de los operadores internacionales de primer nivel, Telia y Global Crossing, con el que se intercambia tráfico IPv4 e IPv6.

2.3. Recursos a compartir

Las diferentes áreas de aplicación de la e-Ciencia necesitan acceder a diferentes tipos de recursos que mayoritariamente son distantes pero imprescindibles para el avance de la investigación. Algunos ejemplos de los recursos existentes en España, sin que sea una lista exhaustiva, se recogen a continuación.

2.3.1. Recursos de cálculo

Mientras que en los últimos años la presencia de centros nacionales de investigación o de servicios en el TOP500²² había ido reduciéndose, en la última edición de la lista España se ha situado como un punto de referencia internacional. Así, *MareNostrum*, del Barcelona Supercomputing Center²³, se ha colocado como el cuarto supercomputador más potente de todo el mundo y el primero entre los europeos. *MareNostrum* proporcionará una capacidad punta de cálculo de unos 40 Tflop/s cuando esté plenamente en funcionamiento a partir de 2005.

¹⁹ <http://www.redclara.net>.

²⁰ <http://www.espanix.net>.

²¹ <http://www.catnix.net>.

²² Hans Meuer, de la University of Mannheim, y Jack Dongarra, de la University of Tennessee, publican desde 1993 esta lista que recopila cada seis meses los 500 supercomputadores más potentes instalados. <http://www.top500.org>.

²³ <http://www.bsc.org.es>.

Hasta ahora, España contaba con diferentes centros de cálculo (autonómicos, universitarios...) que proporcionan servicios de computación de altas prestaciones a la comunidad investigadora. La lista TOP500 muestra la evolución de los recursos computacionales estatales desde 1993. Así, han estado presentes en esta lista los centros autonómicos CESGA, CESCOA y CICA, que hoy en día sólo proporciona servicios de comunicaciones. También han aparecido en la TOP500 los centros de las universidades Jaume I, de Valencia y Politécnica de Catalunya (CEPBA), y los centros especializados como el Instituto Nacional de Meteorología (INM), el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (Ciemat), y el Ente Público de Puertos del Estado (EPPE) (figura 12).

También proporcionan servicios de supercomputación otros centros universitarios como el Centro de Supercomputación Complutense de la Universidad Complutense de Madrid, el Centro de Proceso de Datos de la Universidad Politécnica de Valencia, el Centro Servicios de Informática y Redes de Comunicaciones de la Universidad de Granada, el Centro de e-Ciencia de Cantabria, del CSIC y la Universidad de Cantabria, el Instituto de Biocomputación y Física de Sistemas Complejos, de la Universidad de Zaragoza, el Instituto de Física Corpuscular, centro mixto del CSIC y la Universidad de Valencia, y el CEPBA-IBM Research Institute (CIRI), de IBM y la Universidad Politécnica de Catalunya.

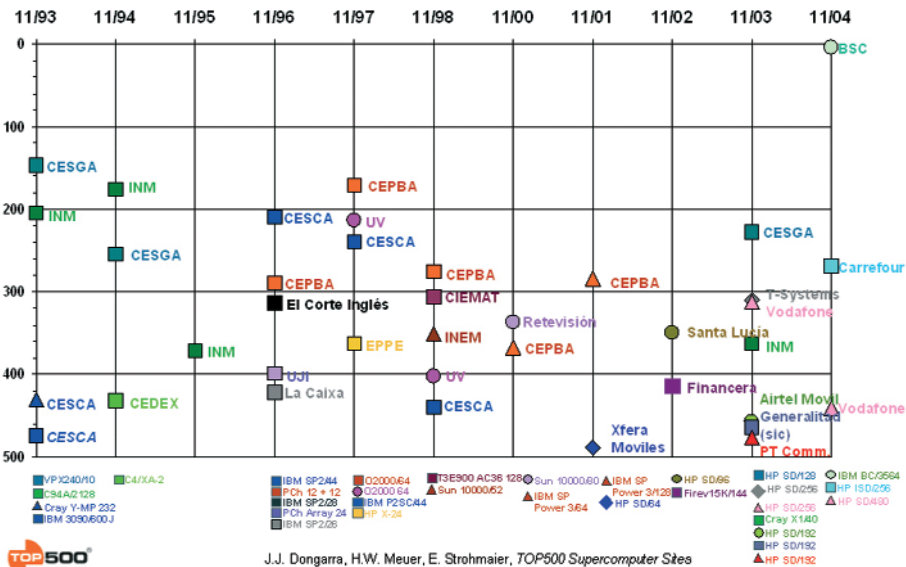


Figura 12. Evolución de los centros españoles en el TOP500

La mayoría de los centros mencionados que aún están en funcionamiento disponen actualmente de máquinas de procesador superescalar, de memoria compartida o de tipo *cluster*. Únicamente el INM utiliza la tecnología vectorial por el buen rendimiento que este procesador proporciona para los cálculos meteorológicos y relacionados con el clima. En total, estos centros disponen conjuntamente de un rendimiento punta de poco más de 4 Tflop/s²⁴.

La falta de inversión en estos centros, que ha limitado enormemente la capacidad de cálculo disponible para la comunidad investigadora, se verá compensada próximamente con la instalación del *MareNostrum*.

2.3.2. Recursos de almacenamiento

Uno de los retos del Grid es garantizar la preservación y la permanencia de los datos e información generados por la comunidad investigadora. Tradicionalmente, estos recursos de almacenamiento estaban instalados en centros de supercomputación, puesto que era donde se generaban grandes volúmenes de datos que era necesario preservar. Sin embargo, hoy en día otros centros especializados disponen también de este tipo de equipos para almacenar los datos generados en instalaciones experimentales, como el acelerador de partículas LHC (Large Hadron Collider) del CERN.

Algunos de los centros que actualmente disponen de equipos de almacenamiento masivo de datos para proyectos en Grid son el Puerto de Información Científica (PIC), el CESGA, el CESCA y algunos centros o departamentos universitarios como el de la Universidad de Cantabria.

2.3.3. Recursos de información

La información es un bien cada vez máspreciado y existe una clara tendencia a su consumo en formato electrónico. Así, bibliotecas digitales y repositorios de información están proliferando debido a las ventajas económicas y de accesibilidad, proporcionando almacenamiento de la información de forma electrónica.

Actualmente, los investigadores españoles publican unos 30.000 trabajos científicos anuales. Las universidades, suscritas a las publicaciones donde sus investigadores participan sin contraprestación económica, invierten 50 M€ en información y, en los últimos años, el aumento tanto de este gasto como del consumo de información ha sido constante (figura 13). Por ello, se está impulsando la creación de repositorios institucionales de publicaciones científicas de acceso gratuito, como ha sido el caso del Parlamento del Reino Unido²⁵.

²⁴ Un Teraflop (Tflop) son 1012 operaciones de coma flotante. El rendimiento se expresa por el número de estas operaciones que se pueden realizar por segundo (Tflop/s).

²⁵ Scientific Publications: Free for all?: Tenth Report of Session 2003-04/House of Commons Science and Technology Committee, vol. I: Report.

<http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm>



Figura 13. La información: elemento clave para el avance científico

Las bibliotecas digitales incorporan información electrónica contratada (publicaciones y revistas electrónicas), información electrónica en acceso abierto generada en España (por ejemplo, tesis doctorales²⁶), y bases de datos referenciales. La finalidad de estas bibliotecas es proporcionar un conjunto nuclear común de información electrónica para la totalidad de la comunidad de un territorio.

Existen también repositorios de interés común, como el directorio Astroweb²⁷, y sistemas de información geográfica, que incorporan georeferenciación de datos, cartografía actualizada y análisis de información, como el proyecto IDEC²⁸ del Institut Cartogràfic de Catalunya.

2.3.4. Otros recursos accesibles

Además de los recursos mencionados, el Grid permite la operación remota o teleoperación de instrumentos como telescopios, microscopios, capturadores de imagen o instalaciones de fusión nuclear como el TJ-II del Ciemat (figura 14). También proporciona accesibilidad a sensores de satélites, redes climáticas o de observación sísmica, hidráulica, etc.



Figura 14. Operación remota de instrumentación

²⁶ <http://www.tdr.cesca.es>.

²⁷ <http://www.vilspa.esa.es/pr/astroweb/astronomy.html>.

²⁸ <http://www.geoportat-idec.net>.

Otro tipo de recursos accesibles, cada vez más en auge, son aquellos que permiten compartición del conocimiento, formación y visualización, como serían salas de trabajo colaborativo (*Access Grid*), herramientas colaborativas avanzadas y centros de realidad virtual (figura 15).



Figura 15. Salas colaborativas (*Access Grid*)

3. Aplicaciones

El desarrollo de la e-Ciencia, en general, tendrá un impacto científico con la explotación eficiente de centros o recursos de excelencia y la existencia de nuevas formas de compartir conocimiento; un impacto tecnológico, ya que permitirá abrir nuevos mercados y nuevas formas de colaboración y desarrollo de proyectos; y un impacto social, pues proveerá de acceso para vencer la brecha tecnológica.

Estos retornos y beneficios serán compartidos por diferentes áreas de aplicación, algunas de las cuales detallamos a continuación, aunque, como se ha comentado anteriormente, no es una lista completa ni excluyente. Estas áreas de aplicación tienen necesidades diversas y grandes retos a afrontar que, de forma individual, apuntamos a continuación.

3.1. Astronomía y Espacio

Astrofísica teórica y cosmología, astronomía y astrofísica observacional y exploración espacial del sistema solar están comprendidas en Astronomía y Espacio, un área que cuenta con una comunidad investigadora formada por 500 científicos, distribuidos en unos 50 departamentos en más de 20 universidades. Existen seis grandes observatorios²⁹, siete centros de excelencia³⁰, dos instalaciones espaciales³¹ y varios centros industriales nacionales.

El uso de Internet en Astronomía y Espacio tiene una larga tradición. Como resultados relevantes específicos de este área cabe destacar el establecimiento de un estándar de intercambio de datos, FITS³², utilizado en todas las misiones y observatorios astronómicos. Se han implementado archivos distribuidos y se han creado servicios de catálogos astronómicos de referencia, que sirven de base para la interoperabilidad de archivos y sistemas³³. También se ha puesto en marcha la teleoperación de observatorios, una nueva forma de observación con telescopios terrestres que tiene su origen en los proyectos espaciales³⁴.

²⁹ Centro Astronómico Hispano Alemán (CAHA), Instituto de Radioastronomía Milimétrica (IRAM), Observatorio del Roque de los Muchachos del Instituto de Astrofísica de Canarias (ORM/IAC), Observatorio de Sierra Nevada del Instituto de Astrofísica de Andalucía (OSN/IAA), Observatorio del Teide del Instituto de Astrofísica de Canarias (OT/IAC) y Observatorio Astronómico Nacional (OAN).

³⁰ Centro de Astrobiología (CAB), Instituto de Astrofísica de Andalucía (IAA), Instituto de Astrofísica de Canarias (IAC), Institut d'Estudis Espacials de Catalunya (IEEC), Instituto de Física de Cantabria (IFCA), Instituto Nacional de Técnica Aeroespacial (INTA) y Laboratorio de Astrofísica Espacial y Física Fundamental (LAEFF).

³¹ INTA/NASA y ESAC/ESA.

³² Flexible Image Transport System (FITS). <http://www.cv.nrao.edu/fits>.

³³ El catálogo SIMBAD del Centro de datos estelares de Estrasburgo (CDS), motor principal para la interoperabilidad en Astronomía, contiene más de tres millones de objetos catalogados con más de ocho millones de identificadores de objetos. <http://cdsweb.u-strasbg.fr>.

³⁴ European Southern Observatory (ESO). <http://www.eso.org>.

Usos y necesidades

Los resultados indicados son algunos en los que Internet ha jugado un papel fundamental en el pasado. Mirando hacia el futuro, la introducción de la tecnología Grid supondría cambios tanto cuantitativos como cualitativos. Así, tendría aplicación en los siguientes dominios:

- La operación y el soporte de grandes recursos observacionales, incluyendo grandes telescopios robóticos, observación remota, observaciones coordinadas entre varios telescopios, así como nuevas formas de implementar misiones espaciales con operación distribuida en varios centros.
- El acceso a servicios de observatorios y archivos de datos como el Observatorio Virtual. Los archivos han jugado siempre un papel esencial en la Astronomía. En un principio se concibieron como repositorios aislados; el siguiente paso fue combinar varios archivos y catálogos para unificar los mecanismos de acceso y obtener la información independientemente de su localización física. El nuevo concepto se ha llamado Observatorio Virtual, en el que los datos son proporcionados al usuario en un hipercubo, cuyos componentes se van poblando con la información existente en diferentes archivos, de forma transparente, y del que se obtiene información realizando proyecciones o cortes.
- La elaboración de modelos y simulación. La astrofísica teórica requiere tratamientos numéricos elaborados que demandan grandes recursos de cálculo para la elaboración de modelos y simulación de escenarios, como la formación y evolución de galaxias y la evolución del universo a gran escala, y la magnetohidrodinámica.
- Desarrollo de sistemas específicos para análisis, visualización y calibración de datos astronómicos.

Retos a afrontar

La adopción del Grid supondrá la creación de nuevos estándares de interoperación e intercambio de datos, la consolidación del concepto de Observatorio Virtual, la integración de modelos astrofísicos y observatorios virtuales, la definición de “observación” como servicio de Internet, y la extensión del concepto Internet interplanetario.

Retornos y beneficios

En el área de Astronomía y Espacio, la e-Ciencia tendría un gran impacto en las formas de colaboración y de producción científica. Así, permitiría la explotación eficiente de los recursos observacionales existentes, la posibilidad de establecer centros de excelencia en servicios especializados que complementen las funciones y los servicios de los observatorios terrestres y espaciales existentes, coordinados con centros similares en otros países, la definición de líneas de investigación en la “Nueva astronomía” basada en el análisis masivo de datos e integración de modelos y datos, y el trabajo en redes de investigación con infraestructura común y nuevas formas de compartir conocimiento.

3.2. Biomedicina y Ciencias de la Salud

Este área comprende todas las disciplinas que estudian los procesos biológicos, la medicina clínica y la salud pública³⁵ (figura 16). Ésta persigue una realimentación continua entre el mundo de la clínica y el mundo de la experimentación en el laboratorio. La integra una comunidad formada por 69 redes temáticas de investigación cooperativa que agrupan 290 instituciones, 1.591 grupos y 11.331 investigadores, así como 4 redes temáticas interdisciplinares (CSIC) y 8 grupos del Instituto Nacional de Bioinformática, etc.

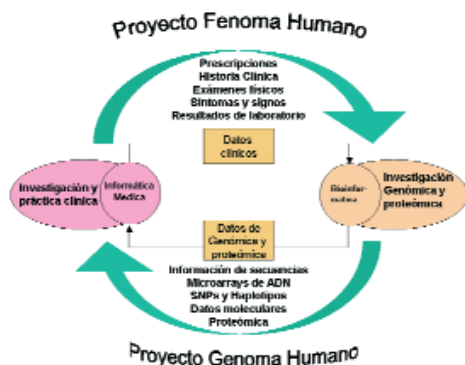


Figura 16. Realimentación en Biomedicina y Ciencias de la Salud

El objetivo principal del desarrollo de la e-Ciencia en Biomedicina y Ciencias de la Salud consiste en dotar a la comunidad científica de un entorno más productivo, dando respuesta a los principales desafíos en materia de procesamiento, almacenamiento y visualización de datos biomédicos de creciente complejidad y volumen³⁶.

³⁵ The Wellcome Trust. Mapping the Landscape: National biomedical research outputs 1988-95. <http://www.wellcome.ac.uk/en/old/AWTpubREPmtl.html>.

³⁶ Proceedings of UK e-Science All Hands Meeting 2003. 2-4th September, Nottingham, UK. <http://www.nesc.ac.uk/events/ahm2003/AHMCD>.

Usos y necesidades

El desarrollo de la e-Ciencia y la aplicación de tecnologías basadas en Grid presentan una serie de peculiaridades en Biomedicina, asociadas a las especiales características de este área³⁷. Algunos requisitos son la suficiente potencia de cálculo para analizar datos científicos; el acceso a fuentes dispersas de conocimiento; el uso remoto de tecnologías de alto rendimiento y recuperación de datos; la disminución en la latencia y realización en tiempo real de procesos de análisis de imágenes; la gestión de datos muy ricos semánticamente, que necesitan ser contextualizados y requieren el uso de metadatos; los derechos de acceso a datos sensibles; y la resolución máxima en ciertas imágenes para ser fiables como herramienta diagnóstica, lo que limita su compresión y la trazabilidad de las operaciones realizadas.

Algunas de las aplicaciones que se han comenzado ya a beneficiar de las ventajas de los planteamientos en e-Ciencia son genómica y proteómica computacional y estructuras y modelado molecular para el diseño de fármacos; el modelado y simulación de interacciones, procesos celulares (*e-cell*) y fisiopatológicos desde las perspectivas de la biología de sistemas (CELLWARE³⁸); análisis de imágenes anatómicas y diagnóstico funcional (NEUROGRID³⁹); planificación y soporte en tiempo real de procesos quirúrgicos (GEMSS⁴⁰, CrossGrid⁴¹); investigación multidimensional sobre modelos animales de enfermedades humanas (BIRN⁴²); e historia personal integrada de salud distribuida, agregación de datos para estudios poblacionales y epidemiológicos y portales e-Ciencia.

Retos a afrontar

Entre los retos científicos, se pretende promover el concepto y el uso de tecnología Grid en la comunidad científica biomédica; el acceso, integración y análisis de datos heterogéneos y distribuidos a distintos niveles de agregación; los estudios a gran escala de patologías complejas integrando datos genéticos, imágenes, historias clínicas y factores ambientales (CLEF⁴³, INFOGENMED⁴⁴); el desarrollo de algoritmos eficientes para el análisis de datos complejos; la gestión de metadatos y desarrollo de ontologías comunes y sistemas de representación semántica estandarizados, y la recuperación de información, minería de datos y textos en biomedicina.

³⁷ J. Montagnat, V. Breton, P. Kunszt. Biomedical applications for the Grid.
<http://www-unix.globus.org/ftppub/liming/biomed-apps.docs>.

³⁸ <http://www.bii.a-star.edu.sg/research/sbg/cellware/index.asp>.

³⁹ <http://www.gridbus.org/neurogrid>.

⁴⁰ <http://www.ccrl-nece.de/gemss>.

⁴¹ <http://www.eu-crossgrid.org>.

⁴² <http://www.nbirn.net>.

⁴³ <http://www.itri.brighton.ac.uk/projects/clef/index.html>.

⁴⁴ <http://www.infofenmed.net>

Los retos tecnológicos comprenden el archivo y curación de grandes volúmenes de datos; el acceso y procesamiento de datos remotos; la compartición, selección y agregación de recursos distribuidos geográficamente; el desarrollo de aplicaciones y gestión y planificación de recursos; la priorización de trabajos en situaciones de emergencia; los sistemas intuitivos de visualización de modelos y resultados de investigación, y aspectos relacionados con la seguridad como la autenticación de usuarios y la conformidad con la Ley Orgánica de Protección de Datos de Carácter Personal (integridad, protección, confidencialidad...).

Retornos y beneficios

El desarrollo de la e-Ciencia en Biomedicina y Ciencias de la Salud no sólo consiste en una infraestructura sobre la que se pueden ofrecer servicios avanzados, sino que promueve una cultura basada en la colaboración y fomenta el intercambio de conocimiento. De este modo, facilita el trabajo colaborativo, la explotación óptima de bases de datos, el avance de la educación, la escalabilidad y versatilidad en los sistemas, la estandarización e interoperabilidad de los datos, la realización de experimentos *in silico* mediante el modelado y la simulación, la interacción entre comunidades tradicionalmente aisladas entre sí y el abordaje de grandes desafíos, como el plegamiento de proteínas, los modelos de células y enfermedades y la evolución de epidemias.

3.3. Ciencia y Tecnología de Materiales

Comprende el estudio de materiales estructurales y funcionales, el crecimiento y síntesis de materiales, la caracterización de materiales y cristalografía, el estudio y diseño de nuevos materiales, los procesos industriales, y la nanociencia y la nanotecnología. Más de 500 investigadores trabajan en universidades y centros del CSIC, así como en el Centro Nacional de Aceleradores de Sevilla. También está en construcción el futuro Sincrotrón del Vallés.

Usos y necesidades

La simulación de materiales abraza simulaciones de propiedades y procesos en materiales, tanto a un nivel macroscópico (estructurales, térmicas...) como microscópico (desde el *ab-initio* hasta los métodos de mecánica estadística, Dinámica Molecular, Monte Carlo, etc.).

Estas simulaciones tienen una gran necesidad de potencia de cálculo, así como de almacenamiento de un enorme volumen de datos (bases de datos globales de simulación, resultados de simulaciones de Dinámica Molecular y accesibilidad de resultados para el resto de la comunidad).

Además, la caracterización experimental de materiales como la cristalografía, si bien tienen escasa necesidad de potencia de cálculo, generan un gran volumen de información.

Retos a afrontar

Entre los retos científicos se encuentra la resolución de problemas de alto interés industrial como son la reactividad química en superficies (catálisis), el diseño de materiales con propiedades predefinidas, el diseño de materiales para la degradación de residuos y productos tóxicos, para la nanoelectrónica y biomateriales, así como el almacenamiento de energía, por ejemplo a través de pilas de combustible (hidrógeno).

Los retos tecnológicos en Ciencia y Tecnología de Materiales comprenden la interoperabilidad de códigos de simulación, la compatibilidad de información entre diferentes códigos de simulación y la definición de estándares para el intercambio de datos (metadatos), la implementación de utilidades de visualización remota de datos de simulación, y la seguridad (autenticación, acceso a datos sensibles, etc.).

Retornos y beneficios

El acceso a una gran potencia de cálculo a través del uso en paralelo de recursos posibilitaría la realización de proyectos “Grand Challenge” a través de la ejecución de trabajos simultáneos en diferentes plataformas disponibles.

La disponibilidad de una base de datos global de resultados de simulación permitiría múltiples usos potenciales, como el análisis y comparación de resultados de distintas simulaciones, la extrapolación a nuevos materiales, y la elaboración de modelos ajustados a los resultados de la simulación. Además, también se tendría acceso a bases de datos experimentales (estructurales, espectroscópicas y de procesos).

3.4. Ciencias de la Tierra

Este área de aplicación comprende la meteorología, que incluye la predicción del tiempo a corto plazo y la predicción estacional (interacciones océano-atmósfera); el sistema Tierra, que trata simulaciones paleoclimáticas y del cambio climático antropogénico, oceanografía (simulación del océano superficial y profundo) y simulación de la biosfera; el modelado del sistema de aguas superficiales y subterráneas, y los estudios medioambientales, que incluyen la simulación de la difusión de contaminantes en atmósfera, mares y aguas dulces, el modelado de transporte de solutos y los modelos de reactividad con catalizadores (difusión de hidrocarburos en sólidos microporosos).

Esta comunidad está integrada por 300 investigadores distribuidos en más de 40 grupos establecidos en departamentos universitarios y en organismos de las administraciones públicas central y autonómica. Además, muchos organismos públicos de investigación, como el Centro de Estudios y Experimentación de Obras Públicas (CEDEX), el Consejo Superior de Investigaciones Científicas (CSIC), el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (Ciemat), el Instituto Español de Oceanografía (IEO), el Instituto Nacional de Meteorología (INM), y el Instituto Nacional de Técnica Aeroespacial (INTA), entre otros, son usuarios directos.

Usos y necesidades

Las principales necesidades de este área son la operación y el soporte de grandes recursos observacionales; el almacenamiento de resultados tanto de experimentos globales, regionales y locales, como de simulación de procesos de larga duración y alta resolución; y el desarrollo y la adaptación de modelos de predicción y de simulación de escenarios futuros, como son la predicción con muy alta resolución espacial del tiempo, la predicción de la difusión de contaminantes y compuestos tóxicos en atmósfera, océanos y aguas dulces, la simulación de escenarios de cambio climático futuro, la predicción operativa oceánica y la predicción operativa de vientos y otras variables para el aprovechamiento y explotación de energías renovables. Además, también se requiere la visualización de datos.

Retos a afrontar

Entre los retos a afrontar por las Ciencias de la Tierra se encuentran diferentes modelados: los de procesos físicos e interacciones dentro del sistema climático en escalas de alta resolución, hasta 1 km, para predicción meteorológica y medioambiental; el modelado hidrológico y oceanográfico, y el de corrientes marinas en zonas de interés nacional. Otro importante reto es el desarrollo y la consolidación de bases de datos virtuales distribuidas.

Retornos y beneficios

El desarrollo de la e-Ciencia permitiría la realización de experimentos *ensemble* para la predicción del tiempo y la simulación del clima, los resultados de los cuales se podrían incorporar a la investigación internacional en esta materia. Así, se potenciaría la participación española en proyectos internacionales.

También se posibilitaría la intercomparación de escenarios resultantes de conjuntos de experimentos de simulación, facilitando la comparación de resultados de distintas simulaciones, el análisis y la identificación de carencias, y la mejora de modelos para perfeccionar la validación. Además, se facilitaría la accesibilidad a bases de datos experimentales.

3.5. Física

La Física engloba la Física de Altas Energías, la Fusión Nuclear y otras áreas como los Sistemas Complejos y la Física Estadística, la Astrofísica, el Espacio, los Materiales, Física-Química, y Fluidos. Unos 800 investigadores y técnicos trabajan en España en Física de Altas Energías, Fusión Nuclear y Sistemas Complejos. Estos grupos de investigación están distribuidos en grupos en las universidades, el CSIC y en centros dedicados como el Ciemat y el IFAE. La colaboración con otros grupos de investigación y la necesidad de tener acceso a grandes instalaciones experimentales distantes son características de esta aplicación.

Usos y necesidades

El desarrollo de la e-Ciencia y la aplicación de tecnologías basadas en Grid permitirían la adquisición, almacenamiento y acceso a grandes volúmenes de datos de instalaciones experimentales, para su uso y reprocesado durante un largo periodo por colaboraciones internacionales de hasta 2.000 participantes.

La Física también tiene elevadas necesidades de simulación, y por tanto de supercomputación. Estas simulaciones comprenden desde colisiones de partículas en todos los canales físicos hasta de sistemas complejos (simulaciones numéricas de dinámica caótica en 3D). Además, para el análisis, calibración y visualización de datos es necesaria la realización de complejos algoritmos sobre terabytes de información (por ejemplo, en la búsqueda del bosón de Higgs es necesario seleccionar 100 colisiones de entre 1 billón).

Retos a afrontar

La e-Ciencia plantea en la Física diferentes retos. Por un lado, la gestión de los recursos distribuidos de una manera eficiente que satisfaga las necesidades de colaboración a través de la compartición de recursos de infraestructura, la interoperabilidad, la seguridad (autenticación, autorización, contabilidad), la implantación de un esquema organizativo (centros de procesamiento, almacenamiento, simulación) y las conexiones de red end-to-end entre centros. Por otro lado, se encuentra el acceso, la simulación y el procesamiento de un elevado volumen de datos, la aplicación de potentes algoritmos de análisis con respuesta rápida, y la consecución de un entorno de colaboración y participación remota (repositorios avanzados para desarrollo de *software* específico, discusión y publicación de resultados, movilidad, técnicas de videoconferencia y visualización avanzadas, y la operación remota de centros experimentales).

Retornos y beneficios

A través de la agregación de recursos, la escalabilidad, la disponibilidad y la redundancia, se lograría satisfacer las demandas de computación de grandes colaboraciones experimentales aprovechando en lo posible una infraestructura distribuida. También permitiría abordar retos computacionales planteados por el análisis físico (mejora de la simulación de procesos que requieren un elevado tiempo de cálculo, de las colas de discriminación en la búsqueda del bosón de Higgs, y del reajuste de los parámetros de control de un experimento de fusión, entre otros).

También se optimizaría la contribución de los participantes, mediante procedimientos comunes y coordinados en las tareas de instalación y operación de la infraestructura y la creación de un entorno de colaboración, movilidad y operación remota de recursos.

3.6. Ingeniería

Las ingenierías clásicas comprenden la investigación en procesos físicos de interés aplicado, la simulación de procesos industriales y el diseño precompetitivo. Esta comunidad está formada por investigadores distribuidos en nueve universidades, cinco centros de investigación públicos y ocho centros de desarrollo industrial.

Características específicas

Este área se caracteriza por la complejidad de los fenómenos, que no pueden simplificarse demasiado. Consisten en simulaciones grandes, de al menos 10-100 millones de puntos, fuertemente acopladas y que generan resultados individuales grandes (10 GB-Terabytes). La realización de estas simulaciones requiere de potentes superordenadores. En Ingeniería existe una fuerte integración internacional y se han establecido comunidades virtuales, no sólo entre la comunidad científica sino que también existe una importante cooperación con la industria, de la que procede la financiación para algunos proyectos. En este área, algunos de los datos generados requieren un tratamiento de confidencialidad, por lo que la seguridad es un aspecto crítico.

Retos y necesidades

En Ingeniería, actualmente, los grandes retos son retos mundiales. Mientras que la combustión y la turbulencia son procesos industriales básicos, esta última es también uno de los grandes retos históricos de la física. La mecánica de

fluidos atmosférica y oceánica es básica para las Ciencias de la Tierra, así como el cambio climático. En este sentido, la multidisciplinariedad y la cooperación son esenciales para afrontar estos grandes retos. Por tanto, es necesario facilitar la cooperación internacional entre universidades y centros así como entre investigadores e industria. También se debe potenciar y mejorar el acceso a grandes instalaciones, así como a bancos de datos existentes que actualmente se comparten por intercambio físico de discos. Para ello, se debe contar con redes rápidas hasta el usuario y de un amplio nivel de seguridad contra pérdida de datos.

La siguiente tabla muestra algunos ejemplos de grandes retos computacionales, el tamaño de la simulación en puntos de la malla, y la previsión del año en el que se podría abordar el trabajo.

Problema	Puntos malla	Año previsible
Turbulencia (problema físico)	10^{12}	2020
Turbulencia grandes escalas (problema industrial)	10^{13}	2025
Combustión (problema físico)	10^{15}	2030
Física atmosférica (e.g., nubes, tornados)	10^{18}	2050

3.7. Química

La Química comprende, por un lado, el estudio de la estructura y la reactividad química, con aspectos como la cinética y dinámica de reacciones, la química organometálica y la catálisis homogénea, y la química atmosférica, y, por otro lado, la investigación en superficies, sólidos y nanopartículas, que engloba interacciones magnéticas en complejos inorgánicos, magnetismo en superconductores de alta T_c (temperatura crítica), y cristales moleculares, entre otras investigaciones. Integran la comunidad de químicos alrededor de 300 investigadores distribuidos en unos 30 departamentos de 25 universidades e institutos del CSIC.

Usos y necesidades

Este área de aplicación se caracteriza por la necesidad de una gran capacidad de cálculo intensivo de altas prestaciones (memoria y disco) y de herramientas de visualización y análisis de resultados. Requiere tener acceso al *software* estándar de química computacional y disponer de un acceso rápido a bases de datos cristalográficos. Entre sus necesidades se encuentran también la paralelización masiva de los códigos más usuales, la elaboración de modelos de sistemas químicos complejos y la realización de simulaciones basadas en un gran número de datos.

Retos a afrontar

Algunos de los retos de la Química a afrontar en los próximos años son la reducción de los subproductos generados en los procesos químicos industriales, la mejora de la calidad del aire urbano, la creación de nuevos materiales de alto valor estratégico y tecnológico, la catálisis enzimática, los motores moleculares y el reconocimiento molecular.

3.8. Tecnologías para la Sociedad de la Información

Para las Tecnologías para la Sociedad de la Información (TSI), la e-Ciencia es más un fin que un medio, que intenta resolver problemas complejos con gran impacto económico y social. Así, Internet puede ser considerada una infraestructura computacional distribuida a escala mundial programable que, por tanto, puede abstraerse desde varios puntos de vista. Las TSI trabajan en jerarquías de esas abstracciones. Más de 15 grupos de investigación de las áreas de sistemas distribuidos, supercomputación y programación concurrente, entre otras, conforman este grupo que representa en torno al 9,5% del antiguo Programa TIC y alrededor del 21% del nuevo Programa Nacional en Tecnologías Informáticas (TIN).

Áreas y retos a afrontar

Los grandes retos en e-Ciencia para las TSI serán el diseño de arquitecturas Grid, el *middleware*, la gestión de información y datos, el desarrollo de *software* y la gestión del conocimiento. Por un lado, se necesita extender la arquitectura Open Grid Services Infrastructure (OGSA) y su implementación, así como implementar nuevos sistemas operativos Grid sobre la red o modificar los existentes, y monitorizar rendimiento, paralelización, gestión de recursos...

Es necesario que el *middleware* desarrollado en proyectos de investigación pase a ser un *software* de alta calidad. Además, se requiere un marco general para la implementación de una autoridad de registro y autenticación. Otro aspecto clave es la gestión de información y el manejo efectivo de grandes volúmenes de datos heterogéneos. También es necesario desarrollar *software* fiable y confiable (certificación), así como nuevos modelos, lenguajes y entornos de desarrollo para programar *grids* a todos los niveles de abstracción y conseguir interoperabilidad semántica. Se requiere paralelización automática y compiladores potentes e “inteligentes” capaces de tomar decisiones, como puede ser la ejecución remota o local de una tarea. En cuanto a la gestión del conocimiento, ésta comprende el *workflow* y la colaboración basados en el conocimiento, el soporte a las organizaciones virtuales, la tecnología de agentes para gestión de recursos, el manejo del descubrimiento del conocimiento, herramientas y servicios de *datamining*, modelado predictivo y descriptivo, y la web y el Grid semánticos.

Desafíos científico-tecnológicos

Los retos científico-tecnológicos consistirán en hacer el Grid más invisible para el usuario, pasar del Grid computacional al Grid del conocimiento, compatibilizar la privacidad y la transparencia, afrontar cuestiones técnicas como escalabilidad, modularidad, interoperabilidad y seguridad, entre otras, conseguir *grids* para sistemas móviles y empotrados, pasar del concepto de “plug & play” al de “connect & share”, conseguir una arquitectura sobre metasistemas operativos, e incrementar el ancho de banda y optimizar su gestión (redes y tecnologías ópticas, redes troncales de muy altas prestaciones, mayor movilidad y ubicuidad de las comunicaciones...).

Desafíos técnicos

Además, existen retos técnicos como la definición del rol del *software* libre y de fuente abierta, así como de los modelos adecuados de licencias de *software* para aportar un marco legal de explotación. También se debe conseguir un grado aceptable de calidad del *software* en todos los niveles, así como crear estándares abiertos de interoperabilidad para las infraestructuras y participación activa en la estandarización (W3C⁴⁵, OASIS⁴⁶...).

⁴⁵ En octubre de 1994, Tim Berners-Lee, inventor de la web, fundó el World Wide Web Consortium (W3C) en el Laboratory for Computer Science del Massachusetts Institute of Technology en colaboración con la European Organization for Nuclear Research (CERN), donde nació la web. Este consorcio contó con el soporte de la Defense Advanced Research Projects Agency (DARPA) norteamericana y de la Comisión Europea. El W3C fue creado con el fin de conducir el World Wide Web hasta su máximo potencial a través del desarrollo de protocolos comunes que promuevan su evolución y aseguren su interoperabilidad. <http://www.w3.org>.

⁴⁶ OASIS es un consorcio global no lucrativo que se encarga del desarrollo, la convergencia y la adopción de estándares e-business. <http://www.oasis-open.org>.

Retornos y beneficios

Los beneficios del Grid reportan sobre sus aplicaciones. Así, el desarrollo del Grid aporta a la competitividad industrial una nueva economía del conocimiento, modelos de negocio Grid, análisis de riesgos financieros, detección de fraude en sistemas de telecomunicación, aplicaciones de simulación (sismos, energía, diseño de aviones, implantes clínicos, etc.), *e-learning*, *eBusiness on-demand* y e-administración, entre otras. Hoy en día, empresas españolas como Telefónica I+D, Atos Origin y Gridsystems tienen proyectos Grid.

3.9. Algunos ejemplos clave

En áreas como Astronomía y Espacio, Biomedicina y Ciencias de la Salud y Física encontramos algunos ejemplos en los cuales se refleja la imposibilidad de abordarlos si no se dispone de la tecnología Grid.

Large Synoptic Survey Telescope (LSST)

El proyecto *Large Synoptic Survey Telescope (LSST)* consiste en proporcionar mapas de todo el cielo cada tres noches, con observaciones continuadas de 10 segundos cada 15 segundos. Estos 10 segundos serán suficientes para alcanzar una magnitud que permita detectar cualquier posible cambio en el cielo hasta niveles no igualados hasta hoy. Esto se va a conseguir con un telescopio, el LSST, que generará un caudal ingente de datos que habrá que procesar en tiempo real y poner a disposición de la comunidad internacional. Ésta, a su vez, debe responder con propuestas para seguir objetos o fenómenos, para alguno de los cuales el tiempo de respuesta es clave.

El flujo de datos de este telescopio supone 1.200 MB/segundo o 30 TB en una noche (10 horas). Datos que han de ser llevados en tiempo real a los centros de análisis para estudios diversos, como el examen de imágenes sucesivas para la búsqueda de fenómenos transitorios. Tanto el envío como el procesado de datos consumen recursos ingentes. Si el LSST se instalase en el Observatorio del Roque de los Muchachos, con un ancho de banda actual de 2 Mbps, se tardaría unas 6 horas en enviar una sola imagen consumiendo todos los recursos del ancho de banda (esto obviamente aislaría además al resto del observatorio) y haría imposible la finalidad científica del LSST.

¿Dónde entra la e-Ciencia? En la necesidad de poner en contacto las observaciones que se hacen en observatorios remotos, con los centros de procesado de datos (CPD), con los centros de archivo de datos, como el Observatorio Virtual, y con los usuarios finales. El caso concreto del Observatorio Virtual es paradigmático de

la e-Ciencia. Se trata de un proyecto coordinado a nivel mundial que está desarrollando las herramientas necesarias para que los astrónomos accedan a los datos astronómicos, estén éstos donde estén, o provengan del observatorio que provengan, incluyendo tanto datos de satélites como de telescopios en tierra. Para el usuario, el lugar de residencia de los datos es transparente. En el caso español, los observatorios de Canarias y de Villafranca serán esa fuente de datos que surtirá al Observatorio Virtual.

Canarias es uno de los lugares que se están considerando para la instalación del LSST. Un factor decisivo para que esto suceda son las comunicaciones entre el observatorio y la red europea de alta velocidad, Géant. Los observatorios, antes remotos, deben estar hoy día al alcance de la terminal de cualquier astrónomo.

Severe Acute Respiratory Syndrome (SARS)

En abril del 2003, la Organización Mundial de la Salud (OMS) anunciaba la colaboración sin precedentes que había permitido identificar en un tiempo récord un nuevo patógeno, un miembro de la familia de coronavirus, como el causante de la SARS. Según la OMS, la rapidez en su identificación había sido el resultado de la estrecha colaboración internacional de 13 laboratorios de 10 países.

Después de que la OMS lanzase una alerta mundial el 12 de marzo de 2003, laboratorios punteros de Canadá, Francia, Alemania, Hong Kong, Japón, Países Bajos, Singapur, Reino Unido y Estados Unidos, y posteriormente de China, se organizaron para crear una red de colaboración que permitió un mes más tarde identificar el coronavirus causante de esta grave enfermedad, con la colaboración de la OMS. El trabajo consistió en la compartición en tiempo real a través de una web segura de imágenes de microscopía electrónica, secuencias genéticas, aislados de virus y datos de muestras de pacientes, así como la realización de análisis en paralelo de muestras, hasta llegar a identificar el agente causante.

Como afirmaba el Dr. Klaus Stöhr, virólogo de la OMS y coordinador de esta red, “en este mundo globalizado, este tipo de colaboración es el único camino para abordar las enfermedades emergentes”.

Large Hadron Collider (LHC)

El más grande y potente acelerador de partículas, el LHC, está actualmente en construcción en el CERN (Conseil Européen pour la Recherche Nucléaire). La entrada en funcionamiento del LHC está prevista para el 2007 y será usado por más de 6.000 investigadores de universidades y laboratorios de todo el mundo. Los requisitos computacionales de los experimentos que

usarán el LHC son enormes: cada año se generarán entre 12 y 14 Petabytes de datos, equivalente a más de 20 millones de CDs, y el análisis de estos datos necesitará una potencia equivalente a más de 70.000 PCs de los más rápidos hoy en día.

El objetivo del proyecto LCG (LHC Computing Grid Project) es satisfacer estas necesidades computacionales sin precedentes a través de la implementación de un grid mundial que integrará en una organización virtual la capacidad de centros de todo el planeta. En España, participan las universidades Autónoma de Madrid, de Barcelona y de Santiago de Compostela, el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, el Institut de Física d'Altes Energies, el Instituto de Física de Cantabria, el Instituto de Física Corpuscular del CSIC y el Puerto de Información Científica.

4. Algunos proyectos en desarrollo

Como hemos comentado anteriormente, hace más de 30 años Leonard Kleinrock ya apuntaba que probablemente veríamos la propagación de los servicios de computación que, como la luz o el agua, llegarían hasta casas y oficinas. En 1998, Ian Foster y Carl Kesselmann publican “The Grid: Blueprint for a Future Computing Infrastructure”⁴⁷ sobre *grids* computacionales, a la par que se iniciaba la Globus Alliance con el fin de organizar la investigación y el desarrollo para crear tecnologías básicas para el Grid.

En poco más de cinco años, tanto en Estados Unidos como en Europa se han puesto en marcha diferentes proyectos Grid. Algunos de los proyectos europeos en desarrollo, de los cuales todos tienen participación española, son brevemente explicados a continuación. Una lista más detallada se puede encontrar en la publicación digital Primeur⁴⁸.

CrossGrid

Este proyecto pretende desarrollar aplicaciones interactivas en el entorno Grid, incluyendo desarrollo *middleware*, *testbed* distribuido y aplicaciones. Las herramientas utilizadas se basan en Globus y DataGrid. CrossGrid⁴⁹ está impulsado por el programa IST (2002-05) y está coordinado por Cyfronet⁵⁰. En cuanto a la participación española, el CSIC es el socio principal y coordinador del apartado de testbed, y también colaboran en el proyecto el CESGA, la Universidad de Santiago de Compostela (USC), la Universidad da Coruña (UDC) y la Universidad Autónoma de Barcelona (UAB). CrossGrid cuenta con una financiación de 6,70 M€ por parte de la UE.

European Data Grid

European Data Grid (EDG)⁵¹ tiene el objetivo de construir la próxima generación de infraestructura de computación, que permita cálculo intensivo y análisis de bases de datos compartidas a gran escala, desde cientos de terabytes a petabytes, entre comunidades científicas ampliamente distribuidas. La herramienta que utiliza es Globus. Este proyecto, impulsado por el programa IST en el 2001 y coordinado por el CERN⁵², cuenta con una financiación de 10 M€. Entre diversos socios europeos, participa en el proyecto el Institut de Física d’Altes Energies (IFAE)⁵³.

⁴⁷ Foster, I. and Kesselman, C. (eds.), *The Grid: blueprint for a future computing infrastructure*, Ed. Morgan Kaufmann Publishers, 1998.

⁴⁸ <http://www.hoise.com/primeur>.

⁴⁹ <http://www.eu-crossgrid.org>.

⁵⁰ www.cyfronet.krakow.pl.

⁵¹ <http://www.eu-datagrid.org>.

⁵² <http://www.cern.ch>.

⁵³ <http://www.ifae.es>.

Enabling Grids for E-science in Europe

Enabling Grids for E-science in Europe (EGEE)⁵⁴ es un proyecto impulsado por la Comisión Europea (2004-06) para desarrollar una infraestructura Grid en Europa. EGEE está coordinado por el CERN, y la participación española está integrada por PIC, CESGA, INTA, UPV, CSIC y RedIRIS. El objetivo de este proyecto es desarrollar una infraestructura Grid en Europa disponible 24 horas al día y cimentada en el desarrollo de una red Grid consistente, robusta y segura; la mejora y el mantenimiento del *software* para proporcionar un servicio estable a los usuarios; y la captación de nuevos usuarios provenientes de la industria y de la ciencia, además de asegurar que reciben soporte y formación de alta calidad. EGEE cuenta con una financiación por parte de la UE de 32 M€.

GAIAGrid

Impulsado por la Agencia Espacial Europea (ESA) en el 2003, GAIAGrid⁵⁵ tiene el objetivo de establecer un Grid computacional para facilitar el acceso a la base de datos de GAIA a los proveedores de algoritmos. En este proyecto, que cuenta con la participación de la UB, el CESCA y GMV, se utilizan las herramientas Globus y GridAssist.

IRISGrid

IRISGrid⁵⁶ es una red temática desarrollada por la comunidad investigadora española e impulsada por RedIRIS desde el 2002. El objeto de la red es coordinar y organizar a nivel académico y científico a los grupos de investigación interesados en esta tecnología, tanto en su desarrollo, implantación y aplicaciones, así como crear la infraestructura Grid nacional que permita su uso tanto a nivel de aplicabilidad en diferentes ámbitos como a nivel de desarrollo e innovación, y organizar acciones para la participación en proyectos del 6º Programa Marco. IRISGrid utiliza la tecnología de Globus Toolkit y cuenta con la participación de 40 grupos de investigación.

Portal de Ciencia Colaborativa

Es una iniciativa puesta en marcha con el objetivo de crear, desarrollar y aplicar un sistema en red para la comunidad investigadora española, que permita un acceso compartido a sus recursos. Ocho universidades (Politécnica de Valencia, de Alcalá, de Salamanca, Politécnica de Catalunya, de Sevilla, Autónoma de Barcelona, de Deusto y Oberta de Catalunya) y tres instituciones científicas (Centro Nacional de Investigaciones Cardiovasculares, Parque Científico de Madrid y Oficina de Cooperación Universitaria) participan en este Portal, que cuenta con un presupuesto de 6 M€.

⁵⁴ <http://public.eu-egee.org>.

⁵⁵ <http://astro.estec.esa.nl/gaia>.

⁵⁶ <http://irisgrid.rediris.es>.

5. Conclusiones

Así como hoy en día pocos cuestionarían los beneficios que Internet ha aportado en todos los ámbitos (ciencia, tecnología, administración, ciudadanía...), la tecnología Grid aportará un paso más en las prestaciones que la Internet proporciona hoy día.

Países como el Reino Unido, que cuenta con un programa de e-Ciencia desde el año 2000, y Alemania, que acaba de anunciar la puesta en marcha de su programa D-Grid, están trabajando en la puesta en marcha de programas e-Ciencia tanto para establecer proyectos en diversas áreas de aplicación como para desarrollar el *middleware* necesario en el país.

Las experiencias en e-Ciencia de otros países han de examinarse críticamente y han de servir de estímulo para desarrollar un programa de e-Ciencia en España, ya que demorar su puesta en marcha alejaría a España de los países que están liderando el desarrollo de la e-Ciencia. Es importante estar trabajando en proyectos Grid para poder participar en el proceso de estandarización y tomar parte en la uniformización de la tecnología.

La comunidad investigadora española afronta hoy en día grandes retos en diversas áreas de aplicación que requieren de un acceso a múltiples recursos, así como disponer de una infraestructura que facilite la compartición de conocimiento y la colaboración. Además de disponer de una gran capacidad de cálculo y de almacenamiento de datos, el uso de instrumental especializado, el acceso a recursos de simulación y visualización, las consultas a bases de datos y el acceso a aplicaciones colaborativas, entre otros recursos, son imprescindibles para abordar estos retos científicos.

Si bien hay áreas como la astronomía en las que se ha utilizado con éxito la teleoperación de instrumental, existen recursos que únicamente pueden ser utilizados de forma local y/o por la comunidad investigadora que forma parte de la institución a la que pertenece el recurso. El desarrollo de la e-Ciencia permitiría, por un lado, evitar desplazamientos y, por el otro, abrir ciertos recursos locales a todos los investigadores, independientemente de la comunidad autónoma o de la unidad de investigación a la que pertenezcan.

Por lo tanto, el desarrollo de la e-Ciencia es fundamental para determinar la capacidad científica y tecnológica en una economía globalizada. Este desarrollo para España se debe orientar en dos ejes: por un lado, acciones horizontales válidas para todas las aplicaciones, es decir, recursos, redes y *middleware* que deberían encauzarse a través de la acción estratégica del Plan Nacional, y por el otro, acciones verticales para cada área específica que dependerán del Programa Nacional correspondiente.

Además, también se debe asegurar la disponibilidad de recursos (computacionales, de almacenamiento, de información, etc.) para la e-Ciencia, fortalecer la conectividad hasta el usuario final con la expansión adecuada de las redes autonómicas y las de campus (o equivalentes), y estimular la participación de la industria española en el desarrollo de la e-Ciencia.

Se debe crear en RedIRIS una estructura organizativa para coordinar los recursos integrantes de la e-Ciencia e identificar y gestionar los elementos. Es necesario establecer una red nacional en la que participen centros de apoyo a la e-Ciencia con el fin de facilitar el acceso transparente a los recursos y la autenticación de los usuarios, garantizar el ancho de banda necesario, dar soporte a la implementación Grid para los grupos de I+D+I, apoyar el desarrollo del *middleware* necesario, y actuar como centros de información y difusión, con un énfasis especial orientado a la transferencia tecnológica.

Además, se debe impulsar la recopilación de repositorios de información electrónica para difundir la investigación universitaria y facilitar la colaboración entre grupos a través de reuniones virtuales, seminarios y conferencias distribuidas (en tiempo real o mediante acceso fuera de línea a través de una videoteca científica) y salas de trabajo colaborativas.

El desarrollo de un programa de e-Ciencia facilitaría la participación de la comunidad científica española en grandes proyectos internacionales de investigación, lo cual también permitiría su participación en el desarrollo de estándares internacionales.

Para finalizar, para lograr estos objetivos es conveniente crear un Consejo Asesor o Comisión de Seguimiento que tenga la misión de impulsar y supervisar el desarrollo de la e-Ciencia en España.

6. Acciones y recomendaciones

Se recomienda al Ministerio de Educación y Ciencia (MEC) que lance urgentemente la acción estratégica incluida en el PNID 2004-2007:

1. Nombrando un gestor de la acción con un perfil preferentemente interdisciplinar.
2. Para afrontar problemas ya identificados como interesantes.
3. Tomando en consideración los proyectos y actividades ya existentes y, por tanto, ya financiados.

Asimismo, se propone la creación de un Comité Asesor y/o de Gestión para impulsar y supervisar el desarrollo de la e-Ciencia en España.

Dada la relevancia internacional de este tema, también se debería establecer una representación para la coordinación en foros tales como el eIRG (*e-Infrastructures Reflection Group*⁵⁷) y el *Computing and Networking Working Group* del ESFRI (*European Strategy Forum for Research Infrastructures*⁵⁸). A nivel global, para un desarrollo equilibrado de la e-Ciencia, es importante potenciar la disponibilidad de recursos (comunicaciones, computación, almacenamiento, información, etc.).

Dada la urgencia del desarrollo de la e-Ciencia se considera de interés que la FECYT, de forma complementaria, también nombre un coordinador interno apoyado por un Grupo de Expertos y facilite los encuentros, tanto a nivel nacional como internacional, para realizar acciones concretas de estructuración de plataformas nacionales; de formación por áreas y difusión a través de talleres de conectividad itinerantes; y organice una jornada internacional sobre estas cuestiones.

RedIRIS podría actuar como coordinador para establecer una plataforma Grid nacional a corto plazo. También debería identificar y gestionar los elementos necesarios comunes: *Public Key Infrastructure* (PKI), *Authentication and Authorisation Infrastructure* (AAI), *end-to-end* (e2e), directorios, buscadores, etc. RedIRIS podría afianzar tecno-logías, inventariarlas y ofrecérselas a los grupos; proponer dentro de cada campo la creación de experiencias piloto que podrían generalizarse y extenderse, y debería fortalecer la conectividad hasta el usuario final con la expansión adecuada de la red nacional y sus conexiones internacionales, contando con la colaboración de las redes autonómicas y las de campus (o redes equivalentes).

⁵⁷ <http://www.e-irg.org>. El principal objetivo de la iniciativa e-Infrastructure es dar soporte a la creación de un marco (político, tecnológico y administrativo) para un uso compartido de los recursos electrónicos distribuidos en Europa de una forma fácil y rentable. El e-IRG fue fundado para definir y recomendar buenas prácticas para los proyectos Grid paneuropeos. Está integrado por delegados de los gobiernos de todos los países europeos.

⁵⁸ http://www.cordis.lu/era/esfri_home.htm.



MINISTERIO
DE EDUCACIÓN
Y CIENCIA



www.fecyt.es