Propuesta para la Creación de un Programa de e-Ciencia
Área Temática de Middleware

RESUMEN

El objetivo principal de este documento es proponer los aspectos principales que se deberían incluir en el área temática de *middleware* dentro de la propuesta para la creación de un Programa de e-Ciencia. En primer lugar se describe de forma general el estado actual de la tecnología, y se analizan las diferentes líneas de investigación para cada uno de los niveles del *middleware* relevantes en un Grid. A continuación se indica como se contempla el desarrollo de tecnología Grid en el Plan Nacional de I+D+I actual, y se enumeran los proyectos de investigación en *middleware* Grid en España. Por último se realiza una propuesta de las líneas de investigación que se deberían incluir en el área temática de *middleware*.

HOJA DE ESTADO DEL DOCUMENTO

Registro de Cambios				
Título: Área Temática de Middleware				
Referencia:				
Versión	**Revisión**	**Fecha**	**Autor**	**Razón del cambio**
1	0	21/4/03	Ignacio Martín Llorente	Creación del documento
1	1	28/4/03	Álvaro Fernández Casani	Información sobre investigación en IFIC, mención de proyectos middleware tipo DataGrid y CrossGrid y propuesta de nuevo nivel con servicios alto nivel para el desarrollo de aplicaciones
1	2	28/4/03	Antonio F. Gómez Skarmeta Diego Sevilla Ruíz	Información sobre investigación en UM, sección sobre servicios básicos y parte de componentes en sección sobre servicios alto nivel
1	3	28/4/03	Julio Ortega Lopera	Información sobre investigación en UGr
1	4	28/4/03	Gabriel García García	Información sobre investigación en UPV y sección sobre perspectivas del VI programa marco
1	5	28/4/03	Francisco Fernández Rivera	Información sobre investigación en USC y mención de proyectos middleware tipo DataGrid y CrossGrid
1	6	28/4/03	Félix García Carballeira	Información sobre investigación en UCIIIM y parte sobre Grid de datos en sección sobre servicios alto nivel
1	7	28/4/03	Ramón Doallo Biempica	Información sobre investigación en UDC
1	8	28/4/03	Miquel Angel Senar Rosell	Información sobre investigación en UAB, modificación global del documento y sección sobre servicios locales
1	9	28/4/03	Rosa M. Badia	Información sobre investigación en UPC y parte de herramientas de ayuda en la sección sobre herramientas Grid
1	10	28/4/03	Rubén S. Montero	Parte de herramientas específicas en la sección sobre Herramientas Grid
2	0	28/4/03	Ignacio Martín Llorente	Integración de todos los comentarios
2	1	29/04/03	Antonio Fuentes y Diego López	Integración de servicios de autenticación, autorización y localización.
2	2	5/05/03	Álvaro Fernández Casani	Ampliación de líneas de investigación en la sección sobre servicios alto nivel y cambio figura
2	3	5/05/03	Miquel Angel Senar Rosell	Correcciones sobre
3	0	5/05/03	Ignacio Martín Llorente	Integración de todos los comentarios

COORDINACIÓN

- Miguel Angel Senar (Departamento de Informática; **Universidad Autónoma de Barcelona**): MiguelAngel.Senar@uab.es
- Ignacio Martín Llorente (Departamento de Arquitectura de Computadores y Automática; **Universidad Complutense de Madrid**): llorente@dacya.ucm.es

INVESTIGADORES INVOLUCRADOS EN LA REVISIÓN

- Antonio F. Gómez Skarmeta y Diego Sevilla Ruíz (Departamento de Ingeniería de la Información y las Comunicaciones; **Universidad de Murcia**): skarmeta@dif.um.es
- Gabriel García (Departamento de Sistemas Informáticos y Computación; **Universidad Politécnica de Valencia**): ggarcia@dsic.upv.es
- Álvaro Fernández (IFIC; **CSIC**): alvaro.fernandez@ific.uv.es
- Eduardo Huedo Cuesta (CAB; **INTA**/CSIC): huedoce@inta.es
- Félix García Carballeira (Departamento de Informática; **Universidad Carlos III de Madrid**): fgcarbal@inf.uc3m.es
- Rosa M. Badía (Departamento de Arquitectura de Computadores; **Universidad Politécnica de Cataluña**): rosab@ac.upc.es
- Ramón Doallo Biempica (Departamento de Electrónica y Sistemas; **Universidad de la Coruña**): doallo@udc.es
- Francisco M. Fernández Ribera (Departamento de Electrónica y Computación; **Universidad de Santiago de Compostela**): fran@dec.usc.es
- Julio Ortega Lopera (Departamento de Arquitectura y Tecnología de Computadores, **Universidad de Granada**): julio@atc.ugr.es
- Antonio Fuentes (**RedIRIS**): antonio.fuentes@rediris.es
- Rubén Santiago Montero (Departamento de Arquitectura de Computadores y Automática; **Universidad Complutense de Madrid**): rubensm@dacya.ucm.es

CONTENIDOS

1 TECNOLOGÍAS DE COMPUTACIÓN EN RED 3
2 TECNOLOGÍA GRID 3
3 NIVELES MIDDLEWARE 4
3.1 SERVICIOS LOCALES 4
3.2 SERVICIOS GRID BÁSICOS 5
3.3 SERVICIOS GRID ALTO NIVEL 5
3.4 HERRAMIENTAS GRID 7
4 EL GRID EN EL ACTUAL PLAN DE I+D+I 9
5 PROYECTOS MIDDLEWARE EN ESPAÑA 9
6 PERSPECTIVAS DE PARTICIPACIÓN EN EL VI PROGRAMA MARCO 11
7 PROPUESTA 12

1 Tecnologías de Computación en Red

El objetivo de esta sección es proporcionar una visión global de la tendencia actual de las diferentes tecnologías que permiten aprovechar de forma conjunta los recursos disponibles en sistemas interconectados por red. Los siguientes modelos de computación en red aportan mecanismos para aprovechar al máximo los recursos distribuidos que generalmente se encuentran infrautilizados:

- **Cluster Computing**: Diseño de un cluster dedicado de equipos como alternativa a la adquisición de un equipo multiprocesador. Su ventaja fundamental es la mejor relación coste/rendimiento. Sus inconvenientes son: dificultad de programación y mantenimiento. Los clusters suelen estar gestionados por sistemas que se encargan de ejecutar las aplicaciones de los usuarios sobre las distintas máquinas en función de diferentes criterios de planificación fijados por el sistema. Estos sistemas de gestión pueden ser sistemas integrados de planificación como MOSIX (www.mosix.cs.huji.ac.il) o gestores de colas batch como PBS (versión libre) y PBS-Pro (versión comercial) de Veridian Systems (www.openpbs.org), LSF de Platform Computing (www.platform.com), SGE de Sun Microsystems (www.sun.com/Gridware) o Condor de la Universidad de Wisconsin-Madison (www.cs.wisc.edu/condor).

- **Intranet Computing**: Unión de la potencia computacional desaprovechada en los recursos hardware distribuidos en una red de área local (un único dominio de administración). Su principal ventaja es que puede proporcionar rendimientos semejantes a los ofrecidos por los sistemas de alto rendimiento con un coste económico casi nulo. La mayoría de los gestores de colas para clusters suelen ofrecer soluciones para unir múltiples clusters independientes dentro de una red local y mover los trabajos desde los clusters más ocupados a los más desocupados. Algunos de ellos también ofrecen facilidades para usar de forma oportunista recursos individuales que no están integrados en un cluster. También existen empresas como GridSystems, Avaki, Entropía o United Devices que comercializan software de Intranet Computing específico para aplicaciones paramétricas.

- **Internet Computing**: Aprovechamiento de la potencia de los recursos distribuidos por Internet siguiendo el modelo cliente/servidor. Actualmente casi todas estas herramientas se limitan a ejecución de aplicaciones paramétricas. Su ventaja es el gran rendimiento que se puede obtener. Sus principales inconvenientes son debidos al bajo ancho de banda y a la escasa seguridad en Internet. Avaki, Entropía o United Devices mantienen versiones de sus herramientas que permiten su uso en Internet. Probablemente el ejemplo más típico de Internet Computing es el proyecto seti@home (setiathome.ssl.berkley.edu). Algunos gestores de colas como Condor ofrecen también la posibilidad de aprovechar recursos sobre Internet ofreciendo mayor seguridad y confiabilidad aunque requiere que todas las máquinas involucradas usen dicho gestor.

El uso de las tecnologías descritas anteriormente posibilita el aprovechamiento eficiente de los recursos dentro de una misma organización. Algunas de ellas como SGE, LSF o Condor permiten incluso unir diferentes departamentos u organizaciones pero con la condición de que sea su software el que gestione los recursos internos. Sin embargo, ninguna de estas tecnologías permite unir dominios de administración diferentes manteniendo la política de seguridad de cada centro y las herramientas de planificación ya en uso. Por otro lado, los interfaces y protocolos básicos que utilizan las herramientas anteriores no están basados en estándares abiertos, condición imprescindible para que la tecnología Grid se extienda, y en pocos años sea tan habitual como actualmente es la tecnología Web.

2 Tecnología Grid

La necesidad de aprovechar los recursos disponibles en los sistemas informáticos conectados a Internet y simplificar su utilización ha dado lugar a una nueva forma de tecnología de la información conocida como *Grid Computing*. Esta nueva tecnología es análoga a las redes de suministro eléctrico: la idea es ofrecer un único punto de acceso a un conjunto de recursos distribuidos geográficamente (supercomputadores, clusters, almacenamiento, fuentes de información, instrumentos, personal...). De este modo, los sistemas distribuidos se pueden emplear como un único sistema virtual en aplicaciones intensivas en datos o con gran demanda computacional.

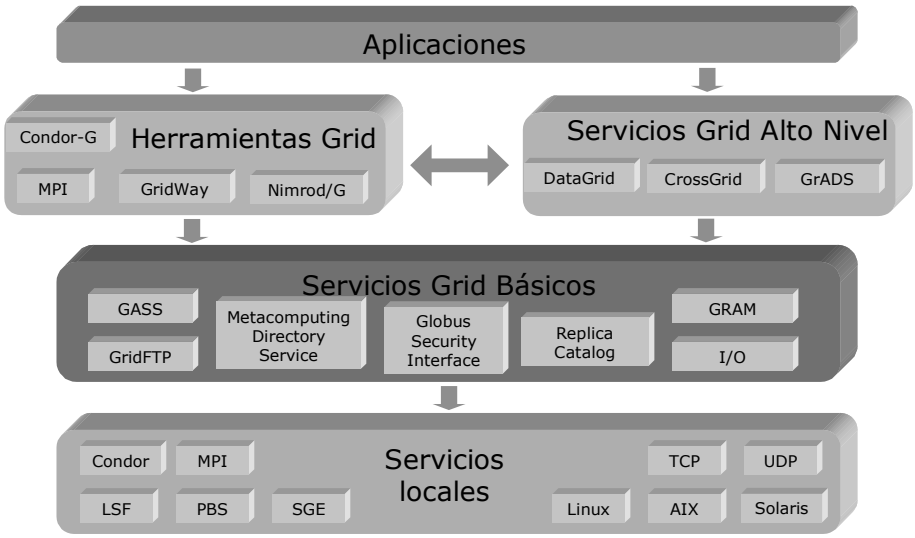
Las tecnologías descritas en la sección anterior son casos especiales de un nuevo paradigma de computación distribuida que en poco tiempo estamos seguros que va a revolucionar no sólo la computación de altas prestaciones sino Internet en general. Esta nueva tendencia denominada *Grid Computing* supone un cambio radical en la colaboración de sistemas conectados a Internet y en particular en la computación de altas prestaciones debido a su enorme potencial respecto al intercambio y gestión de recursos. Un sistema Grid se caracteriza por carecer de un control centralizado, estar basado en estándares abiertos y proporcionar calidad de servicio (www.Gridtoday.com/02/0722/100136.html). **Es importante resaltar que la tecnología Grid no pretende sustituir**

las tecnologías anteriores ya que su ámbito de aplicación es diferente. El objetivo de la tecnología Grid es unir de forma segura los recursos de diferentes dominios de administración, respetando sus políticas de seguridad y herramientas de gestión internas. Además, su objetivo es unir todo tipo de recursos y no únicamente capacidad de procesamiento y almacenamiento.

3 Niveles Middleware

Dentro de los sistemas *Grid* podemos diferenciar tres niveles de investigación relacionados con *middleware*:

- Servicios locales
- Servicios Grid básicos
- Servicios Grid alto nivel
- Herramientas Grid



3.1 Servicios locales

En este nivel se incluye el desarrollo de herramientas para explotar los recursos en la Intranet del centro investigación. En este campo existen en la actualidad muchos servicios disponibles para realizar un uso eficiente de los recursos locales. Podemos destacar:

- Herramientas de monitorización de los diferentes recursos físicos del cluster (Ganglia: ganglia.sourceforge.net, Hawkeye: www.cs.wisc.edu/condor/hawkeye),
- Gestores de colas batch como los mencionados en la sección 1,
- Librerías de programación paralela (MPI: www-unix.mcs.anl.gov/mpi/, OpenMP: www.openmp.org/ o PVM: www.csm.ornl.gov/pvm/pvm_home.html),
- Herramientas de depuración y monitorización de aplicaciones (TotalView:www.etnus.com/Products/TotalView/, Paradyn:www.cs.wisc.edu/~paradyn/, Vampir:www.pallas.com/pages/vampir.htm, Paraver: http://www.cepba.upc.es/paraver).

Las **líneas abiertas de investigación** más representativas relacionadas con los servicios locales se centran en los siguientes aspectos:

- Escalabilidad de servicios y herramientas. A medida que los clusters locales van aumentando en tamaño, se hace también necesario adaptar las herramientas y servicios para que puedan usarse también sobre sistemas de miles de componentes.
- Automatización de políticas de gestión, configuración, control e instalación que garanticen el funcionamiento de clusters de gran tamaño sin la necesidad de una intervención humana constante y que puedan reaccionar a condiciones dinámicas.
- Mecanismos de tolerancia de fallos, que garanticen, por ejemplo, la ejecución de las aplicaciones o el funcionamiento de los servicios ante la presencia de fallos temporales en distintos elementos del cluster.

3.2 Servicios Grid Básicos

El *Grid Forum* (www.Gridforum.org) es el organismo encargado de crear los estándares de servicios y protocolos necesarios para crear la infraestructura o tecnología *Grid*. A pesar de que existen otras tecnologías *Grid* como Legion (www.cs.virginia.edu/~legion/), Unicore (www.unicore.org) o MOL (www.uni-paderborn.de/pc2/projects/mol/), la mayoría de los proyectos *Grid* actuales se están construyendo basados en los servicios y protocolos proporcionados por el ***Globus Toolkit*** (www.golbus.org). La tecnología Globus ha sido seleccionada como estándar de facto por las 12 compañías (Compaq, Cray, SGI, Sun Microsystems, Entropia, IBM, Microsoft, Platform Computing y Veridian en Estados Unidos; y Fujitsu, Hitachi y NEC en Japón) más importantes del sector de computación de altas prestaciones (http://www.globus.org/developer/news/20011112a.html). La próxima versión de Globus denominada **OGSA** (*Open Grid Services Architecture*) muestra una clara convergencia hacía la tecnología de Web Services en el campo de e-Business. El Globus Toolkit 3 (GT3), del que ya existe una versión alfa, 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 Grid 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 estos servicios y protocolos como elemento básico:

- GRAM (Globus Resource Allocation Manager): Este servicio proporciona un API para solicitar el inicio de ejecuciones en un recurso de computación y para gestionar estas ejecuciones una vez invocadas
- MDS (Metacomputing Directory Service): Este servicio proporciona un API para averiguar la estructura y estado de los recursos
- GSI (Globus Security Infrastructure): Este servicio proporciona una solución de autenticación global
- GASS (Global Access to Secondary Storage): Este servicio proporciona mecanismos de acceso y APIs para almacenar ficheros en diferentes sistemas

A partir de estos servicios básicos de middleware, las **líneas de investigación abiertas** se encuentran orientadas mejorar los protocolos y servicios básicos aportados por Globus para adaptarlos a nuevas tecnologías o ambientes. Por ejemplo:

- Añadir a Globus el soporte para IPv6 y por añadidura las funcionalidades de autoconfiguration, multihoming y seguridad.
- Mejorar la infraestructura de clave pública de Globus para hacerla más flexible e integrar extensiones actuales de las PKI para gestión de autorizaciones y credenciales.
- Implantar un esquema de autenticación y autorización más flexible y acorde con los estándares actualmente en uso en la infraestructura middleware de las redes académicas, con el objetivo de facilitar el uso de estos recursos a los usuarios finales del Grid e integrarlo en el uso normal de los recursos Internet

El movimiento hacia los Web Services que está experimentando Globus hace que también haya que replantear de nuevo los servicios ofrecidos, unido a la novedad del mundo de los servicios Web, que abre a la vez un campo de investigación muy interesante y potencialmente con mucho futuro.

3.3 Servicios Grid Alto Nivel

A pesar del tremendo esfuerzo realizado por la comunidad científica, la ejecución y 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 todas las todos los pasos involucrados en la ejecución

de un trabajo, a saber: descubrimiento y selección de recursos, inicialización, envío, monitorización, migración y finalización. Por lo tanto **es necesario desarrollar 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 ....), así como a las demandas dinámicas de la aplicación (tiempo máximo de ejecución, presupuesto, necesidad de hardware/software específico,...).

En este nivel se incluye el desarrollo de middleware que, sobre el middleware básico proporcionado por Globus, proporciona servicios adicionales para aplicaciones y usuarios. Se trata de una capa de servicios reutilizables de más alto nivel que puedan hacer uso de los existentes para ofrecer servicios más flexibles en áreas como planificación, gestión de datos, visualización o colaboración. Por ejemplo, *trading* inteligente de componentes basado en especificaciones (mejoras al lenguaje de especificación RSL de Globus), introducción de mecanismos de calidad de servicio (QoS) para ofrecer servicios con ciertas garantías de servicio (soportando así servicios en tiempo real tales como las aplicaciones colaborativas), gestión de *scheduling* con prioridades, etc.

Entre los proyectos en curso cabe resaltar *European Data Grid* (www.eu-dataGrid.org), *Grid Physics Network* (www.griphyn.org), GrADS (nhse2.cs.rice.edu/grads/) y *CrossGrid* (http://www.crossGrid.org/), que tratan de desarrollar, entre otros, métodos eficientes de computación y acceso a datos distribuidos por medio de servicios basados en componentes Globus.

Por ejemplo, una aportación significativa del proyecto DataGrid es la implantación de un servicio *Resource Broker* que se encarga de la **gestión de trabajos**, buscando recursos del Grid y ejecutando el trabajo en los mismos. Este *broker* se encarga de llevar un seguimiento del trabajo, lanzándolo en otros recursos si alguno falla, hasta la finalización del mismo. Esto facilita el lanzamiento de trabajos al usuario, que sólo se ha de encargar de especificar los requerimientos del mismo y no de los recursos donde ejecutarlo. Además proporciona información sobre el estado del trabajo, haciendo de repositorio central de información y de los datos producidos por el mismo. El proyecto CrossGrid prevé hacer modificaciones sobre el *Resource Broker* anterior, para soportar algunas características hasta ahora no soportadas hasta ahora como el lanzamiento de aplicaciones MPI sobre varios 'sites', o mejorar el diseño del sistema pasando de un *Resource Broker* centralizado a una arquitectura distribuida basada en Agentes de *Scheduling*. Otra de los puntos fuertes de CrossGrid es el desarrollo de un portal Web sobre el cual los usuarios podrán lanzar sus trabajos, mejorando la interfaz con el sistema Grid subyacente.

Por tanto, los siguientes **aspectos pendientes por resolver** serían:

- Desarrollo de servicios de descubrimiento y selección de recursos
- Desarrollo de servicios de planificación y adaptación a las condiciones dinámicas de un Grid y de un trabajo
- Desarrollo de servicios de inicialización, envío, monitorización, migración y finalización de trabajos

Para la mayoría de las aplicaciones a ejecutar en entornos Grid el acceso a los datos es tan importante como el acceso a los recursos de cómputo, puesto que la mayoría de las aplicaciones científicas y de ingeniería requieren el acceso a grandes volúmenes de datos (terabytes o petabytes). En otras ocasiones los datos que generan las aplicaciones exceden de los recursos locales disponibles. Además, muchas de las aplicaciones futuras como lo entornos colaborativos virtuales también requerirán el acceso a datos ampliamente distribuidos.

Un **Grid de datos** proporciona un entorno en el que poder manipular y acceder a datos almacenados en sistemas ampliamente distribuidos. Su objetivo está orientado al intercambio y procesamiento de información de forma segura y eficiente, para lo cual hay que desarrollar servicios que permitan agrupar diferentes sistemas de almacenamiento locales donde poder almacenar, replicar e incluso fragmentar los datos. Los requisitos fundamentales que se imponen en un Grid de datos son:

- Espacio de nombres global.
- Técnicas de reducción de latencia.
- Consistencia en el acceso a los datos.
- Mecanismos de acceso común para localizar y acceder a los datos.
- API para el acceso a los datos.

Muchos de estos requisitos son similares a los de un sistema de ficheros local, sin embargo, presentan una problemática diferente, como son la necesidad de almacenar grandes cantidades de datos, existencia de diferentes protocolos de acceso a los datos locales (GridgFTP, NFS, HTTP-Webdav, etc.), problemas de seguridad y posibilidad no solo compartir datos, sino también distribuir y fragmentar los datos a través de los diferentes recursos de almacenamiento locales.

La resultados que se han obtenido hasta la fecha en los Grids de datos se puede resumir en los siguientes puntos:

- Desarrollo de servicios de directorios que permiten la búsqueda de datos en entornos ampliamente distribuidos.
- Empleo de técnicas de replicación como forma de mejorar las prestaciones en el acceso a los datos. Estas técnicas, sin embargo, no son apropiadas en entornos colaborativos o cuando lo que se quiere es fragmentar los datos a través de diferentes recursos de almacenamiento.
- Desarrollo de protocolos de acceso a datos como GridFTP.
- Desarrollo de API para acceder a ficheros remotos, como el servicio GASS (*Global Access to Secondary Storage*) que proporciona *Globus Grid toolkit.*

Quedan, sin embargo, **aspectos pendientes por resolver**:

- Conseguir un verdadero espacio de nombres global, donde sea sencillo localizar los datos.
- Desarrollar técnicas de almacenamiento de altas prestaciones que permitan mejorar el acceso a los datos en un Grid, como por ejemplo, el empleo de técnicas de E/S paralela y distribución de datos a través de los diferentes recursos de almacenamiento
- Desarrollar servicios para integrar diferentes protocolos y sistemas de almacenamientos locales.
- Desarrollar interfaces de acceso a datos adecuadas para computación de altas prestaciones, como por ejemplo, MPI-IO para entornos Grid.
- Explotar las posibilidades de mecanismos de búsqueda e indexación distribuidos, que empleen las mismas tecnologías Grid para ayudar en la localización de recursos.
- Integrar estos procedimientos con los mecanismos de autorización, de manera que todas las interacciones sean susceptibles de personalización
- Mecanismos de tolerancia a fallos, que garanticen el acceso a los datos en presencia de fallos en el Grid.

En general**, la investigación en servicios de alto nivel se desarrollará en dos líneas**:

- Servicios de utilidad para todo el conjunto de aplicaciones (servicios horizontales), como por ejemplo, servicios de gestión de trabajos y datos, reserva anticipada de recursos, accounting distribuido, suscripción a eventos, tolerancia a fallos, etc.
- Servicios específicos para un dominio de aplicación (servicios verticales), por ejemplo, servicios de visualización 3D para los ámbitos científicos que los requieran, servicios de simulación distribuida genérica, etc.

Estos servicios se pueden ofrecer a través de Servicios Web o utilizando cualquier otro mecanismo RPC como CORBA. En este ámbito, la tecnología de componentes surge como una tecnología de desarrollo muy adecuada. El desarrollo basado en componentes (iniciado, aunque en un estado primitivo todavía, por la nueva versión de Globu 3) permite que los desarrolladores de aplicaciones creen componentes que pueden ser utilizados por cualquier otra aplicación. El desarrollo de aplicaciones se convierte entonces en una actividad de conexión de los distintos componentes que están disponibles para el desarrollo de las mismas y que están distribuidos por todo el Grid. La gestión de qué componentes existen en cada centro de computación se puede realizar de forma automática por herramientas añadidas a las básicas de Globus, además de otras herramientas de desarrollo de aplicaciones basadas en el descubrimiento automático de componentes, la conexión y la puesta en ejecución. El uso de componentes para programación Grid se ha iniciado en diferentes proyectos, como el Common Component Architecture (cca-forum.org), GridCCM (www.irisa.fr/paris/Gridccm/welcome.htm), la especificación Lightweight CCM, Globus en su versión 3 y el propio proyecto PIRAMIDE con el modelo de componentes CORBA *Lightweight Components* (CORBA-LC).

3.4 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 las componentes básicas y de alto nivel.

Librerías y Herramientas para Tipos Específicos de Problemas

Tradicionalmente, existe un conjunto de aplicaciones cuyas demandas computacionales no pueden ser satisfechas por los supercomputadores actuales. La resolución de estos problemas requiere un estudio de cómo el Grid puede mejorar los paradigmas de supercomputación actuales, la investigación de nuevos paradigmas y el desarrollo de herramientas de alto nivel y API's (Application Programmer Interface) que permitan a los científicos e ingenieros expresar de forma sencilla los problemas en un entorno Grid. En particular, ejemplos de problemas que potencialmente pueden beneficiarse del uso de Grids son:

- **Aplicaciones distribuidas de Alto Rendimiento** (High Performance Computing, HPC); cuyos requisitos computacionales únicamente puede satisfacerse mediante la unión de múltiples supercomputadores. Ejemplos de estas aplicaciones son, la dinámica de fluidos computacional, la simulación numérica de procesos físicos complejos, o simulaciones meteorológicas. Entre los proyectos actualmente en desarrollo cabe destacar: MPICH-G2 (www3.niu.edu/mpi/), y Cactus (www.cactuscode.org)

- **Aplicaciones de Alta Productividad** (High Throuput Computing, HTC); problemas que requieren del análisis de todas las posibles soluciones en un espacio de parámetros (problemas NP-Completos). Aunque actualmente estos problemas se resuelven reduciendo el espacio de parámetros mediante alguna heurística (enfriamiento simulado, algoritmos genéticos,...), el Grid ofrece una infraestructura computacional más adecuada para resolver estos problemas de alta productividad débilmente acoplados. Entre los proyectos actualmente en desarrollo cabe destacar: Nimrod/G (www.csse.monash.edu.au/~rajkumar/ecoGrid/), MW (www.cs.wisc.edu/condor), AppLeS (grail.sdsc.edu/), GridWay (www.dacya.ucm.es/asds)

- **Aplicaciones de Ejecución Auto-Adaptativa**: La aparición de la tecnología Grid ha dado lugar a un nuevo paradigma de aplicaciones capaces de adaptar su ejecución de acuerdo a sus requisitos dinámicos. Por ejemplo, los métodos numéricos de refinamiento adaptativo de malla, aumentan de forma sistemática la resolución de la malla numérica en aquellas zonas del dominio con elevados errores de discretización. De esta forma no es posible conocer a priori la cantidad de memoria que necesitará la simulación. Las aplicaciones auto-adaptativas son capaces de buscar recursos adicionales a medida que evoluciona su ejecución para satisfacer sus requisitos. Entre los proyectos actualmente en desarrollo cabe destacar: Cactus Worm (www.cactuscode.org), GridWay (www.dacya.ucm.es/asds)

Otros ejemplos de entornos especializados incluirían, que permite el la resolución de problemas que sigan un modelo de master-worker, Netsolve (http://icl.cs.utk.edu/netsolve), herramienta basada en una arquitectura cliente/servidor para el uso de librerías matemáticas sobre entornos distribuidos, o DAGman (www.cs.wisc.edu/dagman), servicio para controlar aplicaciones compuestas por múltiples trabajos que exhiben relaciones de dependencia entre ellos.

Herramientas de ayuda al desarrollo de aplicaciones

Durante el ciclo de desarrollo de las aplicaciones, es necesario realizar prototipos, depurarlos, sintonizarlos, etc. Cuando las aplicaciones van a ser ejecutadas en un entorno en Grid, la complejidad de este proceso es mucho mayor. Herramientas de desarrollo tales como depuradores, y sintonizadores de la eficiencia son esenciales para que los usuarios puedan entender porque su aplicación no esta dando los resultados o la eficiencia esperados. Otro tipo de herramientas altamente útiles son aquellas que permiten de manera automática obtener programas ejecutables en Grid.

En la actualidad existen pocas herramientas que realicen las tareas indicadas previamente. Algunos ejemplos existentes son:

- P2D2: depurador "Grid-enabled" (NASA)
- Guard: depurador paralelo "relativo" (Monash)
- Ygdrasil toolkit: permite integración a nivel de línea de comando de depuradores como Ladebug, gdb, etc (HP)
- SimGRID: simulador de planificaciones de aplicaciones distribuidas (UCSD, USA)
- Dimemas: simulador de la eficiencia de programas MPI ejecutados distribuidamente (CEPBA-UPC)
- Paraver (CEPBA-UPC), Vampir (Pallas): Herramientas de visualización de la ejecución de aplicaciones distribuidas.

A pesar de que existen algunas herramientas, la mayoría de ellas están en un estado preliminar de desarrollo. **Es necesario extender las funcionalidades de dichas herramientas, robustecer las versiones actuales y lo mas importante hacer conocer estas herramientas a los posibles usuarios.** El campo en el que claramente existe un vacío es el de modelos de programación que faciliten el desarrollo de aplicaciones en GRID de manera que el usuario pueda especificar la funcionalidad de su aplicación sin necesidad de saber con detalle la topología ni características del Grid en el que se ejecutará.

4 El Grid en el Actual Plan de I+D+I

Grid Computing es prioridad temática dentro del Área de Tecnologías de la Información y de las Comunicaciones del Plan Nacional de I+D+I 2000-2003:

1.6. Computación de altas prestaciones
1.6.5. Grid Computing

Además también tiene relación con las siguientes prioridades:

1.6. Computación de altas prestaciones
1.6.4. Herramientas y entornos para el desarrollo de aplicaciones.
1.5. Sistemas en red
1.5.2. Sistemas basados en la Web. Servicios Web. XML.

También está relacionado con la prioridad temática del Programa Nacional De Física De Partículas y Grandes Aceleradores

2. Tecnologías en relación con el proyecto LHC Computing GRID del CERN
2.1 Desarrollo de tecnologías para cálculo en la red (Proyecto LHC Computing GRID del CERN). Se priorizarán los proyectos GRID coordinados con los del Programa Nacional de TIC.

5 Proyectos middleware en España

Nombre del Grupo	Dirección	Nivel Middleware	Proyectos	Responsable	Más información
Arquitectura de Sistemas Distribuidos y Seguridad	Departamento de Arquitectura de Computadores y Automática; Universidad Complutense de Madrid	Herramientas Grid	GridWay	Ignacio Martín Llorente	www.dacya.ucm.es/asds
Tecnología Grid	Centro de Astrobiología; Instituto Nacional de Técnica Aeroespacial	Herramientas Grid	CABGrid	Ignacio Martín Llorente	www.cab.inta.es/~CABGrid
Grupo de Actividades Grid del IFIC	Instituto de Física Corpuscular de Valencia (CSIC)	Servicios Grid Alto Nivel	CrossGrid, Atlas Grid Computing	Jose Salt Cairols	www.alpha.ific.uv.es/Grid
Sistemas Inteligentes y Computación Distribuida	Facultad de Informática. Universidad de Murcia	Servicios Grid Básicos y Alto Nivel	PROYECTO PIRAMIDE: Proyecto IntegRAción de tecnologías MIddleware y Del web como soporte a la infraEstructura Grid	Antonio F. Gómez Skarmeta Diego Sevilla Ruíz	ditec.um.es/~dsevilla/ccm
Circuitos y Sistemas para el Procesamiento de la Información	Departamento de Arquitectura y Tecnología de Computadores. E.T.S.I.Informática (Universidad de	Servicios Grid Básicos	phiPC (TIC2000-1348)	Julio Ortega Lopera	atc.ugr.es/~julio/anuncio.htm

(CASIP)	Granada)				
Grupo de Redes y Computación de Altas Prestaciones	Depto. de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia	Servicios Grid Alto Nivel y Herramientas Grid	GRID-IT	Vicente Hernández García	(Propuesta sometida al Plan Nacional y pendiente de evaluación)
Arquitectura de Computadores	Dep. de Electrónica y Computación. Univ. Santiago de Compostela. 15782 Santiago	Servicios Grid Alto Nivel	Proyecto CrossGrid	Francisco Fernández Rivera	http://www.ac.usc.es
Grupo de Arquitectura de Computadores	Departamento de Informática. Universidad Carlos III de Madrid	Servicios Grid Alto Nivel y Herramientas Grid	Técnicas de almacenamiento de altas prestaciones en entornos Grid Computación de altas prestaciones en clusters de estaciones de trabajo heterogéneas	Félix García Carballeira	arcos.inf.uc3m.es
Arquitectura de Computadores	Dep. de Electrónica y Sistemas. Univ. de A Coruña . 15071 A Coruña	Servicios Grid Alto Nivel y Herramientas Grid	Proyecto CrossGrid Proyecto GaliGrid	Ramón Doallo Biempica	www.des.udc.es
Arquitectura de Ordenadores y Sistemas Operativos	Escola Tecnica Superior d'Enginyeria; Universidad Autónoma de Barcelona	Servicios Grid Alto Nivel y Herramientas Grid	CrossGrid	Miquel Angel Senar Rosell	www.eu-crossGrid.org
Arquitectura de Ordenadores y Sistemas Operativos	Escola Tecnica Superior d'Enginyeria; Universidad Autónoma de Barcelona	Servicios Grid Alto Nivel y Herramientas Grid	Estrategias de gestión autoadaptativas para entornos grid	Emilio Luque/Miquel A. Senar	www.caos.uab.es
Arquitectura de Ordenadores y Sistemas Operativos	Escola Tecnica Superior d Enginyeria; Universidad Autónoma de Barcelona	Servicios Grid Alto Nivel y Herramientas Grid	Interoperabilidad de herramientas	Miquel A. Senar/Anna Cortes	www.caos.uab.es
CEPBA-IBM	UPC	Herramientas Grid	Dimemas (DAMIEN)	Jesús Labarta/Rosa M. Badia	www.hlrs.de/ organization/pds/projects/damien/
CEPBA-IBM	UPC	Herramientas Grid	Grid Superscalar	Jesús Labarta/Rosa M. Badia	www.ciri.upc.es/ projects/rd/rd_archivos/gr/Grid.htm

CEPBA-IBM	UPC	Servicios Grid Alto Nivel	eNANOS	Jesús Labarta/Rosa M. Badia	www.ciri.upc.es/ projects/rd/rd_archivos/gr/Grid.htm
RedIRIS	RedIRIS	Herramientas Grid	PAPO	Diego López	www.rediris.es/app/papi

6 Perspectivas de participación en el VI Programa Marco

En febrero de 2001 la Comisión Europea presentó su propuesta para el VI Programa Marco, que abarca el periodo 2002-2006 (www.mcyt.es/vipm/). Este VI Programa Marco se ha planteado en el contexto del documento *"Hacia un Espacio Europeo de Investigación"* de enero de 2000, donde se muestran resultados preocupantes de la ciencia europea en relación con Japón y Estados Unidos, sobre todo si se tiene en cuenta que la investigación y la tecnología producen del 25% al 50% del crecimiento económico y determinan en gran medida la competitividad, el empleo y la calidad de vida. Como resultado del análisis que se realiza en el documento de enero de 2000, se propone la integración de la I+D europea para alcanzar masas críticas, lograr economías de escala y distribuir mejor los recursos. Así, se plantea la necesidad de *(a)* interconectar los centros de excelencia europeos y el desarrollo de *centros virtuales*, *(b)* apoyar la creación, funcionamiento y acceso a las grandes infraestructuras, *(c)* aumentar el número y la movilidad de los investigadores, y *(d)* ejecutar de manera más coherente las actividades europeas y nacionales. Existe por tanto una clara **sintonía entre estos objetivos y el paradigma que plantea el *Grid Computing***, que puede considerarse como la tecnología de base que habilita un acceso transparente a los recursos científicos distribuidos.

Así, el VI Programa Marco se basa en los principios de concentración de esfuerzos, de mayor coordinación entre las actividades de I+D nacionales y europeas, y en la descentralización de la gestión. En abril de 2001 la Comisión dio a conocer los Programas Específicos en los que está dividido el VI Programa Marco, aunque esta propuesta se modificó en enero de 2002, año en el que se realizó una convocatoria de *Expresiones de Interés* (EoI) para conocer la opinión de la comunidad científica europea.

El importe global máximo correspondiente a este VI Programa Marco es de 17.500 millones de euros y se organiza en tres ejes *(X, Y, Z).* El eje *(X)* corresponde al programa específico de *Integración y el Fortalecimiento del Espacio Europeo de Investigación (ERA)*, y refleja el propósito de concentrar la investigación europea en siete áreas temáticas prioritarias y de atender las necesidades de I+D de las políticas europeas (la llamada octava prioridad). En el eje *(Y)*, definido por el programa de *Estructuración del Espacio Europeo de Investigación*, se incluyen las actividades de investigación e innovación, recursos humanos y movilidad, infraestructuras de investigación, y ciencia y sociedad. Por último, en el eje *(Z)* relativo al fortalecimiento de las bases del espacio europeo de investigación se incluyen las acciones de apoyo a la coordinación de actividades, y de desarrollo coherente de políticas nacionales de investigación.

Dentro del programa específico de *Integración y Fortalecimiento del Espacio Europeo de Investigación*, se definen siete áreas temáticas prioritarias: (1) *ciencias de la vida, genómica y biotecnología para la salud*, (2) *tecnologías de la sociedad de la información*, (3) *nanotecnologías y nanociencias, materiales multifuncionales basados en el conocimiento y nuevos procedimientos y dispositivos de producción*, (4) *aeronáutica y espacio*, (5) *calidad y seguridad de los alimentos*, (6) *desarrollo sostenible, cambio planetario y ecosistemas*, y (7) *ciudadanos y gobernabilidad en una sociedad basada en el conocimiento*. Si bien el desarrollo de una infraestructura de Grid es esencial para alcanzar los objetivos planteados por el programa específico, y por tanto tendría implicaciones en todas las áreas temáticas prioritarias, los aspectos específicos relacionados con la configuración de los niveles de *middleware* relevantes en un Grid, y su utilización, se contemplan en el área temática (2).

Los objetivos del *plan de trabajo* de esa área temática de *tecnologías de la sociedad de la información (IST)* para los años 2003-2004 (www.mcyt.es/vipm/) se organizan según incidan en el nivel de componentes, sistema, o aplicaciones. Una de las prioridades del programa IST (Information Society Technologies) de la Comisión Europea para la construcción del ERA (European Reasearch Area) es el desarrollo de los Grids Computacionales y Grids del Conocimiento. Para ello, se realizará un importante esfuerzo en la implementación de test-beds en infraestructuras de investigación y el desarrollo de las tecnologías Grid. Para cumplir este objetivo la Comisión sugiere la utilización de las tecnologías Grid en varios objetivos estratégicos del programa IST:

Objetivo Estratégico	Convocatoria[1]	Presupuesto (mill. €)	*Objetivos*
Networked businesses and governments	1ª	75	*Tecnologías para la interoperabilidad.*
ehealth	1ª	70	*Desarrollo de nuevas herramientas software para ayudar a los profesionales de la medicina a tomar las mejores decisions para la prevención, diagnóstico y tratamiento.*
Open development Platforms for software and services	2ª	55	*"Open Platforms", middleware y lenguajes que soporten estándares para la interoperabilidad.*
GRID-based Systems for solving complex problems	2ª	45	*Arquitecturas, diseño y desarrollo de los Grid de nueva generación.* *Capacitar las tecnologías aplicadas.*
Improving Risk management	2ª	30	*Tecnologías Grid para formar parte de sistemas integrados para la prevención y la gestión de riesgos.*
Research networking test-beds	*2ª*	*25*	*Promoción de Centros de Excelencia (Grid Technology centres)*

Por otra parte, teniendo en cuenta los beneficios y los problemas implicados en el desarrollo de la tecnología Grid que se pretende impulsar desde el VI Programa Marco, existen muchos otros objetivos del área IST donde el trabajo de investigación en Grid tiene incidencia. Algunos de esos objetivos se indican, entre otras, en las secciones:

- Ancho de Banda para todos (2.3.1.3)
- Sistemas móviles más allá de la tercera generación (2.3.1.4)
- Hacia un entorno de seguridad global (2.3.1.5)
- Negocios y gobiernos en red (2.3.1.9)
- e-Salud (2.3.1.11)
- Aprendizaje asistido por la tecnología y herencia cultural (2.3.1.11)
- Plataformas para el desarrollo abierto de software y servicios (2.3.2.3).
- Sistemas embebidos (2.3.2.5)

Es importante resaltar las posibilidades de cooperación con proyectos existentes en el VI Programa Marco que puedan enlazarse con las actividades desarrolladas a nivel nacional. El VI Programa Marco permite la posibilidad de conexión con clusters de proyectos, Proyectos Integrados o Redes de Excelencia en marcha. La iniciativa de infraestructuras integradas de investigación ha permitido la generación de varias propuestas (EGEE, DEISA) cuyo objetivo es dotar de una infraestructura para el desarrollo de aplicaciones GRID. Por otro lado otras propuestas específicamente centradas en la aplicación de GRID, como la propuesta de red de excelencia HEAVEN (HEALth GRID VENture) cuyo objetivo es la creación de una comunidad virtual de investigadores en el área de GRID aplicada a salud, permitirían, en caso de ser financiadas, la sinergia entre diferentes aplicaciones, infraestructuras y desarrollos.

7 Propuesta

A pesar de haberse realizado un gran esfuerzo durante los últimos años, la madurez del *middleware* actual es relativa. Todavía queda bastante camino por recorrer hasta contar con *middleware* que permita una explotación eficiente y sencilla de un entorno Grid. Proponemos las siguientes líneas de actuación:

1. Servicios locales
    - Herramientas de monitorización,
    - Gestores de colas batch
    - Librerías de programación paralela
    - Herramientas de depuración y monitorización de aplicaciones

2. Servicios Grid Básicos
    - Mejora de protocolos y servicios básicos

[1] La primera llamada del 2003 se cierra el 24/04/03 y la segunda el 15/10/03

3. Servicios Grid de Alto Nivel
    - Servicios de gestión de trabajos y datos para todo el conjunto de aplicaciones (servicios horizontales)
    - Servicios específicos para un dominio de aplicación (servicios verticales)
    - Uso de componentes

4. Herramientas Grid
    - Herramientas para aplicaciones de alta productividad
    - Herramientas para aplicaciones de alto rendimiento
    - Herramientas para aplicaciones adaptativas
    - Herramientas de ayuda al desarrollo de aplicaciones
    - Herramientas para aplicaciones intensivas en datos

Por otro lado, no debemos olvidar que es necesario estar en contacto con los proyectos orientados al desarrollo de aplicaciones en las diferentes áreas temáticas.