GATHERING AND MANAGING GENOTYPE AND PHENOTYPE INFORMATION ABOUT RARE DISEASES PATIENTS

Rafael Mendonça[1], Pedro Lopes[1], Hugo Rocha[2], Jorge Oliveira[2], Laura Vilarinho[2], Rosário Santos[2] and José Luís Oliveira[1]
[1]*University of Aveiro, DETI/IEETA, 3810-193 Aveiro, Portugal*
[2]*Centro de Genética Médica Jacinto Magalhães, INSA, 4099-028 Porto, Portugal*

Keywords: Rare diseases, Genetic mutations, Reference centers, LSDB.

Abstract: Information technology is increasingly present in medicine, and is emerging as a crucial tool both in clinical monitoring and knowledge dissemination, increasing the success rate in diagnosis and subsequently during treatment. A particular sub-group of diseases are designated rare or orphan due to the small number of people suffering from these illnesses. They are sometimes disabling, reducing the quality of life of patients and affect all relatives around them. The aim of this project was to build a Web information system to support practice and clinical research in rare diseases, and to facilitate the collection of scientific information, diagnosis, treatment and patient support. Through this system, one can optimize specialized medical resources, computing resources and data quality, so that the information is available for consulting and supporting future decisions. This developed system also allows health professionals to share information that will be important for enhancing the quality and technological advancement in this area of public health.

1 INTRODUCTION

The European Organization for Rare Diseases (EURORDIS) estimates that there are approximately 5000 to 8000 rare diseases in Europe, affecting about 6% to 8% of the population, and 80% of these cases have a genetic origin. Due to the reduced incidence of each individual disease, it is difficult for patients to find support, both at the clinical and psychologically level (Aymé et al., 2007). Some of these chronic diseases seriously affect patients' quality of life and cause serious damage and disability in social terms.

The existence of a small number of patients for each rare disease affects the creation of a minimum number of cases required for appropriate studies (Schieppati et al., 2008); (Burke, 2002).

The Portuguese National Programme for Rare Disorders, approved in 2008, strives towards a concerted system of patient management, which relies heavily on a platform designed to gather information from the different health providers, for this, we have built a Web-based platform to integrate all this information obtained from studies, in order to have a database with information about these rare diseases and patients. It will support a network of different public entities in the Portuguese health system, named reference center, which detains a wealth of data on patients with rare diseases, composed of multidisciplinary teams and available resources for the diagnosis, treatment and medical monitoring of patients. This organizational structure permits the best procedures to be available, simplifying processes and minimizing the resources needed. Those involved in this information system administer and register/retrieve clinical information about the studies performed.

Concomitant with the systematic, widespread collection of patient data, locus-specific databases (LSDBs) could be generated, which will not only facilitate the molecular diagnostic service in the national context, but also contribute to the global effort of the Human Variome Project (HVP), that aims to collect, curate and distribute all human genetic variation affecting health (Kohonen-Corish et al., 2010); (Patrinos et al., 2011).

Mendonça R., Lopes P., Rocha H., Oliveira J., Vilarinho L., Santos R. and Oliveira J..
GATHERING AND MANAGING GENOTYPE AND PHENOTYPE INFORMATION ABOUT RARE DISEASES PATIENTS.
DOI: 10.5220/0003791204230426
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2012), pages 423-426
ISBN: 978-989-8425-88-1

2 OVERVIEW

The importance of genetic records in diagnosis and results of these rare diseases has taken medicine to a level where research is extremely important to Databases emerging with information about human genome such as the HGMD (Human Gene Mutation Database) (Krawczak et al., 2000). The data are public and accessible by all the research community, and it is important to collect and publish this information in a biomedical application.

Gene mutation databases are dominant in the area of health care enabling representation of diseases. However, there is a lack of detail in clinical information, which indicates the vulnerability of these patient groups (Schneeweiss et al., 2005). So there is a need to provide the clinical tools necessary to register and personalize studies about all rare diseases. The point of reference in this area has been the Orphanet, a web platform directed to the general public, health professionals and patients, to inform about orphan drugs and rare diseases (Nabarette et al., 2006).

3 INFORMATION COLLECTION

In scientific communication, it is essential to use terms transmitting clear and concise medical information. Many conventions, when known and used by all, facilitate understanding among users. In this paper, we highlight the OMIM, *Online Mendelian Inheritance in Man*, (Hamosh et al., 2005) and HGNC, *HUGO Gene Nomenclature Committee* (Bruford et al., 2008). The generic information is obtained building the match between diseases and corresponding genes of this sources. This is essential to establish a correct connection and utilization, eliminating possible errors in data entry and process simplification.

This information solution presents the user with a simple way to select and choose the disease and genes to be studied. It allows the search for disease groups or by complete name. For instance, in the Muscular dystrophy group, OMIM presents three distinct variants: *Muscular dystrophy, limb-girdle type 1A, type 2D* and *type 2E*.

3.1 LSDB

In contrast to general mutation databases, an LSDB is a repository of genetic sequence variants associated with an individual gene. Besides the mutation data, these databases have important information regarding the gene itself and have been referred to as knowledge bases. They are usually used by a group of researchers in collaboration, with expertise in a particular gene or phenotype, and provide a valuable tool for analysis of gene expression and phenotype, both in normal conditions and disease (Claustres et al., 2002).

Given the need for an LSDB for the registration of genetic mutations in this application, the system chosen was an open-source platform LOVD (Leiden Open Variation Database) (Fokkema et al., 2005), a solution that can be installed locally and used by many professionals in this area, which allows registration of genetic mutations, linking them to patients who have been diagnosed. It is important to include this application in this study because it will allow collection of information about genes and their variations with the purpose of registering and visualizing mutations in the genetic code. The database structure follows the recommendations of HGVS (Human Genome Variation Society), which allows the nomenclature used to be generalized among all users (Fokkema et al., 2011).

4 ARCHITECTURE

The communication between client and server is performed using the HTTP protocol, using a web browser to manage users and the reference center network of this application. This Web-based platform provides a set of components/services: patients, genetic mutations, studies, form templates and administration. The core has three different tiers: JSPs pages and JavaScript code to construction/use the forms; the Stripes framework and Java as the code source of this application; and the data persistence using JPA (EclipseLink) to connect to de database server (Figure 1).

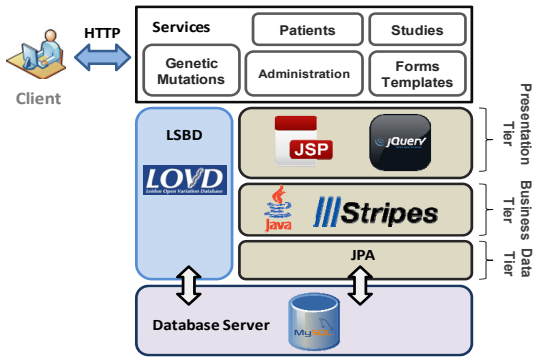

Figure 1: Platform architecture.

The application is hosted on a Web server in combination with the necessary LSDB platforms (LOVD) to store genetic information.

When a reference center is created, the application provides an initial setup and all the patients and genes selected as the targets of the studies are stored. The LOVD platform is immediately available and configured with all information needed to start work in analyzing genetic mutations.

4.1 Data Protection

The main goal of this structure is to protect the information and permit access only to accredited members. So this application divides users between members, when members are in different roles assigned by the responsible for each reference center. This approach implies that we have the master professionals who configure the form templates and the diseases, and choose the users.

Each reference center has its own independent data collection and a set of available components; the application works as a site of the sites.

The authentication used is a simple web platform when the username and password (encryption in MD5) are stored in the database. The authentication between the system and LOVD interface uses cookies. Every reference center website can access all patient records, or add a case that does not yet exist.

5 RESULTS

The application contains generic disease information taken from a database, available to reference center members who are able to access such information. In addition, information is recorded from studies of patients and this can be screening, diagnosis, confirmation of diagnosis, treatment and clinical monitoring.

For better characterization of the different diseases, it was necessary to construct a forms engine, which provides the forms so members can use and later record this information allowing visualization.

It is important to highlight that these forms can be configured according to the needs of the studies and the disease in question. The forms contain the title, a short description and a set of fields. The fields can be of different types: *textBox*, *multitext*, *selectBox*, *radioButton* or *checkBox*. With these fields we aim to respond to the overall needs of form personalization. This mechanism of construction to form templates supports functionalities such as *drag-and-drop* and *sortable*, which makes the construction more intuitive and user-friendly. If the reference center requires genetic records, it is possible to connect between the patient and the genetic records made in their LOVD platform.

So the information collected in all patient studies can be complemented as well as genetic analysis carried out by experts trying to solve and create a standard in each disease, giving support and diagnosis as accurately as possible.

This application has with baseline a Web front end, for two different parts of the system: the management of the centers of reference network; and healthcare support for rare diseases with an LOVD platform to connect patients to genetic results, to collect all clinical information found.

The administrator has permission to initiate the creation of the reference center and can be an entity such as a National Health Service, which should manage the activity of reference centers and can send invitations to new users of this system.

The patient register is unique and only the user with access is able to visualize this information and use these tools. So this unique register permits all information to be centralized and available for future decisions, and prevents data replication. Each patient has a set of the information like studies and genetic mutations (Figure 2).

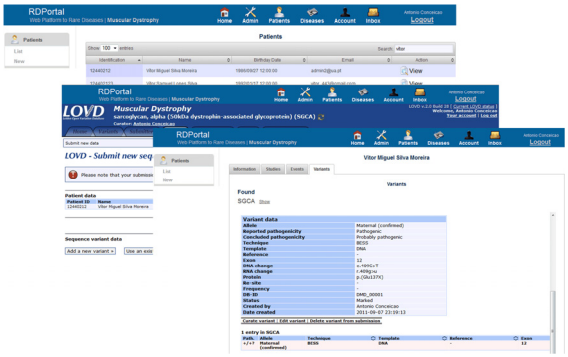

Figure 2: User interface using the reference center platform.

The studies part presents data about medical progress, treatment, procedures and health monitoring. This contains the entire history of all studies made, allowing a specific search for a date or a diagnosed disease, for example. It also allows clinical professionals to add a new study, and choose and collect all the relative information, with the support of the template forms, and if desired, to

highlight any associated suspicion or confirmation of a rare disease.

The genetic mutation part contains all the variants found and the associated gene, with a link that connects to the LOVD interface, showing all registered information about mutations. This information contains all details about Exon, DNA change , RNA change or Protein.

6 CONCLUSIONS

Over time, medical records are stored and used among different healthcare institutions. This problem is even more acute in rare disease due to the lack of information about diseases, symptoms, diagnostics and genetic causes. As such concentration in a single reference center all this information is of paramount importance for physicians, diagnostic labs, patients and their families and for research proposals.

In this paper we have presented a Web-based information system that allows gathering and managing all information about these patients along their lives – from genotype to phenotype. Through a common Internet connection, users have access to information that they can manage according to their profile in the system and to their role in each study. It is possible to manage, in an integrated way, clinical information (patient data, diagnostics, procedures, …) and genotype data, such as patient mutations. The platform's customization allows configuring the system according to each disease requirements. The incorporation of LOVD is also a great asset, considering its wide adoption in the field, since it can be also used outside our platform in an independent yet complementary way.

REFERENCES

Aymé, S. and Schmidtke, J. 2007. Networking for rare diseases: a necessity for Europe. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz,* 50**,** 1477-1483.

Bruford, E. A., Lush, M. J., Wright, M. W., Sneddon, T. P., Povey, S. and Birney, E. 2008. The HGNC Database in 2008: a resource for the human genome. *Nucleic acids research,* 36**,** D445.

Burke, W. 2002. Genetic testing. *N Engl J Med,* 347**,** 1867-75.

Claustres, M., Horaitis, O., Vanevski, M. and Cotton, R. G. H. 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Research,* 12**,** 680.

Fokkema, I. F. A. C., den Dunnen, J. T. and Taschner, P. E. M. 2005. LOVD: Easy creation of a locus specific sequence variation database using an "LSDB in a box" approach. *Human mutation,* 26**,** 63-68.

Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J. and den Dunnen, J. T. 2011. LOVD v. 2.0: the next generation in gene variant databases. *Human mutation*.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. and McKusick, V. A. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research,* 33**,** D514.

Kohonen-Corish, M. R., Al-Aama, J. Y., Auerbach, A. D., Axton, M., Barash, C. I., Bernstein, I., Beroud, C., Burn, J., Cunningham, F., Cutting, G. R., den Dunnen, J. T., Greenblatt, M. S., Kaput, J., Katz, M., Lindblom, A., Macrae, F., Maglott, D., Moslein, G., Povey, S., Ramesar, R., Richards, S., Seminara, D., Sobrido, M. J., Tavtigian, S., Taylor, G., Vihinen, M., Winship, I. and Cotton, R. G. 2010. How to catch all those mutations--the report of the third Human Variome Project Meeting, UNESCO Paris, May 2010. *Human mutation,* 31**,** 1374-81.

Krawczak, M., Ball, E. V., Fenton, I., Stenson, P. D., Abeysinghe, S., Thomas, N. and Cooper, D. N. 2000. Human gene mutation database—a biomedical information and research resource. *Human mutation,* 15**,** 45-51.

Nabarette, H., Oziel, D., Urbero, B., Maxime, N. and Aymé, S. 2006. Use of a directory of specialized services and guidance in the healthcare system: the example of the Orphanet database for rare diseases]. *Revue d'épidémiologie et de santé publique,* 54**,** 41.

Patrinos, G. P., Al Aama, J., Al Aqeel, A., Al-Mulla, F., Borg, J., Devereux, A., Felice, A. E., Macrae, F., Marafie, M. J., Petersen, M. B., Qi, M., Ramesar, R. S., Zlotogora, J. & Cotton, R. G. 2011. Recommendations for genetic variation data capture in developing countries to ensure a comprehensive worldwide data collection. *Human mutation,* 32**,** 2-9.

Riikonen, P. and Vihinen, M. 1999. MUTbase: maintenance and analysis of distributed mutation databases. *Bioinformatics,* 15**,** 852-9.

Schieppati, A., Henter, J. I., Daina, E. and Aperia, A. 2008. Why rare diseases are an important medical and social issue. *The Lancet,* 371**,** 2039-2041.

Schneeweiss, S. and Avorn, J. 2005. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of clinical epidemiology,* 58**,** 323-337.