MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

INBIOMED, The Spanish Cooperative Research Thematic Network on Biomedical Informatics

Santiago J. de Ory[a], Isabel H. Gimeno[a], Verónica Hurtado[a], Antonio Estruch[b], Jose A. Heredia[b], F. Martin Sanchez [a]

[a] *Biomedical Informatics Unit, Institute of Health "Carlos III", Madrid, Spain*
[b] *Departament of Technology, Universitat Jaume I, Castellón, Spain*

Abstract

Biomedical research needs to process and integrate large amounts of data, obtained from different biomedical disciplines, to research complex, multigenic diseases where the interaction of both genetic and environmental factors play an important role. It is essential to study the interactions among all the health information levels - population, disease, patient, organ, tissue molecular and genetic-. Biomedical Informatics is the discipline that provides models, methods and computer tools to integrate genetic, clinical and environmental information. In this paper we describe the successful results of INBIOMED, the Spanish Cooperative Research Thematic Network on Biomedical Informatics funded by the Spanish Ministry of Health, which is a perfect example of this line of work. INBIOMED members developed a technological platform to manage, integrate, model and analyze clinical, genetic, epidemiological and image data, for the research of any complex pathology. They also worked in the quality, security, communication and training aspects, to disseminate Biomedical Informatics in Spain. The work performed by INBIOMED was recognized as Excellent by an International Committee of Experts, who recommended its continuity and strengthening.

Keywords:

ICT in Biomedicine, Biomedical Informatics, INBIOMED

Introduction

The publication of the results obtained by the Human Genome Project has produced huge advances in biomedical research in the last years. This information has contributed to the development of new disciplines aimed at researching all factors that have an influence in diseases. Most diseases are caused by genetic and/or environmental factors, and to completely understand their causes it is essential to know patients genetic characteristics and life habits as well as the environmental agents to which they are exposed [1].

However, these data are obtained from very different biomedical disciplines, which often use them exclusively. The integration of these data is essential to be able to research complex, multigenic diseases in which environmental factors have a strong influence. It is also important to study interactions among all the information levels - population, disease, patient, organ, tissue and molecular or genetic-.

In this sense Biomedical Informatics, a recent discipline between Medical Informatics and Bioinformatics, has a relevant role in the integration of these new data and providing new approaches to old health problems, as you can see in figure 1. It does this by providing the models, methods and informatics tools to integrate information about genetics, clinics, and environmental aspects. Thus, Biomedical Informatics facilitates the development of Genomic Medicine, a new scientific discipline that aims towards the development of personalized treatments and therapies for groups of patients with common genetic characteristics[2].

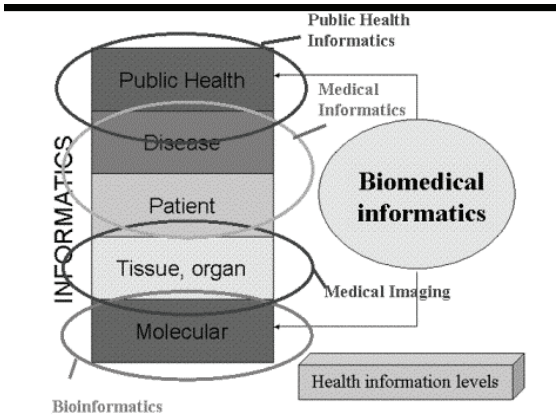

Figure 1- Integration of the health information levels

Biomedical Informatics has increased in the last years worldwide. Each year new initiatives appear in Europe mostly through the EU Framework Programmes and in the USA through the National Centres for Biomedical Computing funded by the US National Institutes of Health. All these initiatives support research in Biomedical Informatics, which is one of the most promising areas to diagnose and treat diseases in the future.

In this paper we describe the work and results obtained by INBIOMED, the Spanish Network of Cooperative

Research on Biomedical Informatics. It was one of the first technological projects in this line of work promoted by the Spanish Scientific Research, Development and Technological Research Plan from 2003 to 2006. This Plan followed the directives of the Europe Union Sixth Framework Program, which consisted of promoting cooperative research and initiatives combining elements from different disciplines. The Network developed a technological platform to store, manage, integrate, model and analyze clinical, genetic, epidemiological and image data that could be used to research any complex pathology. INBIOMED members also developed useful informatics tools, procedures and methods that have been included in the platform (SNPs and genes text mining systems, processing of biomedical images, data analysis of arrays, clinical decision support system, automated questionnaires to evaluate the diet and other environmental factors such as smoking habits). They also developed more specific tools targeted towards solving some of the problems encountered by the "bio-user" nodes. INBIOMED also made important efforts in quality, security, and communication aspects, to communicate and disseminate its results to the scientific community and other stakeholders. Moreover, due to the shortage of training offer in this area in Spain, INBIOMED members have organized Biomedical Informatics Specialization Training Plans and courses directed to researchers.

INBIOMED was formed by 13 biomedical research groups from 11 centres geographically distributed around Spain. INBIOMED coordinated relevant groups in Biomedical Informatics together with technological and pathology research groups that have a multidisciplinary character (Genomic Epidemiology, Pharmacogenetics, Molecular and Image Diagnosis), where the integration and analysis of clinical, genetic, epidemiological and imaging data is needed.

These groups were:

- Polytechnical University of Madrid. Faculty of Computer Science. Medical Informatics Group.
- Municipal Institute of Medical Research. Research Unit on Clinical and Molecular Epidemiology of Cancer.
- Municipal Institute of Medical Research. Research Unit on Biomedical Informatics.
- Sanitary Corporation "Clinic". Medical Informatics Group.
- University of "La Coruña". Faculty of Computer Science. Department of Information and Communication Technologies. Laboratory of Artificial Neural Networks and Adaptative Systems. Medical Informatics and Radiological Diagnosis.
- University of "Santiago de Compostela". Faculty of Pharmacy. Department of Pharmacology. Pharmacogenomics in Drug Research and Development Unit.
- University "Jaume I." Department of Languages and Informatics Systems. Research Group on Integration and Systems Re-engineering.
- University "Jaume I." Department of Technology. Engineering and Design. Informatics Engineering Faculty.
- University of Valencia. Faculty of Medicine. Department of Preventive Medicine. Research Unit on Molecular and Genomic Epidemiology.
- University Hospital "Virgen de las Nieves." Department of Computer Science.
- University of "Castilla-La Mancha." Department of Electric, Electronic and Automatic Engineering.
- Polytechnical University of Valencia. Biomedical Informatics Group. Faculty of Computer Science.
- Institute of Health "Carlos III". Biomedical Informatics Unit.

To achieve its objectives INBIOMED groups were organized in nine different work committees, as you can see in figure 2: Training, Quality and Ethics, Communication, Knowledge Representation, Image Analysis, Structured Data Analysis, Text Analysis, Technological Platforms and Bio-Users (Pharmacogenetics and Neuropsychiatry, Genomic and Cardiovascular Epidemiology and Cancer Epidemiology). Each Committee was composed by different research groups and was directed by a coordinating Committee, responsible for its production and quality. All the research groups were coordinated by the Department of Medical Bioinformatics of the Institute of Health Carlos III.

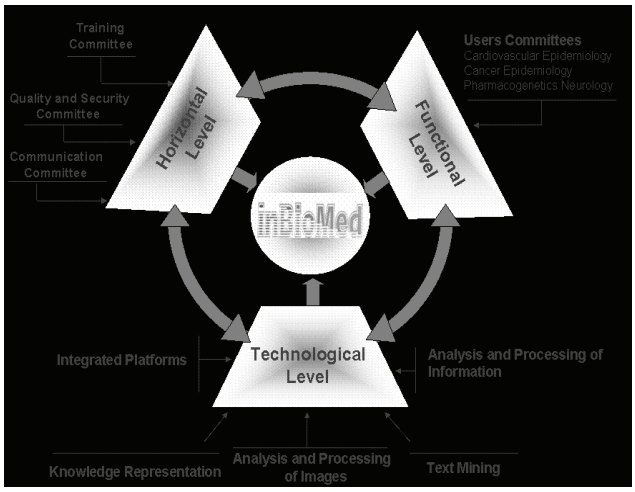

Figure 2- INBIOMED Work Committees

The structure of the cooperative work was defined in such a way that the technological platform aimed to provide solutions to the "bio-user" nodes. The platform was developed by the technological groups and updated with new solutions implemented following the proposals of the "bio-user" nodes.

Methods

The INBIOMED platform

After the creation of the INBIOMED network its members realized that the project had some drawbacks that had to be analyzed. The research groups had problems when they needed to cooperate or share information with other research groups. All the information generated in the laboratories or the gathering or collection of clinical data was very difficult to manage because of their volume. In addition, the Internet provides a huge amount of public relevant information, but in very different formats and the same problem occurred with the analysis and data processing tools.

The solution happened to use an infrastructure based on SOA (Service Oriented Architecture) that makes easy the interaction between all the research groups and the integration of their processes and data. It has a common information model to make a standard way of using the available analysis tools and getting results. To meet these needs, INBIOMED deve- loped a platform based on the development of three layers, as you can see in figure 3:

a) **Client applications** (user interface): This layer implements only the User interface with their input forms and the results presentation, delegating the responsibility to query, store and process the information as much as possible in the other two layers.

Taking into account the Web and Desktop targets, two client applications have been developed so that most of the researchers could use the platform:

1. Windows Desktop *Inbiomed Nexus Workbench* [3] to have the richest client user interface available.
2. Web based *Inbiomed Nexus Site*, it is developed under web technologies using ASP.NET 2.0 exploiting Web-Parts technology that allows the user personalize his framework.

The most interesting functionalities ready to be used are: complete management, shared relational data exploration and query, semantic query on ontology, integrated applications (but others can be developed), design and execution of Workflows, embedded IronPython, execution of IQL scripts and a lot more.

This platform has integrated the last web technologies and it is able to use ontology as data sources in both interfaces. Even those user interfaces have used an ontology to integrate diffe- rent data bases as only one. An example of this are hospital minimum basic data set (MBDS), Each hospital has his own codification language but in order to get a good and complete result many times you have to make several queries so with this integration you only have to query once to the different MBDS.

b) **Web Services Layer** (logic): This layer provides the logic of the platform, providing an easy way to query and process the information. In order to improve the cooperative spirit of INBIOMED, every research group contributes by making available his node to the others.

The main platform site implements a special web service to receive requests, to query and process data in a language oriented to batch execution called IQL [3]. It has, also, to allow the management of users, data sources, access to the other web services…

The IQL engine executes every received request in a separated thread and in a sequential way. All the data manipulation is done by using an ADO.NET data type called DataSet [4]. It manages temporal and persistent variables. For complex tasks it is possible to declare procedures containing other IQL sentences.

The platform provides support for cooperative process management and coordination through the ability to design and run workflows. INBIOMED workflow activities allow sha- ring the data repositories and executing processing tasks hel-ping human and computer based interactions in cross-groups applications to success. Also, it offers pertinent information about workflow execution progress supporting long-time transactions.

c) **Data Sources Layer** (storage): all the database servers are located in this layer. It contains the information sources accessible through the platform (public or private). Data warehouse technology plays a fundamental role guaranteeing that the information stored in the databases and finally published to the platform is well consolidated and oriented to be queried.

The main goal of the platform is to support efficient heterogeneous and distributed information sharing. The INBIOMED platform grants direct access to published relational data sources, but also solves the structural and semantic heterogeneity conflicts with the use of ontologies that define not only concepts and their relations but also mappings of concepts and entities to the published physical relational data.

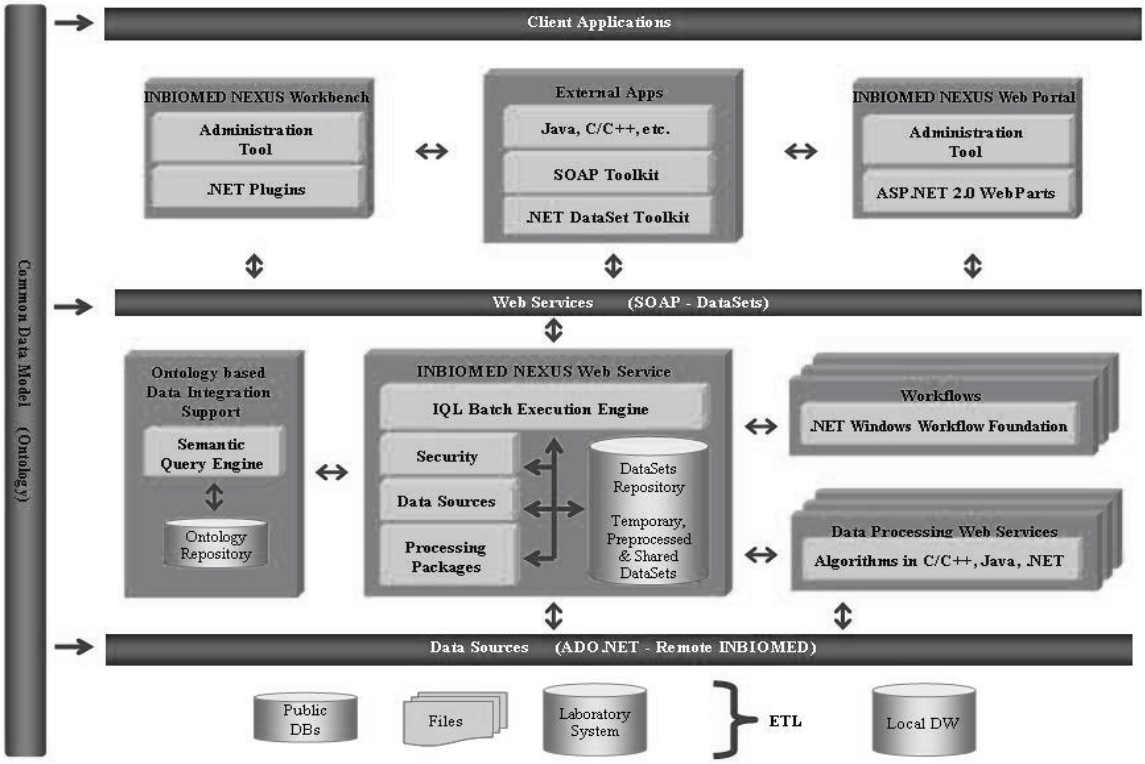

Figure 3 - INBIOMED Platform Layers

Tools

INBIOMED technological groups developed useful informa- tics tools to be used by the "bio-user" nodes and that were included in the platform. These tools include SNPs and genes text mining systems, tools to process, analyze and visualize biomedical images, clinical decision support systems, automated questionnaires, and others. The most relevant tools developed by INBIOMED are the next:PANGEA, a system to share clinical registers and obtain electronic clinical history of the patients.

- ZEHUTI, an algorithm for the calculation of the conceptual similarity between two texts by the projec- tion of concepts into an ontology.
- Bussub, a virtual amplicon retrieval software.
- OSIRIS, a tool for the retrieval of articles from MEDLINE related to the sequence variants reported for a human gene [5].
- SeqPacker and Enzyme-Digest: tools for the analysis of nucleotides sequences.

2 DNASqueezer: a tool to compress DNA and RNA sequences.

- Blast2go: a tool for annotation, visualization and analysis in functional genomics research[6].
- SOC: a distributed system for the decision of clinical diagnostic
- SIREN Project: a computerized system to make inquests in Genomics Epidemiology.

Results

INBIOMED's main objective was to solve the needs of the "bio-user" groups. Each "bio-user" group had different pro-blems which required different solutions and strategies. These are detailed below by researches areas:

Cancer epidemiology

This task was leaded by Dr. Nuria Malats, head of the Research Unit on Clinical and Molecular Epidemiology of Cancer, Municipal Institute of Medical Research.

In response to the needs of the group, INBIOMED has solved in this way:

- In data analysis the researches have developed data mi-ning techniques and comparison of different data mining strategies in data analysis.

- In image visualization and analysis, it has been processed and stored biomedical images that the research in this specific area has created.
- In information retrieval, the members of this group obtain the automation and integration of heterogeneous clinical and pathologic information from hospital databases.
- In SNP searching area it has been produced an utomated search and selection of genetic polymorphisms and the use of OSIRIS to identify articles related to polymorphisms identified in genes of interest.

Pharmacogenetics in neurology

This task was leaded by Dr. Mabel Loza, chief of the Department of Pharmacology. Pharmacogenomics in Drug Research and Development Unit of the University of "Santiago de Compostela". This group had some specific needs related to schizophrenia disease that were solved as described below:

- The SNP searching was solved by automated identification of articles related to polymorphisms in genes of interest and the automation of this SNPs search was facilitated by the use of some informatics tools, like OSIRIS, SNPator and PupaSNP [7]
- A Web format database was created in order to process, store and interchange the images that were generated from several types of formats.
- The management of the Galician Computerized Biobank was improved to use its data and samples.
- The development of algorithms to set genetic expression profiles applied in microarrays assays. By this way the effectiveness predictive model of treatments with neuromedicines can be used in vitro.
- INBIOMED has developed ontology based methods that can resolve the problems associated with syntactic and semantics gaps across data sources. This ontology was created to define keywords used in schizophrenia's pharmacogenomics research and by this way to makes possible to set class relations between these keywords.
- To classify basic and clinical scientific publications according to their conceptual similarities, INBIOMED has developed the ZEHUTI algorithm.

Cardiovascular epidemiology

This task was leaded by Dr. Dolores Corella, chief of the University of Valencia. Faculty of Medicine. Department of Preventive Medicine. Research Unit on Molecular and Genomic Epidemiology. This group presented a different set of needs from the other "bio-user" groups, which are described below.

Informatics systems were developed to facilitate the efficient automatic recruitment and follow up of patients (CITAGEST), the management of biological samples (ALIGERA), the epidemiologic data collection by designing computerized questionnaires, and finally the automation and validation of information about food consumption and its conversion to nutrients by SIREN and VALFRECO.

This area of interest also shared similar needs with the other "bio-users" groups like SNPs search, information retrieval and visualization and analysis of images.

Conclusion

In this paper we have described the work done in INBIOMED, the Spanish Cooperative Research Thematic Network on Biomedical Informatics. This project was a successful example of cooperative research between different disciplines to study complex diseases. Along the three years lifetime of this project an important number of collaborations among the different nodes have taken place. The existence of Biomedical Informatics groups has allowed the setting up of a common ground and a common language between technological and biomedical research groups. The work performed by INBIOMED was recognized as Excellent by an International Committee of Experts, who recommended its continuity and strengthening.

Acknowledgments

The present work has been funded by INBIOMED Research Network.

References

[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". October 2003. pp 13-19. Ed. INBIOMED. Editors: Martín Sánchez, F., López Alonso. V. (ISBN 609-1770-3). 289 pgs.

[3] Estruch A. and Heredia J. A. Technological platform to aid the exchange of information and applications using web services. International Symposium Biological and Medical Data Analysis 2004, ISBN 3-540-23964-2, ISSN 0302-9743, Springer

[4] NET Framework Class Library. DataSet Class.

[5] Juan M García-Gómez, Cesar Vidal, Javier Vicente, Luis Martí-Bonmatí, Montserrat Robles. Medical decision support system for diagnosis of soft tissue tumors based on distributed architecture. Proceedings of the 26th Annual International Conference of the IEEE EMBS. ISBN 0-7803-8440-7 pp 3225-3228. San Francisco 2004.

[6] Conesa A, Gotz S, Garcia-Gomez JM, Talon M, Robles M. Blast2GO: a universal tool for annotation and visualization in functional genomics research. Bioinformatics 2005 21: 3674-3676. ISSN 1367-4803

[7] http://pupasnp.bioinfo.ochoa.fib.es/.

Address for correspondence

Santiago Jiménez de Ory
Medical Bioinformatics Unit
Institute of Health "Carlos III"
Ctra. Majadahonda a Pozuelo, Km. 2.
28220 Majadahonda, Madrid, SPAIN
email: s.jimenez@isciii.es