On the Complexity of Requirements Engineering for Decision Support Systems: The CID Case Study

Ruth Raventós, Stephany García, Oscar Romero,
Albert Abelló and Jaume Viñas

Department of Service and Information Systems Engineering.
Universitat Politècnica de Catalunya, Barcelona.
{oromero,aabello}@essi.upc.edu
http://www.essi.upc.edu

Abstract. Chagas Disease is classified as a life-threatening disease by the World Health Organization (WHO) and is currently causing death to 534,000 people every year. In order to advance with the disease control, the WHO presented a strategy that included the development of the Chagas Information Database (CID) for surveillance to raise awareness about Chagas. CID is defined as a decision support system to support national and international authorities in both their day-by-day and long-term decision making. The requirements engineering to develop this project was particularly complex and the Pohl's Framework was followed. This paper describes the results of applying the framework in this project. Thus, it focuses on the requirements engineering stage. The difficulties found motivated the further study and analysis of the complexity of requirements engineering in Decision Support Systems and the feasibility of using said framework.

Key words: Decision Support Systems, Business Intelligence, Requirements Engineering

1 Introduction

The World Health Organization (WHO), a United Nations (UN) agency founded in 1948, is the directing authority for health matters, responsible to provide leadership on global health matters by setting standards and providing technical support to monitor the health trends [1]. In October 2010, WHO launched the First Report on Neglected Tropical Diseases (NTD) [2], which included the "Chagas disease" among others. Chagas Disease (also known as Human American Trypanosomiasis) is classified as a life-threatening disease caused by a parasite named 'protozoan parasite'. According to the report, Chagas is nowadays usually found in 21 Latin American countries, where the disease is transmitted to humans by the faeces of infected triatome bugs. The WHO report has estimated seven to eight million people infected worldwide, although the concern is that the number of infected people keeps increasing year by year.

By being categorized as NTD, not much attention is drawn by the Chagas disease. In fact, worldwide population is not aware that statistics indicate that NTDs have already caused substantial illness for more than one billion people of our worlds poorest countries. Due to the worrying situation, in 2010 the World Health Assembly (WHA) decided to approve the Chagas disease: control and elimination resolution (SDC10). But it was not until 2013 that the Tricycle Strategy of the Programme on Control of Chagas disease, was presented (VN). The presented strategy had the objective to advance in the disease control by creating a Chagas Information Database (CID) system for surveillance to raise awareness on the Chagas disease. In particular to facilitate the access to information from different sources related to Chagas disease, to then exploit the data by means of dashboards, and analytical tools (e.g. visualize disease statistics, maps and diagrams for transmission routes; infested patients).

The CID project involved specialized stakeholders from very different domains and organizations such as health specialists and national authorities from (or linked to) WHO, entomologists from different research centers of Argentina and software analysts from Universitat Politècnica de Catalunya (UPC-BarcelonaTech), who had complementary yet non-trivially integrable perspectives of the Chagas disease. For instance, entomologists set their focus on studying the triatome bugs (the main disease vector) and how to interrupt this transmission by such vectors. Doctors, in turn, focused on patient diagnosis and treatment and the WHO monitors all relevant aspects of the disease, including provided infrastructures and drug delivery in each country, and provided the multi-perspective expertise needed to glue all of them together. In consequence to this, software analysts were expected to deal with the highly-complex domain of the Chagas disease to begin collecting the requirements before creating the CID system.

From the very beginning, the project was acknowledged as a highly complex one. First of all, the WHO Programme did not have a pre-determined or defined proposal of the information and surveillance system they need. In addition, the interdisciplinary knowledge among the principal stakeholders doctors, entomologists, requirement analysts and database experts - to gather together to reach consensus on the crucial project issues was a true challenge. Several meetings were held in order to let business and technical analysts understand the basics of Chagas: i.e., diagnostic, frequently used drugs, vector control and health systems.

In addition to the understanding among the different stakeholders mentioned above, some other problems were found, mainly related to data, during the requirement engineering process.

In order to achieve the goal of eradicating Chagas (CID's ultimate goal), we needed to integrate data corresponding to the different facets (i.e., medical, vectorial, etc.) of the Chagas disease. Its main difficulty came from the heterogeneity (either in the storage system or semantics) in the multiple data sources and how to cross-analyze all such information adding the geographical and temporal dimensions and treat them at different granuralities (of analysis).

Even though the Chagas disease initially appeared in South America today, due to migratory flows, it is scattered around the whole world. Thus, information can come from any country and users can speak any language. This clearly raises a linguistic problem that had to be overcome by a multilingual interface.

Another issue was the quality of data. To decide what mechanisms would be put in place to check as many constraints as possible to ensure the quality of the data introduced.

As mentioned, the CID project is primarily a DSS. Thus, all relevant data related to Chagas must be gathered, properly stored and managed (this means integrated and cleaned) in order to allow the WHO to analyse and cross data from disparate, yet relevant, from different points of view.

Therefore, in order to systematically tackle the complexity of the project, during the requirements engineering phase we followed the Requirements Engineering Framework of K. Pohl [3] to understand, agree and document all the requirements of the system to be built in compliance with the relevant documentation formats and rules.

This framework has become a well-established for requirements engineering mainly because of the following reasons:

- The first and main reason was that the framework defines the major structural blocks and elements of a Requirements Engineering process (e.g. Elicitation, Negotiation, Goals, etc.).
- It provides a comprehensible and well-structured base for the fundamentals, principles and techniques of Requirements Engineering.
- It is not adhered to any specific methodology, neither to any type of software project.
- In addition, this framework consolidates various research results and has been proven to be successful in a number of organizations for structuring their Requirements Engineering process.

The rest of the paper describes the Requirements Engineering process of the CID Project by applying the Pohl's Framework emphasizing the complexity found.

The importance of this project, in addition to its own purpose, is that many of the lessons learnt can be mapped to most decisional systems. For this reason, this case study has been the base of further research about requirements engineering for Decision Support Systems and has been the reason for further research conducted in this area in our research group to propose a new framework for requirements engineering of Decision Support Systems, which is nowadays missing. As previously discussed, to cover this gap we followed the Requirement Engineering Pohl's Framework and here we show by example how each of its constructs can be adapted to decisional systems. All in all, this paper is an excellent case study showing all the complexities found at this stage when developing a DSS.

2 Requirements Engineering of the CID Project

Every single project must begin with the statement of a requirement. In computer science, a requirements refers to the description of how a software product should perform [4]. The IEEE 610.12-1990 [5] defines "requirement" as follows:

1. "A condition or capability needed by a user to solve a problem or achieve an objective"
2. "A condition or capability that must be met or possessed by the system or system component to satisfy a contract, standard, specification, or other formally imposed document"
3. "A documented representation of a condition or capability as in (1) or (2)"

The CID project was promoted by the World Health Organization (WHO) to change the current reality of a neglected disease, Chagas. The initial vision was to develop a software system that would help controlling and managing all the relevant information related to the disease, i.e., detection, treatment, drug availability / delivery, transmission interruption, normative, etc. so, it would be used as support to decide what actions to take to combat the disease.

More precisely, the main objective of this project is to design a data warehouse for the global WHO information and surveillance system to control/eliminate Chagas disease, which should be based on the following components (to be gathered world-wide):

- Current legal, normative, political and economic frameworks about Chagas.
- Information and surveillance systems in place (with information on circulating parasites, presence of infected and non-infected vector insects, hosts and humans).
- Healthcare data (including screening, diagnosis and care measures and coverage, through biomedical and psychosocial approaches).
- Chagas disease affected population (including population at risk of being infected, estimated number of infected and ill people, already diagnosed and treated Chagas disease patients, from congenitally infected newborns to adults)
- Transmission routes (active and interrupted transmission routes and implemented measures to prevent and interrupt them).

According to main objective of the project, it was necessary to perform a specific group of tasks to analyse the needs, goals, stakeholders and context of the required system. It meant to carry out the following tasks:

- The state of the art related to the system. Is there any other similar system out there?
- Data sources from where to extract the needed data (quality and risks).
- Scenarios, use cases and boundary of the system to be built.
- Software requirements (functional and quality requirements) and constraints.
- Conceptual model of the system.
- Estimation of the timing and budget of the project.

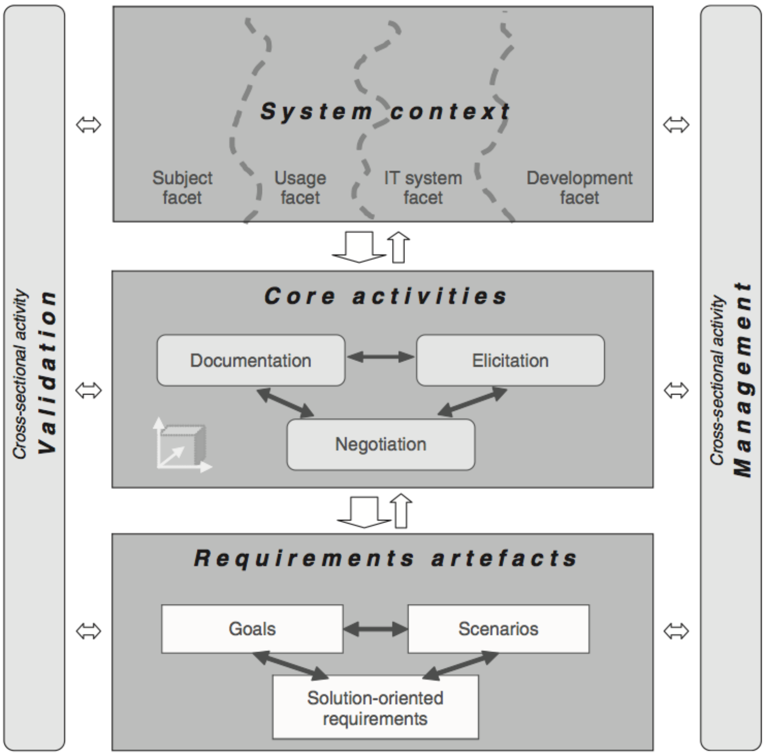

Fig. 1. Requirements Engineering Pohl's Framework [3]

Consequently, the requirements engineering stage was divided in three main tasks (in parenthesis, the amount of hours devoted to each task): state of the art (90h), definition of actors, use cases and system sequence diagrams (150h) and definition of the conceptual model (100h). The project spanned for 3 months.

The main distinction of defining the requirements and developing the design of a system is that the requirements focus on "what" the system should do, rather than "how" it should be done [4]. Meaning that the requirements are centered on defining the problem "what should be developed", while the design specifies "how the system should be developed" [3].

Therefore, Requirements Engineering is a cooperative, iterative and incremental process which aims to ensure the following [3]:

1. All relevant requirements are explicitly known and understood at the required level of detail.
2. Sufficient agreement about the system requirements is achieved between the stakeholders and the IT team involved.
3. All requirements are documented and specified in compliance with the relevant documentation/specification formats and rules.

The "K. Phol's Requirements Engineering Framework" (see figure 1) consists of three main building blocks and two cross-sectional activities which are explained in the following sections of this document.

3 System Context

The requirements in every software system are heavily influenced by the system context. Consequently the requirements for these systems may not be defined without considering the environment where the system is embedded. These include material or immaterial objects such as the following: technical or non-technical systems, people, technologies, business processes and work-flows, laws, already existing hardware and software components, other systems that will interact with the new system, physical laws, safety standards, system users, etc. Therefore, the understanding of the context is a crucial preparation in order to develop a successful requirements specification.

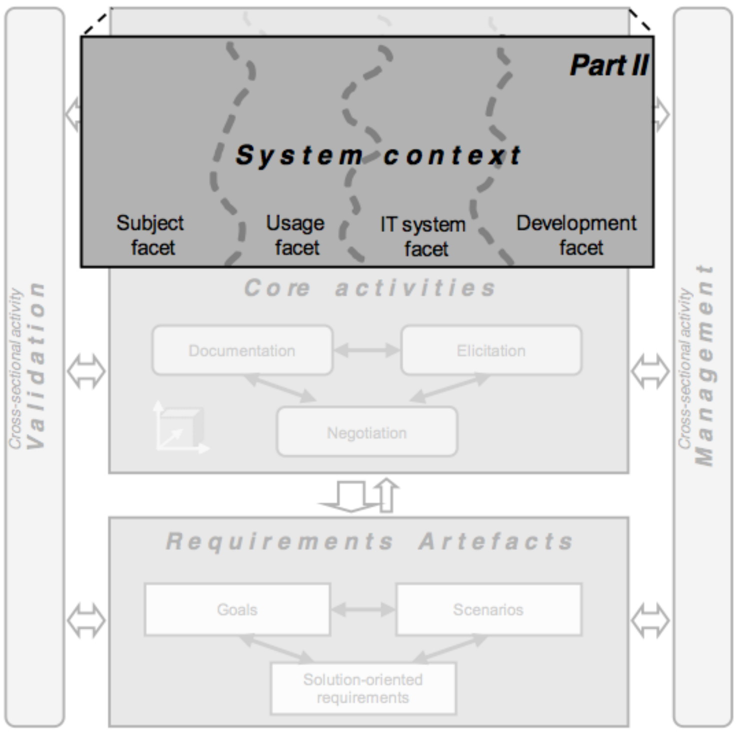

Fig. 2. System Context of Requirements Engineering Pohl's Framework [3]

The system context is structured in the following four different context facets as shown in Fig. 2 which must be considered as the prerequisite phase on the Requirements Engineering process:

1. **Subject Facet**
   This facet is composed mainly of the objects and the events that must be represented in the system; given that the system must process or store information about these. Additionally, this facet also includes aspects that constraint the representation of the system such as data privacy laws disallowing the storage of a certain type of data or accuracy constraints.

2. **Usage Facet**
   This facet which refers to all aspects concerning the usage of the system by people or other systems such as: user groups with specific characteristics, usage goals, desired usage work-flows, or modes to interact with the system through the system's interface.

3. **IT Facet**
   This facet encompasses all aspects related to the operational or technical environment in which the system is deployed. IT facet includes hardware and software components (e.g. software platforms, networks, existing components), IT strategies and policies.

4. **Development Facet**
   Last but not least, the development facet encompasses all the aspects which influence the system's development process, for instance: process guidelines, development tools, quality assurance methods, and other techniques to assure quality in the system. This may always be restricted by law, for instance if there are only certain tools the client requested to be used during the systems' development.

3.1 CID Project Subject Facet

To elicit the information that should be represented in the CID system, first, we needed to understand what is the Chagas disease, its distribution around the World, the signs and symptoms, the transmission, treatment, control and prevention of the disease and its relation with AIDS. All the aspects presented in this section are of extreme relevance to support the WHO in its ultimate objective of eradicate Chagas.

The disease Chagas disease[1] (also known as Human American Trypanosomiasis) is a potentially lifethreatening illness caused by the protozoan parasite, Trypanosoma cruzi. It is found mainly in endemic areas of 21 Latin American countries, where it is mostly transmitted to humans by the faeces of infected

[1] Chagas disease is named after Carlos Ribeiro Justino Chagas, a Brazilian doctor who first discovered the disease in 1909 [6]

triatome bugs. At present about 7 to 8 million people worldwide are estimated to be infected with T. cruzi, including disease non-endemic countries where Chagas disease has been spread through population movements, mainly migration. Chagas disease and its causal agent, T. cruzi, have existed for millions years in the Americas in sylvatic cycles (enzootic disease[2]). Humans arrived in the Americas at least between 26,000 and 12,000 year ago, either overland or across the oceans. Settled communities started to implement agricultural activities and domestication of animals. Over the past 200-300 years, deforestation because of agriculture, livestock rearing and the opening of transportation routes (railroads and highways), cause that Triatoma vectors gradually lost their primary food source: wild-animal blood. Vectors[3] adopted the areas surrounding human dwellings and started to feed on the blood of domestic animals and humans. A new cycle was established and Chagas disease became a zoonosis[4]. Endemic Chagas disease in Latin America began as neglected disease of poor, rural and forgotten human populations. Wild triatomines progressively adapted to the domestic environment, living in cracks in the mud walls and roof of huts. In the second half of the XX century Chagas disease was recognized as an important urban medical and social problem because of progressive urbanization of the rural population in Latin America. In 1960, a WHO Expert Committee estimated that about 7 to 8 million people per year were infected due to blood transmission in Latin America [6].

Distribution Chagas disease occurs mainly in the following 21 Latin America countries: Argentina, Belize, the Bolivarian Republic of Venezuela, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, French Guyana, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, the Plurinational State of Bolivia, Suriname and Uruguay, but no cases have been detected so far in Caribbean countries. Additionally, in the past decades Chagas disease has increased the spread worldwide and it is now detected in the United States of America, Canada, 17 European countries (mainly in Belgium, France, Italy, Spain, Switzerland and the United Kingdom, but also in Austria, Croatia, Denmark, Finland, Germany, Luxembourg, the Netherlands, Norway, Portugal, Romania and Sweden) and some Western Pacific countries, mainly Australia and Japan, see Fig. 3. The disease presence outside disease endemic countries is mainly due to population mobility between Latin America and the rest of the world, but less frequently due to infection through blood transfusion, congenital transmission (from infected mother to child) and even organ transplantation, laboratory accident and child adoption [6].

The above-mentioned population movements have created new epidemiological, economic, social and politic challenges for rich and poor countries where T. cruzi has been spread [6]. In countries with absence of insect vectors, such the European ones, main threats come from infected blood transfusions, organ

[2] a disease of wild animals [6]

[3] Vector that have been transmitting the Chagas disease among animals for 10 million years [6]

[4] A disease that is transmitted between animals and humans endemically [6]

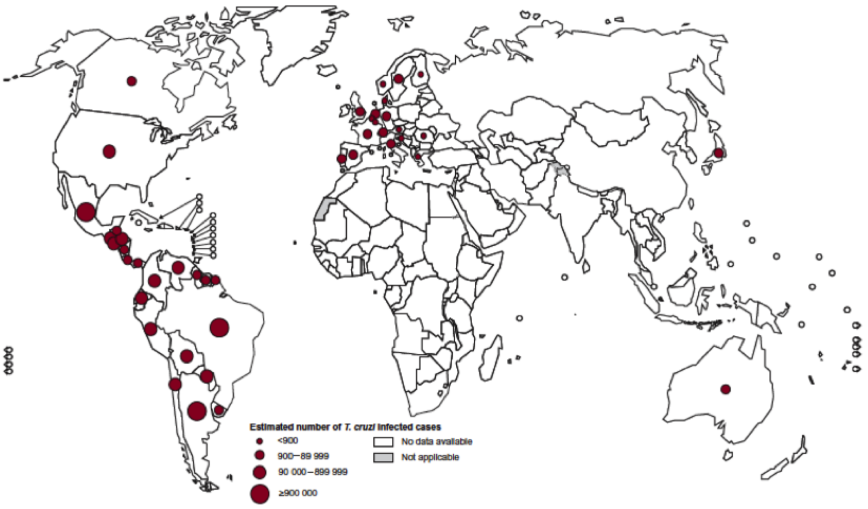

Fig. 3. Global distribution of cases of Chagas disease, based on official estimates, 2006-2010 [6]

transplantations or congenital transmission. On the other hand, the number of cases of infection among people travelling to Latin America for tourism or work reasons has been increasingly reported. Therefore, these countries are progressively realizing that they need to improve their information and surveillance systems at national and supranational levels, increase detection and care of patients with Chagas disease, implement additional controls for blood banks and organ transplantation, and include the differential diagnosis of Chagas disease within travel medicine.

Signs and Symptoms Chagas disease presents itself in two successive phases, an acute phase and a chronic phase.

1. Acute phase: The acute phase lasts for about two months after infection. Historically, with high vector infestation of dwellings and consequent intradomiciliary transmission, the majority of cases were detected before the age of 15 years, with the highest frequency between the ages of 1 and 5 years [6]. Acute Chagas disease, however, can occur at any age and at present, for instance, an important number of cases are related to oral transmission in territories such as the Amazon basin or humid Andes. During the acute phase, a high number of parasites circulate in the blood. Most T. cruzi infected people present mild or absent symptoms, but can also have fever, headache, enlarged lymph glands, pallor, muscle pain, difficulty in breathing, swelling and abdominal or chest pain. And despite most acute infections are unrecognised [6], in less than 50% of people bitten by a triatomine bug, characteristic first visible signs can be a skin lesion or a purplish swelling of the lids of one eye, Fig.4.

Fig. 4. Sign in acute phase of Chagas disease [7]

In most cases, congenital transmission, produced by the transmission of T. cruzi parasites from the infected mother to the child, occurs at the end of the pregnancy or delivery and is a specific acute case. Most infected infants are asymptomatic and have normal weight and vital signs, but some, mainly those born prematurely, may show clinical manifestations, such as low body weight, jaundice, anaemia, hepatomegaly or ocular lesions, among others [6].

2. Chronic phase: The chronic phase of Chagas disease starts when the parasite falls to undetectable levels and general symptoms and clinical or manifestation are initially not present. During the chronic phase, the parasites are hidden mainly in target tissues: the heart and digestive muscle. Current knowledge on all changes produced during the chronic phase is still incomplete, and this initial asymptomatic period of the chronic phase is called indeterminate or latent form. About 50%-60% of infected people will remain in this form lifelong [6]. On the other hand, up to 40% of chronically infected individuals will develop a cardiac, digestive, neurological or mixed form. Cardiac manifestations are the most important clinical consequence of the T. cruzi infection. Up to 30% of patients will suffer from cardiac disorders. These forms occur mainly 10-20 years after the acute phase of the disease and include a broad range of type of damage and clinical manifestations, from mild symptoms to heart failure, arrhythmias and sudden death [6]. The digestive form of Chagas disease has been described in populations south to the Equator line. The infection can lead to dysfunction of the digestive system with dilation of oesophagus and colon (megaoesophagus). Up to 10% suffer from digestive, neurological or mixed alterations [6].

Transmission One of the most important challenges to control/eliminate the Chagas disease is to interrupt its transmission. Chagas disease can be transmitted by six different transmission routes, mainly vectorial, oral, congenital (mother to child) and blood tranfusion routes, but also by organ transplantation and laboratory accidents. The vectorial transmission route has been the most important one. It is mainly found in the 21 endemic countries of Latin America, where most of the insects infected with T.Cruzi are found. These bugs, which typically live in the cracks of poorly-constructed homes of rural or suburban areas, normally hide during the day and feed on human blood when the night falls [6]. In fact, T.Cruzi parasites are mainly transmitted by the contact with faeces and urine of blood-sucking triatomine bugs. The itching caused by the bits leads humans scratching or putting in contact infected faeces with the bite wound, other skin lesions or mucous membranes (mainly in the mouth and eyes), allowing the parasites to enter the circulation system [6]. Another vector-related route of transmission is the oral or food-borne transmission, through the ingestion of food contaminated with triatomine bugs faeces, very prevalent in territories such as the Amazon basin. The other transmission routes T. cruzi can also be transmitted by [6]:

– Blood transfusions using blood from infected donors.
– Vertical transmission (mother to child).
– Organ transplants using organs from infected donors.
– Laboratory accidents.

Treatment In order to kill the parasite, Chagas disease can be treated with either benznidazole [8] or nifurtimox [9]. Both medicines are almost 100% effective in curing the disease if given soon after the infection at the beginning of the acute phase. However, the efficacy of both reduces the longer a person has been infected. Treatment is also indicated for those in whom the infection has been reactivated due to immunedepression, for infants with congenital infection and for patients during the early chronic phase [6]. Treatment should be offered to infected adults, especially those with no symptoms. Anyway, the benefits of medication in preventing or delaying the development of Chagas disease should be weighed against the long duration of treatment and possible adverse reactions [6]. Benznidazole produces adverse reactions in 20%-40% of adult patients, but they usually disappear when the treatment is finished. In a few cases treatment has to be discontinued because of side-effects. On the other hand, children have a higher tolerance of the treatment [10]. Additionally, pregnant women or people with kidney or liver failure should not medicate with neither both benznidazole nor nifurtimox. And nifurtimox is also contraindicated for people who suffer neurological or psychiatric disorders [10].

Control and Prevention There is no vaccine to prevent from T. cruzi infection. Depending on the region, vector control is the most effective method of preventing Chagas disease. Blood screening is necessary to avoid infection through blood and blood products transfusion or organ transplantation. WHO recommends the following approaches to prevention and control [10]:

- Insecticide spraying of houses and surrounding areas.
- House improvements to prevent vector infestation (such as plastered walls, cement floors or corrugated-iron roofs).
- Personal preventive measures such as bed-nets.
- Good hygiene practices in food preparation, transportation, storage and consumption.
- Screening blood donors.
- Testing of organ, tissue or cell donors and receivers.
- Screening of newborns and other children of infected mothers to provide early diagnosis and treatment.
- Use of safety standard protocols to prevent laboratory accidents, especially when dealing with cultures and trypomastigote forms [10]

Chagas Disease and AIDS Reactivation of Chagas disease in chronically infected persons has been associated with immunodeficiency states, such as those related to hematologic malignancies, organ transplantation and corticosteroid therapy [9]. Since the 80s, reactivation of Chagas disease has also been observed in patients with human immunodeficiency virus (HIV) infection. The increased phenomena of population mobility in the last decades, including urbanization and migration to other continents, has permitted the geographical overlap of both HIV and T.cruzi infections, either in endemic and non-endemic countries [11].

3.2 CID Project Usage Facet

Chagas disease initially appeared in South America. However, due to migratory flows, today it is scattered around the whole world. Thus, information can come from any country and users can speak any language. This clearly raises a linguistic problem, and consequently a multilingual interface must be offered, which, from the ICT point of view, is a well know problem that can be easily solved. Nevertheless, two not so obvious problems are also relevant to our case.

Of course, the main source of information will be the governments of the different countries. However, the collaboration of researchers and NGOs (or simply concerned citizens) is also expected. Thus, we need to keep track of who introduced what in the system (also known as data provenance). This will guarantee that we can trace back to the original source any chart or figure. Nevertheless, this is not enough. It is necessary to provide some kind of quality guarantees for any unofficial source. Any data introduced must be backed by some kind of documentation (e.g., journal publication of scientific results). Then, a group of experts designed by WHO will validate the information

Additionally, such kind of sensitive epidemiological data can easily raise concerns in the governments. Therefore, two different versions of the system will be produced: a centralized Web interface, and a desktop application. This will allow the off-line introduction of data, that after the necessary checking by the national authorities can later on be uploaded in the centralized WHO server. For the data in the server, under the responsibility of WHO, the necessary security mechanisms must be provided in order to guarantee confidentiality.

Therefore the critical stakeholders identified (along with their usage goals) were the following:

- World Health Organization: promoter of the project interested in visualize, manage and analyse all the information related to Chagas. The WHO specialists will be the ultimate decision makers and need all relevant data, coming from any country, non-profit organization or research group in order to gain a clear picture of the current status of the disease. According to such picture, the WHO will establish a worldwide roadmap that, again, thanks to the system, will be monitored, analysed (see what impact had) and define the next milestones. All in all, following the natural BI cycle.
- Health Ministry Officers: responsible to provide, modify and delete any Chagas information (about healthcare, transmission interruption, systemic or normative information) related to their respective countries. These stakeholders will provide relevant data about their countries related to Chagas but they will also have access to reports and analysis created for their own uploaded data to let them gain insight of the situation in their countries (e.g., are they meeting the milestones established in the WHO roadmap?)
- Researchers / NGOs: interested in provide information related to healthcare and transmission interruption and also in the visualization and analysis of such information. The ultimate goal of these stakeholders is to perform in-depth analysis by means of developing complex models to simulate and predict certain behaviours. For example, according to climate data, where could the kiss bug be endemically found in a time frame of 5 years?
- Technical groups: who take care of the consistency of the information provided by other stakeholders. They are as well considered as stakeholders because technical groups will continuously monitor the current status of the disease by performing random analysis over the data. For example, one important contribution of such groups are estimations published (see Section 3.3). The difference with the WHO stakeholder is that these groups are supporting / consultant groups of experts for the WHO, whose experts have the last word about the roadmap to be develop for the next years.

Above, the stakeholders are classified according to their position. Below, we alternatively classify them from the point of view of how they would introduce, manage manage, delete, query, analyse or visualize information with the system as follows (note that a position -e.g., researchers- may play several of these roles below):

- Data producers that will introduce relevant information in the system. Thus, they nurture the system with relevant data. From the stakeholders previously identified, these are the national authorities (which we refer to as official data producers) and researchers, NGOs or any other unofficial source (which we refer to as unofficial data producers). One key aspect of the project is to identify how to manage (i.e., gather, homogenize and store) such data.

- Data consumers that will analyze the available information according to their granted access to data. We further classify the data managed by the system according to access grants: open data (i.e., available to anyone), research-related data and sensitive data. The first kind is typically accessed by regular users (e.g., concerned citizens), who want to access pieces of public information. The main indicators / statistics about Chagas will be open for everyone. For the second kind, we refer to them as data analysts who aim at performing advanced analytical queries and gain insight into data (typically, researchers and NGOs). Thus, an authorised account will be needed to access this data. For the third kind, this data is only available for the official data producer (e.g., a national health ministry) and the WHO technical groups.

Related to this data classification, although not stakeholders, we also identify the following relevant actors (further details in Section 5.1):

- Data administrators who will take care of consistency of data inserted in the system (e.g., by applying protocols, granting certifications, etc.). These role is exclusively played by a specific technical group (IT-oriented).
- The system administrator, who is an IT person, that will manage and maintain the system from the ICT point of view.
- Other Information Systems: quite popular nowadays and to be used to also provide relevant information such as ProMed, PubMed and Google Alert information systems.

A detailed description of the data workflows and the interaction of the stakeholders with the system can be found in Section 5.3.

3.3 CID Project IT Facet

We analysed information systems and technologies used in WHO that would help us understand the requirements and challenges involved in the development of the system.

Global Health Observatory The Global Health Observatory (GHO) is part of the World Health Organization (WHO). GHO is the WHOs gateway to health-related statistics from around the world. Therefore, the main goal of GHO is to provide a easy access to:

- Country data and statistics with focus on comparable estimates.
- WHOs analyses to monitor global, regional and country situation and trends.

GHO presents itself in four different sections, GHO theme pages, GHO database, GHO country data and GHO issues analytics reports. GHO theme pages covers global health priorities such as the health-related Millennium Development Goals, mortality and burden of disease, health systems, environmental health, noncommunicable diseases, infectious diseases, health equity and violence and injuries. GHO theme pages present:

- Highlights showing the global situation and trends, using regularly updated core indicators.
- Data views customized for each theme, including country profiles and a map gallery.
- Publications relevant to the theme.
- Links to relevant web pages within WHO and elsewhere.

The GHO database provides access to an interactive repository of health statistics. Users are able to display data for selected indicators, health topics, countries and regions, and download the customized tables in diferent formats such as Excel, CSV, HTML and GHO XML. The GHO country data includes all country statistics and health profiles that are available within WHO. The GHO issues analytical reports on priority health issues, including the World Health Statistics annual publication, which compiles statistics for key health indicators. Analytical reports will address cross-cutting topics such as women and health.

Standards The WHO Indicator and Measurement Registry (IMR) is a central source of metadata of health-related indicators used by WHO and other organizations. It includes indicator definitions, data sources, methods of estimation and other information that allow users to get better understanding of their indicators of interest.

It facilitates complete and well-structured indicator metadata, harmonization and management of indicator definitions and code lists, internet access to indicator definitions, and consistency with other statistical domains.

It promotes interoperability through the SDMX-HD indicator exchange format and allows incorporation of appropriate international standards such as SDMX MCV (Metadata Common Vocabulary), ISO 11179 (Metadata Registry), DDI (Data Documentation Initiative) and DCMES (Dublin Core).

The SDMX-HD is a data enchange format (Statistical Data and Enchange Metadata) based on the data enchange format SDMX. SDMX-HD intends to serve the needs of the Monitoring and Evaluation community. It has been developed by WHO and partners to facilitate the enchange of indicators definitions and data in aggregate data systems.

The SDMX standard describes statistical data and metadata through a data structure definition (DSD), which defines concepts that define dimensions, attributes, codelists and other artefacts necessary to describe the structure and meaning of data. A parallel metadata structure definition (MSD) describes metadata associated with data at observation, series, group, and dataset levels.

Data repository The GHO data repository provides access to over 50 datasets, which can be selected by theme or through a multi-dimension query functionality. The datasets contains information about mortality and burden of diseases, the Millennium Development Goals (child nutrition, child health, maternal and reproductive health, immunization, HIV/AIDS, tuberculosis, malaria, neglected diseases, water and sanitation), non communicable diseases and risk factors, epidemic-prone diseases, health systems, environmental health, violence and injuries, equity among others. In addition, the GHO provides on-line access to

WHO's annual summary of health-related data for its 194 Member states: the World Health Statistics 2012.

GHO has only information about eight of the seventeen neglected tropical diseases. The eight disease are: Buruli Ulcer, Dracunculiasis, Human African Trypanosomiasis, Leprosy, Lymphatic Filariasis, Schistosomiasis, Soil-transmitted helminthiasis and Trachoma.

The information provided by GHO can vary. In some cases GHO provide information about the number of new reported cases for a disease (e.g. Buruli Ulcer) and in other cases can provide country data. Therefore, the information that can be found in the NTDs section is:

- Country data.
- Status of endemicity.
- Annual incidence of cases by country.
- Number of reported cases.
- Population living in endemic areas.
- Population treated.

Country data The Country data table shows countries in rows and indicators in columns for a particular year. The indicators may vary between diseases. Fig.5 shows an example of country data for schistosomiasis.

	Country	Year	SAC population requiring PC for SCH annually	Population requiring PC for SCH annually	Number of people targeted	Reported number of people treated	Age group	Reported number of SAC treated	Programme coverage	National coverage
AFR	Angola	2006	-	-	-	-	-	-	-	-
AFR	Angola	2007	-	-	-	-	-	-	-	-
AFR	Angola	2008	-	-	-	-	-	-	-	-
AFR	Angola	2009	-	-	-	-	-	-	-	-
AFR	Angola	2010	2,620,044	4,722,157	-	-	-	-	-	-
AFR	Angola	2011	2,680,353	4,849,215	-	-	-	-	-	-
AFR	Benin	2006	-	-	-	-	-	-	-	-
AFR	Benin	2007	-	-	-	-	-	-	-	-
AFR	Benin	2008	-	-	-	51,433	SAC	51,433	-	0.69%
AFR	Benin	2009	-	-	-	-	-	-	-	-
AFR	Benin	2010	1,211,805	2,263,785	400,475	364,697	SAC	364,697	91.07%	16.11%
AFR	Benin	2011	1,246,286	2,333,298	-	-	-	-	-	-
AFR	Botswana	2006	-	-	-	-	-	-	-	-

Fig. 5. Country data (Schistosomiasis disease) [12]

Status of endemicity Status of endemicity table shows countries in the rows and year in columns. The value for a concrete year and country could be endemic or non-endemic. It is possible to filter the data by country, year, WHO region and World Bank income group. This table can be downloaded in different formats: CVS, Excel, HTML and GHO-XML. Fig.6 shows the status of endemicity for Trachoma disease.

	Status of endemicity for blinding trachoma
Country	2010
Slovenia	Non-endemic
Solomon Islands	Endemic
Somalia	Endemic
South Africa	Non-endemic
Spain	Non-endemic
Sri Lanka	Non-endemic
Sudan	Endemic
Suriname	Non-endemic
Swaziland	Non-endemic
Sweden	Non-endemic
Switzerland	Non-endemic
Syrian Arab Republic	Non-endemic

Fig. 6. Status of endemicity (Trachoma disease) [12]

Number of new reported cases The number of new reported cases table shows countries in rows and years in columns. The value for a concrete country and year can be a natural number or Data not reported. It is possible to filter the data by country, year, WHO region and World Bank income group. Also, it is possible to download de data in different formats, CSV, Excel, HTML and GHO-XML. Fig.7 shows the number of new reported cases for Buruli Ulcer.

Country		2008	2009	2010	2011	2012
Equatorial Guinea	rted	Data not reported	Data not reported	Data not reported	0	Data not reported
French Guiana	2	8	2	7	3	2
Gabon	32	53	41	65	59	45
Ghana	668	986	853	1048	971	632
Guinea	rted	80	61	24	59	82
Indonesia	rted	Data not reported	Data not reported	Data not reported	Data not reported	Data not reported
Japan	3	2	5	8	10	4
Kenya	rted	Data not reported	Data not reported	Data not reported	Data not reported	Data not reported

Fig. 7. Number of new reported cases (Buruli Ulcer) [12]

Population living in endemic areas The population living in endemic areas table shows countries in rows and years in columns. The value for a concrete country and year can be a natural number or No data available. It is possible to filter the data by country, year, WHO region and World Bank income group. Also, it is possible to download de data in different formats, CSV, Excel, HTML and GHO-XML. Fig.8 shows the population living in endemic areas table for Trachoma disease.

	Population living in trachoma endemic areas				
Country	2006	2007	2008	2009	2010
Brazil	No data available	27000000	27000000	27000000	27000000
Burkina Faso	No data available	3978611	5299721	5299721	7308338
Burundi	No data available	No data available	1871208	1871208	941085
Cambodia	2125447	2125447	2125447	2125447	2416543
Cameroon	No data available	No data available	No data available	5276973	No data available

Fig. 8. Population living in endemic areas (Trachoma disease) [12]

Population treated The population treated table shows countries in rows and years in columns. The value for a concrete country and year can be a natural number or No data available. It is possible to filter the data by country, year, WHO region and World Bank income group. Also, it is possible to download de data in different formats, CSV, Excel, HTML and GHO-XML. Fig.9 shows the population living in endemic areas table for Trachoma disease.

	Population treated for active trachoma				
Country	2005	2006	2007	2008	2009
Djibouti	No data available	No data available	No data available	No data available	72
Egypt	No data available	No data available	No data available	No data available	No data available
Eritrea	No data available	No data available	1364817	1364830	0
Ethiopia	2618488	2600000	6044714	14832830	15695222
Fiji	No data available	No data available	No data available	No data available	No data available
Gambia	18402	401397	143658	151135	111969
Ghana	796378	770124	899580	147122	Case management
Guatemala	No data available	No data available	No data available	No data available	No data available
Guinea	0	0	0	0	0

Fig. 9. Population treated (Trachoma disease) [12]

Map gallery The GHO map gallery includes an extensive list of major health topics. Maps are classified by the following themes: Alcohol and Health, Child

Health, Cholera, Environmental health, Global influenza virological surveillance, Health systems financing, HIV/AIDS, Malaria, Maternal and reproductive health, Meningococcal meningitis, Mortality and global burden of disease (GBD), Neglected tropical diseases (NTDs), Noncommunicable diseases, Road safety, Tobacco control and Tuberculosis (Referncia web GHO Map Gallery).

The GHO map gallery allows searching maps by geographical region, topics and keywords. Fig.10 shows a possible search defining region as world, topic as neglected tropical disease and keyword as Chagas.

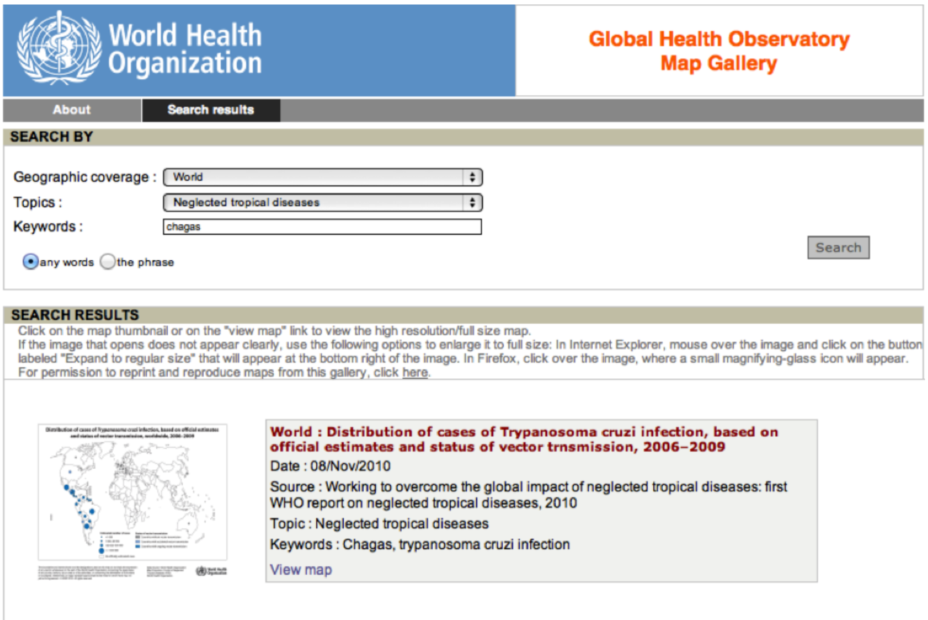

Fig. 10. GHO map gallery search engine [12]

The neglected tropical disease map gallery contains maps about NTDs topics such as distribution of a disease in a concrete year (Fig. 11), new cases reported (Fig. 14), facilities involved (Fig. 13) and prevalence rate (Fig. 12).

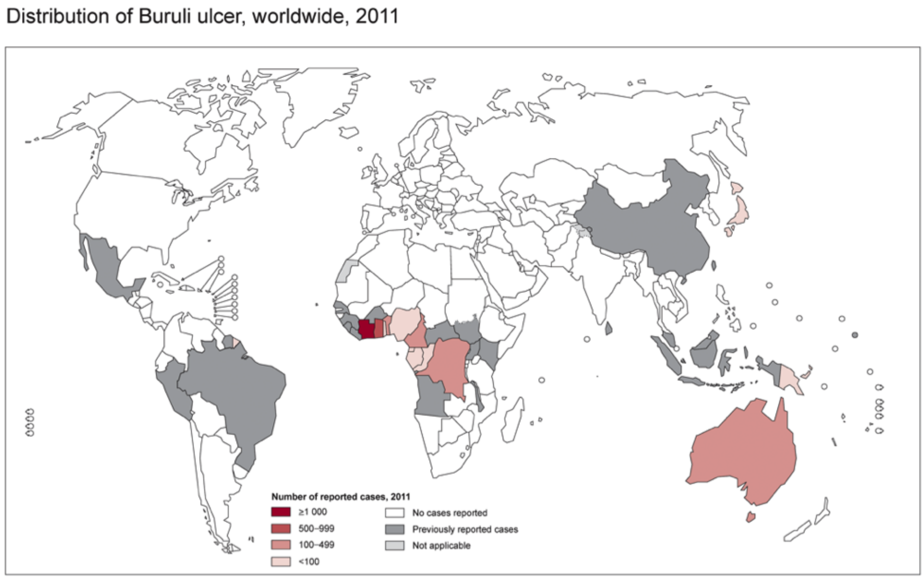

Fig. 11. Distribution of Buluri Ulcer [12]

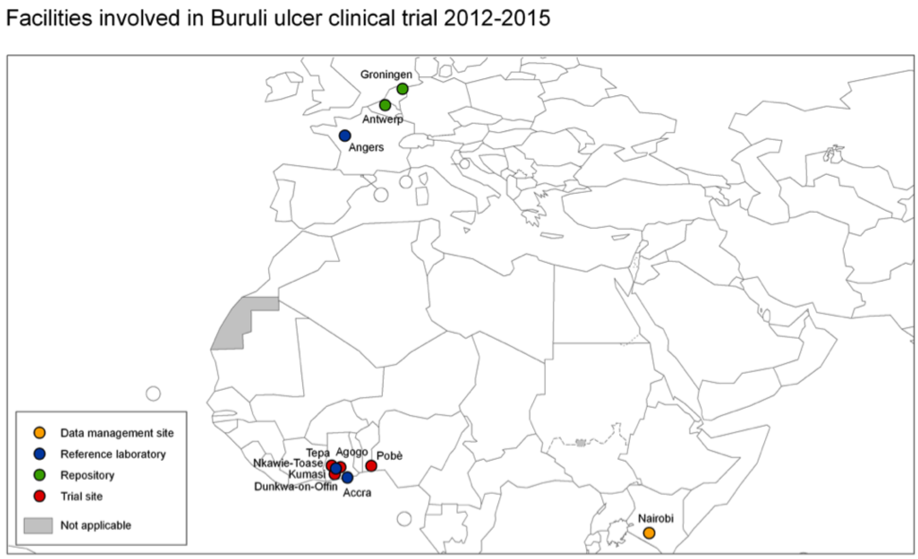

Fig. 12. Facilities involved in Buruli Ulcer [12]

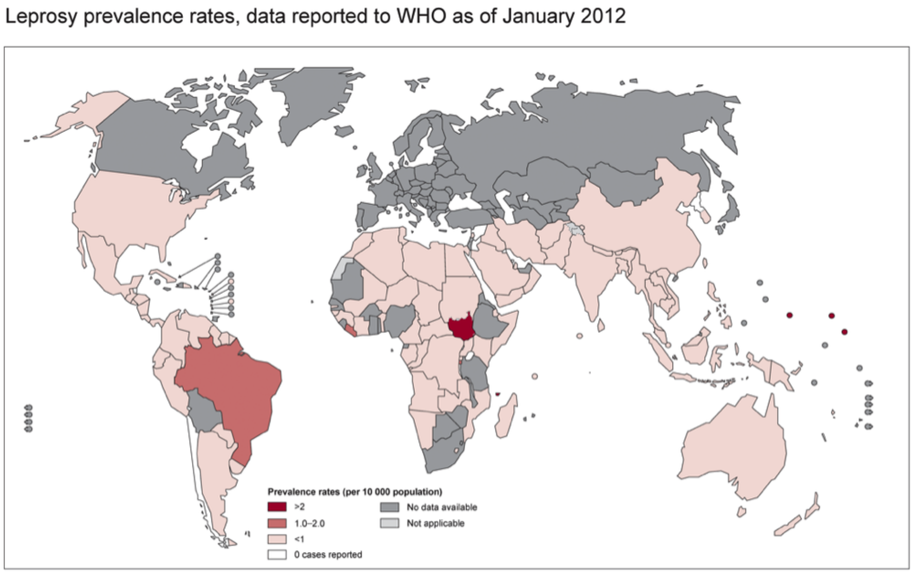

Fig. 13. Leprosy prevalence rates [12]

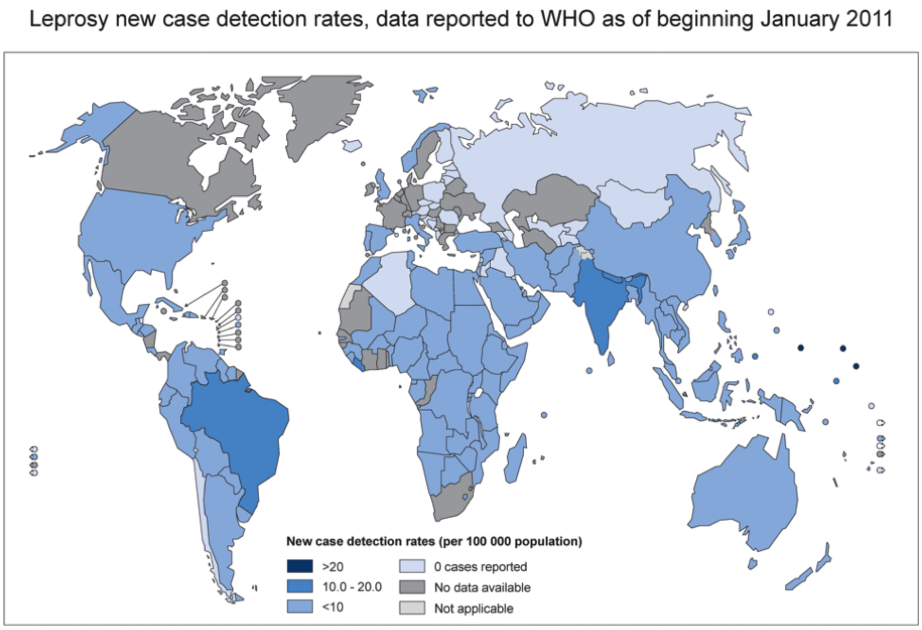

Fig. 14. Leprosy new case detection rate [12]

Country statistics The country statistical pages bring together the main health data and statistics for each country, as compiled by WHO and partners in close consultation with Member States, and include descriptive and analytical summaries of health indicators for major health topics. Each country statistics page has three main sections:

1. Country data and statistics
2. Country profiles
3. Primary metadata

3.4 CID Project Development Facet

Due to the complexity of the project, at this stage it was decided to estimate only the costs involved for the requirements part of the project and to postpone the estimation of the cost of the whole project until the end of the requirements engineering process.

4 Core Activities

This block of the Pohl's Framework is composed of the three main core activities Documentation, Elicitation and Negotiation as shown in Fig.15.

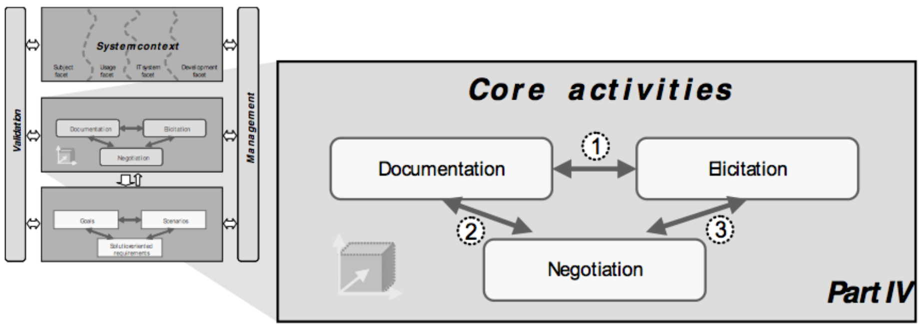

Fig. 15. Core Activities of Requirements Engineering Poh'ls Framework [3]

These are described below:

1. **Elicitation** The goal of this activity is to elicit the requirements at all the levels of the system. These requirements may exist in different forms, for instance: in ideas, intentions, requirements models, in existing systems, etc. In addition, there are many sources (stakeholders, documents, existing systems) that have to be elicited in order to gather all the possible requirements to support the acceptance of the system.

2. **Negotiation** Each stakeholder has different wishes and needs than may be in conflict with each other; reason why negotiation aims to achieve an agreement among them. Conflicts must be resolved involving the most relevant stakeholders while also these may be seen as an opportunity to create also new ideas.

3. **Documentation** The main aim of this activity is to document the elicited requirements according to the rules. According to Klaus Pohl, documentation is important in order to establish a common reference, promote communication between stakeholders, support training for new employees and to preserve expert knowledge. Rules for documentation may be either:
   - General Rules (e.g. define layout, headers, version history),
   - Documentation Rules (e.g. rules to ensure quality of template for the use of elicitation and negotiation), or
   - Specification Rules (more restrictive than documentation rules e.g. rules for subsequent development activities like the use of requirements specification language).

   In order to avoid lexical ambiguity, the author suggests glossaries may be used along with synonyms and related terms.

In general, these activities were not much different of any traditional elicitation, negotiation and documentation activities. However, there are two main differences that must be considered when dealing with DSS. On the one hand, the heterogeneity of the stakeholders involved. In general, in DSS, and specifically, in the CID project, the point of view of each stakeholder is drastically biased by previous experiences and tend to simplify other factors that are not releveant to him / her. In our case, we realised that not even the WHO had a clear view of all the factors affecting Chagas. This difficulty of understanding each other was not only a challenge for IT people, but also for other actors involved. For example, entomologists and doctors (among whom we must distinguish specializations) also had problems to combine their visions. Due to this complexity we identify the most relevant techniques and methods to elicit and negotiate in DSS projects (see Sections 4.1 and 4.2). On the other hand, the real difference comes from the artifacts produced during these activities (see Section 5.3 for further details).

4.1 CID Project Elicitation

Once described the system context of the project, software analysts with the help of health and vectorial experts trawled the work to learn and understand the requirements of the project, i.e. elicit the requirements of the system.

In addition to brainstorming the elicitation techniques used were the following:

1. Interviews [13]: software analyst performed several interviews to health and vectorial experts. Most of them were exploratory interviews, i.e. conversations by means of which the interviewer elicited information about the opinion or view of the interviewee with respect to some issue. The results of

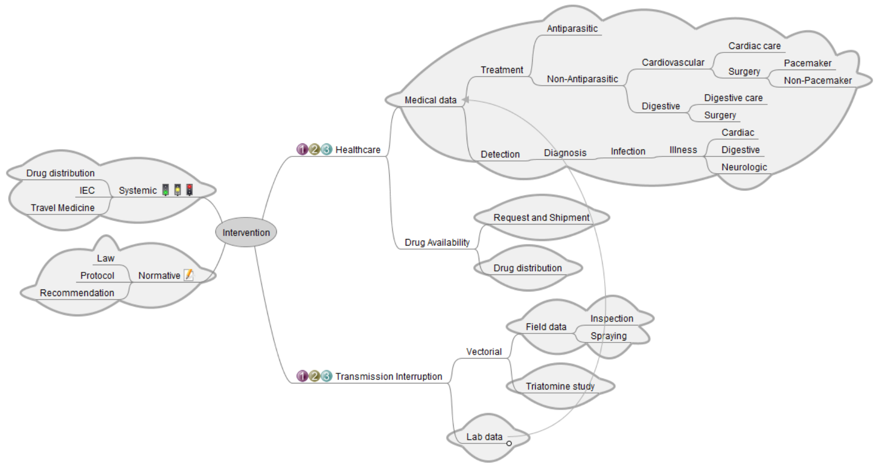

Fig. 16. Mind map produced to identify the main factors affecting Chagas in the intervention package (result of one brainstorming session)

such interviews were qualitative. For example, software analyst interviewed a health expert in co-infection to understand all the variables to take into account when defining the part of the conceptual model related to diagnosis.

2. Workshops [14]: around 4-5 people joined during several 2-3 days sessions to refine the requirements. Some of the typical different assistance techniques used were brainstormings, iterative goal and scenario definitions, discussions and work in sub-groups and presentation of the results in the plenum. An example of the agenda for one of the workshops was:

   9:00 - 9:30 Welcome session and presentation of the goals of the workshop: information required about Diagnosis, Drug Availability, Vectorial, Treatment, Systemic and Normative Intervention.
   9:30 - 10:00 Brainstorming of the goals of the workshop
   10:00 - 10:45 Categorisation of the brainstorming results,
   10:45 - 13:00 Grouping and refinement of the goals, assignment of goals to subgroups,
   13:00 - 14:30 Lunch break
   14:30 - 18:00 Work in sub-groups: scenario definition for the elicitated goals
   18:00 - 19:00 Presentation of the intermediate results by each sub-group.

3. Questionnaires [13]: questionnaires were suitable when physical interviews or workshops were not feasible. Basically, lists of questions from the software analyst to the health or vectorial experts were send by email. Initially, software analysts tried to ask specific questions related to the conceptual model being defined. However, it was required to change to more user friendly documentation (basically, forms) for a better understanding of the questions.

4. Observation [15]: this technique consists on an observer eliciting requirements by studying stakeholders or existing systems. Even though there is no a current information system to eradicate Chagas, similar projects have been developed for other diseases within the WHO context.

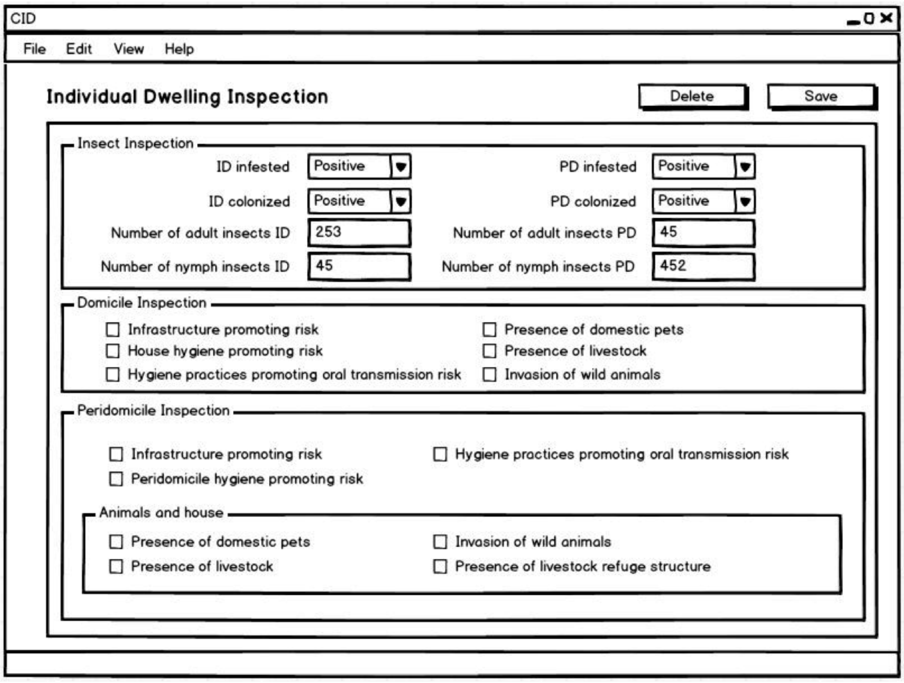

Fig. 17. Mockup of the collection of information for Individual Dwelling Inspection

The elicitation techniques described above were also supported by the following assistance techniques [3]:

1. Brainstorming [16]: Brainstorming was basically performed during the first workshops for defining the main goals of the project and clarification of ideas and concepts.
2. Mind mapping [17]: We used the FreeMind tool [18] to represent the context of the project and to group the different use cases that were needed to develop in the whole project as shown in Figure 16.
3. Prototyping [19]: this technique was completely necessary to support the definition of the use cases and the conceptual model of the project. Conceptual models in UML were difficult to understand for Vectorial and Health experts, but the creation of mockups (with the Balsamiq Mockup tool [20]) facilitated the feedback from them. We created them for each subpart of the conceptual model. An example is shown in Figure 17.

4.2 CID Project Negotiation

In this project, the main conflict was to reach a consensus between the value obtained to collect as much information as possible to the maximum level of detail from the countries to the actual ability to collect such information. For example, at the very beginning the health expert asked that the system collected details of each dwelling inspection on a vectorial control. However, to ask for so much information and at this level of detail entailed the risk that the system would never be used or that the information entered would not be accurate enough. The use of mockups, as mentioned above, were indispensable to resolve these conflicts and reach consensus. Finally, priority was given to get the minimum information necessary to obtain the maximum impact to combat Chagas.

This conclusion is extensible to any DSS. In any DSS the trade-off of value added versus difficulty to gather the needed data must be analyzed in depth and negotiated. At the negotiation stage, it must clearly be defined what is the value added and weight it against the difficulty to obtain such information. For example, ideally, some stakeholders were asking for several parameters at house level when gathering information about infestation. However, after showing some mockups of the forms the field teams should fill it was clear for all of us that it was unfeasible and, in practice, most teams would refuse to fill all that information per house (due to its time-consuming nature). It was more advisable to gather data at a certain aggregation level, with few (but highly relevant) parameters. In this precise example, this was negotiated and agreed by presenting a mockup (that of Figure 17) and modify it until reaching consensus.

4.3 CID Project Documentation

Several documents were generated during the whole project: for describing the system context, the goals, the scenarios and the specification of the solution-oriented requirements. The latter were documented following the IEEE 830-1998 software requirements specification standard .

Table of Contents

Appendixes
Index
The requirements were defined at a level of detail that permits designing and implementing the software system that meets the defined requirements. It includes the definition of the function systems, the conceptual model, and the software design of the system.

5 Requirement Artifacts

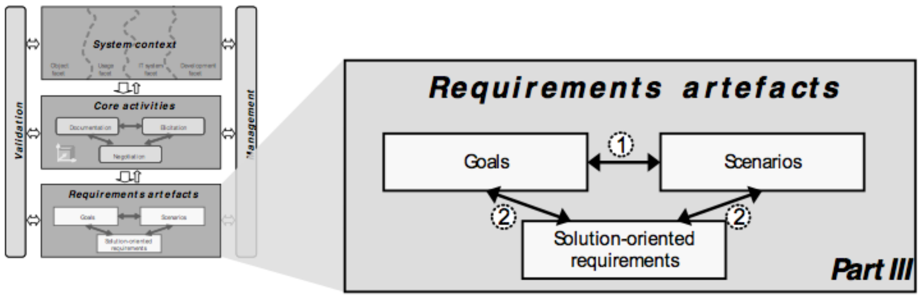

Fig. 18. Requirements Artifacts of Requirements Engineering Poh'ls Framework [3]

The term "requirements artifacts" stands for a documented requirement using a specific format. In this section we differentiate between three different requirements artifacts according to Klaus Pohl, namely goals, scenarios and solution-oriented requirements [3] and shown in Fig. 18:

– **Goals:**
  In Requirements Engineering domain, the stakeholder's intentions about the objectives are documented as goals stating what exactly is expected or required from the system.

– **Scenarios:**
  As mentioned above, goals are used to document the stakeholder's intentions. Therefore, scenarios are further used to document intentions by illustrating if the scenario satisfies or fails to satisfy the goal by representing the interactions between the system and its actors.

– **Solution-Oriented Requirements:**
  As mentioned above, goals and scenarios are the basic foundations for developing solution-oriented requirements. In contrast to goals and scenarios, solution-oriented requirements are created to specify a deeper required level of details. By joining these three together requirements, we may define the reference software engineers to implement the system.

5.1 CID Project Goals

To define the goals of each actor we classified the actors that would interact to the system in three types of users and three types of systems as shown in Fig. 19.

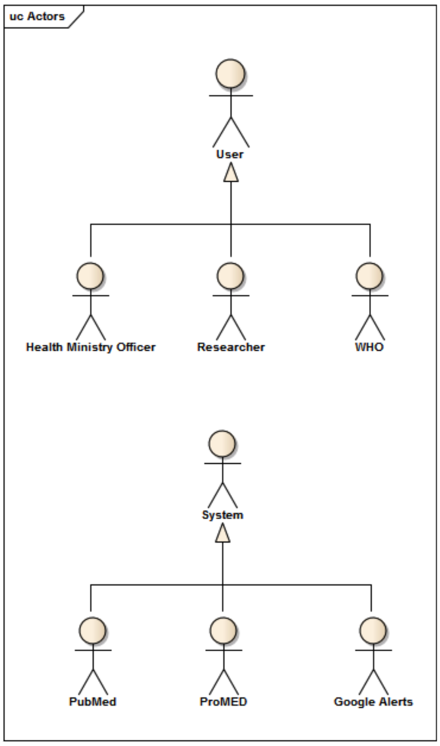

Fig. 19. Actors of the CID system

For each one, we defined in natural language his or her goals.

The main goals of the WHO actor are to visualise, manage, analyse and import/export information of the system. On the one hand, since WHO must have access to the entire information system database, he or she should be able to consult, but not modify, any type of information related to any country. Additionally, WHO actor will be able to manage assessments, regulations and certifications. On the other hand, he or she can generate and visualise maps and diagrams with any type of information gathered in the system. Finally, he or she will be able to import and export any type of information provided by countries or researchers by uploading it with an Excel file.

The HMO actor, in a similar way as the WHO actor, should interact with the system in order to visualise, manage and import/export information. Health Ministry Officers should be allowed to provide, modify, and delete information related to his own country, therefore he or she will not be able to consult, modify and delete information from other countries. Information he or she can manage ranges from healthcare, through transmission interruption, to systemic and normative. In particular, he or she can manage the following information: treatments, diagnosis, insecticide applications, triatomine bug studies, inspections, regulation and systemic information. In addition, he will be able to generate and visualise maps and diagrams with the information he provided to the system. Finally, he will be able to import and export any type of information by uploading it with an Excel file.

The Researcher actor, in a similar way as the other actors, interacts with the system in order to visualise, manage, import/export, and analyse information. Researchers can provide information related to healthcare, and transmission interruption. In particular, they can manage the following information: treatments, diagnosis, insecticide applications, triatomine bug studies and inspections. In addition, they will be able to generate and visualise maps and diagrams with the information they provided to the system. Finally, they will be able to import and export any type of information by uploading it with an Excel file.

Actors may represent roles played by human users, external hardware, or other subjects such as information systems. The ProMED, PubMed and Google Alert information systems are also considered actors of the system. All these three actors interact with the system notifying it when new cases of Chagas disease appear.

5.2 CID Project Scenarios

Once defined the main goals of the actors, the next step was to look for the systems scenarios by means of use cases. We needed several iterations with the WHO experts before we arrived to the final use cases version. Furthermore, all the use cases defined were classified in four groups: management of tables, visualization of information, exportation and importation of information, and analysis of information.

The following list presents all uses cases for each of the previous groups:

- **Management of tables**: Manage Regulations, Manage Assessments, Manage Certifications, Manage Triatomine Study, Manage, Dwelling Triatomine Study, Manage Collective Triatomine Study, Manage Insecticide Application, Manage Collective Insecticide Application, Manage Dwelling Insecticide Application, Manage Inspections, Manage Dwelling Inspections, Manage Collective Inspections, Manage Diagnosis, Manage Individual Diagnosis, Manage Collective Diagnosis, Manage Collective Oral Diagnosis, Manage Treatments, Manage, Individual Treatments, Manage Collective Treatments, Manage Drug Distribution, Manage Systemic Information, Manage Associations, Manage IEC, Manage Research, Manage Drug Distribution System, Manage Lab Quality Control and Manage Health Economic.
- **Visualization of information**: Visualise Diagrams and Visualise Maps.
- **Import/Export information**: Export Information and Import Information.

– **Analysis of information**: Analyse Data

Following we show as example of one of the Use Cases defined:

Use Case Name
Manage Individual Diagnosis
Active Actor
Researcher
Trigger
Researcher indicates manage individual diagnosis
Preconditions
The Researcher must be identified and authenticated.
Stakeholders and interests

1. Researcher: manage (create, modify or delete) individual diagnosis information he provided.

Main Success Scenario

1. User indicates to manage individual diagnosis
2. System presents all registers of available individual diagnosis gathered into the system
3. User selects an existing individual diagnosis register
4. System presents the individual diagnosis information
5. User updates information about the individual diagnosis and then indicates save the individual diagnosis register
6. System updates the register

Extensions

3a. Create a new individual diagnosis register

1. User indicates to create a new individual diagnosis register
2. System presents a form in order to fill in with the individual diagnosis information
3. User enters information about the individual diagnosis and confirms the creation of the individual diagnosis register

Return to step 6 of Main Success Scenario

5a. Delete individual diagnosis register

1. User indicates to delete the individual diagnosis register
2. System asks for confirmation
3. User confirms deletion of the individual diagnosis register

3a. No confirmation

1. User cancels the action

Return to step 4 of Main Success Scenario

Return to step 6 of Main Success Scenario

Outcome

An individual diagnosis register has been created, updated or deleted with the information provided by user

5.3 CID Project Solution-Oriented Requirements

Solution-oriented requirements define the data perspective and the functional perspective perspective on the system. The data perspective focuses on defining the data/information to be managed by the system. We defined this information, the conceptual model, by means of UML class diagrams [21]. The aim of said conceptual model was to define and represent real-world concepts, that means, define relationships between different concepts of Chagas disease and WHO, and describe the meaning of terms and concepts used by health specialists about Chagas disease. Additionally, the conceptual model of the information system was structured in different packages to make easier the comprehension of all its concepts. Therefore, the following packages were defined: Healthcare package, Transmission Interruption package, Systemic package, Normative package, Geographic and Temporal package, Wildlife package and the Bibliographic package.

An example of a fragment of the conceptual model is shown in Fig.20

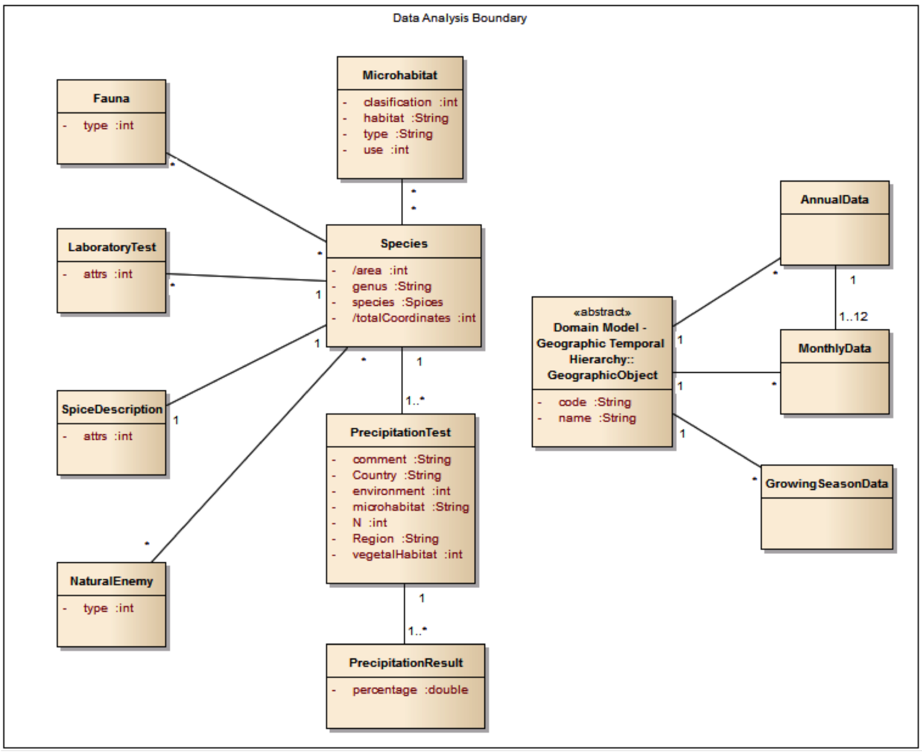

Fig. 20. Wildlife package's conceptual model

The functions of the systems were described by refining the UML use case diagrams: for the most relevant functions of the system, sequence diagram with operation contracts in OCL [22] were specified.

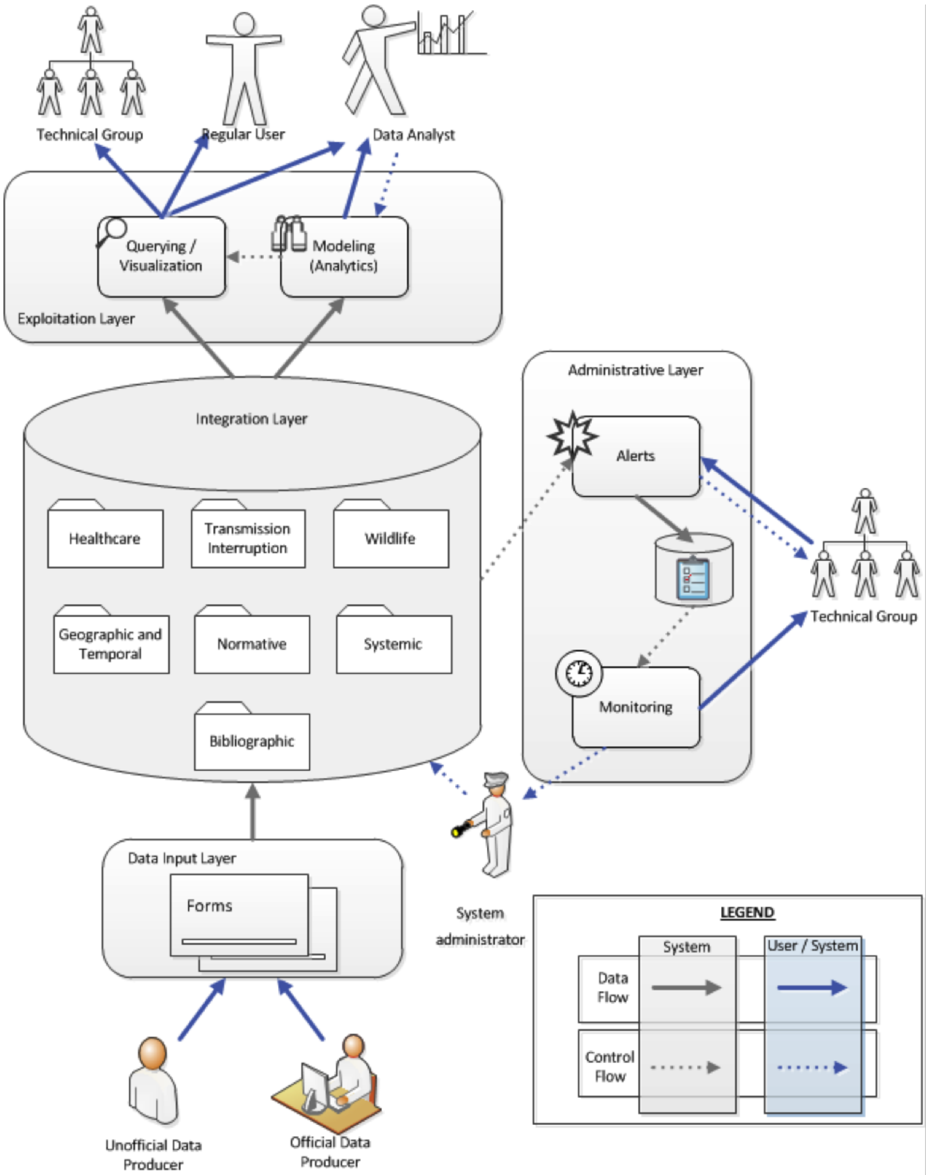

Fig. 21. Functional Architecture

Regarding the design, the system functional architecture is described in Fig.21 and it is divided in 4 main layers, namely:

- The data input layer through which relevant data related to the Chagas disease is introduced in the system,
- The integration layer (or data warehouse), which shows a single, unified view of the data gathered,
- The administration layer, which monitors the system to detect relevant events, and
- the exploitation layer, which is meant to query and perform advanced analytics over the gathered data.

This architecture also refines the use cases discussed above. The data input layer is meant to let end-users introduce their data into the system. Data gathered is properly identified by the person / group / organization uploading such data and arranged in 5 different subsystems within the data warehouse: healthcare, transmission interruption, wildlife, normative and systemic.

- Healthcare: This subsystem contains information about healthcare indicators, and is divided in three categories: patients diagnosis and treatment, and drug delivery. For the first two, data is gathered either at the individual level (i.e., data is related to an anonymized person and indicators such as age, gender,nationality or phase of the patient, among others, are stored) or collective level (i.e., data related to a geographic area and aggregated measures such as number of infected, coinfected or evaluated people, among others, are stored). For the latter, the system stores data related to medicines and their distribution.
- Transmission interruption: Although Chagas disease can be transmitted by different routes, this subsystem focus on the vectorial transmission and contains information related to dwelling inspections, dwelling insecticide applications and triatomine bug studies. The three of them can be reported either at the individual or collective level. Individual data reported refers to a specific dwelling and indicators such as the number of adult kiss bugs and nymphs found or bugs infested with the T. Cruzi parasite, among other indicators, are stored. Collective data is related to a geographic area and aggregated measures such as number of inspected, infested, or colonized dwellings, among others, are stored).
- Wildlife: This subsystem stores data about flora, fauna and climatology. In order to foresee to what regions the Kiss Bug could spread in the near future information about all known species which can transmit the Chagas disease as well as climatological factors are stored.
- Normative: This system stores data about regulations, certifications and assessments. Regulations are typically provided by official users, whereas certifications and assessments are introduced by WHO.
- Systemic: Data about the activities and procedures that countries are implementing in order to control and eliminate the Chagas disease. This includes

data such as recognized associations, research capacity, data on IEC (Information, Education, Communication), etc.

Relevantly, any piece of data introduced in the system is always characterized with a geographical and a temporal object, which are managed by the geographical and temporal subsystem. Finally, the system is prepared to allow researchers introduce complementary data to the official data uploaded by countries. Research data can be backed up with external publications (e.g., journal papers) and for these cases the bibliographic subsystem is responsible for storing and managing the bibliography. Any uploaded research data is automatically put in quarantine until validated by the corresponding technical group who can decide to either make it available, once manually checked its quality, or kept in quarantine if its veracity is doubtful. This validation task is performed by means of the monitoring subsystem available under the administration layer, which is also responsible to support predefined monitoring alerts. In short, the technical group may decide to set up some high-level protocols. For this milestone we implement ECA (event-condition-action) rules. In presence of a certain event (e.g., a new infection case is inserted), if a certain condition is given (e.g., this infection case is associated to a country -geographical object-, which had no previous infection cases), an automatic alert is triggered (e.g., an email is sent) to the technical group. The technical groups is responsible for defining which events and conditions must be monitored through the alerts module, which, in turn, will store the ECA rules in the metadata repository. The monitoring module monitors the right piece of information from the data warehouse and an alert is triggered if the ECA rule holds. Monitored data can be consulted by the technical group at any time.

Last but not least, the exploitation layer is responsible for querying / visualizing and performing advanced analytical tasks on the data; the ultimate goal of the whole system. According to the granted access rights of each kind of user a user may query and visualize that portion of data available to him / her. Then, the same module is responsible to render the query result. Alternatively, data analysts need advanced tools, such as R, to create models of different aspects related to Chagas in order to foresee the impact of certain actions / inactions, whose results are eventually rendered by means of the querying / visualization module.

6 Validation and Management

Validation aims to state if the input and output (requirements artifacts) of the core activities satisfy the criteria established. This phase of validation is done by involving different resources, such as: stakeholders, requirement sources (laws and standards) and external reviewers. Validation answers the question: Am I building the right system?.

Apart from Validation, Management is another one of the cross-sectional activities in the Requirements Engineering framework. The following three bullets explain the three sub-activities in the management phase related to the middle framework blocks.

6.1 CID Project Validation and Management

In this project, following principles of requirements validation according to Pohl, we ensured that relevant and appropriate WHO Health, Vectorial and software experts were involved for each aspect of a requirement artefact to be checked during validation; we considered changing the documentation format of the requirements into a format that matched the validation goal and the preferences of stakeholders who performed the validation and; we also established guidelines that clearly determined when or under what conditions an already released requirements artefacts had to be validated again. We mainly used techniques of inspection, desk-check, walkthrough and prototyping when appropriate.

The goal of management in requirements engineering is to (1) observe the system context to detect context changes, (2) manage the execution of requirements engineering activities and (3) manage the requirement artefacts. The first goal has not been a big deal, since as part of their own work, health and vectorial experts have been participating in the worldwide events related to any possible law, policies or medical changes that could affect the project and software specialist have been analysing the emergent technologies that could be used in the implementation of the project. The most difficult task in the management of the execution of requirements engineering activities has been to make-up the agendas and trips of the stakeholders for the workshops. Finally, for the management of the requirement artefacts, we used, on the one hand, TortoiseSVN [23], a free open-source Windows tool for the Apache Subversion control system for controlling the different versions of all documentation created and, on the other hand, to ensure the traceability among the different software models, all of them were defined in the Enterprise Architect, a collaborative modeling, design and management platform based on UML 2.4.1 and related standards from Sparx Systems company [24].

7 Conclusions and Further Research

The CID Project was useful as a case study of the application of the Pohl's Framework to obtain the requirements of a Decision Support System. The complexity found in the CID project was an example of what is expected to deal in most of these projects:

– **Complexity of Decision Support Systems:** Unlike operational systems, Decision Support Systems such as Data Warehousing are a complex and expensive component of the corporate's IT infrastructure involving a conglomerate of extraction, cleansing and transforming systems, Data Warehouse database(s), Data Marts, metadata management system, usually presented in a summarized format for visualization purposes [25]. For instance, the CID project consisted of gathering information from heterogeneous sources such as: hospitals using databases to record the infected patients, researchers using excel forms to record data about the raids and dwellings inspections, and health minister officers managing information about regulations in PDF format; making this a complex process to build the CID system.

- **Integration with diverse data sources:** Decision Support Systems collect data from heterogeneous sources including the operational systems or other external sources. Therefore, procedures and requirements must be specified for the data collection and integration [26]. Plus interfaces and connections to already existing systems (inside or outside the organization) must be considered [25]. During the CID project, heterogeneous sources mentioned in the previous example had to be integrated to present a single unified interface; thus, problems of heterogeneous source integration were present such as: *technical heterogeneity* (integrating different file formats and databases such as Microsoft Excel, PDFs, and SQL Server databases) and *semantic heterogeneity* (identifying and integrating concepts from the sources which semantically correspond to one another, such as "kiss bug" in the U.S.A is semantically equivalent to "vinchuca" in Argentina).
- **Expressing Decision Support Systems requirements:** At early stages of the Decision Support Systems project, the future users of the system have difficulties to express their requirements [27]. In contrast to operational systems which are based on consistent specifications [25]. For instance, in the CID project, the stakeholders aimed to collect as much information as possible with the maximum level of detail (e.g. number of insects per house). However, it is too ambitious to obtain this level of detailed data, plus the system could become too complex. Hence, an agreement among the stakeholders had to be reached by negotiating the right balance of having information with enough level of detail (i.e. aggregated data per certain area) to impact the main objective of elimination of Chagas. As believed by Paul Valéry: *"What is simple is wrong, what is complicated is useless"* [28].
- **Identifying relevant subjects and perspectives of analysis:** The multidimensional model has become the de facto standard for Decision Support Systems that categorize the data as facts (representing a business activity) or dimensions (perspective of the data) being usually textual [29]. While analyzing user requirements, in Decision Support Systems it is crucial to identify the subject of analysis by perceiving the metrics behind the user demands. Also to determine the extent to which a metric can be operated (summarized, counted, etc.) along with the different perspectives from where to analyze the relevance subjects of analysis [26]. As discussed later, this specific analysis-oriented conception requires considering specific data models.
- **Data exploration and summarization:** It is needed to firstly discuss the relevance of exploring data from different points of view, meaning changing the data granularity (data detail level) on demand. To do so, summarizability (assuring the correctness of summary operations) is a must; specifically to validate the correctness of aggregated results for whichever combination

between facts and dimensions [26]. Summarizability is done in order to avoid erroneous conclusions and decisions [30]. Thus, it is important to express the requirements constraint regarding the aggregated data. For instance, in the CID project, visual representations of hierarchical structures in space and time were required, for instance, showing "Infected Dwellings" for Argentina regions over a period 2008-2014 to detect potential infected geographical areas.

- **Fast track of user requirements changes:** Decision Support Systems are in constant evolution and must face changes during its development that may impact on the models built to meet the end-user requirements [26]. The issue is that the relationship between the requirements, elements in the model and data sources are lost in the process since no traceability is specified neither documented [31]. Accordingly, track of all the requirements and changes must be kept in Decision Support Systems during the complete project's life-cycle. For instance, showing requirements dependencies with cross-references analysis to demonstrate the impact of a certain change in a fact or dimension (e.g. traceability matrices) [32].
- **High-quality documentation and Standard Vocabulary:** Decision Support Systems projects use preexisting operational documentation in order to define the data extraction and integration procedures; however, legacy documentation usually lacks quality, which makes the task of extracting the requirements harder [26]. In addition, a vocabulary, is necessary to identify definitions, acronyms and abbreviations for establishing lexis to be shared among all parties related to the project [33]. For instance, during the CID project, health specialists (e.g. entomologists) certainly use medical and domain terminology for their every day communication such as "trypanosomiasis and Triatominae", which software analysts and neither doctors were able to understand, and neither to match these with the heterogeneous sources. For this reason, there is a huge need to bridge the gap between stakeholders and IT people by creating a Vocabulary. However, not much advantage can be taken from the vocabulary in a plain document, but certainly there is a need to translate this vocabulary into a machine-readable format with a cross-reference structure to be used when creating the future schema for the system.

7.1 Further Research

The project clearly motivated to do further research about the current approaches of Requirements Engineering in Decision Support Systems: we needed to review the literature of Requirements Engineering for Decision Support Systems to obtain a good grasp of the main published work. In short, many different aspects of the whole Requirement Engineering cycle have been presented but there is no overall and comprehensive framework such as Pohl's framework for decision support systems. After carefully reading the literature related to the Requirements Engineering framework, we noticed that we would achieve a more understandable classification by organizing the found literature according to this building-blocks proposed in the framework. We believed that the use of this framework would be a good fit for our study due to the following reasons:

1. As mentioned previously, the first and main reason is that the framework defines the major structural building-blocks and elements of the Requirements Engineering process (e.g. Elicitation, Negotiation, Documentation, Goals, etc.).
2. It provides a well-structured base for the fundamentals, principles and techniques of Requirements Engineering.
3. It is not adhered to a specific methodology or a type of software project.
4. In addition, this framework consolidates various research results and has been proven to be successful in a number of organizations for structuring their Requirements Engineering process.

For the aforementioned reasons, the use of this framework would provide a better organization and structure of the found literature (Requirements Engineering for Decision Support Systems).

As a natural continuation of this work, and after all the lessons learnt, we are formalizing our findings in the form of a comprehensive framework, following the lines above describe, within the context of an IT4BI Master Thesis of one of the authors of this paper (S. Garcia) and supervised by other two (O. Romero and R. Raventós). The expected outcome will be a systematic generic approach entitled "Requirements Engineering for Decision Support Systems" (RE4DSS) based on the thorough view of the domain gained during the analysis of the literature.

References

1. Organization, W.H.: About who. `http://www.who.int/about/en/` (2014) [Online; accessed 11-July-2014].
2. Savioli, L., Daumerie, D., Crompton, D.W.T., Peters, P., Organization., W.H.: Working to overcome the global impact of neglected tropical diseases : first WHO report on neglected tropical diseases. World Health Organization Geneva, Switzerland (2010)
3. Pohl, K.: Requirements Engineering - Fundamentals, Principles, and Techniques. Springer (2010)
4. Aurum, A., Wohlin, C.: Engineering and managing software requirements. Volume 1. Springer (2005)
5. Iee, E.: IEEE Std 610.12-1990(R2002). IEEE Standard Glossary of Software Engineering Terminology (1990)
6. Coura, J.R., Viñas, P.A.: Chagas disease: a new worldwide challenge. Nature **465**(n7301_supp) (2010) S6–S7
7. Peters, W., Gilles, H.M., et al.: A colour atlas of tropical medicine and parasitology. Wolfe Medical Publications Ltd. (1977)
8. Feasey, N., Wansbrough-Jones, M., Mabey, D.C., Solomon, A.W.: Neglected tropical diseases. British medical bulletin **93**(1) (2010) 179–200
9. Ferreira, M.S., Nishioka, S.d.A., Silvestre, M.T.A., Borges, A.S., Araújo, F.R.F.N., Rocha, A.: Reactivation of chagas' disease in patients with aids: report of three new cases and review of the literature. Clinical infectious diseases **25**(6) (1997) 1397–1400

10. Organization, W.H., et al.: Working to overcome the global impact of neglected tropical diseases: first who report on neglected tropical diseases. (2010)
11. Vaidian, A.K., Weiss, L.M., Tanowitz, H.B.: Chagas' disease and aids. Kinetoplastid biology and disease **3**(1) (2004) 2
12. : Sustaining the drive to overcome the global impact of neglected tropical diseases: second WHO report on neglected diseases. Advanced proof edn. WHO/HTM/NTD/2013.1. World Health Organization, Geneva (2013) Copies of materials can be requested through: NTD Information Resource Centre E-mail: ntddocs@who.int.
13. Oppenheim, A.N.: Questionnaire design, interviewing and attitude measurement. Bloomsbury Publishing (1992)
14. Leffingwell, D., Widrig, D.: Managing software requirements: a use case approach. Pearson Education (2003)
15. Sommerville, I., Kotonya, G.: Requirements engineering: processes and techniques. John Wiley & Sons, Inc. (1998)
16. Osborn, A.F.: Development of creative education. Buffalo, NY: Creative Education Foundation (1961)
17. Buzan, T.: Mind mapping. Pearson Education (2006)
18. Polivaev, D.: Freemind-free mind mapping software (2008)
19. Sommerville, I.: Software engineering, 7th or 8th edition
20. Guilizzoni, G.: balsamiq mockups. Balsamiq, http://www. balsamiq. com (2010)
21. Uml, O.: 2.0 superstructure specification. OMG, Needham (2004)
22. Warmer, J.B., Kleppe, A.G.: The object constraint language: Precise modeling with uml (addison-wesley object technology series). (1998)
23. Küng, S., Onken, L., Large, S.: Tortoisesvn: A subversion client for windows
24. Architect, E.: Sparx systems (2010)
25. Winter, R., Strauch, B.: A method for demand-driven information requirements analysis in data warehousing projects. In: HICSS. (2003) 231
26. Paim, F.R.S., de Castro, J.F.B.: Dwarf: An approach for requirements definition and management of data warehouse systems. In: Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International, IEEE (2003) 75–84
27. Connelly, R., McNeill, R., Mosimann, R.: The Multidimensional Manager: 24 Ways to Impact Your Bottom Line in 90 Days. Cognos Incorporated (1997)
28. Valéry, P.: Notre destin et les lettres. (1937)
29. Pedersen, T.B.: Multidimensional modeling. In: Encyclopedia of Database Systems. (2009) 1777–1784
30. j. Lenz, H., Shoshani, A.: Summarizability in olap and statistical data bases, IEEE Computer Society (1997) 132–143
31. Maté, A., Trujillo, J.: A trace metamodel proposal based on the model driven architecture framework for the traceability of user requirements in data warehouses. Inf. Syst. **37**(8) (2012) 753–766
32. Toranzo, M., Castro, J., Mello, E.: Uma proposta para melhorar o rastreamento de requisitos. In Pastor, O., Diaz, J.S., eds.: WER. (2002) 194–209
33. Britos, P., Dieste, O., Garcia Martinez, R.: Requirements elicitation in data mining for business intelligence projects. In Avison, D., Kasper, G., Pernici, B., Ramos, I., Roode, D., eds.: Advances in Information Systems Research, Education and Practice. Volume 274 of IFIP - The International Federation for Information Processing. Springer US (2008) 139–150