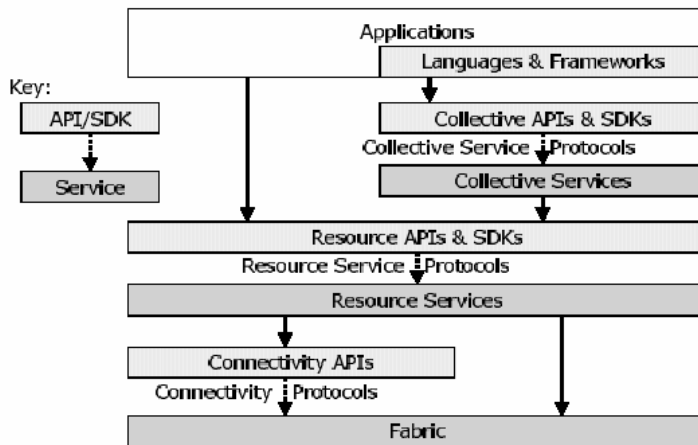


Propuesta de Estructuración de un Programa de e-Ciencia

Esta propuesta¹ esta basada en la experiencia (limitada) en proyectos Grid actuales.

Arquitectura:

Los Grids computacionales definen una arquitectura “genérica” de construcción de aplicaciones (ver figura) junto con una estructura distribuida no solo de infraestructura sino también en desarrollo de middleware y aplicaciones.



(I.Foster, C.Kesselman, S.Tuecke, en Anatomy...)

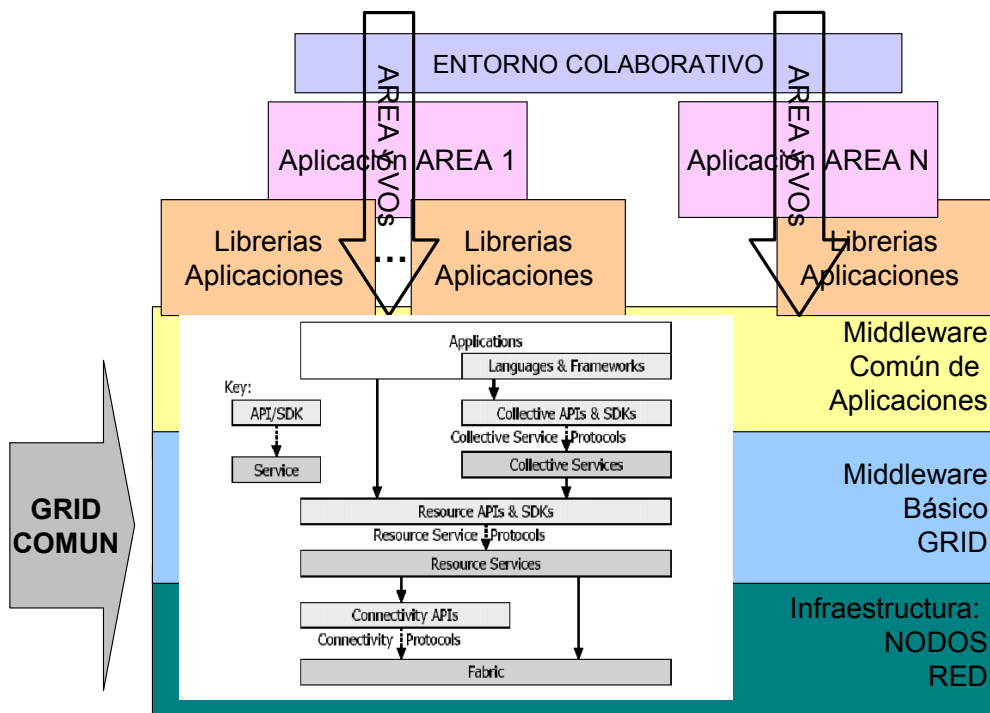
Las fronteras entre las diferentes capas de esta arquitectura, y de su distribución, son las que definen la estructura de un programa de e-Ciencia/Grid.

Uno de los puntos básicos es la idea de una “infraestructura” Grid común, en principio no obligatoria. El punto de unión actual es la red académica nacional (RedIRIS) y europea (Geant), conectando los recursos de computación. Estos recursos pueden ser multidisciplinares (ejemplo: los recursos de computación de un centro de Supercomputación, o del centro de cálculo de una Universidad, o de un cluster de un Centro de Investigación), y en ese caso la existencia de un middleware común permitirá aprovechar de forma mucho mas efectiva estos recursos, sobre los que se establecerán para su uso diferentes Organizaciones Virtuales (VO). Por ejemplo, las comunidades de Físicos de Altas Energías, Astrofísicos, Biólogos Computacionales, podrían compartir los recursos de computación sobre un Grid del CSIC, administrando el acceso y uso de los mismos mediante varias VO en cada área. La existencia y uso de este middleware básico común es uno de los pasos propuestos por proyectos planteados ante el 6PM como EGEE. Pero un entorno diferente (por ejemplo el industrial) podría requerir una infraestructura Grid distinta.

Este proceso puede extenderse a niveles comunes a las aplicaciones, no definidos en el esquema anterior, con puntos como acceso a bases de datos, técnicas comunes de data-mining, visualización, etc. Esta es otra de las ideas discutidas en el caso de Complex Problem Solving using Grids (6PM).

El esquema anterior quedaría así ampliado como indica la figura siguiente:

¹ Esta propuesta es solo un documento de discusión para motivar la misma.



Las áreas de trabajo propuestas de acuerdo al esquema anterior son:

- 1) Puesta en marcha de infraestructura GRID y operación de la misma en testbeds y en modo producción. Disponibilidad de recursos existentes de computación en esquema GRID, incorporación de nuevos recursos, procedimientos de instalación automatizados, “autonomic computing”. Autoridades de certificación. Soporte de VO. Acceso a los recursos de Red. Coordinación con los operadores de red.
- 2) Desarrollo básico de tecnología común para proyectos GRID
Incluye aportaciones originales o refuerzo de líneas existentes en middleware básico y común para desarrollo de aplicaciones
- 3) Desarrollo de aplicaciones en el entorno GRID: librerías comunes y aplicaciones finales en cada área temática.
- 4) Transferencia de tecnología, Cursos, Difusión y Proyectos con Empresas

Organización

- ***Unidad básica: Centros de e-Ciencia***

Una organización a través de Centros de e-Ciencia o similares tiene varias ventajas:

- 1) Posibilidad de co-financiación institucional/local/regional de los proyectos
- 2) Soporte local de usuarios y desarrolladores
- 3) Facilidad de difusión, contactos cercanos con recursos locales y empresas
- 4) Esquema similar al del soporte de Red
- 5) Competencia saludable

Las funciones de los centros son:

- 1) Aportar recursos computacionales en un entorno GRID, y mantener operativos estos recursos en el marco de un GRID nacional
- 2) Promover proyectos en áreas temáticas de interés regional y especialmente en colaboración con empresas.
- 3) Difusión de la iniciativa GRID
- 4) Participar en la coordinación a nivel nacional

La descripción de los centros de e-Ciencia se detalla mas adelante. Su financiación se realiza a través de los Equipos indicados a continuación, y está ligada a la participación en proyectos.

- ***Equipo de Soporte de Infraestructura GRID***

Este equipo estará distribuido entre los centros de e-Ciencia, pero funcionara de forma coordinada. La dotación a aportar desde el MCyT o a traves de fondos institucionales, locales o regionales será:

- a) complementar los recursos humanos de proyectos nacionales o internacionales para mantener operativo el equipo de soporte de infraestructura GRID (típicamente 2 personas por centro)
- b) complementar recursos de computación necesarios en un entorno Grid (ej. servidor “gatekeeper”, servidores de publicación de recursos, etc)

La financiación puede realizarse mediante acciones especiales complementarias de los proyectos realizados en los centros de e-Ciencia correspondientes.

Este equipo operará un testbed nacional. Existirá un centro responsable de Coordinación de Soporte que organizará un servicio de asistencia continua incluyendo un HelpDesk. Funcionará siguiendo un modelo similar al definido por el proyecto europeo EGEE.

Además este equipo coordinará los temas relacionados con la adquisición e instalación de infraestructura de computación, y relación con proveedores.

- ***Equipo de Soporte de Red, Seguridad y Herramientas de Colaboración***

Este equipo de soporte estará coordinado por RedIRIS, y contará con una persona dedicada parcialmente a esta labor en cada centro de e-Ciencia. Entre las tareas de este equipo está la consideración de todos los aspectos relativos a la red, el establecimiento y

operación de la Autoridad de Certificación, y la instalación y mantenimiento de herramientas colaborativas (desde listas de correo, webs, a videoconferencia, salas avanzadas, etc). La financiación de la parte RedIRIS estará definida por el MCyT, y cada centro aportará personal permanente ya existente reforzado con contratos en su caso (de nuevo incluidos en la Acción Especial complementaria a proyectos realizados)

- ***Equipo de Desarrollo de la Arquitectura GRID y elaboración de middleware***

Este equipo reunirá a profesionales del campo de computación distribuida, contando con la información del equipo de Aplicaciones. Se financiará mediante proyectos del programa TIC, coordinados en su preparación a través de este equipo. Las áreas básicas relacionadas abarcan desde middleware básico, herramientas de computación distribuida, acceso a bases de datos, etc.

- ***Equipo de desarrollo e integración de Aplicaciones***

Este equipo coordinará el desarrollo de aplicaciones GRID, promoviendo las mismas a través de proyectos adecuados bien de los distintos programas nacionales, o europeos. El Equipo contará con un asesor por área, y coordinadores de integración del Equipo de Soporte de Infraestructura y de Middleware. Estará abierto a nuevas áreas siguiendo los criterios de programas nacionales y europeos. Cada área contará con una financiación específica para posibilitar la coordinación y en su caso la gestión de la/las VO. correspondientes. En principio esta financiación se solicitará dentro de uno de los proyectos correspondientes en el área. El proyecto se asignará formalmente al centro de e-Ciencia correspondiente al coordinador del mismo.

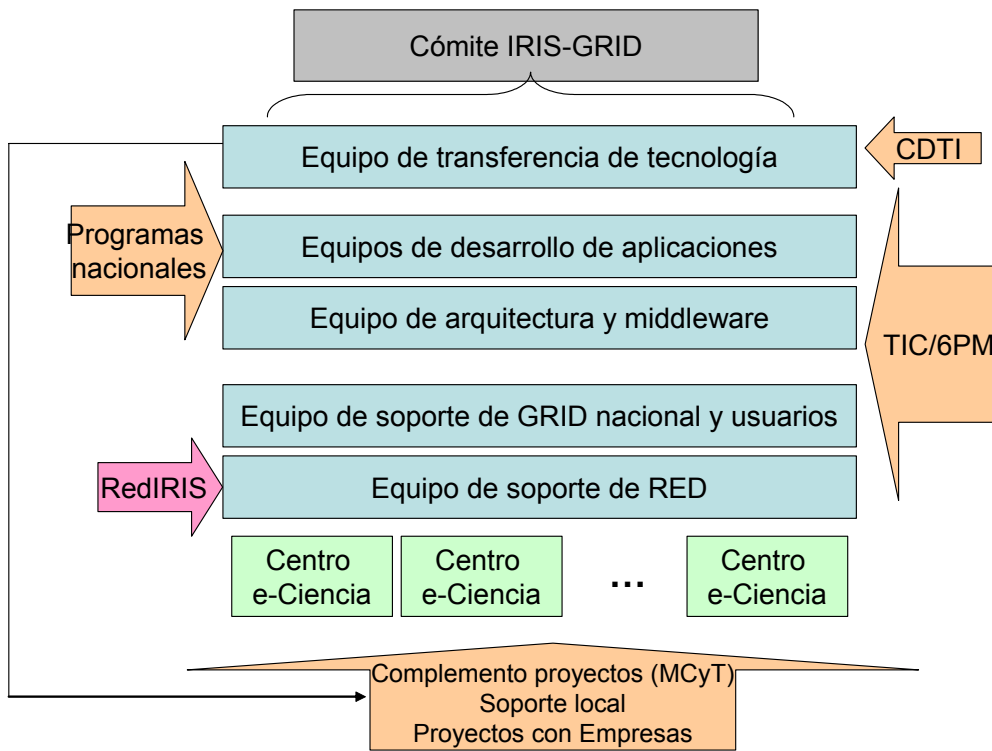
La relación de áreas para proponer proyectos incluye, entre otras, las siguientes: Física de Altas Energías, Área de Salud, Bio-computación, Astrofísica, Meteorología, Medio Ambiente, etc.

- ***Equipo de transferencia de tecnología, difusión y prospectiva en iniciativas internacionales***

Este equipo promoverá la participación en proyectos con empresas, utilizando las diversas herramientas existentes en el MCyT y en contacto con el CDTI. Coordinará las actividades de cada centro de e-Ciencia. Su coordinador se elegirá entre estos.

Una de las áreas de promoción posible de proyectos es la simulación en la industria del Automóvil, Aeronautica, etc.

Todos los miembros de los equipos anteriores estarán adscritos a uno de los centro de e-Ciencia. Cada Equipo contará con un coordinador. La coordinación entre todos los equipos anteriores se llevará a cabo mediante el comité ejecutivo de IRIS-GRID, que contará con presencia de los coordinadores de los centros de e-Ciencia, y de los equipos antes referidos. El comité ejecutivo estará asesorado por un comité técnico y un comité científico. El comité técnico actuará además en la revisión de los proyectos presentados y la evaluación de los realizados, e incluirá expertos de la comunidad GRID internacional. El comité científico asesorará en las áreas de aplicación, y recogerá la opinión de los responsables (gestores de programas, etc.).



ELABORACION DE LAS PROPUESTAS POR AREA TEMATICA

Se propone elaborar para el informe a presentar ante nuestra comunidad y al MCyT hacia principios de Junio, una propuesta elaborada bajo la coordinación explícita de un grupo en cada una de las áreas temáticas.

Las propuestas pueden incluir:

- 1) Motivación de un entorno GRID en el área.
- 2) Estudio de necesidades y proyectos de la comunidad correspondiente.
Informe por grupo del area de proyectos relevantes en programas nacionales/internacionales. Proyectos GRID en marcha a nivel nacional/internacional
- 3) Definición explícita de Use Cases
- 4) Evaluación del posible uso de middleware común
- 5) Desarrollo previsto de middleware específico del área
- 6) Recursos de infraestructura y humanos disponibles, dimensión.
- 7) Definición de posibles proyectos piloto.
- 8) Perspectivas de participación en el 6PM
- 9) Transferencia de tecnología, visibilidad y difusión de los proyectos

Física de Altas Energías: CIEMAT

Área de Salud: Universidad Politécnica de Valencia

Bio-computación: CNB

Química Computacional: Universidad de Valencia

Astrofísica: IAA / IAC

Meteorología: Universidad de Cantabria

Sistemas Complejos: IMEDEA

Medio Ambiente: Universidad de Santiago (?)

...

De modo similar se deben plantear propuestas para los equipos de soporte de Infraestructura (CESGA, CSIC, IFAE), equipo de Red (RedIRIS), equipo de Arquitectura GRID y middleware (UCM/CAB, UAB, UPC, USC, UGR...)

ORGANIZACIÓN DE UN CENTRO DE e-CIENCIA

La propuesta de organización de centros de e-Ciencia será genérica, pero requiere detallar la constitución de los primeros cinco centros (Galicia, Cataluña, Valencia, Madrid, Cantabria) en cuanto a recursos y organización.

Estos centros, aún siendo virtuales, contarán con un coordinador y miembros adscritos. El coordinador será identificado por el MCyT, dentro de una propuesta del comité ejecutivo respaldada por el comité asesor de IRIS-GRID, formado por científicos de reconocido prestigio o experiencia en el área, y confirmado por la institución correspondiente. La adscripción será voluntaria y abierta, bien como usuario o con aportación de infraestructura. La gestión de los recursos, organizada a través de las correspondientes VO, tendrá en cuenta este punto.

Cada centro de e-Ciencia deberá tener actividad en

- i) infraestructura conectada en GRID:
de cara a una participación adecuada en las próximas iniciativas del 6PM, esta infraestructura, mantenida con un nivel de calidad de producción, contará con el equivalente de un mínimo de 50 nodos de computación (dual Xeon 2.4 GHZ o equivalente) y 1TB de disco on-line con soporte de un HSM , backbone Gigabit, conexión a la RedIRIS2 de 155Mbps o superior, desarrollo de middleware genérico y de soporte global de aplicaciones
- ii) promoción de una línea temática de aplicaciones
- iii) desarrollo de proyectos con la participación significativa de empresas.
- iv) administración de recursos y diseminación

Cada nodo de e-Ciencia deberá contar con un volumen de actividad en proyectos GRID de al menos 200keuros anuales. El MCyT podrá complementar los recursos logrados en proyectos europeos o con empresas para garantizar el desarrollo armónico del centro: complementando en infraestructura proyectos con dotación de personal, o viceversa. En particular, la dotación inicial será orientativamente de 100k€ por centro para complementar principalmente la infraestructura existente hasta el nivel indicado²

En cualquier caso cada centro deberá identificar dentro del personal adscrito al mismo:

- Coordinador
- Responsable de infraestructura GRID en producción
- Responsable de recursos de red
- Promotor de proyectos con empresas
- Responsable de diseminación

Todos los responsables participarán a su vez en la organización nacional dentro del comité ejecutivo, comités técnicos, foros industriales y de diseminación de IRIS-GRID

² La extensión de nodos a las áreas Sur, Centro, Baleares, Canarias, etc, seguirá un esquema similar. La dotación de infraestructura necesita de renovación cada 3 años aproximadamente, por lo que la dotación de infraestructura de 500k€ anuales es una estimación razonable. Del mismo modo las acciones sobre redes temáticas requerirán un complemento similar.