

# Integración de Datos Mediante Tecnologías Grid

*Ignacio Blanquer, Vicente Hernández*  
*GRyCAP*

- **Introducir el Concepto de Grid de Datos y su Aplicación como una Tecnología Integradora.**
- **Describir las Diferentes Aplicaciones de Estas Tecnologías en el Ámbito de los Datos Médicos.**
- **Destacar las Ventajas y Características Específicas de los Grids de Datos en Salud.**
- **Comentar Algunas de las Soluciones Actualmente en Desarrollo.**

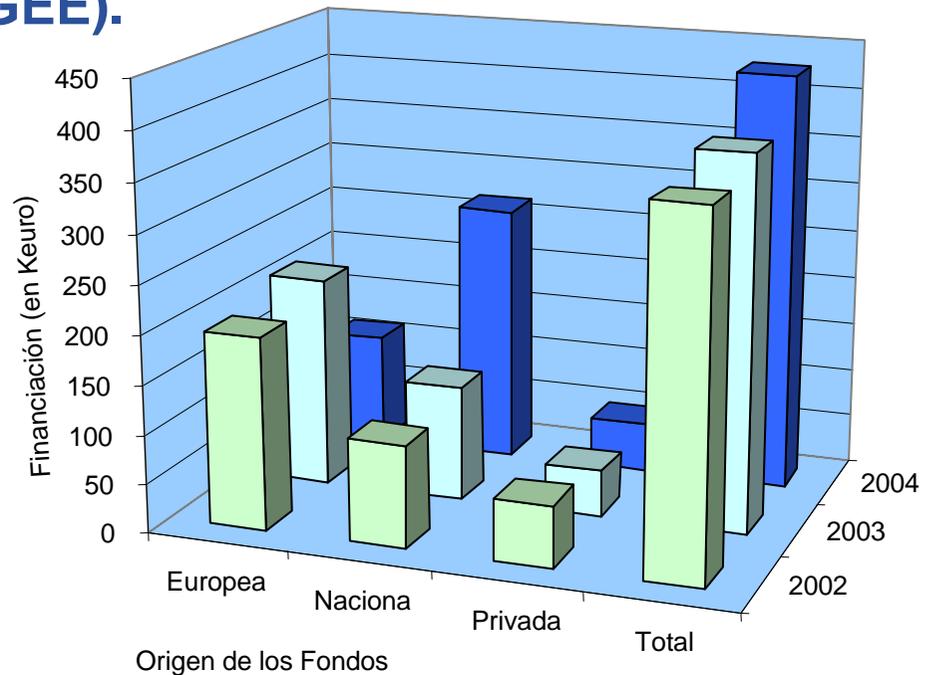


- **El GRyCAP.**
- **Conceptos de Grid, DataGrids y HealthGrids.**
- **Aplicaciones del Grid en el Almacenamiento de Datos Médicos.**
- **Beneficios y Conclusiones.**



- Grupo de Investigación con 20 Años de Actividad y Compuesto por 30 Investigadores.
- Con Experiencia en la Participación en más de 50 Proyectos (17 Europeos, Siendo V. Hernández el Project Manager en 10 de Ellos).
- 7 de Ellos en el Sector Médico (HIPERCIR, DISMEDI, VRSUR, IDEAS, EUTIST-M, TT@MED, EGEE).
- El GRyCAP se Encuentra Integrado en el Instituto de Aplicaciones de las Tecnologías de la Información y Comunicaciones Avanzadas (ITACA) y en el Centro en Red de Ingeniería Biomédica (CRIB).

[www.grycap.upv.es](http://www.grycap.upv.es)



- **El Grid es una Infraestructura que Permite que Comunidades Virtuales Compartan Recursos Heterogéneos y Distribuidos con el Objetivo de Resolver un Problema Común Difícilmente Abordable Mediante Otras Tecnologías y Sin Comprometer la Independencia en la Gestión de los Recursos Individuales.**
  - El Grid es una Arquitectura Software (Middleware) y una Serie de Componentes que Proporcionan Altas Prestaciones, Alta Disponibilidad, Tolerancia a Fallos, Seguridad e Interoperabilidad de Datos y Computadores.
- **Los Grids de Datos (DataGrids) Consideran como el Recurso a Compartir el Espacio de Almacenamiento y los Datos.**

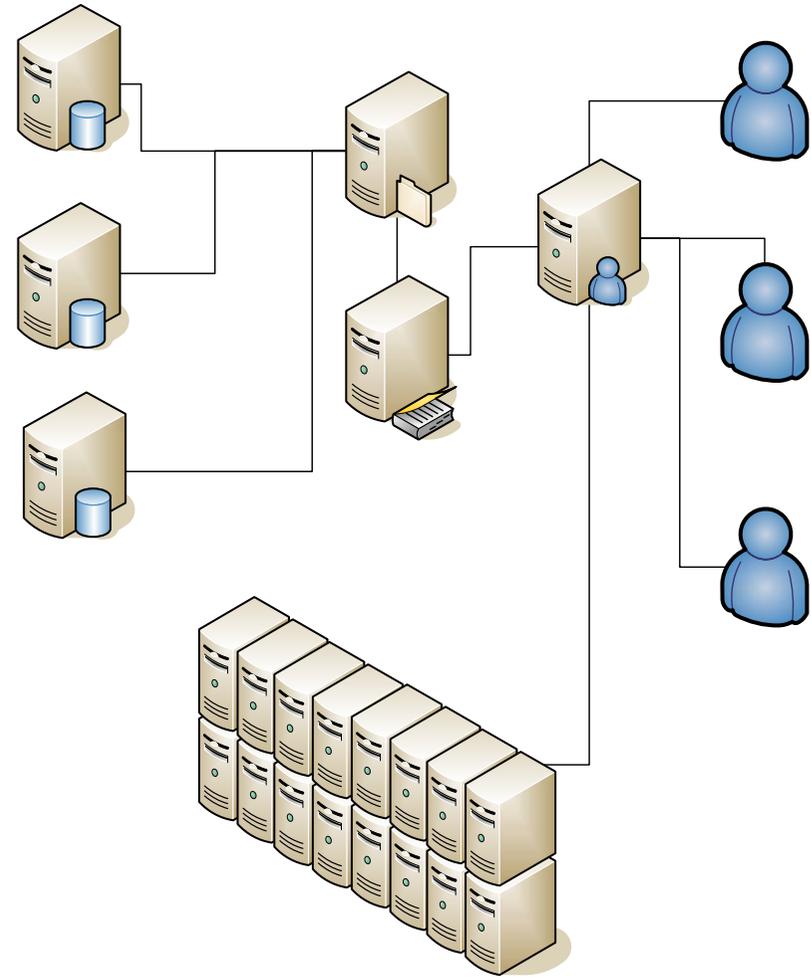


- **Los HealthGrids son Servicios Grid o Componentes Middleware que Resuelven Problemas Específicos del Proceso de Datos Médicos.**
  - Los Recursos en los HealthGrids son Bases de Datos, Capacidad de Proceso, e Incluso Experiencia Médica o Equipos Médicos.
  - En la Actualidad se Están Utilizando para la Asistencia en Epidemiología y en Tratamiento Individualizado del Paciente:
    - Integración de Grandes Bases de Datos y Extracción de Conocimiento.
    - Ejecución de Simulaciones Complejas de Modelos Biomédicos.
    - Procesamiento e Integración de Bases de Datos Genómicas y Proteómicas.
    - Creación de Entornos Colaborativos para Compartir Datos y Herramientas en Investigación.



- **Las Aplicaciones de los Grids de Datos Pueden Considerarse:**
  - Creación de un Repositorio Distribuido para el Almacenamiento Masivo de Datos Médicos.
    - Apropiado para el Almacenamiento de Datos Biomédicos (Signos Vitales, Imágenes, Información Biomolecular).
    - Especialmente Interesante en el Almacenamiento a Largo Plazo.
    - Se Aprovechan los Recursos de Almacenamiento de un Conjunto de Sistemas.
  - Integración de Repositorios Distribuidos.
    - Apropiado Cuando las Bases de Datos ya Existen.
    - Especialmente Interesante para la Extracción de Conjuntos de Datos Relevantes que Requieran un Postproceso.
    - Se Puede Combinar el Uso de Recursos de Procesamiento Intensivo.

- **Gestión de las Organizaciones Virtuales y Seguridad**
  - Organiza Usuarios, Grupos, Permisos de Acceso, Autenticación y Autorización.
- **Gestión del Almacenamiento**
  - Organiza el Almacenamiento y el Acceso a Datos y Recursos Distribuidos, Además de la Catalogación, Replicación y Control del Acceso.
- **Gestión de la Carga de Proceso**
  - Proporciona el Balance de Carga y la Selección de los Recursos Compatibles y Adecuados.
- **Sistema de Información y Accounting**
  - Permite Organizar y Descubrir los Recursos Disponibles en el Sistema, Así como Registrar su Uso.



- **Gestión de las Organizaciones Virtuales y Seguridad.**
  - El Grid No Requiere la Creación de Cuentas Individuales para Cada Usuario en Cada Recurso.
  - Los Usuarios se Identifican a Través de un “User Interface” Mediante Certificados X.509 (EU-GridPMA) y la Creación de un “Proxy” que Representa las Credenciales del Usuario en Cada Recurso (“Single Sign-on”).
  - La Autorización en el Acceso de los Recursos Se Realiza Mediante Listas de Usuarios Gestionados de Forma Centralizada (Actualizada de Forma Periódica Incluyendo Listas de Revocación de Permisos).
  - Los Permisos Se Otorgan a Nivel de Organización Virtual, Aunque los Esquemas Actuales (VOMS) Permiten Definir Subgrupos y Asignar Permisos (Incluso en el Acceso a Datos) Diferentes para Cada Grupo.
  - Las Transacciones se Realizan Sobre Protocolos Seguros Mediante el Uso de Certificados.



- **Gestión del Almacenamiento**
  - Diferentes Orientaciones
    - Almacenamiento Masivo
      - *Los Grids Permiten Utilizar de Forma Cooperativa el Espacio de Almacenamiento de los Diferentes Recursos del Grid.*
      - *Este Espacio de Almacenamiento Se Gestiona de Forma Global en la Forma de un Gran Directorio Virtual.*
    - Integración de Repositorios Individuales
      - *Los Grids Permiten Incluir Interfaces sobre los Catálogos Individuales Para Construir una Gran Base de Datos Virtual.*
  - Los Datos se Gestionan por Organización Virtual, Incluyendo la Gestión de Permisos a Grupos de Usuarios Individuales.
  - Los Grids Permiten la Creación de Réplicas de los Datos Para Aumentar la Eficiencia en el Acceso y Proceso de los Datos.
  - La Gestión de los Datos se Realiza Mediante Catálogos que Contienen Información de los Identificadores, Réplicas y Ubicación Física de Todas las Réplicas.

- **Gestión de la Carga de Proceso**

- Una Importante Aportación de los Grids es el Uso Cooperativo de los Recursos de Proceso de Datos.
- Los Recursos de Proceso se Gestionan de Forma Semi-Centralizada sin Renunciar a la Autonomía de Cada Centro, Accediendo a Través de “Front-Ends” que se Integran en un Nivel Superior.
- Existen Gestores de Carga que Deciden, Sobre los Recursos Disponibles, Cuál es el Más Adecuado Atendiendo a Factores Como Requerimientos, la Proximidad de los Datos o la Carga del Recurso.
- Los Recursos se Pueden Compartir para Varias Organizaciones Virtuales y Ser Accedidos por Diferentes Gestores de Recursos, (Incluso Varias Infraestructuras Grid).
- Adicionalmente a la Gestión Individual, Existen Herramientas de Control del Flujo (TABERNA) que Permiten una Gestión a Mayor Nivel.

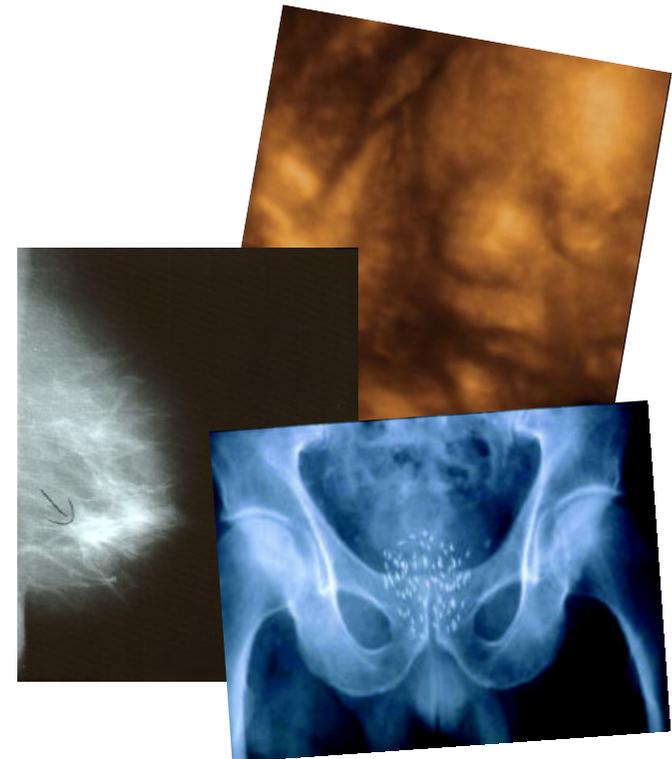


- **Monitorización y Accounting**

- La Selección de los Recursos Más Adecuados para el Proceso o el Acceso a Datos Requiere Conocer el Estado Actual de cada uno de los Procesos.
- La Introducción de Nuevos Recursos Debe Ser lo Más Transparente Posible para los Usuarios y los Administradores.
- Los Diferentes Recursos Publican la Información Necesaria Para Su Descubrimiento y Gestión en un Esquema Jerárquico.
- Además los Recursos Registran Todas las Operaciones Realizadas por los Usuarios Individuales (Seguridad o Explotación).
- Tradicionalmente se ha Utilizado una Jerarquía de Servidores LDAP (GT-MDS), Aunque en la Actualidad se Están Sustituyendo por Sistemas Basados en Bases de Datos Relacionales (R-GMA).

# Proyectos y Ejemplos de la Utilización de Grids en el Ámbito Médico

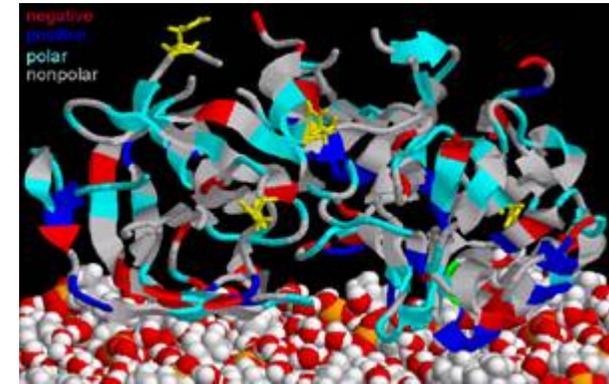
- **Áreas de Aplicación de los Grids en Salud**
  - Imagen Médica
    - Federación de Archivos (BIRN, GRID-IT, EMBRACE).
    - Mamografía (NDMA, MAMMOGRID, eDIAMOND, GPCALMA).
    - Proceso de Imágenes (DataGrid, EGEE-NA4).
  - Terapia
    - Simulación de Radioterapia (GATE, GEMSS-RADPT).
    - Simulación Absorción de Fármacos (GEMSS-Cophit).



- **Áreas de Aplicación de los Grids en Salud**

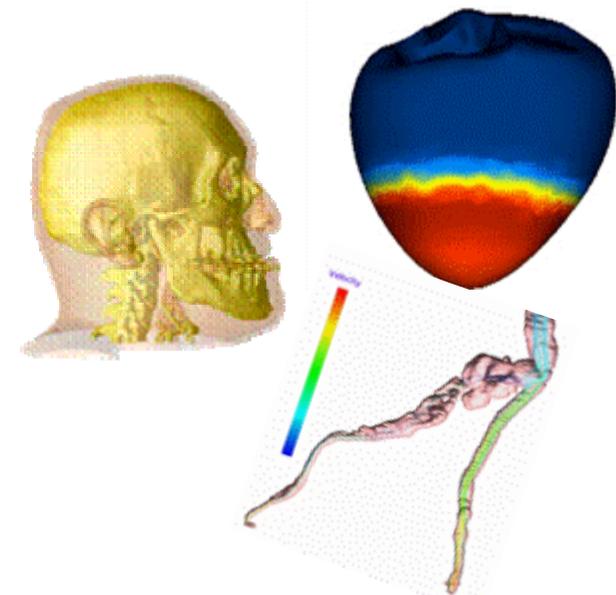
- **Bio-computación**

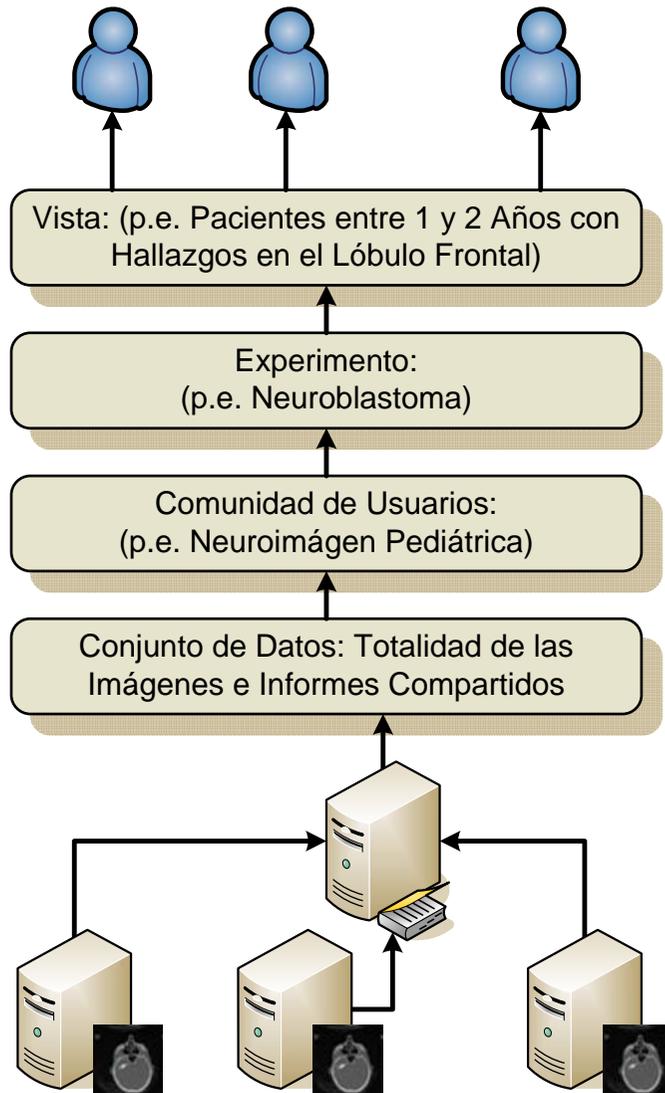
- Simulación de Dinámica de Proteínas y Ontologías Genéticas (BioGRID, GPS@, INFOGENMED, GenoGrid, myGRID...)..



- **Simulación Biomédica y Planificación Quirúrgica**

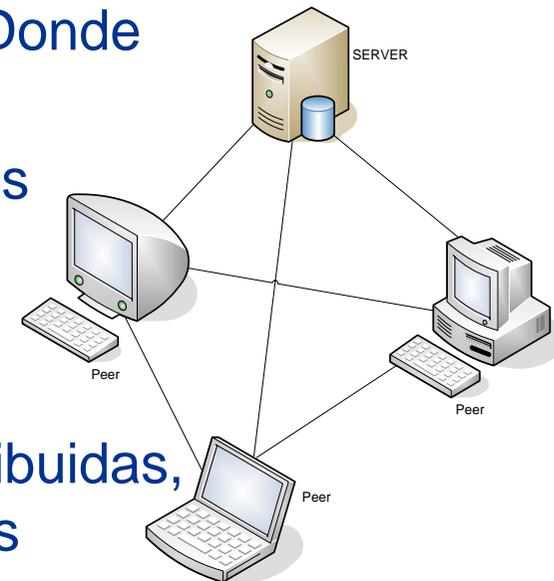
- Vascular (CROSSGRID, GEMSS-BloodSim).
- Maxilo-Facial (GEMSS).
- Actividad Cardíaca (gCAMAEC).



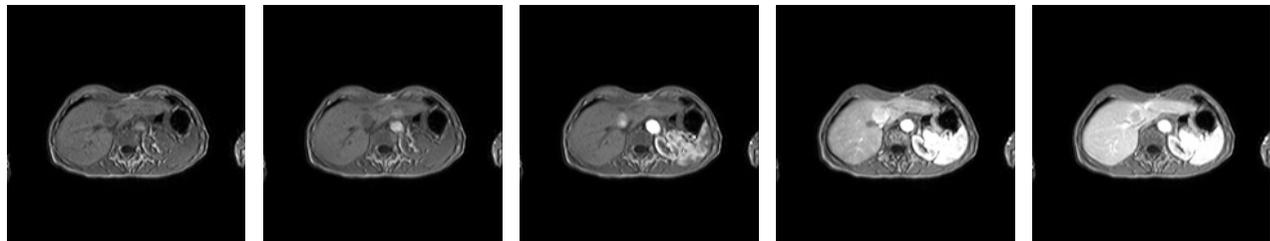
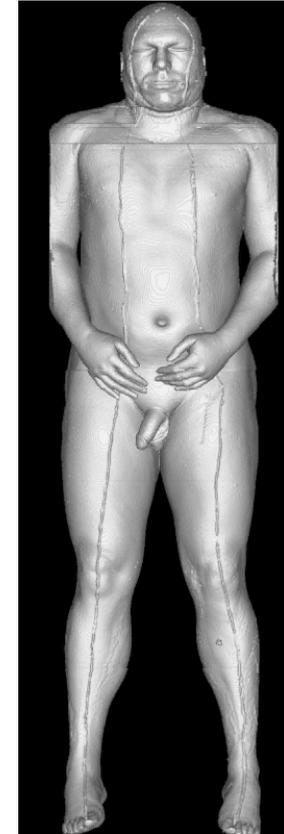


- Virtualiza un Repositorio Global Compuesto por Almacenes Individuales.
- Creación de Diferentes Niveles en los que Se Organizan Datos y Comunidades Virtuales.
- Definición de Ontologías y Esquemas que Permitan Indexar los Datos Semánticamente.
- Integración de Elementos de Pre- y Post- Proceso.
- Seguridad, Pseudo-Anonimización y Soporte a la Protección de la Privacidad.
- Sin Alteración de los Procesos Normales de la Práctica Clínica.

- **Un Entorno Peer-to-Peer para Compartir Imágenes Médicas y Diagnósticos Proporcionando Búsquedas por Contenido.**
  - El Objetivo es Facilitar la Recogida de Casos Relativos a Patologías Infrecuentes o Para Completar una Muestra Suficientemente Grande Para Estudios Epidemiológicos.
  - Aunque estos Estudios se Encuentran Disponibles en los Diferentes Centros, no Existe un Entorno Donde se Pueda Acceder de Forma Integral.
  - La Herramienta Permite Compartir Estudios de Manera Individual, Uniendo Recursos Personales.
  - Implementa Autenticación, Anonimización Automática de los Datos, Búsquedas Distribuidas, Transmisión Progresiva y Valoración de los Usuarios y Diagnósticos.



- **Renderizado de Imágenes Médicas de Muy Gran Dimensión Mediante Grids Corporativos**
  - Dataset del Visible Human Project (1500 Cortes).
  - La Reconstrucción Volumétrica Tiene un Elevado Consumo de CPU, pero También de Memoria.
  - La Tecnología Grid Permite Compartir los Recursos de los Computadores Conectados a una Red (Memoria y Tiempo de Procesamiento).
- **Corregistración en Grid**
  - Almacenamiento Distribuido y la Corregistración Mediante Métodos Elásticos de Secuencias de MRI del Abdomen.
  - Orientado al Estudio de la Farmacocinética.
  - Reducción en un Factor de 12 del Tiempo de Proceso.



- **EGEE “Enabling Grids for E-Science”**

- Crear una Infraestructura Grid en Producción Eficiente y de Calidad a Nivel Europeo.
- Actualmente Engloba 10000 Recursos de más de 27 países (Incluyendo Asia-Pacífico y Norteamérica). Más de 400 Computadores se Dedicar a Biomedicina.
- Orientado al Despliegue de Aplicaciones en Producción para el Soporte a la e-Ciencia en Diferentes Áreas
  - Física de Altas Energías.
  - Biomedicina.
  - Otras: Geofísica, Meteorología, Astronomía.



- **Miembros de la Asociación HealthGrid**

- Vicente Hernández es uno de los Fundadores.
- El Objetivo de la Asociación es la Integración de Grupos e Investigadores en Grid con Interés en la Aplicación de estas Tecnologías sobre la Salud.
- Cuenta con 50 Socios y ha sido la Promotora del Libro Blanco sobre Grid para Salud, Encargado por la EC.
- La UPV ha Sido la Encargada de la Elaboración del Primer Capítulo de Dicho Libro.



- **Las Tecnologías Grid Permiten el Almacenamiento y Proceso de Datos a Gran Escala Utilizando Recursos Estándar ('Comodity').**
- **En el Ámbito Médico Existen Numerosos Ejemplos Altamente Satisfactorios en el Uso de Estas Herramientas para Crear Bases de Datos Federadas.**
- **Además, el Grid Proporciona la Capacidad de Realizar Procesamiento Complejo e Intensivo.**
- **Las Áreas de Aplicación más Extendidas son Imagen Médica y Genómica, Aunque Existe un Gran Potencial en Otras Áreas.**



**Ignacio Blanquer**

**Universidad Politécnica de Valencia**

**Camino de Vera s/n**

**46022 Valencia, Spain**

**Tel: +34-963877356**

**Fax. +34-963877359**

**E-mail: [iblanque@dsic.upv.es](mailto:iblanque@dsic.upv.es)**

