A "pivot" Model to set up Large Scale Rare Diseases Information Systems: Application to the Fibromuscular Dysplasia Registry

Laurent TOUBIANA[a,1], Adrien UGON[a], Alessandra GIAVARINI[b], Jérémie RIQUIER[a], Jean CHARLET[a], Xavier JEUNEMAITRE[b], Pierre-François PLOUIN[b] and Marie-Christine JAULENT[a]

[a] *INSERM, U1142, LIMICS, F-75006, Paris, France;*
Sorbonne Universités, UPMC Univ Paris 06, F-75006, Paris, France;
Université Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France.
[b] *Hypertension unit, department of genetics and rare vascular diseases reference center;*
Hopital Europeen G Pompidou (HEGP);
Paris-Descartes University.

Abstract. The SIR-FMD project is a partnership between the Department of Genetics and Reference Centre for Rare Vascular Diseases at the Georges Pompidou European Hospital in Paris and the Medical Informatics and Knowledge Engineering Laboratory of Inserm. Its aim is to use an ontological approach to implement an information system for the French Fibromuscular Dysplasia Registry. The existing data was dispersed in numerous databases, which had been created independently. These databases have different structures and contain data of diverse quality. The project aims to provide generic solutions for the management of the communication of medical data. The secondary objective is to demonstrate the applicability of these generic solutions in the field of rare diseases (RD) in an operational context. The construction of the French FMD registry was a multistep process. A secure platform has been available since the beginning of November 2013. The medical records of 471 patients from the initial dataset provided by the HEGP-Paris, France have been included, and are accessible from a secure user account. Users are organized into a collaborative group, and can access patient groups. Each electronic patient record contains more than 2,200 items. The problem of semantic interoperability has become one of the major challenges for the development of applications requiring the sharing and reuse of data. The information system component of the SIR-FMD project has a direct impact on the standardisation of coding of rare diseases and thereby contributes to the development of e-Health.

Keywords. Hospital Information Systems, Electronic Health Records, Information Storage and Retrieval, Knowledge Bases, Semantics, Chronic Diseases, Rare Diseases.

---

[1] Corresponding Author.

Introduction

The Department of Genetics and Reference Centre for Rare Vascular Diseases at the Georges Pompidou European Hospital (HEGP) in Paris contributes to the care of patients with various rare arterial diseases including Fibromuscular Dysplasia (FMD). It also conducts research in this field. A partnership between HEGP and the Medical Informatics and Knowledge Engineering Laboratory (LIMICS) at INSERM led to the SIR-FMD project, which aims to implement an information system, using a "pivot" model, for the French Fibromuscular Dysplasia Registry.

Recruitment of patients with renal and/or cervical FMD allows the constitution of a cohort and the collection of biological material; both should facilitate significant progress in understanding the origin and prognosis of this disease.

However, several databases have been created independently by different entities (Hypertension Unit, Department of Genetics) within the same hospital, leading to data dispersion and heterogeneity. These databases are heterogeneous in their structures and contain data of diverse quality. The heterogeneity causes problems of partitioning of the dataset and reduces its overall visibility. The resulting complexity makes it difficult or even impossible to conduct epidemiological studies.

This problem is particularly severe in national programs, such as the establishment of the National Data Bank for Rare Diseases (BNDMR) or "Rare Disease Cohorts" (RaDiCo), which aim to harmonize the collection of data for medical care and research. About 7,000 rare diseases are listed in the Orphanet database, such that the number of patients involved is very large (4 to 6% of the French population). Reference centres in different hospitals throughout France are confronted by the problem of dispersion and data communication.

These observations reveal the need for a generic information system, a single resource for various applications and allowing the reuse of datasets for different studies. We therefore aimed to establish generic solutions for the management of communicating medical data. The project involves the experimental development of a platform offering tools, methods and resources implementing semantic interoperability [1].

The construction of the French FMD registry was a multistep process. From the point of view of information processing, the main aim of this project was to reduce the structure and semantic heterogeneity in existing databases (including international databases). The secondary objective was to demonstrate the applicability of these generic solutions in the field of Rare Diseases (RD) in an operational context.

1. Methods

The methods involved the following methodological steps:

Step 1: Inventory of the existing databases. This required close collaboration between the IT team and an expert in Rare Diseases, an FMD expert in this case. Each item was explained in context and adapted by aligning local terminologies with reference terminologies. Reference terminologies are the terminologies used nationally or internationally to describe biomedical information, *e.g.*: MedDRA for the side-effects of medication; HPO for symptoms; ATC for medical products; ICD10; phenotypic terminology; gene names; CCAM; and Orphanet for rare diseases.

The material is primarily composed of sources containing FMD datasets already available at the HEGP: i) the initial DFM base, containing 400 cases. Each case record contains about 300 items; ii) a document with information about the activities of HEGP. This database contains data validated by the HEGP and includes about 50 items; iii) the "base PROFILE" created via a paper Case Report Form (CRF) which contains data for 100 cases from 12 French centres. Each case record of this database contains about 200 items; iv) the BAMARA database generated by the registration of a limited amount of information collected for all patients treated for one or more rare diseases.

The methods used for this step were needs analysis and conceptual modelling of the domain studied using the Unified Modelling Language (UML).
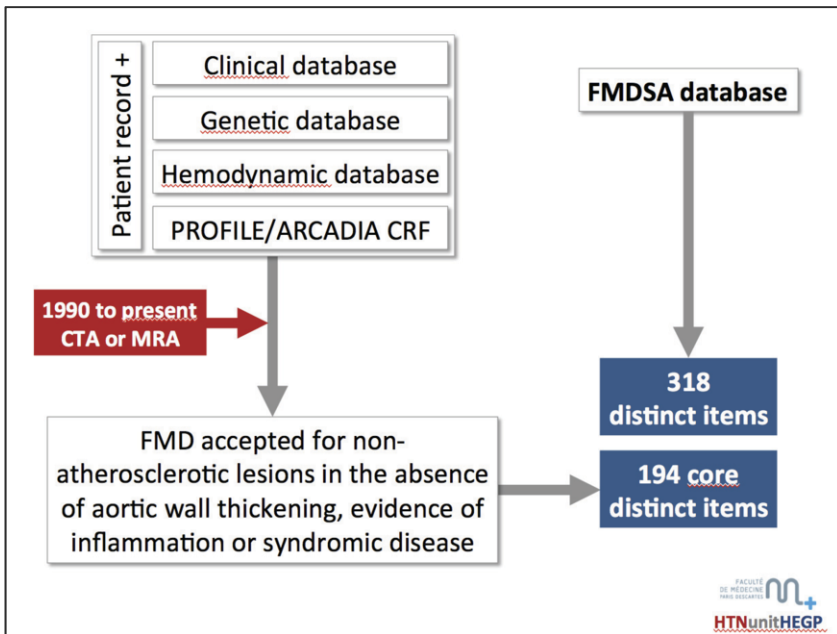
Figure 1. Clinical, haemodynamic and angiographic information were collected from patients with renal artery FMD to generate the initial HEGP database.

Step 2: Implementing the data model of the target database. FMD experts analysed the material described above and reached a consensus on a multidimensional conceptual model for FMD descriptions. This "pivot" model allows interfacing the different types of existing databases.

Step 3: Constitution of the semantic repository project (knowledge base). Not all terminologies are useful in their entirety in the context of a specific DFM and a semantic resource base must be built from terminology. This resource is used to annotate the information described in the database.
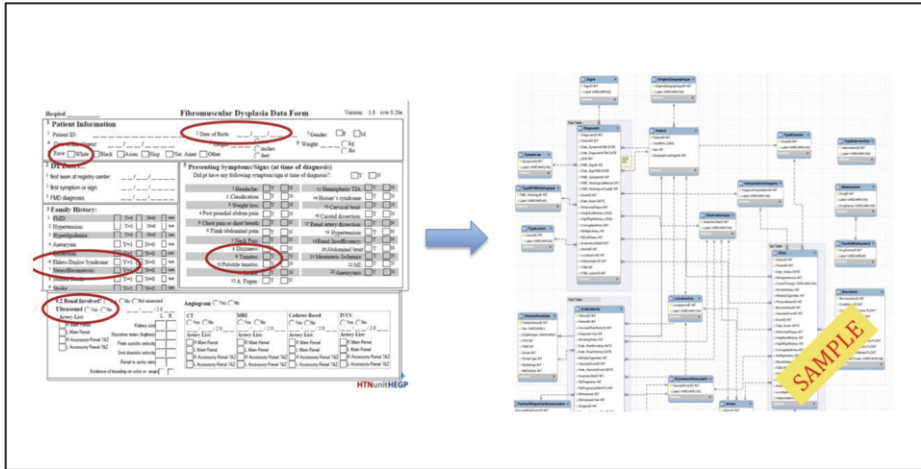
Figure 2. Multidimensional data "pivot" model to interface different existing databases. Implementing the data model of the target database.

Step 4: Implementing lookup tables of correspondence between items of different databases and semantic resources. Development of rewrite rules between databases based on lookup tables. We used semantic web tools to combine data and knowledge, and alignment tools for terminologies [2,3].

Step 5: Functional specifications of a web service for the online access of the target database. A platform was developed with the "MetaSurv" generator [4,5,6], allowing the target database to be made available online. The first functionality of the system is its ability to exploit data from heterogeneous databases. For doing this, IT processes (ETL "Extract, Transform and Load") were developed to implement the rewrite rules to supply the target database with the HEGP dataset. Initially, this platform has been made accessible only by members of the French network; however, it has been designed to be used internationally at a later stage.

2. Results

The SIR-FMD information system was developed in 2013 with the "MetaSurv" generator described elsewhere [6]. It collates a patient record in a standardized representation. On the client side, SIR-DFM relies on existing local Internet networking facilities. Via a web browser, the client connects to the interface, which is connected to the databases. SIR-DFM fulfils several requirements: scalability, portability, reliability, accessibility and cost-effectiveness oriented toward non-proprietary software.

The secure platform has been available as an evaluation platform since the beginning of November 2013. The medical records of 471 patients from the initial dataset provided by the HEGP have been included, and are accessible through a secure user account. Users are organized into a collaborative group, and can access patient groups. Electronic patient records contain more than 2,200 items.

This application is organized into multiple specifications which themselves provide access to intelligent electronic forms. These electronics forms are scalable, i.e. it is possible at any time to add new items that can be used for any other patients.

3. Discussion

Through a multidisciplinary health and IT partnership, we have developed innovative approaches to health information processing, addressing both methodological and application issue. The aims were to deal with practical problems confronting medical research and patient management, and secondly to contribute to advances in knowledge engineering, IT models, decision-support tools and translational computing. Therefore, the SIR-FMD project contributes to the development of the interactions between health professionals and researchers by improving the digital network between competence centres, national reference centres, national databases, patient records, cohorts and other sources of records. The project contributes to the organisation of activity in reference centres for rare diseases. In particular, it facilitates opening up of data sources at an European and international level [7]; indeed, its use of services for semantic interoperability and of ontologies are particularly well suited to multilingualism.

The problem of semantic interoperability has become one of the major challenges for the development of applications requiring the sharing and reuse of data. One of the ways in which the SIR-FMD project contributes to the development of e-Health is that its information system component contributes to the standardisation of coding of rare diseases. Taking semantics into account allows a network of different information systems (care, research, epidemiology) to be created. The applicability and generic nature of this model facilitates the implementation of original annotation services. The management of temporal aspects of the semantic annotations produced constitutes an innovative break with currently available tools.

References

[1] A. Kama, R. Choquet, G. Mels, C. Daniel, J. Charlet, M.-C. Jaulent *An ontological approach for the exploitation of clinical data.* Stud Health Technol Inform. 2013;192:142-6.

[2] S. Schulz, Schober D, Daniel C, Jaulent MC. *Bridging the semantics gap between terminologies, ontologies, and information models.* Stud Health Technol Inform. 2010;160(Pt 2):1000-4.

[3] GA.Komatsoulis, Warzel DB., Hartel FW., Shanbhag K., Chilukuri R., Fragoso G., Coronado S., Reeves DM., Hadfield JB., Ludet C., Covitz PA. *caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability.* J Biomed Inform. 2008 Feb;41(1):106-23

[4] L. Toubiana, M.Cuggia. *Big data and smart health strategies: findings from the health information systems perspective.* Yearb Med Inform. 2014 Aug 15;9(1):125-7.

[5] P. Landais, Messiaen C, Rath A, Le Mignot L, Dufour E, Ben Said M, Jais JP, Toubiana L, Baujat G, Bourdon-Lanoy E, Gérard-Blanluet M, Bodemer C, Salomon R, Aymé S, Le Merrer M, Verloes A; *CEMARA task force. CEMARA an information system for rare diseases* Stud Health Technol Inform. 2010;160(Pt 1):481-5.

[6] L. Toubiana, S. Moreau, G. Bonnard *MetaSurv: Web-platform generator for the monitoring of health's indicators and interactive geographical information system.* Stud Health Technol Inform. 2005;116:989-93.

[7] P.-F.Plouin, L. Toubiana, M.-C. Jaulent, X. Jeunemaitre *French FMD registry: organizational structure, data elements, summary of findings. International Fibromuscular Dysplasia (FMD)* Research Network symposium May15th and 16th 2014 Cleveland, Ohio USA