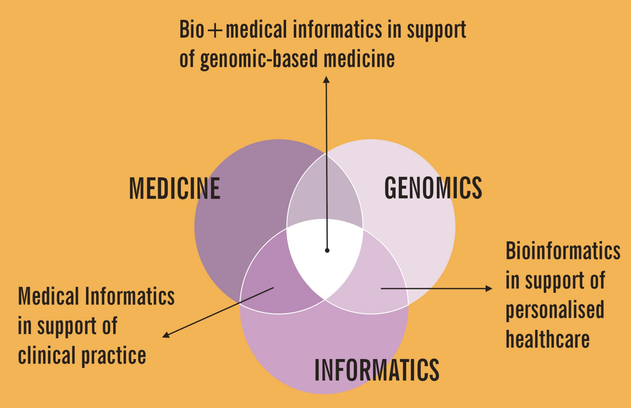
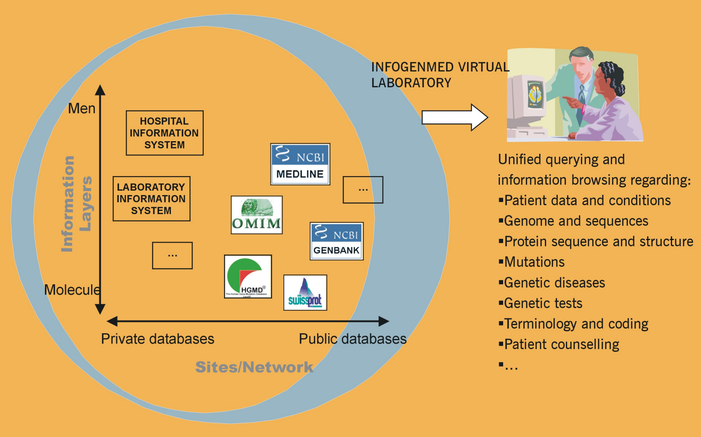
Integration of Genetic and Clinical Information Sources for Health Applications

Ilídio C. OLIVEIRA (1), Artur FURTADO (1), José Luís OLIVEIRA (1), A. Sousa PEREIRA (1), Victor MAOJO (2), F. MARTIN-SANCHEZ (3), Ankica BABIC (4), Manuel SANTOS (5)
(1) University of Aveiro, DET/IEETA, 3810-193 Aveiro, Portugal, (2) Polytechnical University of Madrid (UPM), (3) Institute of Health Carlos III (ISCIII), (4) Linköping University, (5) STAB Vida - GENÓMICA

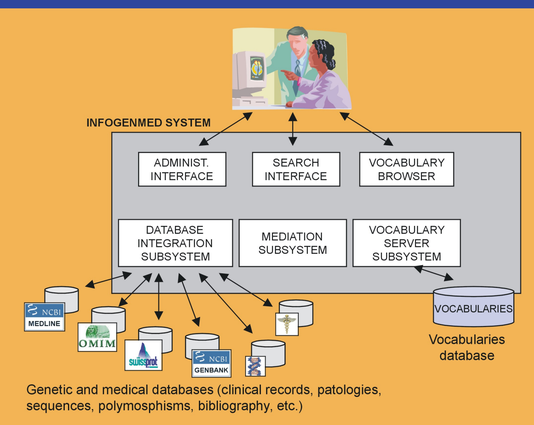
INFOGENMED is building a software environment to access and integrate genetic and medical information for health applications.
Medical information regarding patients' condition and diseases will be easily accessible and related genetic information will also be located, retrieved and presented in an unified, user-friendly way.
This will empower medical practice, research, knowledge development and the collaboration between bioinformatics and medical informatics, paving the way for individualized medicine: as the knowledge of the relationships between human genes and physiopathological states improves and the understanding genetics base of drug action increases, genetic information will become an essential resource at the Practitioner's hand.
Two main issues are involved in this project: the integration of heterogeneous medical and genetic databases over the Internet, and the integration of medical and genetic terminologies in a vocabulary server.



MOTIVATION. Synergies between Bioinformatics and Medical Informatics are introducing a new field: the Biomedical Informatics, enabling Information Technology to support the molecular medicine [1].



OBJECTIVE. To develop a system that integrates multi-layer, multi-site information resources for health applications. Information at genetic level (sequences, mutations,...) must be related to the patient and diseases, extending the data domains available to the Practicioner [2].



METHODS. Users will connect with remote medical and genetic databases through a standard Web browser. Once the information is located at heterogeneous remote databases, a set of tools creates a unification of the contents of those databases, creating a virtual database. Users will have the perception that they are working with a single, local database [3].

How to bring genetic information to the medical practice?

New technologies (e.g.: biochips, bioinformatics), new research approaches (e.g.: proteomics, genomics) and the Internet are revolutionising biomedical research. Most notably, the Human Genome Project is making a main contribution to the knowledge of the relationships between human genes and physiopathological states [4].
The integration of these massive amounts of genetic information in the clinical environment will give rise to a new clinical practice: diagnosis will be more precise and include genetic testing that could be performed at the point of care and therapy methods will include personalized drugs [5].
A new breed of systems and software tools is necessary to convert the enormous amount of data that geneticists and molecular biologists can obtain at their labs in information that physicians and health workers can work with.
The goal of the INFOGENMED project is to build a software environment to access and integrate genetic and medical information for health applications from available from scattered data resources, both in local as in public networks.

Technical challenges towards biomedical data sources integration

The successful unification of heterogeneous clinical and biological databases must address complex problems such as providing adequate search and retrieval engines, link medical terminologies and vocabularies with genomic and proteomics data, and collect and conciliate distributed data fragments under a unified view. These challenging requirements must deal with complex issues:

- Different sources of relevant information are spread over the Internet
- Wide range of formats difficult data interchange
- Codification and terminology is not unified, quality is difficult to discern
- Medical coding systems not ready for managing genetic information
- Bioinformatics tools designed for researchers (not for health practitioners)
- Lack of guidance for the physician

Health practitioners requirements

Unlike the existing tools for biological information management, targeting the research community, INFOGENMED aims to build tools ready to the health practitioners. A survey is being conducted to gather feedback from domain experts; some results already available are:

- Mostly frequently accessed databases (by them) are OMIM, EDDNAL,Orphanet, dbSNP, Ensembl, PubMed, CDC, among others.
- Too much time is wasted in searches and databases hopping. The possibility to search several databases at the same time is a definite plus .
- There are problems with nomenclature mismatches that hinder the seamless use of information sources.
- Tools should be able to interrelate different areas of interest and intuitively help users to navigate through available information. They should present More friendly interfaces enabling to get larger amounts of information.
- Support for collaboration between experts can be a plus, like "on-line conference" to contact with other groups when studying a specific disease (specially in rare diseases).

Together with the health professionals, the project team has been able to define clinical pathways to guide the users through the information space: starting by choosing an appropriate entry point, users will be able to seamlessly retrieve related information, in spite of its location (see illustration below).

Methods and tools under development

The INFOGENMED project is working on the design and development of a novel system to support the linkage and seamless search of scattered databases, including:

- development of methods and tools to locate, access and integrate distributed (medical and genetic) data;
- construction of a vocabulary server for the association of (medical and genetic) terms (ontologies);
- design of a user-friendly interface and development of an "assistant" to help users in resorting to such methods and tools for the benefit of their practice (e.g.: flowchart representation for pathway visualization);

A testbed is also being developed for the validation of the integrated system in the field of rare genetic diseases.

Expected results and benefits

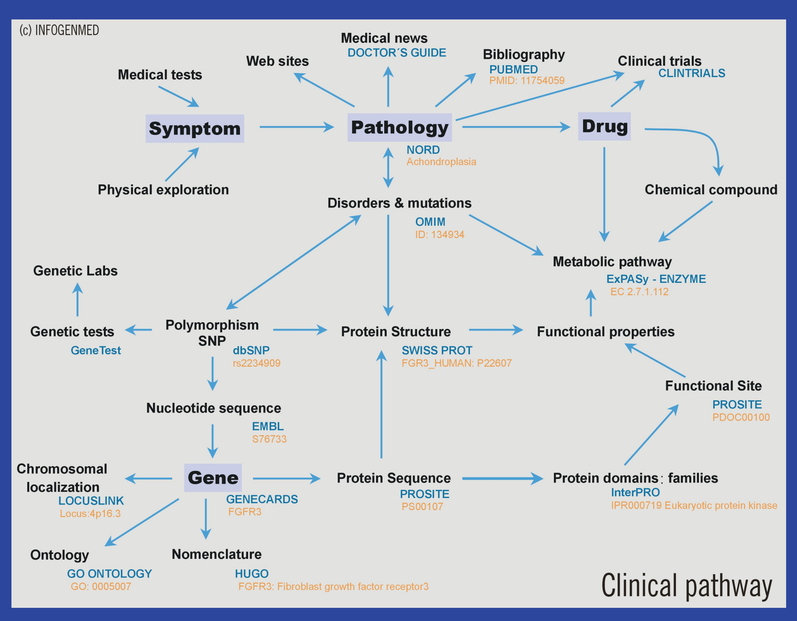
The following results are expected from the project:

- improving existing methods regarding medical/genetics virtual databases;
- a vocabulary server that aims to combine existing terminology systems in Medicine and Genetics
- novel framework for clinicians to locate, search, access, retrieve and use genomic information
- health practitioner oriented, genetic aware, clinical information management tools.

The consortium

The Infogenmed consortium joins institutions with complementary backgrounds:

- IEETA is a leading research organization of University of Aveiro, Portugal. It is particularly renowned for its work in the fields of signal processing, telematics and biomedical engineering. It is the project coordinator and participates in several technical tasks. http://www.ieeta.pt
- Universidad Politecnica de Madrid (UPM) is a leading Spanish university in the fields of engineering and applied informatics. UPM, through the Department of Artificial Intelligence (DIA), is the main responsible partner for designing and implementing the tools being developed in the project. http://www.dia.fi.upm.es
- Instituto de Salud Carlos III (ISC) is the Spanish national institute of health. It provides strong expertise in Bioinformatics and clinical applications of genomics. The role of ISCIII in the project is coordinating the database creation and the clinical testing in hospitals. http://www.isciii.es/biotic
- Linkoping Universitet brings the Swedish experience and expertise in biomedical engineering, bioinformatics and telematics in Healthcare. Linkoping is particularly active in analyzing the confidentiality and security issues of the project. http://www.imt.liu.se
- Genomica is a private Portuguese company exploiting and transferring to wide range of users the latest developments in genomics and genetics. Accordingly, it is mostly concerned with the dissemination and exploitation of project results. http://www.grupostab.com/



References

[1] BIOINFOMED (2003) *Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare*, BIOINFOMED Study White Paper.
[2] F. Martin-Sanchez, V. Maojo e G. Lopez-Campos (2002) "Integrating Genomics into Health Information Systems", *Methods of Information in Medicine*, 41: pp. pp. 25-30.
[3] A. S. Pereira, V. Maojo, F. Martin-Sanchez, A. Babic e S. Goes (2002) "The Infogenmed Project" In Proc. of *Int. Congress on Biological and Medical Engineering*, Singapure.
[4] F. S. Collins e V. A. McKusick (2001) "Implications of the Human Genome Project for medical science", *JAMA*, 285: pp. 540-4.
[5] R. B. Altman (2000) "The Interactions Between Clinical Informatics and Bioinformatics", *J Am Med Inform Assoc*, 7(5): pp. 439-443.

Acknowledgement

The INFOGENMED project is funded by the European Community, under the programme Information Society Technologies (IST) programme, Key Action 1 Systems & Services for the Citizen/Health.

Get involved

To support the specification of the methods and tools, an international user survey is being conducted and several domain experts invited to state their vision. Your feedback is welcome! Pay us a visit at www.infogenmed.net/survey

Project updates and detailed contact information is available at the project web site:

http://www.infogenmed.net