# Updating the BIOINFOMED Study: Recent Outstanding Developments in Biomedical Informatics

**Fernando Martin-Sanchez[1], I. Hermosilla[1], F. J. Vicente[1]**
[1]Instituto de Salud Carlos III, Majadahonda, Madrid, España

**Summary:** In December 2001, the European Commission promoted a Conference in which more than 400 experts analyzed the synergies arisen between Bioinformatics (BI), Medical Informatics (MI) and Neuroinformatics. In November 2002, and in order to contribute to the strategy of the European R&D; policy for the following decade in such areas, the White Paper of the BIOINFOMED project was presented at the international congress BIOINFORSALUD 2002 (Valencia, Spain). In the strategic document entrusted by the Commission, the relations established between BI and MI were analyzed, resulting in a new definition of the discipline of Biomedical Informatics (BMI) that aims to facilitate the developments of Genomic Medicine. To elaborate the White Paper, a committee of 30 international experts coordinated by the Institute of Health "Carlos III" (ISCIII) designed the agenda with 18 lines of research that corroborated the existing synergy. They pointed out those areas in which the efforts had to be prioritized. In this article, we update this study by highlighting recent outstanding developments in this field. The projects presented respond to a careful selection carried out among the numerous initiatives that have arisen in the three years passed between the publication of the White Paper and this article. Some of the projects analyzed in this paper are: the IT infrastructure for biobanks under the Public Population Project in Genomics (P3G) Consortium, a Network of Excellence (INFOBIOMED) in Biomedical Informatics constituted within the European Union Sixth Framework Program for Research and Technological Development, the initiative headed by HL7 to include genetic information in the electronic health record, the proposal of a Human Phenome Project, a Spanish Cooperative Research Thematic Network (INBIOMED) in Biomedical Informatics, the new National Centers for Biomedical Computing (NCBCs), funded by NIH, under the BISTI initiative and the projects related with the simulation and modeling of Human Physiology.

**Keywords:** Biomedical Informatics, Bioinformatics, Medical Informatics, Neuroinformatics

## 1. Introduction

On December 14th, 2001 (http://14dec2001.ramit.be), the Conference "Synergy between Research in Medical Informatics, Bioinformatics and Neuroinformatics" was organised in the Brussels under the Belgian Presidency of the European Union, by the Belgian Federal Ministries of Social Affairs and Public Health and the European Commission – Directorate General Information Society and Directorate General Research. In this meeting about 400 researchers, professors, institution leaders, and industrial representatives gathered to share their vision on the prospect of synergy between medical informatics (MI) and bioinformatics (BI) as well as on the processes required to achieve that goal.

The Brussels Meeting was the kick-off point for the activities related to the EC IST BIOINFOMED Study. The aim of the project was to analyze the relationships and potential synergies between MI and BI. The goals of the project were the analysis of the state of the art in the area of the study, the proposal of an R&D agenda [1] and the identification of the key players as well as the dissemination of the study (http://bioinfomed.isciii.es).

The study gave a general overview of the evolution of each of the disciplines to achieve a better understanding of the possibilities of the synergy between them. Historically, MI and BI have developed separately and only occasionally have researchers of both disciplines collaborated in the past. Although the roots of BI and MI are located in different application domains, these domains increasingly overlap. Results of research in molecular medicine will have an impact on clinical medicine. This shared application domain will provide a natural space to collaborate. Medicine will benefit from the achievements of biological research, and biology will benefit from the use of clinical data for research. The conclusion was that as the domains begin to overlap, both communities increasingly would share a common goal, a common context, for exploring collaboration.

The results from our study crystallized in a White Paper that was presented in the International Congress BIOINFORSALUD held in Valencia (Spain) and in the HealthGrid Conference in 2003 in Lyon (France). The paper was published in a summarized version in February 2004 in the Journal of Biomedical Informatics [2].

## 2. Was our approximation correct?

Although it was rather difficult to anticipate the future needs that the intersection between MI and BI will introduce in health care and research, several new initiatives seem to reinforce the findings of the White Paper. To value for the time the consistency of our ideas, the best method is to determine what initiatives have started in BMI since the White Paper was published and in what thematic areas. The fact of finding projects that are fitted inside someone of the areas proposed in the research agenda implies some degree of consolidation of the discipline of BMI (see Table 1).

The projects chosen respond to a careful selection carried out among the numerous initiatives that have arisen in the years passed between the publication of the White Paper and the time of writing this paper that was presented in EuroMISE 2004 international congress, held in Prague (Czech Republic). The projects analyzed include: the IT infrastructure for biobanks under the Public Population Project in Genomics (P3G) Consortium, INFOBIOMED, a Network of Excellence in Biomedical Informatics recently constituted within the European Union Sixth Framework, the initiative headed by HL7 to include genetic information in the electronic health record, The Human Phenome Project (HPP), a Spanish Network of Cooperative Research (INBIOMED), the new National Centers for Biomedical Computing (NCBC), funded by NIH, under the BISTI initiative and the initiatives related to the simulation & modeling of Human Physiology.

### 3. Updated Information

#### 3.1 P3G

Biobanks or population repositories are one of the proposed solutions for the integration of information obtained from environmental and life style data of populations together with their genetic and clinical data [3]. Recently five large Biobanks, Quebec's CARTaGENE, GenomEUtwin project (involving 8 countries), Estonia's genome project, the Center for Integrated Genomic Medical Research (CIGMR) and Western Australian Genome Health Program (WAGHP) have merged into an International Consortium called Public Population Project in Genomics (P3G, http://www.p3gconsortium.org/index.cfm) supervised by Professor Bartha Knoppers of the "Centre de recheche en Droit Public in Montreal". As it can be read at its web page: *"Its main objective consists of the creation of an open, public and accessible knowledge database"*.

It aims to make available to the scientific community under a single frame all clinical, genetic, lifestyle and environmental data collected from different sources. To carry out this ambitious project, the consortium will harmonize data collection of demographic, social and clinical data from the four repositories. Phenotypes that present common characteristics among them will be standardized and a common nomenclature system will be developed to denominate variations found in the five resources. Both genotypes and phenotypes will be stored in the databases.

The access of all members to the information contained in the databases will be coordinated by P3G, that will develop all the security measures needed, taking into account regulations currently in force in each of the countries of origin. To obtain these objectives it will promote the development of the tools needed as technical support for the population genomic research. This characteristic is in complete agreement with the focus presented in BIOINFOMED (Fig. 1) where technologies enabled the synergy between two areas that search for common nexus: MI and BI (line 17 of 18, see Table 1).

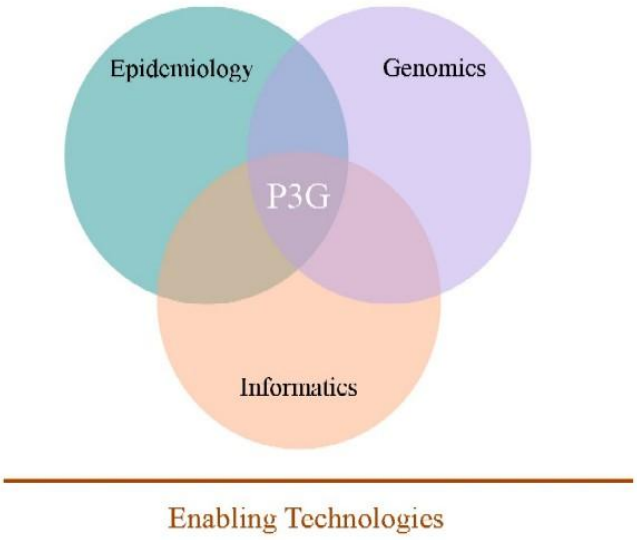


*Fig. 1. Synergies between Informatics, Epidemiology and Genomics in the context of Biobank information infrastructures.*

 

This international initiative will encourage geographical mobility of researchers to facilitate the interchange and transfer of knowledge, abiding by the ethical aspects involved in genomic research, following the philosophy of an international consortium.

Lastly, and as a paradigm of any the biomedical project, the results obtained in the genetic epidemiological research of both diseased and healthy populations will be compared and evaluated, consolidating the ultimate objective of BMI, where we frame this initiative: health.

### 3.2 INFOBIOMED

One of the main achievements of the BIOINFOMED project was the development of a research agenda with eighteen lines of research proposed to meet the gaps that hinder collaboration between BI and MI. There is no doubt that the best corroboration that could be made to consolidate the agenda is to create a network of international groups that will set specific objectives in a given term.

On the 4th of February 2004 took place in Barcelona, Spain, the kick-off meeting of the European Network of Excellence INFOBIOMED (IST-2002-2.3.1.11 e-Health), within the VI framework program of the R & D that counts with the participation of 16 European partners in the next three years [4]. Under the name "Structuring European Biomedical Informatics to Support Individualized Healthcare" and with a budget of 4.8 M Euros, INFOBIOMED is coordinated by Professor Ferran Sanz from Research Group on Biomedical Informatics (GRIB) of the IMIM in Barcelona, Spain.

INFOBIOMED (http://www.infobiomed.org) was born with the purpose of developing tools that will be implemented in the integration of clinical data with genetic data in four research pilots embedded in BMI (see Figure 2). To carry this out, the exchange of methodologies, tools and technologies between BI and MI will be promoted within a European BMI network that will be an open forum of knowledge interchange and dissemination. The training and mobility of the research staff will be a constant in the search of full research capability in this area in Europe, expecting it to last in time.

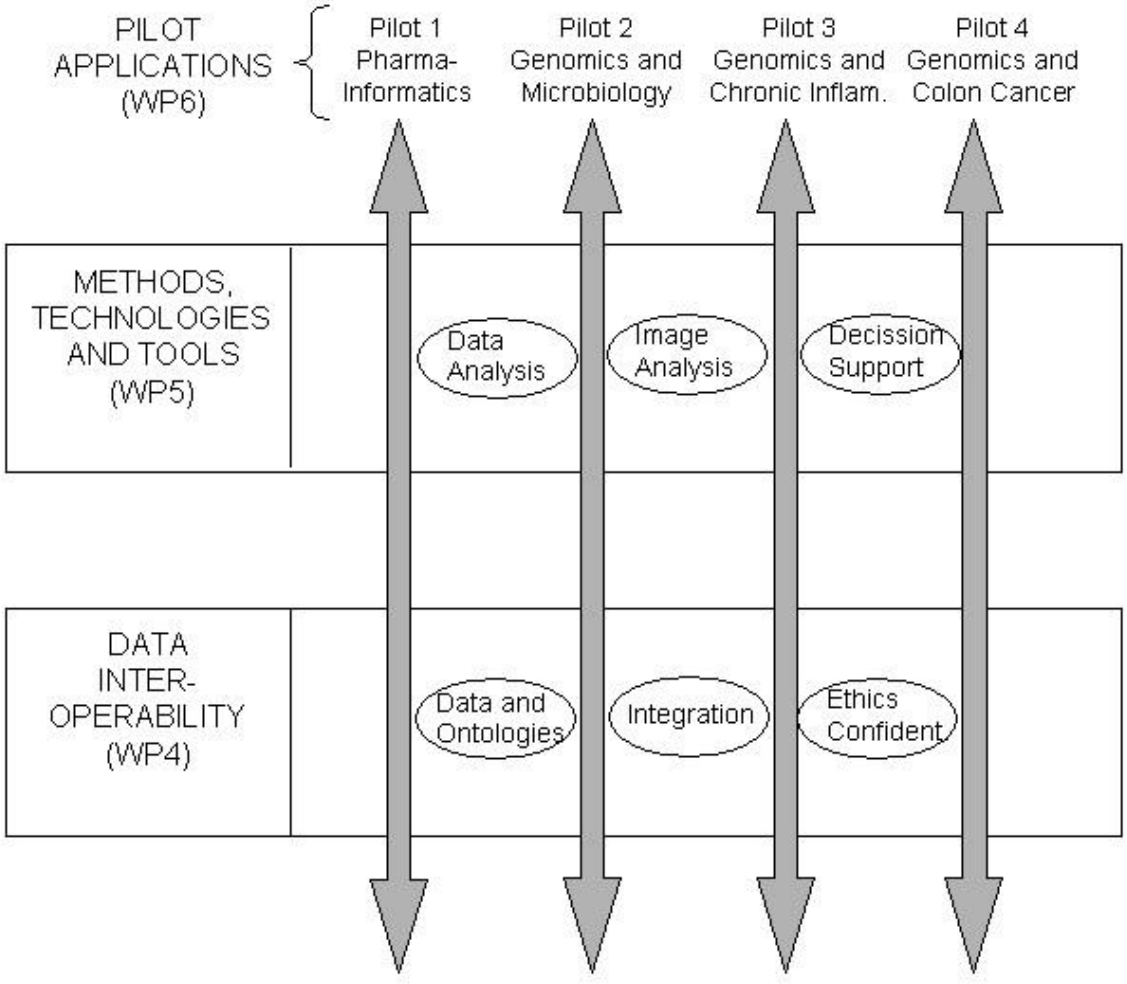


*Fig. 2. Overview of INFOBIOMED Network of Excellence Technical Workpackages.*

The four pilots planned will show the importance of the synergy because it will mean the materialization of the objectives pursued in four large thematic areas, Pharmainformatics, Microbiology, Chronic inflammation and Colon cancer.

### 3.3 HL7 Clinical Genomics Special Interest Group

One of the challenges for daily clinical practice in the coming years is the incorporation of the patient's genetic information in the electronic health record. Health Level Seven (HL7) is an organization responsible for harmonizing protocols or specifications in the area of clinical and management data being one of several ANSI-accredited Standards Developing Organizations (SDOs). HL7 is divided in turn into Technical Committees and Special Interest groups (SIG) and, recently, the HL7 Clinical Genomics Group (http://www.hl7.org/special/Committees/clingenomics/index.cfm) has been created.

Among the main objectives of this last group are to support application of genomics in clinical medicine, to specify use-cases and data requirements, to review existing genomics specifications and to recommend enhancements to HL7 standards to support genomics. This SIG met in Memphis, USA on September 8th – 10th, 2003, to debate about how the Electronic Health Record must support clinical genomics including genetic testing, storage and retrieval of genomic and proteomic information.

This SIG is presided by four people: Jill H. Kaufman (IBM Life Sciences), Scott Whyte (Catholic Healthcare West), Amnon Shabo (IBM Research Lab in Haifa) and Peter L. Elkin, (Professor of Medicine, Director, Laboratory of Biomedical Informatics, Department of Internal Medicine, Mayo Medical School).

They have developed a functional module of EHR system that was balloted and has had its first phase approved for its use in trial mode.

### 3.4 The Human Phenome Project (HPP)

HPP is a proposal of an international project whose main objective would be to obtain phenome databases, this is, the complete phenotypic representation of a species [5]. The phenotype is the morphological, biochemical, physiological and gestural expression of the genotype under certain environmental conditions. To create a phenome, we first have to enumerate the characteristics that make up a phenotype and the relations that could be established between them, which constitute the traits. HPP is an international initiative that would establish the protocols to choose, collect, store, quantify, retrieve and integrate those phenotypic data with their corresponding genotypic data.

The ultimate goal of HPP is to gather and provide knowledge about diseases. Reason why the phenotype defined through the enumeration of its characteristics and traits will be a disease oriented phenotype. To be able to define the phenotypes, several aspects related to the diseases will be researched:

- Phenotypic parameters used for diagnosis that are inherited. HPP will carry out epidemiological studies that will include these phenotypes directed towards inherited diseases.
- Making association studies of endophenotypes, phenotypic characteristics of intermediate heritability, easier to monitor than the direct disease phenotype.
- Quantitative measure of phenotypic parameters in metabolic pathways to improve rational drug design.
- Protocol standardization for measuring phenotypic parameters.
- Comparative phenomics with animal models.
- Protocol standardization for the collection of large volumes of phenotypic data from a sample.

In this same line of work it is also worth mentioning a web-based initiative called PhenoFocus that tries to group all those laboratories, researchers and institution interested in this field (http://www.phenofocus.net/). It is "*an open collection of researchers interested in developing optimal public-domain solutions for computational handling of phenotype data*".

### 3.5 INBIOMED


 

INBIOMED (http://www.inbiomed.retics.net) is a national thematic network of cooperative research in Biomedical Informatics supported by the Biomedical Research Agency (FIS) of Spain [6]. The network has 13 nodes belonging to 11 research centers and universities geographically distributed and groups more than 100 researchers. The INBIOMED network allows the collaboration between researchers coming from several areas of bioinformatics and medical informatics. Other groups are experts in areas like: computer technologies, genomic epidemiology, pharmacogenomics, and molecular and image-based diagnosis. The structure of the cooperative work is defined in such a way that a technological platform provides help to the "bio-users" nodes. The platform is developed by the technological groups and updated, almost weekly, with new solutions implemented following the proposals of the "bio-users" nodes.

Several bio-computational tools have been developed within the INBIOMED network. Moreover, the "bio-users" nodes have seen how their work becomes easier with the procedures and methods provided by the technological nodes. The produced platform includes modules for integration of heterogeneous, distributed databases, gene expression data management, visualization and renderization of 3D images, medical decision support, cell count tools, gel strips analysis procedures and text mining methods among others. All these tools, procedures and methods have been enlarged with another more specific tools to solve some of the problems suggested by the "bio-users" nodes. An overview of the technologies used in INBIOMED is shown in figure 3 below.

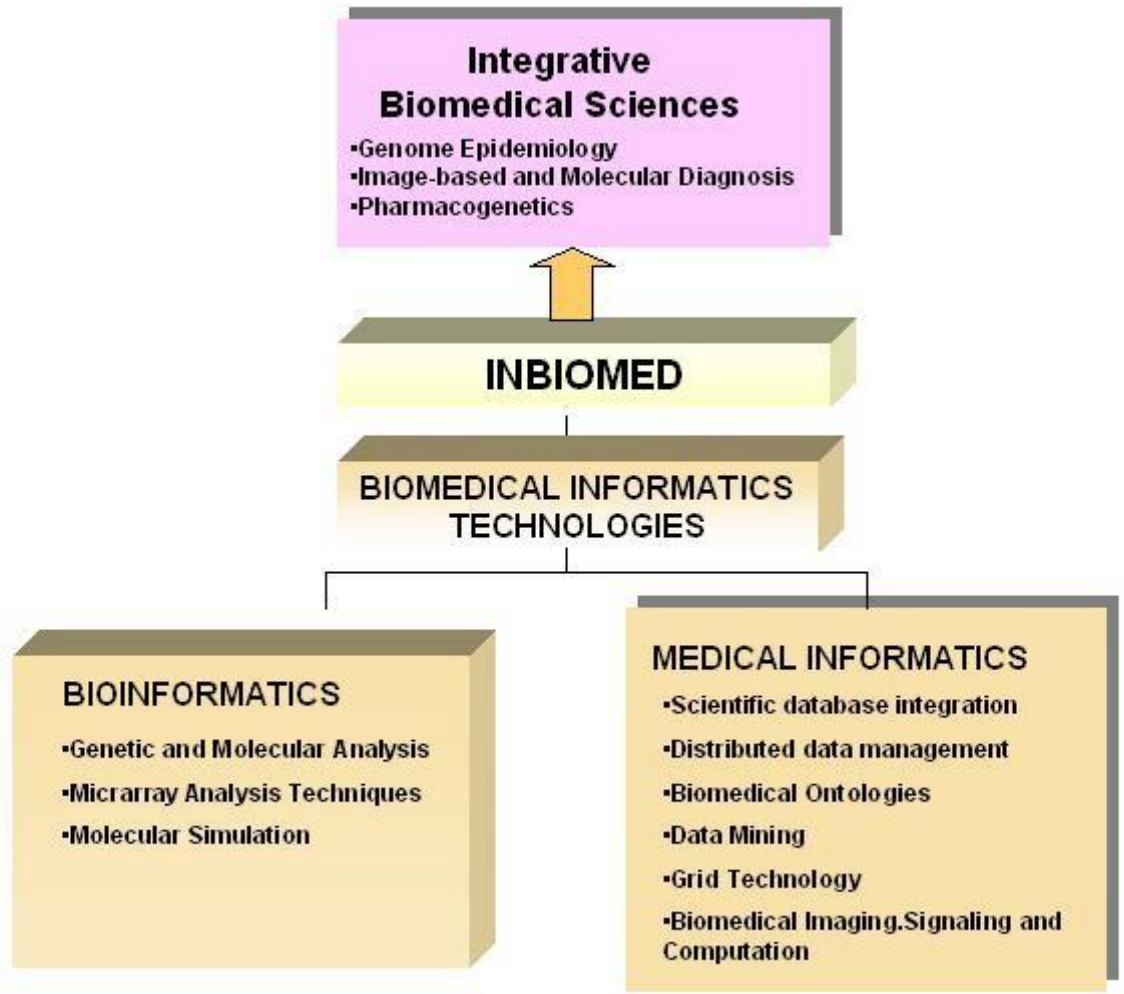


*Fig. 3. Schema of the Biomedical Informatics Technologies used in INBIOMED.*

### 3.6 National Centers for Biomedical Computing (US)

The United States of America has approved the creation of four new national centers, called National Centers for Biomedical Computing (http://www.bisti.nih.gov/ncbc), to develop an international computing framework in biomedical computation. They are funded by the NIH, under the BISTI initiative (http://www.bisti.nih.gov/ncbc/index.cfm). These centers' main goal is to create the core of a computing infrastructure to speed progress in biomedical research. New software programs and other tools will be developed to enable the research community to analyse, model, simulate and share data of human diseases.

The centers are part of the National Institutes of Health RoadMap for Medical Research. Researchers related to the four centers will create new computational tools by means of data collected in both the lab and the clinic. A main goal of the centers is to distribute the developed tools and to train future users.

Research teams of the four new centers consist of experts in computation, biology and behavioral science to collaborate in several projects.

The four new centers awarded in 2004 are Physics-Based Simulation of Biological Structures Center, (http://cbmc-web.stanford.edu/simbios/index.html), National Alliance for Medical Imaging Computing (http://www.na-mic.org/), Informatics for Integrating Biology and the Bedside (http://www.partners.org/i2b2) and Center for Computational Biology (http://www.loni.ucla.edu/CCB/).

### 3.7 Virtual Human Physiology

The Human Physiome project is a worldwide public domain effort that attempts to provide a comprehensive framework for modeling the human and other eukaryotic physiology (http://www.bioeng.auckland.ac.nz/physiome/physiome_project.php). *"It aims to develop integrative models at all levels of biological organisation, from genes to the whole organism via gene regulatory networks, protein pathways, integrative cell function, and tissue and whole organ structure/function relations. Current projects include the development of:*

- *ontologies to organise biological knowledge and access to databases,*
- *markup languages to encode models of biological structure and function in a standard format for sharing between different application programs and for re-use as components of more comprehensive models,*
- *databases of structure at the cell, tissue and organ levels,*
- *software to render computational models of cell function such as ion channel electrophysiology, cell signalling and metabolic pathways, transport, motility, the cell cycle, etc. in 2 & 3D graphical,*
- *software for displaying and interacting with the organ models which will allow the user to move across all spatial scales.*

Within this section we also mention the initiative for the development of the EuroPhysiome project. This possibility is explored in the White Paper recently presented by the European Commission "Towards Virtual Physiological Human: Multilevel Modeling and Simulation of the Human Anatomy and Physiology". This paper reviews the needs and challenges identified to achieve the objective of the "Virtual Physiological Human". They span from making a better use of data, methods and tools to the development of new methods, standards libraries and tools for future research.

## 4. Conclusions

After the analysis of the previous initiatives, we can come to the conclusion that the initial approach of the BIOINFOMED study was correct. All these ambitious projects such as the P3G Consortium, the Network of Excellence INFOBIOMED or NCBCs among others, can be directly mapped to some of the 18 research lines identified in the White Paper (see Table 1). This shows that the analysis and conceptualization done at the time in the project was consistent and durable.

However, we are detecting parallel lines of research, and therefore having no intersection point, of BI towards phenotype and MI towards genotype. This lack of collaboration brings about the danger of falling in the trap of continuously reinventing the wheel, which will take time from better focusing towards a faster and more productive movement forward of research.

*Table 1. Mapping of the project analyzed with the lines of research proposed in the BIOINFOMED white paper.*

| **ENABLING TECHNOLOGIES** | | UPDATE |
|---|---|---|
| 1 | Grid | |
| 2 | Security | |
| 3 | Data communication standards | INFOBIOMED |
| 4 | Knowledge representation to facilitate the virtual integration of heterogeneous clinical and genetic databases | INBIOMED |
| 5 | Data and text mining | |
| MI IN SUPPORT OF FUNCTIONAL GENOMICS | | |
| 6 | Phenotype databases suitable for genomic research | HPP, Phenofocus |

| | | |
|---|---|---|
| 7 | Disease Reclassification | INBIOMED |
| 8 | Pharmacogenomics | INFOBIOMED |
| BI IN SUPPORT OF INDIVIDUALIZED HEALTHCARE | | |
| 9 | Genetics data model for the EHR | HL7-CG-SIG |
| 10 | Clinical guidelines and decision making using genetic information | |
| 11 | Telegenetics | |
| 12 | New methods and information platforms to manage genetic data in clinical Research | |
| 13 | Point-of care data acquisition systems | |
| 14 | Microbial Genomics | INFOBIOMED |
| BMI IN SUPPORT OF GENOMIC MEDICINE | | |
| 15 | Molecular and Functional Imaging | |
| 16 | Modelling and Simulation | NCBC, Simulation of Human physiology |
| 17 | Populational Repositories | P3G |
| 18 | e-Learning | |

**Acknowledgements**

This research has been supported by funding from the EC INFOBIOMED Network of Excellence, the INBIOMED project, Ministry of Health, Spain, and the Ministry of Education, Spain.

**References**

[1] Maojo V., Martin-Sanchez F., Billhardt H., Iakovidis I., Kulikowski C.: Establishing an Agenda for Biomedical Informatics. Methods of Information in Medicine 2003; 42: 121-125.

[2] Martin-Sanchez F. et al.: Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Health Care. Journal of Biomedical Informatics. 2004 Feb; 37(1): 30-42.

[3] Kaiser J. Biobanks. Population databases boom, from Iceland to the U.S. Science 2002; 8;298(5596):1158-61.

[4] Sanz F., Diaz C., Martin-Sanchez F., Maojo V.: Structuring European Biomedical Informatics to Support Individualized Healthcare: Current Issues and Future Trends. Medinfo 2004; 2004: 803-7.

[5] Freimer N., Sabatti C.: The Human Phenome Project. Nature Genetics. 2003 34(1): 15-21.

[6] Lopez-Alonso V., Sánchez J.P., Liébana I., Hermosilla I., Martin-Sanchez F.: A Platform for the Integration and Sharing of Genetic, Clinical and Epidemiological Data Oriented to Biomedical Research. 4th IEEE International Symposium on BioInformatics and BioEngineering (BIBE 2004). 222-226 / IEEE Computer Society Press / May 2004.

.

 