

e-Infraestructuras en Europa

Víctor Castelo Gutiérrez
Secretaría General Adjunta de Informática
Consejo Superior de Investigaciones Científicas (CSIC)
Ministerio de Ciencia e Innovación (MICINN)
Calle del Pinar, 19
28006 Madrid, España
victor.castelo@csic.es

Resumen. Las actuales redes de investigación han propiciado la aparición de la e-Ciencia y de una nueva forma de investigar. Las e-Infraestructuras permiten nuevos retos con nuevas aplicaciones y todo un nuevo abanico de posibilidades. Se presentan las e-Infraestructuras europeas, su conexión global y en particular con Latinoamérica, su coordinación y la evolución que se está produciendo en este entorno.

Palabras Clave: redes, investigación, e-Ciencia, e-Infraestructuras, coordinación, Europa, Latinoamérica, RedCLARA, e-IRG, e-InfraNet, e-IPF, Geant, IBERCIVIS, IBERGRID, PRACE, Grid, HPC, Supercomputación, Computación ciudadana, Gisela, EVALSO, eVLBI, aplicaciones, EGI, ES-NGI, NREN.

1. Las redes cambiando el modelo

Las redes de investigación han dado pasos muy importantes en los últimos años, las redes de fibra óptica permiten disponer de un ancho de banda casi ilimitado, lo que hecho que entremos en una nueva fase en la que se cambia la forma de hacer investigación.

Hay una nueva concepción del mundo de la investigación en la que a modo de ejemplo, podemos ver la trayectoria en barco realizado por la expedición de la Comisión Científica del Pacífico [1], en la que en 1862 una serie de investigadores españoles realizan un periplo por gran parte de Latinoamérica.

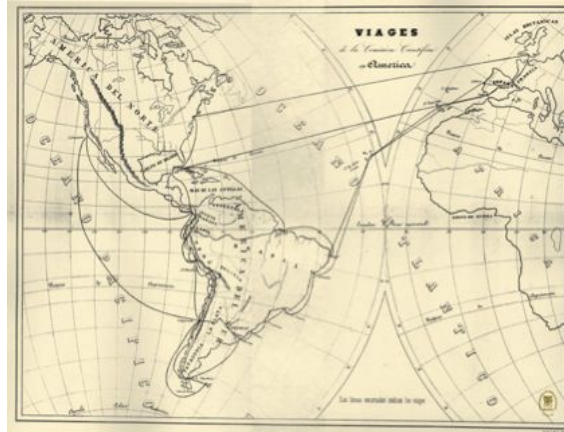


Fig. 1. Trayectoria de la expedición de la Comisión Científica del Pacífico

Toman muestras y hacen fotografías, hacen una captura de datos muy importante, pero lo curioso es que podemos ver cómo el viaje tiene cierta similitud con las formas de los actuales cables submarinos y cómo son utilizados por la Red de Investigación de Latinoamérica, RedCLARA, y su conexión con Europa y otras regiones, en una de sus primeras representaciones.



Fig. 2. RedCLARA y sus conexiones externas

Es una comparación simplista, pero que nos da una idea de cómo se acercan las comunidades científicas y cómo se pueden establecer ahora sistemas que pueden ser compartidos desde ambos lados del Atlántico.

Las redes de investigación actuales disponen de troncales de múltiples de 10 Gbps y están interconectadas entre sí de forma que se genera una especie de Intranet de la Investigación, en la que en el esquema de la figura 3 se observa cómo la interconexión entre recursos o investigadores de una universidad española con otros en RedCLARA o de Internet2 pasa por una serie de enlaces de color rojo controlados, conocidos, y con posibilidades de establecer servicios especiales si son necesarios.

Algo muy distinto a cuando estamos en un ISP comercial que se interconecta con otros a nivel nacional por un punto neutro, como Espanix en el caso español, y por la Internet Global con el resto del mundo, pero usando dominios desconocidos y sobre todo incontrolados por la Comunidad Científica.

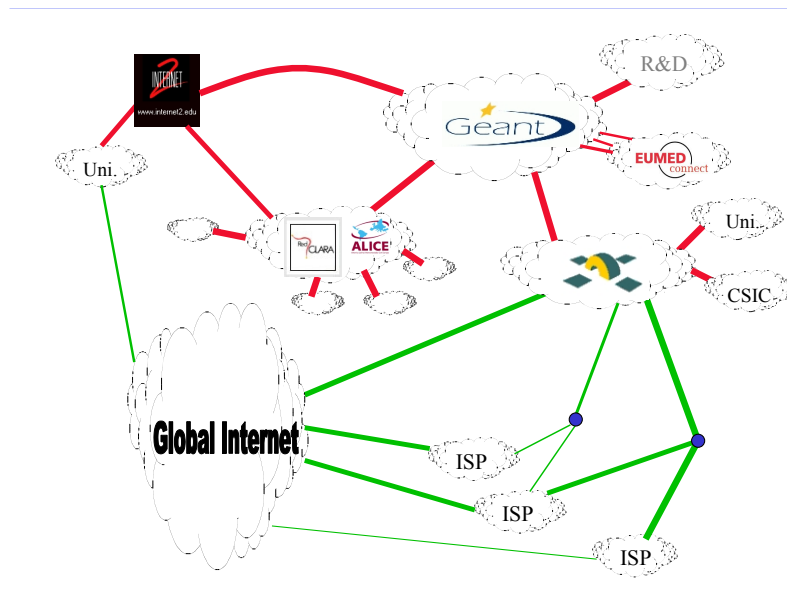


Fig. 3. Esquema de conexiones en la Intranet Global de la Investigación

2. e-Ciencia, e-Infraestructuras

Así pues, en este estado de desarrollo de las interconexiones de nivel de redes de investigación se produce una “potenciación” de la Ciencia, lo que se conoce como e-Ciencia (la “e” de enhanced) en la que los recursos de todo tipo, desde sistemas de cálculo a los propios investigadores se conectan usando esas redes y elementos de software (middleware) que permite conectar las diferentes necesidades de todas la áreas de la ciencia con los recursos.

En Europa la red básica de investigación es Geant [2].

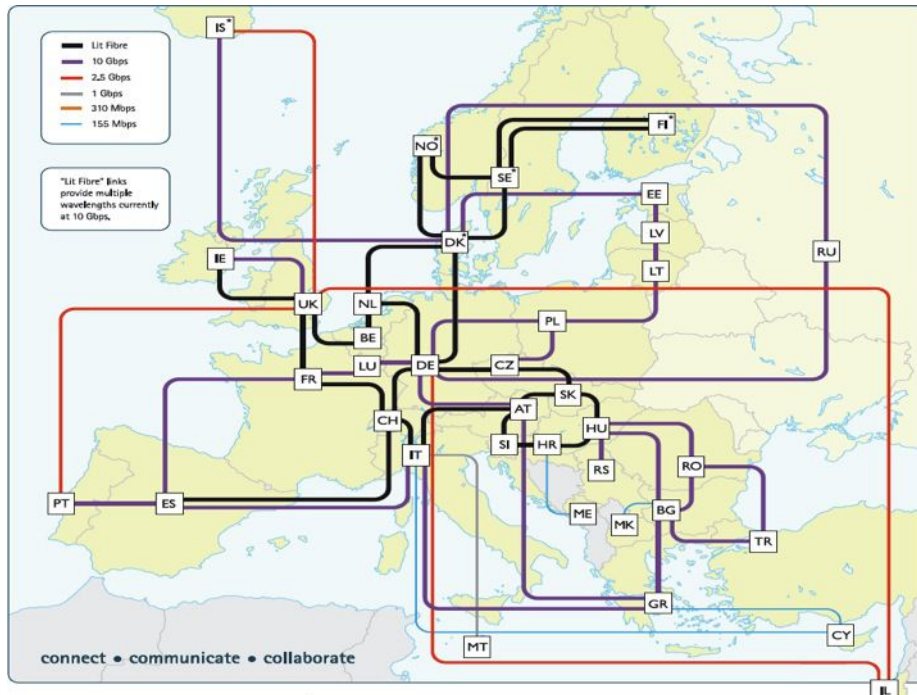


Fig. 4. Red europea de investigación Geant

Geant dispone de enlaces de hasta $N \times 10$ Gbps, y parte de su huella utiliza fibra oscura controlada por DANTE, que es el operador de toda la infraestructura. Siendo DANTE un entidad propiedad de las redes de investigación europeas.

Pero Geant no está aislada y se interconecta en esa Intranet de la Investigación con una serie de enlaces intercontinentales, entre otros su conexión con la RedCLARA de Latinoamérica.

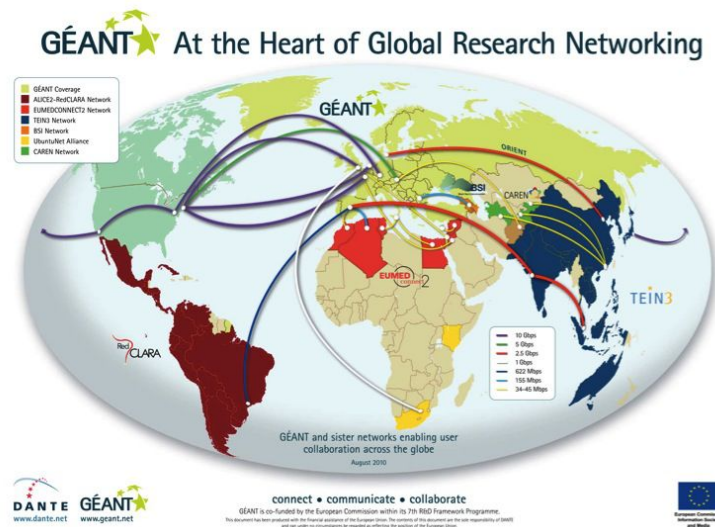


Fig. 5. Conexiones externas de la Red europea de investigación Geant

La red es básica para la conexión de otras e-Infraestructuras: elementos de computación, repositorios de datos, sistemas de control remoto o visualización, etc.

También hay elementos que usan las redes de investigación y utilizan las e-Infraestructuras, se trata de las Infraestructuras científicas (RI), algunas de ellas nacionales y en otros casos, dada su importante dimensión, se ha llegado a una colaboración europea o internacional que permita su construcción. Esto ha llevado a que los países diseñen un roadmap de sus infraestructuras científicas nacionales y se marquen una serie de prioridades en las RI europeas que se han establecido en el roadmap de ESFRI (European Strategy Forum on Research Infrastructures) [3]. El caso español se señala en las referencias [4].

3. Foros e iniciativas europeas de coordinación de e-Infraestructuras

La coordinación de todas la infraestructuras es algo básico, la redes de investigación se coordinan desde DANTE, que ya se ha citado anteriormente, y el Policy Committee de Geant, aparte de otros foros más de investigación y del desarrollo de Grupos de Trabajo y proyectos en los que también colabora la asociación Terena.

Para la coordinación de infraestructuras de investigación se emplea el foro anteriormente mencionado de ESFRI, que genera un roadmap que es revisado anualmente. Desde el punto de vista de las e-Infraestructuras existe el e-IRG (e-Infrastructures Reflection Group) [5] que genera su roadmap, con un visión de futuro, y white papers, como algo más coyuntural. Ambos son revisados cada 6 meses y además se producen informes dedicados o se propicia la creación de task forces para temas concretos. El e-IRG identifica todos los elementos de la e-Ciencia, en todos sus aspectos, desde las infraestructuras a la comunicación con los usuarios, y realiza recomendaciones a todos los niveles de cara a crear un marco europeo de recursos electrónicos distribuido, eficiente y sostenible.

Pero además del e-IRG, con una componente evidentemente técnica, hay otros dos foros, e-InfraNet [6] y el e-IPF. e-InfraNet es un formato de proyecto financiado por la Comisión Europea (ERA-NET), en el que participan organismos financiadores y su objetivo es la coordinación de políticas y programas a alto nivel (con ministerios implicados). Lo fundamental son las convocatorias conjuntas, en una serie de áreas focales: Cloud, Green, y Openess, y la realización de Workshops dedicados. Por último el e-IPF (e-Infrastructures Policy Forum) es el foro de nivel más político, está promovido por la Comisión Europea (muy en relación con e-InfraNet) y tiene como objetivos compartir las visiones y políticas sobre e-Infraestructuras mediante representantes de alto nivel, pero con la idea de no duplicar trabajos de otros foros y con la posibilidad crear de grupos de trabajo para puntos concretos.

Luego existen otros foros de coordinación concretos de algunas e-Infraestructuras y elementos de diseminación, como e-Science Talk [7], proyecto financiado por la Comisión Europea, que entre otras herramientas dispone de un interesante boletín semanal, el iSGTW.

En España existe una Red de e-Ciencia [8] entre cuyos objetivos se encuentran la coordinación de todo tipo de e-Infraestructuras y su relación con los usuarios finales de cara a la mejora en el uso de las e-Infraestructuras. La coordinación de la e-Ciencia entre España y Portugal, y con especial atención a su relación con Latinoamérica, se realiza dentro de Ibergrid [9], que organiza un workshop anual y coordina todo tipo de actividades entre los dos países en temas como redes, Grid, Supercomputación, o repositorios de datos.

4. Aplicaciones de la e-Ciencia

Son muchos los ejemplos que podemos poner de aplicación de la e-Ciencia y de e-Infraestructuras, y muchos de ellos en coordinación con Latinoamérica. El proyecto AugerAccess [10] facilita la conexión del Observatorio Pierre Auger, que consiste en un conjunto de detectores de diversos tipos, para analizar los rayos cósmicos de ultra-alta energía. Los detectores están distribuidos en una superficie de unos 300 km cuadrados en la Pampa Amarilla en Argentina. La conexión del Observatorio a alta velocidad con InnoRed, la Red Argentina de Investigación, y de allí con RedCLARA, permite que investigadores de Europa y de cualquier parte del mundo puedan realizar observaciones, calibraciones y controles en tiempo real de forma remota.

EVALSO [11] es otro proyecto que realiza la conexión de los observatorio de ESO Paranal y Cerro Armazones a Europa mediante las conexiones por Reuna, Red de Investigación chilena, y RedCLARA.

El uso y control remoto de sistemas de fusión nuclear, microscopios electrónicos o sensores submarinos pueden ser otros claros ejemplos de uso de la redes y del middleware adecuado para hacer e-Ciencia. Los mismos investigadores interactúan entre sí con dispositivos que van más allá de la videoconferencia, cada vez con sistemas de más alta resolución, tele-presencia, tele-inmersión o realidad virtual compartida.

Otro claro ejemplo de e-Ciencia es el eVLBI (Very Long Baseline Interferometry) [12], que mediante la conexión en tiempo real de varios radiotelescopios en diferentes partes del mundo permite analizar fuentes de emisión remotas en el espacio lejano incrementando la relación señal/ruido, por interferometría, y obtener muy importantes resultados en muy poco tiempo. Se trata de algo que se hacía antes enviando cintas grabadas localmente para procesarlas posteriormente. En muchos casos se podía producir algún error en la toma de datos o en la calibración que era detectado meses más tarde y que hacía perder un tiempo valioso, algo que en la actualidad es detectado de forma inmediata. Proyectos como EXPRES y NEXPRES han permitido construir la infraestructura de conexiones, del orden de 1 Gbps, además del hardware y software capaz de realizar el tratamiento de la información que se recolecta en JIVE, Holanda.

Se podrían contar más ejemplos desde la modelización del cerebro, el proyecto Blue Brain o el Observatorio Virtual Astrofísico, hasta la transmisión en el momento de su emisión de óperas desde el Liceo de Barcelona, con alta calidad de imagen y sonido y que son presenciadas como una enseñanza reglada en universidades de España y de Latinoamérica.

5. Infraestructura Grid y su evolución

Todo el despliegue de detectores y la toma de datos del LHC no habría sido posible sin el desarrollo del Grid. Mucho antes de que comenzase la explotación del LHC se desarrolla toda la infraestructura capaz de poder almacenar y tratar los 15 Petabytes/año producidos (15 millones de Gigabytes). Se crea una estructura de niveles, Tier0, el propio CERN, Tier1 y Tier2, conectados entre ellos, y que permiten el almacenamiento, su respaldo y el acceso de los investigadores a esta cantidad ingente de información. La estructura y el software se fue desarrollando gracias a proyectos, como EGEE, pero en los últimos años se vio la necesidad de evolucionar a un modelo del tipo del existente en las NRENs: una estructura nacional, NGI (National Grid Initiative), y luego una coordinación a nivel europeo, EGI (European Grid Initiative) [13].

La nueva estructura está en sus primeras fases operativas y se sufraga, aparte de las cuotas de cada uno de los países, con proyectos, como EGI-InSpire, de 4 años y con una contribución de la Comisión Europea de 25 M€. Algo parecido a EGI se está implementado en Latinoamérica mediante el proyecto GISELA, como evolución de EELA y EELA2.

En el caso español la iniciativa de Grid nacional es la ES-NGI [14], promovida por el Ministerio de Ciencia e Innovación, y coordinada por el Instituto de Física de Cantabria, centro mixto entre el CSIC y la Universidad de Cantabria. Participan en la red numerosas instituciones, con más de 5.000 cores y 500 TB de disco, además de 3.000 cores y 3.000 TB dedicados a aplicaciones del LHC. Todo ello coordinado con la NGI portuguesa, Ingrid, en el marco de Ibergrid.

6. Supercomputación

La computación distribuida tipo Grid es utilizable en muchas aplicaciones, pero cuando hay que pasar información entre un equipo y otro para colaborar en el cálculo, incluso a la velocidad de la luz (300.000 Km/s), para 1.000 km tenemos un retardo de 0,0033 segundos, algo muy elevado si lo comparamos con los tiempos de ciclo de una CPU en un equipo actual. Por tanto las aplicaciones que necesitan pasar de forma frecuente información entre equipos requieren infraestructuras de computación de muy alta velocidad, fuertemente interconectadas de manera local, en una sala informática, son los supercomputadores. El estado actual de Supercomputación (High Performance Computing: HPC) hace que la simulación sea una técnica aplicable a casi todas las ramas de la ciencia permitiendo obtener resultados mediante computación en un tiempo record y sin necesidad de realizar experimentos “in vitro”.

La Supercomputación en la práctica dispone de niveles de una forma piramidal en cuanto a la capacidad de los recursos, de forma que en la base de la pirámide pueden estar equipos departamentales o de una universidad, subir hasta equipos regionales y llegar hasta un nivel máximo de un país (Tier1). En la actualidad en Europa, para conseguir un nivel estratégico en el contexto mundial se ha realizado una actuación conjunta, PRACE (PaRtnership for Advanced Computing in Europe) [15] que culmina la pirámide Europea y que en realidad no es un único equipo, sino un conjunto de ellos (Tier0) que de forma continua se van sucediendo

para que se pueda mantener superordenadores europeos competitivos. España es uno de los países que se ha comprometido a financiar y albergar uno de los nodos Tier0. La participación española en PRACE se realiza mediante la encomienda del MICINN al Barcelona Supercomputing Center, Centro Nacional de Supercomputación (BSC-CNS). También el BSC-CNS coordina una serie de centros Supercomputación en España que ofrecen parte de su tiempo de cálculo, en forma de una ventanilla única, a todos los investigadores españoles, se trata de la Red Española de Supercomputación (RES) [16]. Los accesos a PRACE y la RES se realizan mediante los comités de acceso correspondientes.



Fig. 6. Pirámide de recursos de Supercomputación

7. Desktop Computing

Existen otras posibilidades de computación, a otro nivel, pero que en algunos casos pueden resultar interesantes, una de ellas es la Computación Voluntaria o Desktop Computing, en la que ordenadores cedidos por ciudadanos o dispersos por una organización son utilizados para calcular en sus tiempos muertos, cuando están siendo infrautilizados o corren el salvapantallas en momentos de descanso. Se trata de aplicaciones con paralelismo nulo, que pueden segmentarse en tiempos no muy elevados, con pocas necesidades de memoria, de entrada/salida, y sin problemas con licencias de software. El proyecto español Ibercivis [17], que ya está siendo utilizado y coordinado con Portugal y en fase de colaboración con países de Latinoamérica (Brasil, Cuba, Argentina), utiliza tecnología BOINC y dispone de más 25.000 usuarios registrados y unos 8.000 cores diarios.

8. Datos Científicos

Estamos en una época de la Ciencia en la que la cantidad de datos se ha incrementado de forma exponencial. Se está en una fase en la que se está pasando de un modelo en el que cada experimento o área disponía de un sistema aislado en forma de silos, a otro en el que es importante la compartición de datos y de recursos de almacenamiento, y la interoperabilidad.

Los repositorios de datos que se construyen deben cumplir una serie de requisitos como la accesibilidad, inclusión de metadatos, calidad de los mismos, etc. en un ecosistema de repositorios. Diversos foros y proyectos coordinan este tipo de actividad a nivel europeo y a nivel global. La tendencia al Open Access es imparable, promoviendo el acceso libre y sin restricciones a la publicación científica, incrementando el impacto de los investigadores e instituciones.

9. Conclusiones

El gran despliegue de redes de investigación y sus capacidades actuales está propiciando la e-Ciencia. Se ha hecho un gran despliegue pero todavía hay mucho que hacer para llegar a disponer de redes de gran capacidad en todos los países y aumentar la capilaridad en las redes nacionales. Es necesario que se establezcan sistemas de coordinación de la e-Ciencia y las e-Infraestructuras a nivel nacional y regional, procurando involucrar a las NRENs, los entes financiadores y a los usuarios. Todos ello en un entorno que tenga en cuenta la gobernanza y la sostenibilidad. Factores adicionales a tener muy en cuenta son los repositorios de datos, el Green IT y la evolución muchos casos a sistemas distribuidos empleando Cloud, tanto en soluciones privadas como comerciales o híbridas.

Referencias

1. Comisión Científica del Pacífico <http://www.pacifico.csic.es/>
2. GEANT <http://www.geant.net/>
3. ESFRI http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
4. Construyendo la Ciencia del Siglo XXI priorización ESFRI española
5. <http://www.micinn.es/stfls/MICINN>
6. e-IRG <http://www.e-irg.eu/> <http://www.e-irg.eu/publications.html>
7. <http://www.e-irg.eu/e-irg-workshop-madrid-17-june-2010.html>
8. eInfraNet <http://www.e-infranet.eu/>
9. e-Science Talk <http://www.gridtalk.org/e-sciencetalk/>
10. Red española de e-Ciencia Red Española e-Ciencia <http://www.e-ciencia.es/>
11. Ibergrid <http://www.ibergrid.eu/>
12. AugerAccess proyecto <http://www.augeraccess.net/>
13. EVALSO <http://www.evalso.eu/evalso/>
14. e-VLBI <http://www.evlbi.org/evlbi/evlbi.html>
15. EGI <http://www.egi.eu/>
16. ES-NGI <http://www.es-ngi.es>
17. PRACE <http://www.prace-project.eu/>
18. RES http://www.bsc.es/plantillaC.php?cat_id=451
19. Ibercivis <http://www.ibercivis.es>