On the requirements of biomedical information tools for health applications: the INFOGENMED case study

Ilídio C. Oliveira[1], José Luís Oliveira[1,2], Miguel Santos[3], F. Martin-Sanchez[4], A. Sousa Pereira[1,2]
[1] *Instituto de Eng[a] Electrónica e Telemática de Aveiro, Campus Univer. de Santiago, 3810-193 Aveiro*
[ioliv@ieeta.pt]
[2] *Departamento de Electrónica e Telecomunicações, Universidade de Aveiro*
[3] *Genómica / STAB Vida, Oeiras*
[4] *Instituto de Salud Carlos III (ISCIII), Madrid*

Abstract: 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. In this communication we report on the user requirements identified through interviews and questionnaires.

Keywords: Bioinformatics; distributed biomedical data integration; health applications; user requirements.

1 Introduction

Advances in genomics are fostering the understanding of genes structure and functions and the knowledge of relationships between genes and psicopatology states. Most notably, the Human Genome Project is making a major contribution to the understanding of genetic level implications to the human health [1]. The integration of these massive amounts of genetic information in the clinical environment is expected to give rise to a new clinical practice, where clinical diagnosis and treatments will be supported by information at molecular level [2]. *Molecular medicine* is expected to provide:

- More precise diagnosis, including *genetic tests* that could be performed at the point of care to uncover the diseases' causes or predict eventual future complications;
- *Personalized drugs* with a narrower scope and more specific action thus increasing drug efficiency with minimal inconveniences (e.g.: bacteria resistance to drugs);
- Therapy methods acting directly on the genetic code. The *gene therapy* holds the potential to correct mutations and minimize hereditary diseases.

The generalization of molecular medicine requires an increased exchange of knowledge between clinical and biological domains. While the former has a deep knowledge of the *phenotypes*, the last has the practice and the tools to analyse the *genotype*.

The wealth of information available both in biology and in medicine is so impressive that requires specific tools for information management [2] [3]. Present tools, however, seams to lack the necessary methods and features to effectively link genetic and clinical information and, moreover, health applications and existing genetic databases [4]. A new bred of information tools, integrated in the clinical workflows, is needed to search, retrieve and relate genetic and clinical data. This means that patient data, information about pathologies, clinical trials, genetic sequences, proteins, etc, must be accessible to practitioners and, moreover, integrated.

2 The infogenmed project

The unification of heterogeneous biomedical databases under one virtual system could be the first major step towards developing a robust computational model to store, retrieve and analyse information at multiple levels, from "molecule to man". This is the objective of the INFOGENMED project, aiming at build a virtual laboratory to access and integrate genetic and medical information for health applications, effectively bridging the two domains. Using the system, general and genetic practitioners, biologists, chemists, among others, will be able to seamlessly navigate local and remote biomedical databases (Figure 1).

INFOGENMED is a European Community funded project, started in September 2002, with the duration of two years. The goals of the INFOGENMED project are closely related to those of other European efforts to exploit the interactions between medical and bioinformatics [5]. The project

has selected the rare genetic diseases field as a testbed for methods and tools being developed.

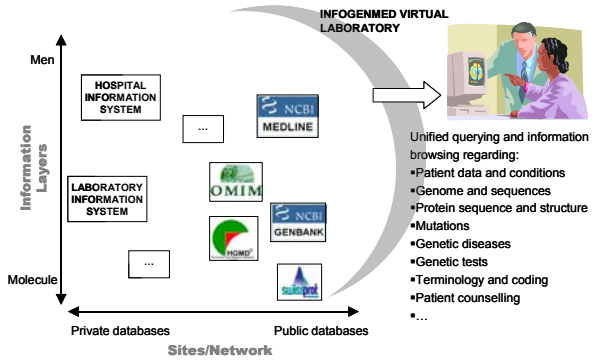

Figure 1: The INFOGENMED system integrates multi-layer, multi-site information resources for health applications.

3 Health applications requirements

A major challenge for biomedical informatics consists on adapting traditional systems to new genetic-based diagnostic and therapeutic tools [6], a challenge that INFOGENMED tries to address. This requires an understanding of biomedical professionals' requirements, an endeavor started at the very beginning of the project. A twofold approach was applied to characterize user needs, including interviews to a restricted Expert Panel and a questionnaire distributed to a wider prospective target.

3.1 Expert Panel inputs to requirements definition

The Expert Panel (gathering 10 professionals) provided valuable contributions to the definition of user requirements at a strategic level, helping the project team to build a vision and validate the feasibility of the envisaged tools. In this sense, a representative set of experts, gathering complementary skills and backgrounds was deemed to be necessary, joining three complementary profiles: medical view, biology view and information technology view.

Some illustrative inputs from this panel includes:

- Too much time is wasted in searches and databases hopping. The possibility to search several databases at the same time through a single point of query and results analysis (*information hubs*) is a definite plus.
- There are problems with *nomenclature mismatches* that hinder the seamless use of current biomedical information sources. Not only this happens between different domains of knowledge (e.g.: clinical and biological) as with databases with semantic related contents (e.g.: no standard nor complete ontology do describe the existing biological data).
- Tools for biomedical data integration should be able to interconnect different areas of interest and intuitively *help users to navigate through information* at different levels. They should present user-friendly interfaces and use both summary and detailed views if information is appropriate.
- Support for *collaboration* between experts can be a plus (e.g.: on-line conference) to contact with other groups when studying a specific disease or therapy.
- *Counselling* will be important to approach genetics-related issues with patients.
- *Genetic testing* is expected to become a near-future generalized practice, for which the information tools, and particularly the Electronic Medical Record, must be prepared [6].

3.2 Questionnaire results

Complementarily to the Experts inputs, the project team is conducting a survey through a structured questionnaire, to collect a broader feedback and, at the same time, promote a greater awareness regarding the project objectives. Current answers to this questionnaire were mainly collected on-line, allowing for international contributors to participate (available at the project web site http://www.infogenmed.net/survey). It was announced in the project web site and disseminated among the partner' contacts. The participation of two Portuguese partners in the consortium has resulted until now in a strong contribution of Portuguese professionals to the requirements definition (19 out of 30 replies). The survey is supported by a minimal on-line demonstrator that enables even unaware users to get a glimpse of the future system possibilities.

The questionnaire focused mainly on the existing information usage patterns and system requirements identification.

Biologist/Biochemist	10
Medical researcher	9
Medical doctor (practitioner)	3
Bioinformatician	2
Computer scientist	2
Other (Pharmacist, etc.)	4
Total	30

Table 1: Respondents distribution according to professional activity.

Some key inputs from currently available results are:

- Respondents report a strong familiarity in medical and genetic data usage (Figure 2) – 76% of then stating they search for medical data in a regularly basis, while genetic data is sought by 70% of respondents. It is interesting to note that half of the respondents use both kinds of data regularly, reinforcing the idea that they are complementary in the context of life sciences (Figure 2). This is an important result, keeping in mind that not all respondents are health professionals (Table 1). Not surprisingly, almost all respondents that actively seek for biomedical information use the Internet for information retrieval.
- When asked to assess the importance of genetics to the present medical practice, the respondents clearly indicated there's a real contribution but, still, somewhat moderate (Figure 3). However, concerning the future role of genetics in medicine, there is a strong conviction that it will play a decisive role.
- Respondents were asked to rate how important they found a given list of requirements (voting from –2 to 2, zero being a neutral vote). It is interesting to note that the top requirement is the ability to link different vocabularies and coding systems (Figure 4), as noted also by the Expert Panel previously. The need for tools that are easy to use was also stressed. Another trend that seams to gain importance among the biomedical community is the availability of tools for free (or at least at low cost). One should not forget the plethora of advanced tools available for free in the bioinformatics field, especially valuable data banks such as the one created by the Human Genome Project.

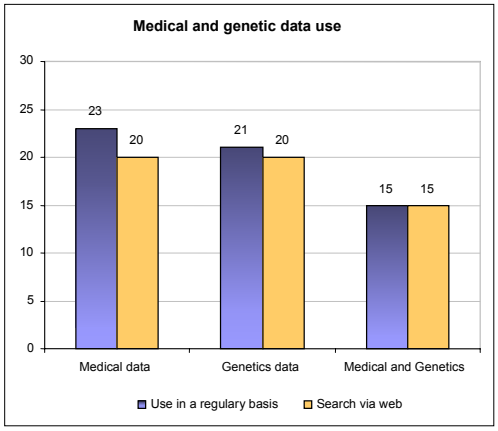

Figure 2: Usage of medical and genetic data.

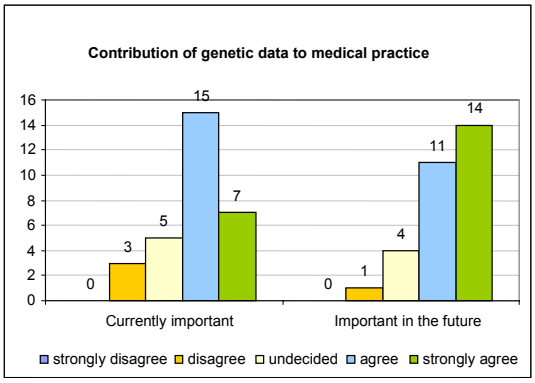

Figure 3: Contribution of genetics to the clinical practice.

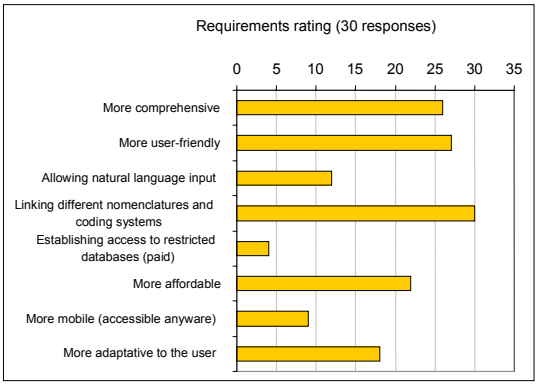

Figure 4: User requirements scoring.

4 Clinical protocols

Users will resort to the INFOGENMED environment to perform *clinical* pathways, i.e., starting from a familiar entry point (e.g.: Symptom, Organ, Drug, etc), navigate into deeper levels of detail (e.g.: Disorders & mutations, Enzyme, etc), walking through a net of biomedical relevant links. It is the system that autonomously retrieves and relates data for the several concepts engaged in a conceptual clinical pathway.

The illustrated protocol can be read as follow (Figure 5):

- There are several entry points to the protocol; **symptoms**, **organ** or **bibliography**. All lead to **pathology,** which is another main entry point.
- The pathology may be due to a **mutation**, i.e., change in the nucleotide sequence of a DNA molecule (information available at OMIM) or to a **polymorphism or SNP** (differences in DNA sequence among individuals, groups or populations) (dbSNPs). These are detected through a **genetic test (**GeneTest) carried out in the **genetic labs** (GeneClinics).

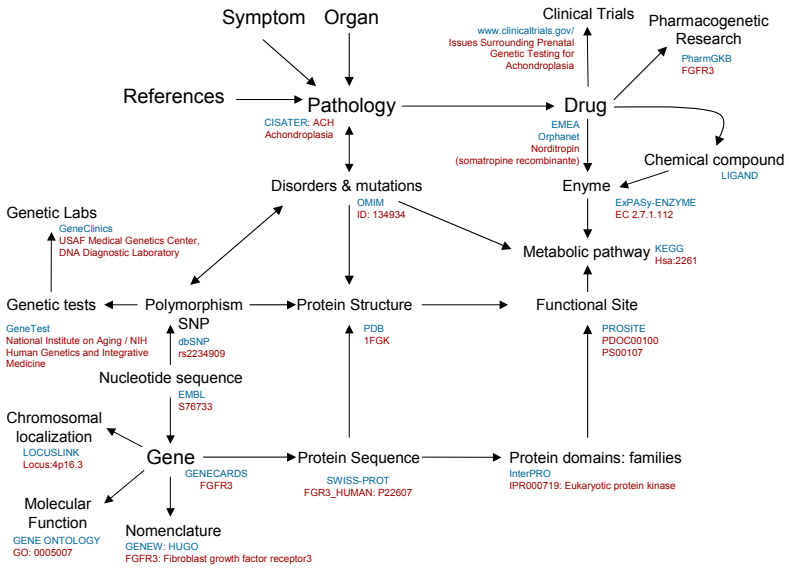

Figure 5: Clinical pathway illustration. Databases' names are shown in blue: access numbers and query terms are shown in red.

- SNPs are within a **nucleotide sequence** (EMBL) which in turn is in a **gene** (Genecards). This gene has a name (Genew, HUGO), a **chromosomal localization** (LOCUSLINK) and a **molecular function** found within Gene Ontology (GO).
- The gene codes for a **protein**, a **sequence** of amino acids (SWISSPROT).
- The sequence determines the **structure** of the **protein** (PDB). The protein is classified into **protein domains: families** (InterPRO).
- It has a **functional site** (PROSITE).
- Proteins have **enzymatic** properties (EXPASY-ENZYME) in **metabolic pathways** (KEGG).
- Enzymes are therapeutic targets for **drugs** (European Agency for the Evaluation of Medicinal Products, EMEA).
- Drugs are **chemical compounds** (LIGAND) that are developed through **pharmacogenetic research** (PharmGKB) and can be used in **clinical trials** (Clinical Trials).
- Most of the entries described can directly link to bibliography (Pubmed).

5 Conclusions

Medical and bioinformatics need to cooperate to build tools able to access and relate clinical and genetic data, empowering present and future medical practice. The previously highlighted user requirements are perfectly aligned with INFOGENMED goals: to facilitate the seamless integration of medical and genetic information, through an easy to use, Web-based environment, tailored to the health professionals. A measure of the success of the system will be it ability to support the execution of clinical pathways, autonomously retrieving and linking information for semantic related concepts.

Acknowledgement

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

References

[1] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, 2001.

[2] A. Bayat, "Science, medicine, and the future: Bioinformatics," *British Medical Journal (BMJ)*, vol. 324, 2002.

[3] M. Palakal, S. Mukhopadhyay, and J. Mostafa, "An intelligent biological information management system,",, 2002.

[4] BIOINFOMED, "Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare," BIOINFOMED Study White Paper, 2003.

[5] V. Maojo, I. Iakovidis, F. Martin-Sanchez, J. Crespo, and C. Kulikowski, "Medical Informatics and Bioinformatics: European Efforts to Facilitate Synergy," *Journal of Biomedical Informatics*, vol. 34, 2001.

[6] F. Martin-Sanchez, V. Maojo, and G. Lopez-Campos, "Integrating Genomics into Health Information Systems," *Methods of Information in Medicine*, vol. 41, pp. pp. 25-30, 2002.