

Propuesta para la Creación de un Programa de e-Ciencia

Área Temática de Salud

Resumen

Este documento pretende fomentar la discusión en la preparación de la propuesta para el área temática de la salud en el programa de e-Ciencia de RedIRIS. El documento constituye un primer borrador de los puntos a destacar en la propuesta final.

1. Motivación de un Entorno GRID en el Área de la Salud

1.1. Situación de la TIC en el área de la salud

La informatización de los servicios asistenciales es un proceso complejo y lento pero inexorable. Numerosos servicios se encuentran en la actualidad total o parcialmente informatizados (administración, laboratorio, radiodiagnóstico,...) en gran parte de la sanidad pública y privada española. La necesidad de disponer de un sistema seguro, ágil, robusto y eficiente para el almacenamiento, proceso y transmisión de información relacionada con la salud es absolutamente aceptada por toda la comunidad usuaria, científica y empresarial.

Sin embargo, existen numerosas barreras que limitan la velocidad con la que se están adaptando los sistemas, más aún si se compara con otros procesos del mundo empresarial. La naturaleza multimedia de la información de la salud (imágenes, texto, señales, vídeos, procesos, etc.), la complejidad de su tratamiento (proceso de imágenes, análisis de señales, extracción de información en textos, etc.) y su dificultad de transmisión (numerosos formatos incompatibles, gran volumen, etc.) constituyen un factor limitador técnico importante. Más aún, otros factores, como la privacidad de la información, la disponibilidad de los recursos o el manejo del enorme volumen de datos históricos, suponen un freno aún mayor.

Las tecnologías GRID suponen una importante oportunidad para la solución de numerosos de estos problemas.

1.2. Justificación de GRID como solución

El mundo de la salud constituye un ejemplo perfecto para la implantación de un entorno GRID. La existencia de una comunidad virtual de personal médico proveniente de varios centros (asistencia primaria, especializada, urgencias) que cubren un área asistencial se adapta a la estructura GRID. La información de los pacientes se encuentra distribuida y su acceso y proceso como un conjunto global es absolutamente deseable. Además, el acceso y proceso al gran volumen de datos médicos puede ser abordable desde un conjunto de recursos GRID.

La seguridad, autenticación y fiabilidad que ofrece la arquitectura GRID responde a los requerimientos que el proceso de este tipo de información necesita.

1.3. e-Ciencia en Salud

La madurez actual de los entornos GRID es relativa. Si bien estos entornos se encuentran en un estado admisible para la comunidad científica, aún existen numerosos conceptos que deben solucionarse para el despliegue del GRID en salud a nivel generalizado.

Por una parte, se necesita adaptar las aplicaciones para que funcionen en un entorno GRID, con un énfasis especial en la seguridad y confidencialidad de la información. Muchas aplicaciones GRID en salud compartirán similares requerimientos, por lo que es importante definir una capa de 'GRID en salud' que ofrezca la funcionalidad común a las aplicaciones.

Además existen problemas de índole tecnológico, como los requerimientos de infraestructura local en los centros asistenciales que necesitan ser analizados detenidamente.

Por tanto, la aplicación de GRID en salud requiere de un importante esfuerzo investigador, y este proyecto pretende ofrecer el entorno adecuado para que éste se desarrolle.

2. Estudio de Necesidades y Proyectos por Regiones

La I+D+I de las TIC en el ámbito de la salud implica a una gran variedad de entes de naturaleza diversa, como centros tecnológicos, usuarios médicos y empresas proveedoras. Es por tanto importante que el ámbito de este proyecto trascienda de los centros de investigación y abarque a todos entes implicados en la medida de lo posible.

Los centros hospitalarios, usuarios de la tecnología, son a la vez grandes proveedores de contenido. Esto puede implicar que se requiera analizar en determinados casos la conexión entre las redes informáticas hospitalarias y Red IRIS. En el caso de redes hospitalarias públicas, existen importantes dotaciones de infraestructura muy apropiadas para este proyecto (p.e. red ARTERIAS de la Comunidad Valenciana). En el caso de redes hospitalarias privadas, existe una importante tendencia a la conexión de alta velocidad entre los centros pertenecientes a grupos asistenciales privados (como el grupo NISA).

El diseño, planificación e implementación de un Middleware básico para aplicaciones de GRID en salud requerirá disponer de gran experiencia en aplicaciones biomédicas y de aplicaciones GRID. Es necesario por tanto identificar centros y experiencias en el área biomédica que puedan proporcionar la información que se necesita, al mismo tiempo que coordinarse con otras iniciativas similares que puedan estar desplegándose en otros países.

2.1. Proyectos en el Área de la Salud con Interés en GRID

La comunidad investigadora de las TIC en salud es muy numerosa en España. Recientemente, y para coordinar las acciones de investigación, se han puesto en marcha diversas redes de excelencia bajo el Fondo de Investigación Sanitaria del Instituto de Salud Carlos III que pretenden facilitar la cooperación entre los centros especialistas en proceso de imágenes, telemedicina o tecnología sanitaria.

Además de estas acciones, existen numerosas iniciativas cuya cooperación puede proporcionar una gran cantidad de aplicaciones susceptibles de beneficiarse de las tecnologías GRID (aplicaciones computacionalmente intensivas, con grandes requerimientos de almacenamiento de datos, colaborativas, etc.).

2.2. Proyectos en GRID

La definición de un middleware de GRID en salud requiere, además de la experiencia relativa a las aplicaciones de las TIC en salud, el conocimiento de los middleware GRID genéricos. Los entornos actuales no están absolutamente maduros, y se espera que evolucionen en los próximos años notablemente. Este carácter dinámico de la tecnología básica debe ser tenido en cuenta para asegurar que los desarrollos que se realicen durante el proyecto sean de aplicación tanto al principio como al final del proyecto.

Este conocimiento requerirá del contacto con los comités de desarrollo y estandarización y los proyectos de middleware genéricos existentes.

2.3. Perspectivas de participación en el 6PM

Si bien el proyecto pretende la creación de un GRID en salud a nivel nacional, las características intrínsecas de la tecnología GRID permiten su extensión a áreas de aplicación más grandes, como el entorno europeo.

La apuesta que la Comisión Europea está realizando sobre las tecnologías GRID en el VI Programa Marco es notable. No sólo existen líneas prioritarias para proyectos específicos de GRID, sino que en muchas de otras áreas hay una recomendación del uso de estas tecnologías. Más precisamente, la línea prioritaria de 'e-Health' recomienda el uso de esta tecnología en la medida de lo posible.

Existen numerosas acciones en el VI Programa Marco cuya conexión parece necesaria. Existen varias propuestas de infraestructura (EGEE por ejemplo) cuyo objetivo es desplegar una infraestructura de GRID y avanzar en el desarrollo del middleware, con interés en las áreas de Física de Altas Energías y Salud. Por otro lado, existen otras propuestas de redes de excelencia en el área de GRID aplicado a la salud, como la Propuesta de Red de Excelencia 'HEALTH GRID VENTure' (HEAVEN). Es importante destacar que tanto las propuestas de infraestructura como en la red de excelencia HEAVEN tienen entre sus objetivos el contacto con desarrolladores de aplicaciones GRID en varias áreas, como la salud.

3. Definición Explícita de 'Use Cases'

3.1. Información Básica

El área de la Salud, como se ha comentado, constituye un área multidisciplinar en la que la información es de carácter multimedia y los procesos que sufre son variados y complejos. Es por tanto necesario antes de describir los casos de uso, el definir con una mínima precisión los tipos de datos y procesos que intervienen de forma genérica (y sin pretender ser exhaustivos) en el área.

3.1.1. Tipos de Datos

Los datos relativos a la salud pueden estructurarse en función del objetivo de las aplicaciones. Atendiendo a éste objetivo, se pueden clasificar en tres niveles: población, paciente y órgano. Si bien se consideran más niveles inferiores (celular y biomolecular), éstos se caracterizan por una problemática más específica y se tratarán en el área de bio-computación.



Datos a nivel población

El interés de los datos poblacionales es extraer conocimiento que corresponda de forma genérica a grandes grupos de pacientes, con el objetivo de diagnosticar, planificar una terapia o definir las características endémicas de una determinada población.

La información relevante a éste nivel la constituyen conjuntos que resumen o totalizan la información de los pacientes representativos. No son relevantes a este nivel los datos que caracterizan de forma única la información (datos personales), si bien lo son los datos que permiten su agrupamiento (área de residencia, sexo, rango de edad). A nivel médico, es relevante disponer de información sobre patologías y terapias, así como de los datos que han conducido a un diagnóstico (imágenes de radiodiagnóstico, señales vitales, análisis).

Datos a nivel paciente

La información relevante a nivel de paciente es la contenida en la historia clínica. La historia clínica informatizada constituye un gran reto en la gestión hospitalaria actual. Numerosos esfuerzos se han realizado en la interconexión de datos médicos (HL7, DICOM, Vital, etc.).

La información relativa a los pacientes adolece de los siguientes problemas:

- Se encuentra distribuida, fruto de la visita a lo largo de la vida de un paciente a diferentes centros de atención sanitaria.
- Es confidencial, por lo que su acceso debe estar restringido a nivel electrónico, al menos al mismo nivel que a nivel impreso.
- Tiene un gran volumen, por lo que su transferencia considerarse con cuidado para evitar la congestión del tráfico.
- Es multimedia, por lo que su almacenamiento requiere de un tratamiento especial.

Son datos a este nivel:

- Información demográfica: Datos personales generales y médicos del paciente.
- Información radiológica: Imágenes radiológicas, medicina nuclear.
- Historial clínico: Episodios, diagnósticos, tratamientos, alergias, intolerancias.
- Información analítica: Análisis de muestras, medidas bioeléctricas.

Datos a nivel órgano

La información relevante al nivel de órgano la constituye aquella que refleja el estado o caracteriza a un órgano o sistema funcional. Esta información puede ser parte de la información del paciente o bien constituir modelos y parámetros funcionales de órganos.

La información a nivel de órgano puede considerarse:

- Imágenes anatómicas y funcionales de órganos.
- Señales y muestras.
- Modelos computacionales del comportamiento funcional de órganos-

3.1.2. Tipos de Procesos

Dada la naturaleza de los datos que intervienen en el GRID en salud expuesta anteriormente, es importante describir los procesos que intervendrán en su gestión para poder definir más claramente los casos de uso que aparecen.

Procesos que requieren computación de altas prestaciones

La simulación de los modelos funcionales de órganos o los modelos de comportamiento de pacientes y poblaciones requiere de una gran cantidad de recursos dedicados a la resolución conjunta de un único problema. Ejemplos de este tipo de procesos son: la simulación multimodal de la actividad funcional de sistemas orgánicos, como el vascular o el respiratorio; la identificación de unidades anatómicas o funcionales en imágenes 3D; la identificación de arritmias y la localización precisa de los focos en señales cardíacas; etc.

Procesos que requieren de alta productividad

El carácter experimental de la investigación en salud requiere en muchos casos realizar numerosas simulaciones variando un conjunto de parámetros. De los resultados de la experimentación se ajustan experimentalmente tratamientos o diagnósticos.

El GRID puede proporcionar una gran cantidad de recursos para la ejecución eficiente, robusta y simultánea de diferentes experimentos.

Procesos que requieren el acceso a grandes volúmenes de datos

El proceso de las bases de datos médicas es importante tanto para la consolidación de la información perteneciente a poblaciones, la integración de toda la información perteneciente a un paciente concreto o la extracción de conocimiento a partir de técnicas de minería de datos. Es importante permitir la búsqueda de información a partir de conceptos complejos y no indexados, como similitud de imágenes, muestras, diagnósticos, etc.

Procesos de colaborativos o de comunicación igual a igual.

La información médica relativa a pacientes concretos se almacena de forma organizada en los centros asistenciales. Esta información está sujeta a una serie de normas que garantizan la confidencialidad e integridad de los datos.

Sin embargo, la investigación en cualquier área de salud requiere disponer de casos representativos de una determinada patología o colectivo. Estos casos pueden anonimizarse de forma que sea irreconocible su origen, y su compartición puede ser de gran utilidad a la comunidad científica. Si bien

estos casos podrían hacerse disponibles a partir de las grandes bases de datos distribuidas, es habitual que los facultativos dispongan en repositorios locales colecciones de datos seleccionados por su interés.

La posibilidad de compartir de manera eficiente y segura estos datos a través de Internet permitiría facilitar la colaboración entre especialistas y el acceso a una mayor cantidad de datos de calidad. La comunicación mediante arquitecturas P2P resulta especialmente apropiada, ya que se aumenta el número de réplicas, se reducen los tiempos de descarga y se trabaja sobre datos preseleccionados.

4. Evaluación del Posible Uso de Middleware Común

La tecnología GRID básica en la que se soporta el middleware específico de la salud deberá elegirse teniendo en cuenta su eficiencia y aceptación. La tecnología GRID todavía no se encuentra en su total madurez y por tanto la capa de servicios específica de la salud deberá ser lo suficientemente independiente para poder adaptarse a cambios en la estructura básica del Middleware.

Actualmente, OGSA (Open Grid Service Architecture) proporciona una estructura suficientemente conceptualizada. La decisión deberá tomarse teniendo en cuenta las iniciativas europeas con las que se pretenda establecer conexiones.

5. Desarrollo Previsto de Middleware Específico del Área

La puesta en marcha de aplicaciones piloto permitirá identificar requerimientos, componentes y procesos que sean comunes a gran número de aplicaciones de GRID en salud. Estos componentes y requerimientos constituyen la capa intermedia que permitirá desarrollar más rápidamente las aplicaciones GRID en el área de la salud.

Entre estos componentes, se encontrarán:

- Módulos de interfaz, capaces de cargar, transmitir y convertir los datos provenientes de diferentes fuentes en diferentes formatos.
- Módulos de proceso de imágenes, capaces del filtrado, segmentación, proyección 3D, reformateo de planos, etc.
- Módulos de cálculo intensivo, capaces de resolver problemas numéricos específicos, como análisis mecánicos de prótesis, dinámica de fluidos, transmisión de potenciales eléctricos, etc.

La ilustración 1 muestra la relación entre las aplicaciones en salud y el GRID genérico. Esta ilustración muestra como ejemplo tres potenciales componentes de esta capa intermedia.

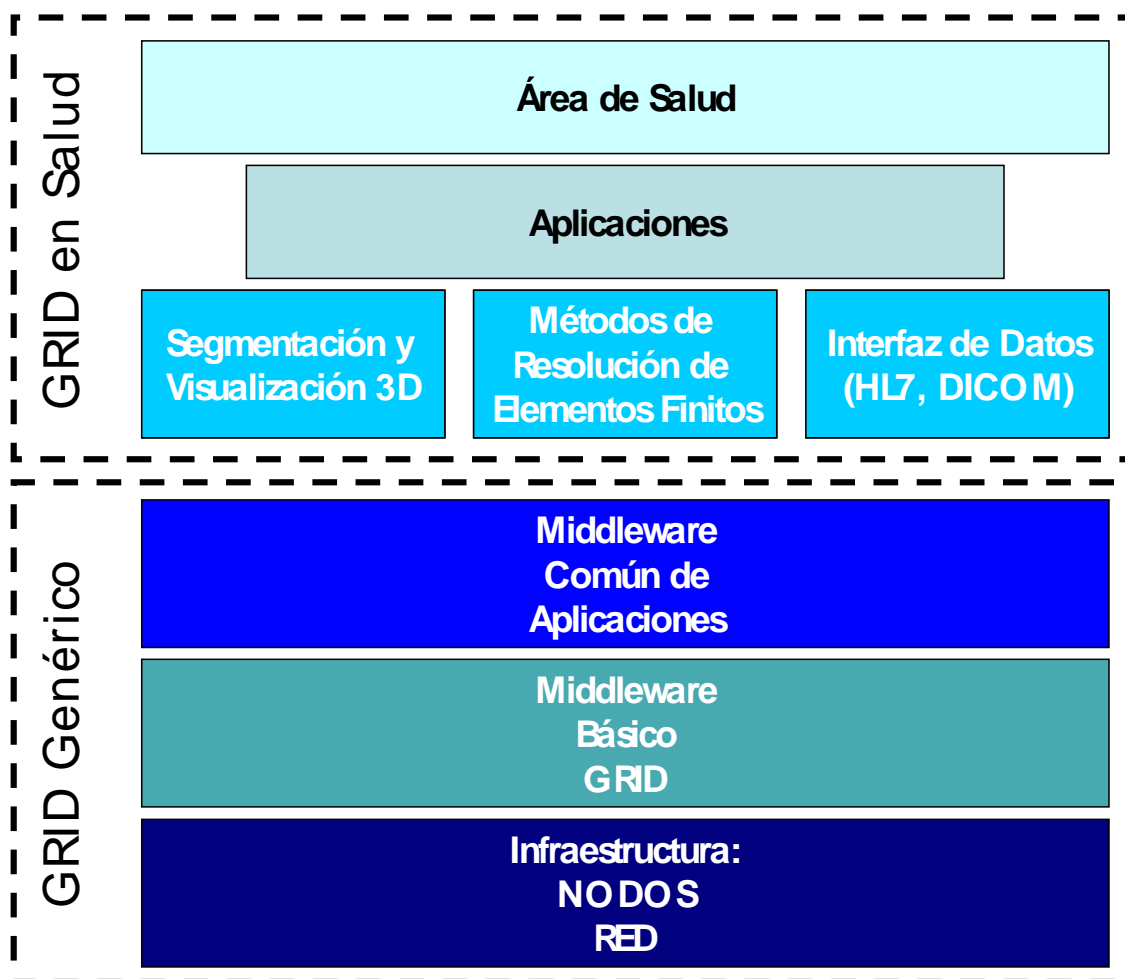


Ilustración 1: Esquema de un GRID en salud.

6. Recursos de Infraestructura y Humanos Disponibles, Dimensión

Un programa de e-Ciencia en salud es intrínsecamente multidisciplinar, tanto en lo relativo a personal (personal médico, ingeniería, informática, telecomunicaciones, químico, etc.) como en lo relativo a los centros (centros de investigación, laboratorios, hospitales, etc.).

Esta característica, unida a los requerimientos técnicos de un programa de e-Ciencia, conlleva a la necesidad de infraestructuras de comunes entre centros de investigación y proveedores de información (hospitales principalmente). La estructura GRID de cómputo puede ubicarse en centros de investigación y de servicio, pero la conexión de los centros proveedores de información con el entorno GRID debe permitir grandes transferencias de datos. Los requerimientos variarán en función de los proyectos piloto que se quieran implantar.

7. Definición de Posibles Proyectos Piloto

El proyecto pretende definir una capa de servicios genéricos para las aplicaciones GRID en salud. Esta capa se construirá a partir del análisis de requerimientos de aplicaciones y estándares existentes o en desarrollo identificados durante la etapa exploratoria del proyecto. A partir del análisis de estas aplicaciones se determinarán servicios, componentes, interfaces y

formatos que darán lugar a la capa mencionada. Sobre esta capa se desarrollarán experiencias piloto que demuestren los beneficios de la tecnología y de los resultados del proyecto.

Potencialmente, las aplicaciones piloto que se desarrollarán abordarán:

- Proceso de imágenes: Identificación de órganos y entidades funcionales, proyección 3D, generación de modelos, herramientas de diagnóstico asistido por computador.
- Simulación funcional de órganos y sistemas: Dinámica de fluidos, transporte de potenciales eléctricos, deformaciones visco-elásticas, análisis estructural.
- Acceso inteligente a grandes volúmenes de datos: Extracción de conocimiento y minería de datos sobre bases de datos poblacionales, búsqueda de patrones en imágenes y señales, localización de similitudes en diagnósticos y pruebas.
- Trabajo colaborativo: Compartición segura de datos médicos, videoconferencia, asistencia remota.

8. Transferencia de Tecnología, Visibilidad y Difusión de Proyectos

El proyecto persigue la puesta en marcha de un programa de e-Ciencia, pero con el objetivo de que este entorno sea beneficioso tanto para la comunidad científica como para el mundo empresarial.

La disponibilidad de un sistema de cómputo con una potencia pico del orden del Teraflop y una capacidad de almacenamiento de varios Terabytes permitirá abordar problemas tipo gran reto que actualmente son inabordable, máxime para los centros que no disponen de suficiente infraestructura para abordar problemas de tamaño medio.

Por tanto, es de especial interés para mantener la viabilidad del proyecto una vez terminada la duración propuesta, el disponer de un gran número de usuarios y clientes que puedan justificar la continuación de las operaciones. Más aún, el acceso a gran escala justifica la participación de empresas que puedan beneficiarse de este concepto tanto como usuarios como servidores específicos.

Para conseguir un gran impacto y un gran número de usuarios se plantea la difusión de resultados a tres niveles:

- Foros científicos. Atendidos principalmente por investigadores que desarrollan nuevas aplicaciones en el área de la salud, con el objetivo de atraer nuevas aplicaciones al área GRID.
- Foros de salud. Atendidos principalmente por usuarios (facultativos, enfermería, gerencia, etc.), pretenden fomentar el uso de las aplicaciones GRID, a la par de involucrar a los centros usuarios en el entorno GRID como proveedores de contenido.
- Foros empresariales y tecnológicos. Atendidos principalmente por empresas, con el objetivo de desarrollar y experimentar modelos de negocio que puedan beneficiarse de la tecnología GRID.

La difusión consistirá en:

- Elaboración de una imagen corporativa acorde con el proyecto global.
- Preparación de material de difusión impreso y electrónico.

- Mantenimiento de un sitio Web con la información actualizada, disponiendo de áreas públicas y privadas.
- Organización de presentaciones y demostraciones en eventos.