

CEMARA: a Web dynamic application within a n-tier Architecture for Rare Diseases

Claude MESSIAEN¹, Loïc LE MIGNOT¹, Ana RATH³, Jean-Baptiste RICHARD¹,
Eric DUFOUR², Mohamed BEN SAID¹, Jean-Philippe JAIS¹, Alain VERLOES⁴,
Martine LE MERRER⁵, Christine BODEMER⁶, Geneviève BAUJAT⁵, Marion
GERARD-BLANLUET⁴, Eva BOURDON-LANOY⁶, Rémi SALOMON⁷, Ségolène
AYME³, and Paul LANDAIS¹ for the CEMARA task force

¹ *Paris Descartes University, UPRESS EA 4067 and Service de Biostatistique et
d'Informatique Médicale, APHP - Necker-Enfants Malades Hospital* ; ² *SELIC*;
³ *Orphanet and INSERM SC 11* ; ⁴ *Department of Genetics, APHP- Robert DEBRE
Hospital*; ⁵ *Department of Genetics*, ⁶ *Dermatology unit and* ⁷ *Paediatric Nephrology
unit, Necker-Enfants Malades Hospital, Paris - France*

Abstract: Rare diseases include a group of conditions characterized by a prevalence lower than 5 per 10,000 in the community. In France, any rare disease affects less than 30,000 patients and often much less. Three to 4% of children and 6% of the population in Europe are affected. It is a true public health stake since most diseases do not have any curative treatment. In France, the Ministry of Health has initiated a National Rare Diseases Plan. Twenty five out of 132 labelled Reference Centres (RC) decided to share a common Information System named CEMARA. It is dedicated to collect continuous and complete records of all patients presenting with a rare disease, and their follow-up. The main objective of CEMARA is to contribute to the missions of the RC regarding the registration and description of their activities, coordination of the network of their correspondents, organization of the follow-up of rare diseases, and analysis of the epidemiological patterns. A description of CEMARA is provided as well as its cooperation with Orphanet and Genatlas, and a presentation of 11803 current records collected by more than 300 health care professionals belonging to more than 70 sites.

Keywords: Dynamic Web interface; Decisional Information System; Thesaurus; Rare Diseases;

Introduction

Rare diseases encompass a group of conditions characterized by a low prevalence. In Europe, a disease is "rare" when less than one out of 2,000 people suffers from a specific clinical syndrome. In France, rare diseases might affect 30,000 patients. For

¹Corresponding author : Pr Paul LANDAIS – Service de Biostatistique et d'Informatique Médicale – Hôpital Necker-Enfants Malades 149 rue de Sèvres 75743 Paris, France email : landais@necker.fr

the European Community, it means up to 230,000 patients. According to the definition, 5000 to 7000 rare diseases are currently identified. Three to 4% of children and 6% of the population in Europe is affected. It is a true public health concern since most diseases do not have any curative treatment. In France, the Ministry of Health has initiated a National Plan for Rare Diseases; 132 Reference Centres (RC) [1] have been labelled for a group of diseases or, even a given disease such as “constitutional bone disorders” or “juvenile arthritis”, respectively. Among their missions, the RC are involved in the epidemiological monitoring of pathologies they are in charge of, in order to better assess their distribution on the territory, the existing management networks, as well as social, educational and familial repercussions of these pathologies. An Information System named CEMARA (CENTres MALadies RAres) was set up in order to collect continuous and complete records of all patients presenting with a rare disease, and when applicable, their follow-up.

The main objective of CEMARA is to contribute to the missions of the RC regarding the registration and description of their activities, coordination of the network of their correspondents, organization of the follow-up of rare diseases, and analysis of the epidemiological patterns.

Material and Methods

CEMARA professional network

Presently, 25 centres are members of the CEMARA network in France. Each centre is composed of one or more clinical units, managed by a coordinator and distributed on the whole French territory. A medical correspondent is designated for each site (a map is available at <https://cemara.org/presentation/show.jsp?sm=membres>). Close to 300 members are registered for 70 sites. All gave their consent and signed the confidentiality charter of the network. CEMARA obtained its agreement in accordance with the French Protection Act.

Thesauri: conception and constitution

For the diagnosis labels of the patient records, a classification was set up with the health professionals corresponding to their needs. For rare diseases, the International Classification of Diseases 10th version (ICD10) does not provide the necessary support since its granularity is not appropriately designed for such a target, and many rare genetic disorders are not listed in. Moreover, the structure of ICD10 is not adapted to a field where nosology evolves very rapidly (for example following new genetic discoveries leading to splitting existing clinical diagnoses) and where 300 new syndromes are reported every year.

We thus linked our work to Orphanet [2] that designed specific thesauri dedicated to rare diseases. Orphanet has a know-how in the domain and is member of the European Rare Disease Task Force². It provides an ontological support to experts for each field of CEMARA. The specificity of this collaboration provides a shared ontology for each thesaurus of CEMARA with a lexical uniqueness for the labelling of each disease including the management of synonyms and eponyms associated to some diseases or disorders (for instance De Toni-Debré-Fanconi syndrome and proximal tubulopathy). Furthermore, the associated nosography/nosology helps health professionals from different fields to share common terms. Orphanet can also benefit from a feedback through CEMARA by studying, at steady rate, the associations of diagnosis with keywords and/or notes entered by the physicians,

²Rare Disease Task Force : <http://www.rdtf.org/> [last connection on Feb 28, 2008]

which can lead to evolutions of the thesaurus. Regarding gene mutations we use GENATLAS, which contains relevant information with respect to gene mapping and genetic diseases [3]. A specific link was also developed towards this database. OMIM codes will be also provided.

Information System (IS): Design and Implementation

The IS of CEMARA collates in a standardized representation a minimal patient record elaborated by paediatric and adult health professionals. CEMARA aimed at fulfilling several requirements: scalability, portability, reliability, accessibility and cost effectiveness oriented toward non-proprietary software. The architecture of CEMARA is based on a n-tier architecture. Via a web browser, the client tier connects to the middle tier, which is connected to several databases: the production database, the geographical dictionary database, and the thesaurus database. A data warehouse and a geographical information system allowing queries and representation are in progress. Their framework is close to an already available application for end-stage renal disease [3]. The middle tier supports client services through Web containers and business logic services through component containers. Business logic in the middleware is interfaced with a SGBD dependant handler which supports the transactions toward the database. At the client side, CEMARA relies on existing local Internet networking facilities and on a widely spread computer configuration in medical settings. Access via a personal digital assistant is also available.

³GENATLAS <http://www.dsi.univ-paris5.fr/genatlas/> [last view Feb 28, 2008]

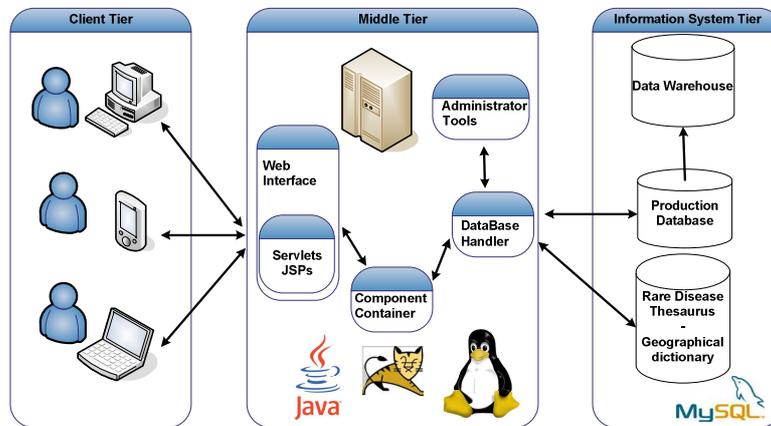


Figure 1 - N-tier architecture for CEMARA

Maintenance and evolutions are made centrally, which reduce deployment costs and delays. The structure of the production database matches the structure of the CEMARA network and respects privacy, confidentiality and security [4]. A specifically developed double-entry function prevents identity doubles [5].

The setting-up began on May 23rd, 2007, initiated by fifteen centres. Coordinators and users were trained, through centralized and on-site sessions. A clinical research assistant organized the training and technical assistance.

CEMARA also includes institutional collaborations with: the Ministry of Health, the National Institute for Health Surveillance, the “Haute Autorité de Santé” and the National Health Insurance Fund concerning the design of national protocols for diagnoses, treatment and follow-up, as well as for cost analyses, associations of patient families for improving patient care, university hospitals, for the coordination and connection of professionals through the CEMARA network.

Interoperability

CEMARA was conceived in order to communicate with other sources of information: the use of XML, as an exchange format, permits a greater flexibility and better capacities to exchange data with other information systems such as French Medical Insurance system or Hospitals Information systems. It also allows importation of former databases (Figure 2).

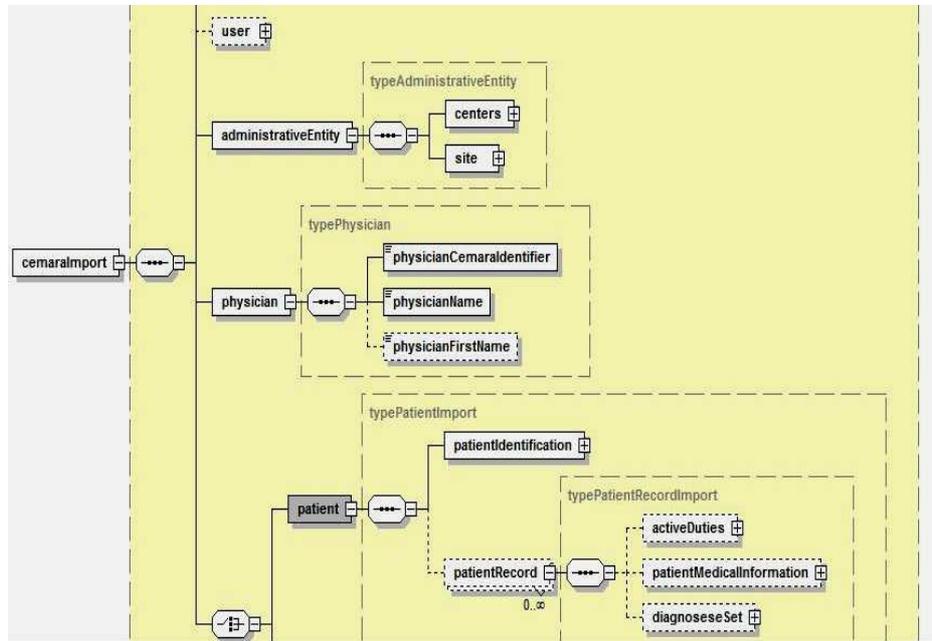


Figure 2- CEMARA: Part of an XML data import scheme

Data entry sheets

All the centres share a common core of information that includes identification data (for the index case and relevant family members), diagnosis, context and medical activity. Patient identification is based on: name, surname, birth date, death date, place of residence... Modalities of medical activities and modes of recruitment are recorded. Optional information is gathered about the patient's medical history e.g. antenatal and/or neonatal information. One or more diagnoses are labelled using the corresponding thesaurus, as well as the status of this diagnosis (probable, confirmed, on, in process) a chromosomal description of the anomaly if required, and additional keywords for atypical signs and symptoms, or for the patients with an unknown diagnosis. Keywords were prepared by the steering committee. Diagnosis selection use Asynchronous JavaScript And XML (AJAX) queries. Complementary information for specific diseases are described in the so called “petals”, attached to the core data, and focused on specific data collection.

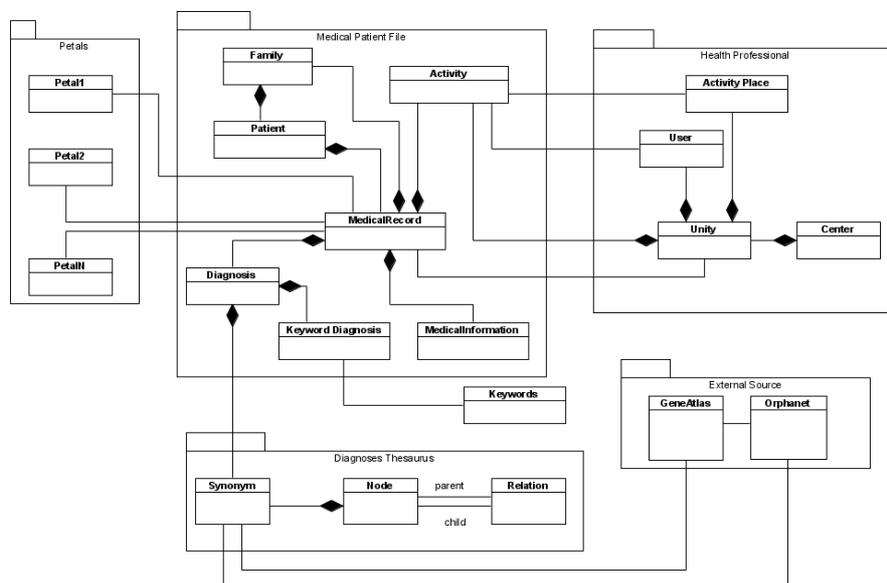


Figure 3 – UML scheme of CEMARA

Results

More than 300 health care professionals belonging to more than 70 sites of the 25 RC contributed to the data collection. Currently, the data set gathers 11,803 records and includes 6076 male patients (340 fetuses) and 5536 females (345 fetuses). For 191 fetuses gender was non-determinable. Median age was 6.8 years. Patients with rare diseases were mainly referred to centres by paediatricians (4168) and medical specialists (2918), while for foetus by gynaecologists (612) and by centres for prenatal diagnosis (273). These patients were mainly enrolled via outpatient clinics (9134) or hospital wards (689). Foetal cases were described either during pregnancy (519), or after spontaneous miscarriage or termination of pregnancy (446). As expected, the number of patients per diagnosis is limited, even for the most prevalent diagnoses. Moreover, roughly 1/3 records are classified as “unknown diagnosis”, meaning that no diagnosis has been reached, which is a common occurrence for patients with multiple malformations, whose syndrome do not fit with currently delineated entities.

Discussion

The EC program for rare disease, initiated in 2003, aimed at improving information and knowledge for the development of public health [6]. In France, the rare diseases plan played a key role to promoting the management of these diseases [7]. The RCs have to standardize their information and structure its collection. It triggered the development of CEMARA. The experience shared with Orphanet and Genatlas enabled us to extend the expertise gathered with the French national REIN program [2] we developed for end-stage renal disease. It strengthens the durability of our application. Furthermore, the application is based on open source software, which allows lower development costs. The conception of an evolutionary design conveyed by its modularity ensures its scalability. Within nine months, close to 12,000 rare disease cases were recorded. It shows the involvement of the professionals and the utility of this application. Users of databases often claim that they don't have a real feedback on the data they feed. Therefore, we provide every three months reports on activity and diagnoses. Each participant can use a multicriteria search tool to explore his case-mix. We also prepared advance epidemiological analysis computed with R [8] on anonymous and aggregated data using our experience derived from the SIGNE, a web geographical information system [9]. CEMARA allows assessing the geographical distribution of existing cases for providing the appropriate offer of care. It allows better describing new cases of rare diseases and organizing their follow-up, and identifying the medical and psycho-social needs. The use of a web-based interface intends to help medical recording, management and reporting of rare diseases. It aims to contributing to the public health adaptation to the offer of care with

a better knowledge of diseases, thus providing a support to health care decision-making in this complex and specific domain.

Acknowledgments

The coordinators of the 25 CEMARA centres are warmly acknowledged for their participation to the program as well as the members of all the centres

References

-
- [1] http://www.alliance-maladies-rares.org/IMG/pdf/Liste_Centres_de_Reference.pdf
 - [2] CEMARA et Orphanet : une collaboration au service des centres de référence,
<http://www.orpha.net/actor/Orphanews/2007/070321.html> [last view Feb 28, 2008]
 - [3] Landais P, Simonet A, Guillon D, Jacquelinet C, Ben Saïd M, Mugnier C, Simonet M. SIMS REIN: a multi-source information system for end-stage renal disease. *C R Biol.* **325** (2002), 515-28.
 - [4] Rindfleish TC. Privacy, Information Technology and Health Care. *Communications of the ACM* August **40** (1997), 93-100
 - [5] Le Mignot L, Mugnier C, Ben Saïd M, Jais JP, Richard JB, Le Bihan-Benjamin C, Taupin P, Landais P. Avoiding doubles in distributed nominative medical databases: optimization of the Needleman and Wunsch algorithm. *Stud Health Technol Inform.* **116** (2005), 83-8.
 - [6] http://ec.europa.eu/health/ph_projects/2004/action1/action1_2004_10_en.htm [last view Feb 28,2008]
 - [7] http://www.sante.gouv.fr/htm/actu/maladie_rare/plan.pdf
 - [8] The R Project for Statistical Computing: <http://www.r-project.org/> [last view Feb 28, 2008]
 - [9] Richard JB, Toubiana L, Le Mignot L, Ben Said M, Mugnier C, Le Bihan-Benjamin C, Jais JP, Landais P. A Web-based GIS for health care decision-support. *AMIA Annu Symp Proc.* (2005), 365-9.