



# Comunicación

# 240

## **TIC EN INVESTIGACIÓN BIOMÉDICA: LA PLATAFORMA DE INTEGRACIÓN Y ANÁLISIS DE DATOS INBIOMED**

### **Fernando Martín Sánchez**

Jefe del Área de Investigación en Bioinformática  
Instituto de Salud Carlos III

### **Santiago Jiménez de Ory**

Área de Investigación en Bioinformática  
Instituto de Salud Carlos III

---

## Palabras clave

*TIC en Biomedicina, Informática Biomédica, Servicios Web, Integración de datos, workflows, ontologías.*

## Resumen de su Comunicación

*La integración de información derivada del Proyecto Genoma Humano junto con los datos de las historias clínicas electrónicas de los pacientes, proporciona nuevas oportunidades para el avance de la medicina. Las TIC juegan un papel central en la investigación biomédica y están llamadas a contribuir decisivamente a mejorar la prevención, diagnóstico y tratamiento de las enfermedades en un futuro cercano.*

*Como ejemplo de iniciativa TIC en este campo realizada en la Administración española se describe la Red Temática de Investigación Cooperativa en Informática Biomédica INBIOMED. Esta red ha sido financiada (2003-2005) por el Ministerio de Sanidad y Consumo y fue coordinada por los firmantes de esta comunicación. En este período 13 hospitales, universidades y centros públicos de investigación (más de 100 técnicos e investigadores) han desarrollado un sistema de almacenamiento, integración y análisis de datos clínicos, genéticos y ambientales que va a facilitar la investigación de enfermedades complejas.*

*La plataforma INBIOMED consiste en una infraestructura de información distribuida, en la que se han empleado estándares léxicos y tecnológicos, para hacer posible la integración de los datos de los distintos grupos de investigación. El sistema elegido para dar soporte a este modelo es un almacén de datos federado a través de servicios. La arquitectura de la plataforma está compuesta por cuatro niveles: aplicaciones cliente, servicios, almacén de datos y fuentes de datos. La comunicación de las aplicaciones con el sistema se realiza a través de servicios Web utilizando el protocolo SOAP. En la comunicación se describe con detalle la arquitectura y las principales aplicaciones que ya están accesibles desde esta plataforma.*

# TIC EN INVESTIGACIÓN BIOMÉDICA: LA PLATAFORMA DE INTEGRACIÓN Y ANÁLISIS DE DATOS INBIOMED

## 1. Introducción

La secuenciación del genoma humano supone una gran oportunidad para el desarrollo de nuevas disciplinas que permiten avanzar en el conocimiento de las causas de las enfermedades. Se está constatando un interés creciente en el desarrollo y aplicación de herramientas y métodos de tratamiento de la información genética que pueden aplicarse en medicina.

Gran parte de las enfermedades se producen como respuesta a factores genéticos y/o factores ambientales. Por lo tanto, para comprender en su totalidad el desarrollo de una patología, debemos conocer no sólo las características genéticas del paciente, sino también los factores ambientales a los que está expuesto y sus hábitos de vida [1].

Los análisis genéticos a los pacientes serán más frecuentes en un futuro cercano, y la información contenida en los mismos, combinada con los datos de las historias clínicas digitales, permitirán la realización de estudios que hagan posible la Medicina Genómica [2]. Esta nueva disciplina, donde se integra información proveniente de múltiples recursos, permitirá desarrollar terapias y tratamientos personalizados para grupos de pacientes con características genéticas comunes.

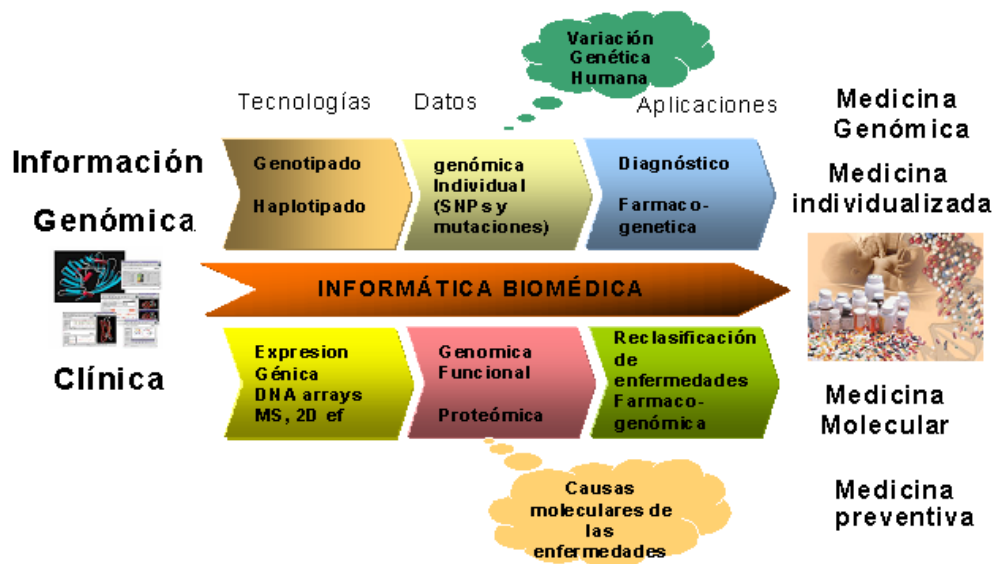


Figura 1: Medicina Genómica

La Biomedicina comprende todas las áreas que estudian los procesos biológicos, la medicina clínica y la salud pública. Tras la secuenciación del genoma humano, esta área se ha visto influenciada por los nuevos conocimientos que se ven traducidos en una mejor caracterización de las enfermedades y en el desarrollo de diagnósticos, tratamientos y programas de prevención adaptados a cada paciente o población.

El volumen de la información a manejar proveniente de todos estos ámbitos puede llegar a ser intratable por las técnicas clásicas que se utilizaban hasta el momento. La informática puede dotar a la comunidad científica Biomédica de entornos más productivos, desarrollando herramientas que faciliten y optimicen las investigación realizada. Además, la gran cantidad de datos a almacenar y su visualización de forma comprensible para los investigadores biomédicos suponen un reto para la informática que muchos están dispuestos a asumir.

Las TIC en Biomedicina abordan la adquisición, almacenamiento, integración, y análisis de datos biomédicos. Tratan de proporcionar los modelos, métodos y herramientas informáticas que permitan la integración de la información genética, clínica y sobre factores ambientales para contribuir decisivamente a mejorar la prevención, diagnóstico y tratamiento de enfermedades en los próximos años.

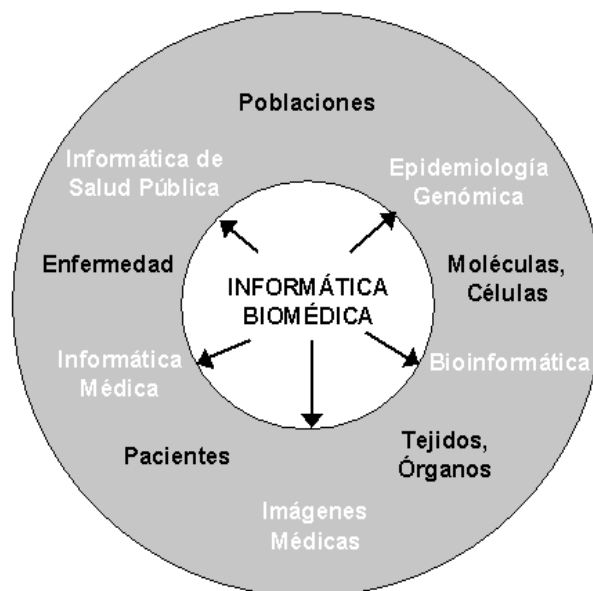


Figura 2: Las TIC en Biomedicina como nexo de unión entre disciplinas

## 2. Perspectiva Internacional

Durante los últimos años, el auge de las TIC en biomedicina ha sido imparable a nivel internacional. Cada año surgen nuevas iniciativas, tanto en Europa como en Estados Unidos, para financiar la investigación en informática biomédica, ya que ha sido señalada como una de las áreas más prometedoras a corto plazo, con una aplicación práctica muy importante en el tratamiento y diagnóstico de enfermedades.

En Europa, la financiación destinada a TIC en Biomedicina ha ido en aumento en los últimos Programas Marco de I+D. Asimismo, las áreas a las que ha ido destinada esta financiación se han ido definiendo y concretando a lo largo del tiempo. Mientras en las primeras convocatorias los proyectos financiados se centraban en el desarrollo de software para personal clínico y servicios y sistemas de telemedicina, en los Programas Marco 4º y 5º la financiación se ha volcado en proyectos de redes y sistemas de información en salud.

Dentro de las iniciativas europeas financiadas por los Programas Marco, el Área de Bioinformática del Instituto de Salud Carlos III ha participado en diversos proyectos. BIOINFOMED [3,5] e INFOGENMED [4] se

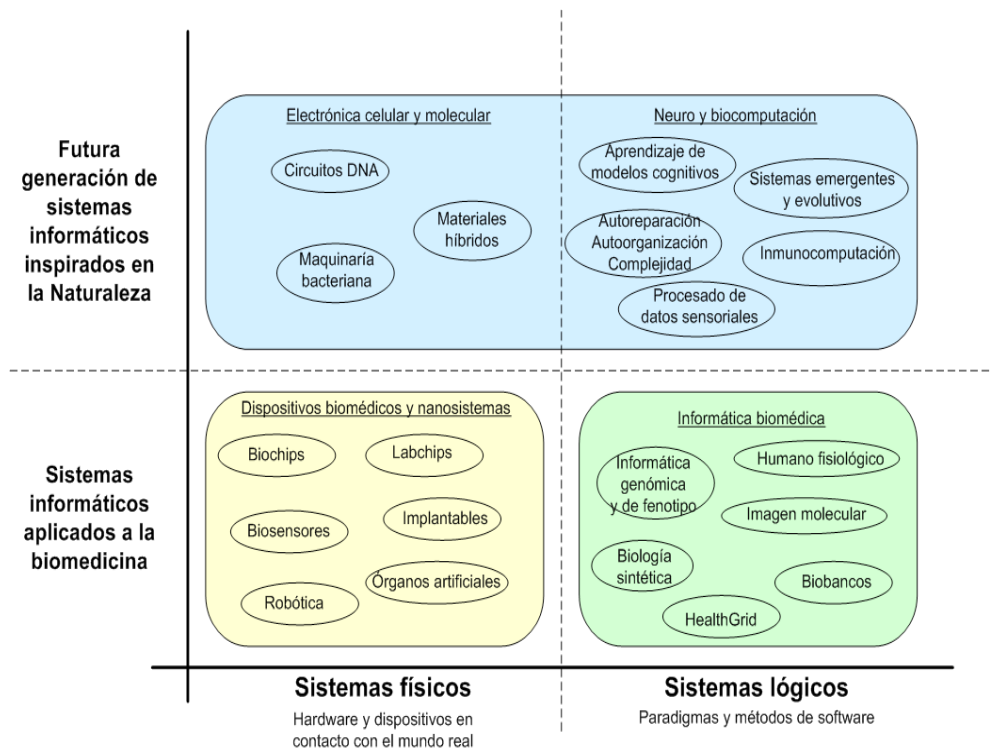
financiaron dentro del 5º Programa Marco y han finalizado. El proyecto europeo INFOGENMED [6], diseñó un entorno para acceder e integrar información médica y genética con la intención de mejorar la práctica clínica.

En el 6º Programa Marco las actividades financiadas en las que participa el Área de Bioinformática y Salud Pública son diversas. INFOBIOMED [7] es una Red de Excelencia que trata de crear una comunidad de informática biomédica estable en Europa y desarrollar aplicaciones pilotos en áreas tan diversas como la farmacoinformática, la genómica y la microbiología, la genómica y la inflamación crónica y la genómica y el cáncer de colon. Dentro de esta red trabajan 16 nodos pertenecientes tanto a centros de investigación públicos como privados, universidades y empresas de 10 países de la Unión Europea [8].

SYMBIOmatics [9], proyecto también financiado dentro del 6º Programa Marco, pretende documentar el estado del arte de las TIC en Biomedicina en Europa para identificar las líneas de trabajo futuras a explorar. Está previsto que para finales de 2006 se presenten todos estos resultados en la redacción de un Libro Blanco, con el fin de potenciar la investigación en esta área.

El propósito del último proyecto financiado en el que participa el Área de Bioinformática y Salud Pública, BEYOND-THE-HORIZON [10], es proporcionar un visión de las tendencias emergentes en las TIC. En concreto, se está coordinando el área temática "BIO-ICT Synergies".

Ante la próxima convocatoria del 7º Programa Marco, el Área de Bioinformática y Salud Pública está colaborando en diversas actividades prospectivas. Una de las más prometedoras es el workshop "Extending the scope of IST at the Crossroads with Life Sciences", que tuvo e lugar en 2004 y definió cómo las TIC pueden converger con la investigación y aplicación de la ciencias de la vida. Esta tormenta de ideas se concretó en un informe cuyos principales resultados puede verse en la Figura 3.



**Figura 3:** Desarrollos para la informática biomédica

A nivel internacional una de las iniciativas más importantes para la investigación en TIC en Biomedicina son los National Centres for Biomedical Computing financiados por el National Institute of Health de Estados Unidos.

### 3. Retos para las TIC

Como resultado de los trabajos e investigaciones llevadas a cabo en el área de TIC en biomedicina surgen diversos retos va a concentrar la mayor parte de los esfuerzos en los próximos años.

La gran cantidad de datos con los que se trabaja y que se intenta analizar y estudiar hace necesarias estructuras de computación suficientemente potentes. Esta necesidad de manejar gran cantidad de datos ha supuesto la irrupción con fuerza de la computación GRID. Este tipo de computación proporciona un entorno de información distribuido, seguro, potente e interoperable, que permite analizar los datos de forma fiable y mucho más rápidamente que con la computación "tradicional". Tal es la importancia del GRID en la investigación biomédica que han aparecido iniciativas como HealthGrid [11], dedicada a GRID en biomedicina. Esta asociación organiza cursos, workshops y congresos en los que se exponen los avances en la materia y se forma a nuevos investigadores.

Los biobancos [12] representan otra iniciativa que ha surgido en la investigación biomédica. La idea de los biobancos es realizar estudios poblacionales para determinar la influencia de factores ambientales y genéticos en el desarrollo de las enfermedades. Por lo tanto, se requiere un estudio pormenorizado de genes, datos clínicos y exposición a factores ambientales. El análisis de estos datos permitirá comprender las distintas interacciones entre los factores genéticos y ambientales en algunas enfermedades. Además, toda la información resultante facilitará la toma de decisiones clínicas, tanto respecto a la salud pública como al diagnóstico y tratamiento de cada paciente.

La tecnología debe enfrentarse también al procesamiento de las imágenes producidas por la investigación biomédica. Estas imágenes tienen una características que hacen necesario desarrollar nuevas herramientas tecnológicas para su interpretación y análisis. El nuevo reto para el área del procesamiento de imágenes en biomedicina es el de la imagen molecular y funcional [13]. La investigación en este campo acerca cada día más la posibilidad de visualizar en vivo procesos genéticos y celulares, lo que aumentaría considerablemente el potencial diagnóstico.

La posibilidad de esta integración y análisis conjunto acerca a la informática a una nueva área denominada Biología Computacional [14]. La Biología de Sistemas analiza cómo se expresan los genes y cómo interaccionan a nivel de célula, tejido, órgano o incluso todo el organismo.

Las TIC en Biomedicina también facilitan la investigación biomédica modelando y simulando los distintos sistemas y procesos que tienen lugar en el cuerpo humano, habiendo recibido esta área el nombre de "Virtual physiological Human" y constituyéndose en una de las líneas de investigación más importantes en los próximos años.

### 4. Iniciativas en España

Como ejemplo de iniciativa de TIC en Biomedicina realizada en España podemos citar la Red INBIOMED [15], uno de los primeros proyectos tecnológicos que responde a las tendencias actuales de investigación cooperativa fomentadas por el Plan Nacional de Investigación Científica, Desarrollo e Investigación Tecnológica para el período 2003-2006, y que se atiende a las directrices del 6º Programa Marco de la Unión Europea.

La Red INBIOMED [16], financiada por el Fondo de Investigación Sanitaria (FIS) del Ministerio de Sanidad y

Consumo y coordinada por el Área de Bioinformática y Salud Pública del Instituto de Salud Carlos III, es una Red Nacional Temática de Investigación Cooperativa (RTIC) en Informática Biomédica, en funcionamiento desde el año 2003 hasta finales del 2005. En ella han participado 13 grupos de investigación biomédica adscritos a hospitales, universidades y centros de investigación de 6 Comunidades Autónomas (Andalucía, Castilla-La Mancha, Cataluña, Galicia, Madrid y Valencia). Los grupos que componen la Red son tanto de carácter tecnológico (bioinformática, informática médica o análisis de imágenes) como de investigación biomédica, y han coordinado su actuación para cumplir los objetivos generales de la Red.

El proyecto de la Red INBIOMED surgió con la idea de crear en España un sistema de integración de datos que facilitara la investigación en Biomedicina de enfermedades complejas a partir de los datos estudiados por diferentes disciplinas.

Con este planteamiento se desarrolló una plataforma que permitiera almacenar e integrar datos clínicos, genéticos, epidemiológicos e imágenes de diversas fuentes y favoreciera la aplicación de herramientas informáticas para el análisis de toda la información procedente de esta integración, para ser empleada en la investigación de cualquier patología compleja.

La plataforma INBIOMED [17] consiste en una infraestructura de integración de datos distribuida, sobre la que los distintos grupos que componen la Red pueden hacer públicas sus propias fuentes de datos o aplicaciones de interfaz de cliente. En el desarrollo de esta estructura es imprescindible el uso de los estándares léxicos y tecnológicos, para hacer posible la intercomunicación o integración de datos e información de los distintos grupos de investigación y desarrollo.

El sistema elegido para dar soporte a este modelo es un almacén de datos federado a través de Servicios. De este modo, cada grupo puede utilizar el sistema de forma independiente de la Red y poner a disposición de los demás grupos los datos o herramientas que estime conveniente.

La arquitectura de la plataforma, tal y como se puede ver en la Figura 4, está compuesta por cuatro niveles: aplicaciones cliente, servicios, almacén de datos, y fuentes de datos.

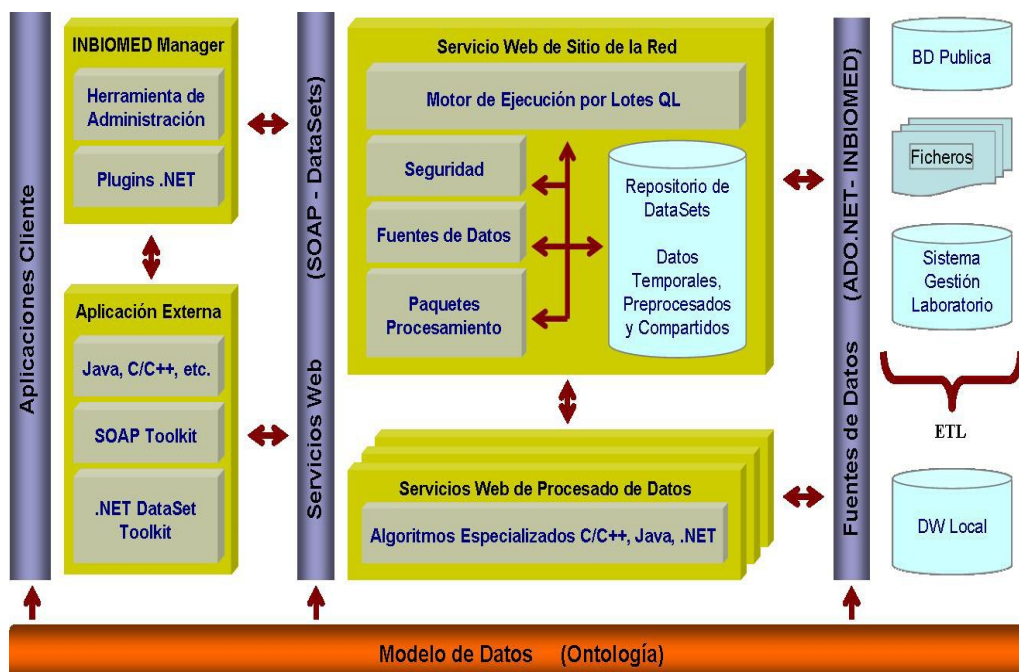


Figura 4: Arquitectura de la plataforma INBIOMED

Las aplicaciones cliente son las aplicaciones independientes de la plataforma, que se integran en el sistema a través de los servicios. La comunicación de las aplicaciones con el sistema se realiza a través de servicios Web utilizando el protocolo SOAP (HTTP para el transporte de mensajes y XML para el formato de estos mensajes).

El nivel de servicios se divide en capas de aplicación, red y datos y sirven, respectivamente, como interfaz de las aplicaciones que deseen comunicarse con el sistema, para mantener repositorios de información y para mantener y controlar los accesos a los datos de cada grupo.

En el almacén de datos los grupos de la Red migran sus estructuras de información para adaptarlas al modelo de la plataforma.

Las fuentes de datos serán las aplicaciones y bases de datos que cada grupo utilice en su investigación y gestión. Pueden ser tanto internas (uso privado de cada grupo) como externas (de acceso público a través de Internet).

Los servicios y las fuentes de datos son encadenados en flujos de trabajo que ocultan la complejidad de los procesos al usuario final. La plataforma ofrece una capa de integración semántica, que ofrece el uso de ontologías ya existentes como GO (Gene Ontology) o UMLS (Unified Medical Language System), o desarrolladas por la propia Red INBIOMED (CBMD- Conjunto básico mínimo de datos) para anotar los datos y recuperar más eficientemente la información requerida.

## 5. Líneas de trabajo futuras

En la actualidad, aunque se han producido avances significativos en la investigación biomédica a nivel de integración de datos, modelado a varias escalas o sistemas de redes, quedan varios retos científicos pendientes que afectan a la aplicación de las TIC en Biomedicina.

Para afrontar la resolución de estos problemas es necesario un enfoque integrado que permita su estudio mediante la combinación de elementos provenientes de distintas disciplinas. Una de las últimas tendencias, y con un gran impacto desde el punto de vista científico y tecnológico, es lo que se conoce como el desarrollo de las Tecnologías Convergentes – NBIC [Nanotecnología, Biotecnología, Tecnologías de la Información y Ciencias del Conocimiento] [18], cuyo esquema se puede ver en la Figura 5. Se espera que la fusión, interacción y sinergia de estas Tecnologías, tradicionalmente separadas, conducirán a desarrollos tecnológicos absolutamente innovadores y a la mejora de la calidad de vida de las personas.

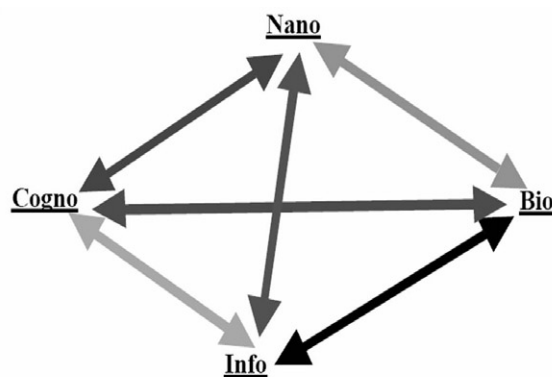


Figura 5: Tetraedro NBIC



Por último, destaca también el desarrollo de Tecnologías de la Información y la Comunicación inspiradas en las Ciencias de la Vida [19]. La interacción entre ambas ramas producirá, según este punto de vista, novedosas utilidades y métodos de procesamiento de información y redes más adaptativas y robustas. En la Figura 6 se puede ver esta forma de interacción, que conforma un bucle: los métodos y herramientas computacionales se usan para estudiar datos y sistemas biolómicos y, a su vez, los sistemas de la biomedicina pueden servir como modelo para nuevas redes y sistemas computacionales.

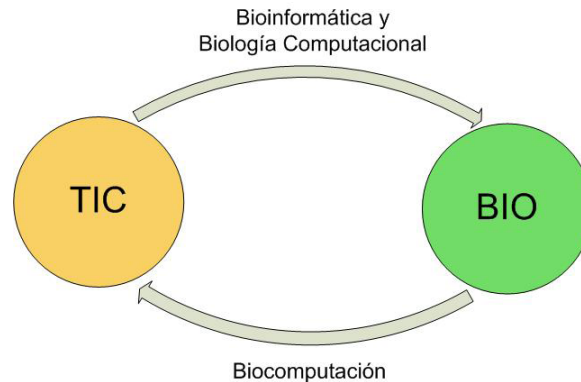


Figura 6: Interacción entre TIC y las Ciencias Biomédicas

## 6. Bibliografía

- [1] Freimer N, Sabatti C. (2003) "The Human Phenome project." *Nat Genet* 2003; 34:15-21.
- [2] Martín Sánchez F., López Alonso, V., Sánchez Merino J. P. (2004) "Definiendo la Informática Biomédica: Objetivos y Líneas de Investigación". In *Informática Biomédica. Recopilación de Ponencias de la Jornada Científica celebrada en Madrid, Octubre 2003*. pp 13-19. Editorial INBIOMED. Editores: Martín Sánchez, F., López Alonso, V. (ISBN 609-1770-3). 289 pgs.
- [3] BIOINFOMED <http://bioinfomed.isciii.es>
- [4] INFOGENMED <http://infogenmed.ieeta.pt>
- [5] Martín-Sánchez F., Iakovidis I, Norager S. (2003) "Synergy between Research in Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare." EC-IST 2001-35024 BIOINFOMED Study. "Prospective Analysis of the Relationships and Synergy Between Medical Informatics and Bioinformatics". White Paper.
- [6] Babic A., Maojo V., Martín-Sánchez F., Santos M., Sousa A. (2005) "The INFOGENMED Project: A Biomedical Informatics Approach to Integrate Heterogeneous Biological and Clinical Information." En *ERCIM News. Especial: Biomedical Informatics*. n° 60.
- [7] INFOBIOMED <http://www.infobiomed.org>
- [8] Maojo V., Martín-Sánchez F., Barreiro J.M., Díaz C., Sanz F. (2004) "Biomedical Informatics: From Past Experiences to the Infobiomed Network of Excellence." En el libro "Biological and Medical Data Analysis". Springer, 418-426.
- [9] SYMBIOMatics - <http://www.symbiomatics.org>
- [10] BEYOND-THE-HORIZON - <http://www.beyond-the-horizon.net>
- [11] Breton V, Dean K, Solomonides T et al. (2005) "The Healthgrid white paper". *Studies in Health Technology and Informatics*, 112, 249-321
- [12] Hansson MG, Levin M (2003) "Biobanks as resources for health." Uppsala University.
- [13] Bogdanov, A.A, Licha, K. (2005) "Molecular Imaging: An Essential Tool in Preclinical Research, Diagnostic Imaging, and Therapy". Series: Ernst Schering Research Foundation Workshop Vol. 49.
- [14] Weston AD, Hood L (2004) "Systems biology, proteomics, and the future of health care: towards

---

predictive, preventative, and personalized medicine.” *Journal of Proteome Research*, 3, 179-196.

[15] INBIOMED <http://www.inbiomed.retics.net>

[16] Martín Sánchez, F., López Alonso, V., Sanchez Merino J.P., Liebana Marcos, I. (2004) “Red Temática de Investigación Cooperativa en Informática Biomédica, INBIOMED.” Especial sobre la Red INBIOMED publicado en I+S. *Informática y Salud. Revista de la Sociedad Española de Informática y Salud*, 46: 7-13. ISSN: 1579-8070. Editorial MIC.

[17] Heredia Álvaro, J. A. (2004) “Plataforma INBIOMED: Objetivos y Arquitectura”. In *Informática Biomédica. Recopilación de Ponencias de la Jornada Científica celebrada en Madrid, Octubre 2003*. pp 21-29. Editorial INBIOMED. Editores: Martín Sánchez, F., López Alonso, V. (ISBN 609-1770-3). 289 pgs.

[18] Roco M.C., Bainbridge W.S. (2002) “Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science.” National Science Foundation.

[19] Sipper M (2002) “Machine Nature: the coming of age of bio-inspired computing”. McGraw-Hill, New York