3ª ed

MIM

Geographic Information Systems applied to patient distribution for Family Health Teams

João Rolando Azevedo

MESTRADO EM
INFORMÁTICA MÉDICA
2º CICLO DE ESTUDOS

October|2011

U. FMUP U. FC

3ª ed

MIM

Geographic Information Systems applied to patient distribution for Family Health Teams

João Rolando Azevedo

MESTRADO EM
INFORMÁTICA MÉDICA
2º CICLO DE ESTUDOS

ORIENTADORI

Alberto Freitas

October|2011

“Man needs difficulties; they are necessary for health”

“Carl Jung”

Abstract

Background: Family focused environment in Primary care is considered to be the future and help is required to implement new methods to perform it. One theory consists in dividing patients accordingly to geographic clusters.

Aim: To study and implement methodologies for the distribution of patients in Health Units, and develop a tool to aid in this process.

Methods: A health unit was selected to recollect and process bio-geographic data of patients. A manual division was executed and various methods were tested. An information system was developed in order to help in the division and to compare the manual with the automatic process.

Results: The original data contained a significant percentage of errors (29%). This led to the cross validation of addresses, a process that took months. Only after, various patient division techniques could be applied. One technique showed itself as having the most advantages. A robust GIS system was developed.

Discussion: The analysis took a significant amount of time. The method of dividing the patients proved itself appropriate to this situation, and could probably be applied in many urban locations. The obtained GIS provided time saving and better data comprehension.

Conclusion: Technologies in general and the system developed in particular can help patient allocation and represent a breakthrough in time-saving

Keywords

"GIS (Geographic Information Systems)"; "Public health"; "Community Health"; "Nursing"; "Family Nursing"; "Family Nursing by Geographic Area", "Primary Care Physicians", "Family Physician (Doctor)"

Resumo

Introdução: Os Cuidados de Saúde Primários são focalizados nos cuidados à Família. Nestes domínios, por vezes é necessário implementar novas metodologias. Uma das teorias existentes consiste em dividir os pacientes de acordo com áreas geográficas

Objectivos: Estudar e implementar metodologias para a distribuição de pacientes em Unidades de Saúde e desenvolver uma ferramenta para ajudar neste processo.

Métodos: Uma Unidade de Saúde foi seleccionada de forma a recolher e processar dados bio-geográficos de pacientes. Foi executada uma distribuição manual numa 1ª fase e vários métodos foram testados. Um sistema de informação foi desenvolvido par ajudar a distribuição e a comparação entre métodos manuais e automáticos.

Resultados: Os dados originais revelaram ter uma percentagem significativa de erros (29%). Isto levou a uma validação cruzada de endereços, um processo que demorou meses. Apenas depois, várias técnicas de distribuir pacientes puderam ser testadas. Uma das técnicas revelou-se como possuindo mais vantagens. Um Sistema de Informação Geográfica robusto foi desenvolvido.

Discussão: A análise demorou um tempo significativo a ser realizada. O método de distribuição de pacientes escolhido demonstrou ser apropriado para esta situação e, provavelmente, poderá ser usado noutras localizações de âmbito urbano. Sistema obtido permitiu ganhos de tempo e uma melhor compreensão e processamento da informação.

Conclusão: As novas tecnologias em geral e o sistema desenvolvido em particular, podem ajudar a alocação de pacientes em áreas geográficas e representam uma mais-valia em termos de poupança de tempo.

Palavras-chave:

"SIG (Sistemas de Informação Geográficos)"; "Saúde Publica"; "Saúde comunitária"; "Enfermagem"; "Medicina"; "Enfermagem de Família"; Médico de Família"; "Enfermeiro de Família por área geográfica".

Acknowledges

To professor Alberto Freitas for its guidance in making this thesis and all the effort.

To all professors within the Master's degree in Medical Informatics for all help and collaboration.

To the "ACES Porto Ocidental" for requesting our help and all the collaboration given, especially Dr. Rui Medon and Dra. Maria José Ribas. To the unit that was created "Unidade de Saúde Familiar Bom Porto" for all the help and patience, specially Dra. Susana Vilas Boas.

To the Department of Sciences of Information and Decision in Health (CIDES), the Center of Investigation in Technologies and Systems of Information in Health (CINTESIS) and project HR-QoD (PTDC/SAU-ESA/75660/2006) funded by the Science and Technology Foundation for making possible the realization of an information system.

To Carlos Oliveira, student in the informatics engineering department of ISEP for making the actual software in collaboration with Medical Faculty and myself.

To my master's colleague and friend, Juliano Gaspar, for all the incentive and help. Without him, this work would not be possible.

To all my friends that endured all my social absences due to this work, specially Nuno Pinheiro, Camilo Reis and Paulo Pedroso.

Special thanks to Bruno Pires for making time and possibilities to fulfill this work.

To all my work colleagues for all the patience in my absences and dark circles from saying up all night.

To my family, specially my mother for never doubting me and my capabilities.

.

List of Acronyms

ACES Heath Center Groups (Agrupamentos de Centros de Saúde – in Portuguese)

AJAX Asynchronous Javascript and XML

API Application Programming Interface

API Application Programming Interface

ARS Regional Health administration (Administração Regional de Saúde – in Portuguese)

CIDES Department of Sciences of Information and Decision in Health (Departamento de Ciências da Informação e Decisão em Saúde – in Portuguese)

CSS Cascading Style Sheets

FMUP Faculty of Medicine of University of Porto (Faculdade de Medicina da Universidade do Porto - in Portuguese)

GIS Geographic Information System

GPS Global Positioning System

HTML HyperText Markup Language

ISEP Engineering Superior Institute (Instituto Superior de Engenharia – in Portuguese)

JS Javascript

KML Keyhole Markup Language

PHP Hypertext Preprocessor

SINUS Health Information System (Sistema de Informação da Saúde – in Portuguese)

SPSS Statistical Package for the Social Sciences

SQL Structured Query Language

TXT Text File

USF Family Health Unit (Unidade de Saúde Familiar – in Portuguese)

WHO World Health Organization

Articles Published

Articles published:

Gaspar, J., **Azevedo, J**., Leal, J., and Cruz-Correia, R. (2010). GEODADIVAS: Geographic Information Systems for Blood Donation Management in Portugal. Paper presented at: HealthInf 2010 (Valência). – Available in annex 2.

Gaspar, J., **Azevedo, J**., Leal, J., and Cruz-Correia, R. (2010). Using Open Source to Create a Geographical Information System for Blood Donations. Paper presented at: HealthInf 2010 (Valência). – Available in annex 3.

Gaspar, J., Leal, J., **Azevedo, J**., and Cruz-Correia, R. (2010). GEODÁDIVAS - Sistema de Informação Geográfica para Gestão de Dádivas de Sangue em Portugal. In ABO (Portugal), pp. 33-43. - Available in annex 4.

Abstracts presented:

Leal J, **Azevedo J**, Gaspar J, and Cruz-Correia R. GEODÁDIVAS - Sistema de Informação Geográfica open source para Gestão de Dádivas de Sangue em Portugal. APIH – VII Congresso Nacional da Associação Portuguesa de Imunohemoterapia. 2009, Tomar, Portugal.

*Oliveira, C., **Azevedo, J**., Gaspar, J., Marreiros, G. and Freitas, A. (2011). GeoPrimaryHealth - Sistemas de Informação Geográfica na Qualidade em Saúde: Desenvolvimento de um Protótipo. In CNES 2011 - 12ª Conferência Nacional de Economia da Saúde (Lisboa, Portugal), pp. 1. - Available in annex 5.

*This abstract resulted in a poster that received the prize for best poster in the conference.

Articles accepted for publication:

Azevedo, J., Oliveira, C, Gaspar, J. and Freitas , A. (2011). Geographic Information Systems applied to patient distribution for Family health teams in primary health care. HEALTHINF 2012 (Vilamoura, Portugal)- Available in annex 6.

Articles submitted waiting for Acceptance:

Azevedo, J., Oliveira, C., Gaspar, J. and Freitas, A. (2011). Distribution of Patients by health Professional using GIS: a Preliminary Study**.** Submitted to HEALTHINF 2012 (Vilamoura, Portugal). - Available in annex 7.

Table of Contents

Figure index

Table Index

1. Introduction

Whenever we face ourselves with a new technology in Health, we tend to thrive with both excitement and doubt. Excitement for the capabilities that new and modern informatics applications can represent in a complex and important environment, such as Health care providing. Immediately we visualize new breakthroughs and new discover that push civilization forwards. And doubt because the natural resistance to change is inevitable (Arkowitz, 2002) and also because some new technologies can introduce more error than before (Ash et al., 2003).

Nevertheless, the human being, and the inquiring minds in particular, always try to assist workflows and procedures by trying to apply new knowledge to existing one. Helping this situation is the increasing production of data by health environments (Chaudhry et al., 2006).

Most studies take place in Hospitals, however this one takes place in Primary Care, where proximity is the word of honor (Foy et al., 2001) and where patient rely on their Doctor and Nurse in order to represent the liaison for all other specialties and to respond to their everyday needs in terms of non-complex health problems.

Simple aspects such as visualizing patient addresses in maps or deciding which information is relevant can be a challenge to both investigators and Health professionals.

This following work will address this issues and hopefully will represent an enhancement is this field of knowledge.

1.1 Aim

The main aim of this study is to analyze and implement methodologies for the distribution of patients in Health Units.

Combined with data, problems in Patient allocation between Health Professionals are always an interesting subject that, wen well addressed, can contribute to Health improvements. In fact, this is the final objective of this work, that someone can see its work facilitated and improved and also to ameliorate the Health Care providing.

A better understanding of existing data in Health environments is also one of the goals of this work. It is intended to help understanding nowadays problems or to serve as a base for future work.

1.2 Motivation

The initial drive of this dissertation is related with my daily work. Being a nurse involved in the new Primary Health Care reform and having to deal with its implications, questions of accessibility of patients to the Health Units and better resource sharing are always present.

Patient distribution among Health Professionals is also a concern because in Primary Care, patients have to be distributed by Physicians and nurses forming a list. The ways to do this are always complex and with ethical implications.

Having this problems in mind, an opportunity emerged when the Clinical Council of "*Agrupamento de Centros de Saúde Porto Ocidental (ACES)"* made a formal request to the Faculty of Medicine, in which I was currently frequenting the masters in Medical Informatics, to help it dividing patients for a new Health Unit. The new Health Unit was being projected to the city of Porto and had the ideal conditions to perform this kind of studies because they would start from zero. The unit would begin without any past records or patients allocated in any list. This meaning that a large group of patient without any Health Care team would be divided and allocated to both a Physician and a Nurse (being their Family Physician or Doctor and Family Nurse).

In the end, we can state that this study was the cross over between a necessity and an opportunity.

1.3 Dissertation structure

This dissertation starts with a background study of the state of the art on the subjects to be treated. It comprehends Primary care and its implications to Health, followed by the presentation of some of the current positions on patient distribution. After that geographic Information Systems (GIS) are addressed. Finally some considerations on Data in Health environments are made.

After this it can be divided in 3 different parts:

1. The first one involves the data retrieved and its processing to form a starting point to the following study;
2. The second comprehends the elaboration of an Information System fed by the data processed above;
3. Finally, in the third part a questionnaire was elaborated to recollect health Professionals Opinion on more features for the Information System.

All 3 parts have this structure:

- Objective presentation;
- Methodology followed;
- Results;
- Discussion;
- Conclusion.

Finally, considerations for future work and references used are stated.

2. Background

2.1 Primary Health Care

Primary Health Care is a key element in any Health System. It is in the front line, being the first contact with the population. It must be accessible when needed and provide care along all life cycle. Health promotion, self-empowerment and disease prevention are the pillars to a successful Primary System (Atun, 2004).

“Primary Health Care happens every day: when common people learn or do something useful to they’re health and to people close to them”(Sakellarides, 2006)

Studies show, that in countries where an emphasis is given to Primary Care, global health costs are diminished(Atun, 2004). Apparently this is due to the fact that Primary Care professionals use less expensive technology and are focused on prevention rather than treatment(Franks et al., 1992).

In them we can find: Physicians, Nurses, Psychologists, Pharmaceuticals, Dieticians, Nutritionists, Social Services staff, administrative staff, managers, supporting staff, other health professionals, etc.

Health asymmetries’ and the effects of income disparities are reduced in regions that invest in Primary Care (Shi et al., 1999). More, morbidity and mortality are reduced in a significant number(Stewart et al., 1997) and hospitalization rates decrease to 14/100000 and 11/100000in both acute and chronic diseases respectively(Gulliford, 2002).

For all the above, the importance of investing in Primary Care demonstrates itself vital.

2.1.1 Professionals' distribution in Portugal

Table 1 illustrates the distribution of clinical and non-clinical staff in Portugal in 2007 and the number of house visits performed by each category(DGS, 2007).

It is clear that nurses represent the majority of human resources available and have a tremendous potential in home environment. However we see that the number of nurses working in Primary care is still insufficient when compared with hospital environment.

Table 1: Distribution of professionals by working place and category and number of house visits performed (DGS, 2007)

	clinical staff			non clinical		
Workplace	Physicians	Nurses	Others	Administrative	Auxiliary staff	Others
Primary care	7033	7309	1113	6554	400	558
Hospitals	17067	30969	9089	10346	20998	4715
number of home visits done	142381	2354591				13589

One of the interesting aspects is the number of home visits done by professional. Nurses are the leading ones with a number that exceeds 2 million per year. In this context, family approach is facilitated and Nurses do perceive their role as family centered professional (MCSP, 2008).

Changes in demographic aspects such as population ageing, increasing number of chronic diseases, etc., and also economical cut backs can and will influence greatly the way care providers work.

In this context, nurse care takes a vital role in helping people overcome the challenges that increasing age and prevalence of co-morbilities represent.

2.2 Primary Care Reform

In 2005 (MS, 2005) the Portuguese Government restructured the Health Care Institutions. The basic unit of this level of care is the Health Center divided into: Family Health Units, Personalized Care Units and Continuous Care Units. The first ones presented a new group of changes, responsible for a responsibility driven work. . In 2007, for the first time in Portuguese History, professionals could choose their team and incentives where promised to efficiency and quality of care (MS, 2007). In this new format, Home Care is prioritized. Professionals are encouraged to have a pro-active behavior and seek out the patient in his home environment.

In fact, Physicians, Nurses and Clinical Secretaries formed a cohesive group and where forced to write their own plans to the unit to be created. Health indicators had to be measured after, in order to evaluate the professionals work. In the beginning of the year, a negotiation takes place between the central administration and the group. They set targets, always in conformity with good practices and quality standards

In the last years, an attempt has been made in European Union to change the way Health Care is provided.

The shift characterizes itself by a deviation from hospital cares to a community and primary level. Studies in this department show that in order to save coasts and implement a prevention drive, an important bet in primary care must be done (WHO, 2000). In this new format, Home Care is prioritized. Professionals are encouraged to have a pro-active behavior and seek out the patient in his home environment.

The final expected result is than rather than the patient going to the health center after a problem occurs, primary care professional call him before something bad happens.

In 2007, the Ministry of Health of the Portuguese Government stated that the methodology to be adopted in Primary Care Nursing was: Family Nurse by Geographic Area (MS, 2007). In article 9 we can read: “to each nurse should be entrusted with patients within 300 to 400 families by specific geographic area”. In practice, this number corresponds in an average of 1500 to 1750 individual patients. This also can be applied to family Physicians lists.

The Primary Care Reform gives Family Nursing a huge autonomy and responsibility in order to insure quality answers to the community needs.

2.3 Family Health Teams

In the last years, an attempt has been made in European Union to change the way of Primary Care providing.

The shift characterizes itself by a deviation from hospital cares to a community and primary level. Studies in this department show that in order to save coasts and implement a prevention drive, an important bet in Primary Care must be done (WHO, 2000).

Nursing, as a profession, is inextricably family related. However, this being a separate disciplinary area is relatively new (ICN, 2002). Physicians in this scenario work as a cohesive team with Nurses and have the same Community oriented work.

Recently, WHO published a report ”Health XXI” in which it focused the importance of family nurse, a concept relatively new and at the same time surprisingly knowable. In this paper the roles of nurses in family environment where presented (WHO, 2006):

- Continuous care;
- From birth to tomb;
- Promotion and protection of health;
- Disease prevention;
- Rehabilitation;
- Palliative care.

The way of accomplishing this had the following processes (WHO, 2006):

- Developing world strategies that guaranties an efficient number of professionals and a good level of education;
- Legal requirements updated and adequate;
- Stimulate the nurse professional to work in a more inter-dependent and autonomy way, thus improving their potentiality;

- Based on this new paradigm, a new educational program is being proposed in the way that theory and practice can meet.

So, the figure of "family nurse" is created. He should integrate the heath providing philosophy within an institution. "Wherever the nurse works, his focus is the family- its health, its capacity to grow, take care of itself and contribute to the community" (Joel and Stallknecht, 2000). WHO also suggests that Physicians adopt the same way of Health Care providing because they have the same postulates.

A family oriented professional role is defined by (OE, 2002):

- Identify a population throw a organized file;
- Taking responsibility for the care of nursing needs of the defined population;
- Shared responsibility for the hole care giving (involving other professionals);
- Meet the health state of its population.

Through the early detection of problems the Family Team (both Family Physician and Family Nurse) can serve as a sort of middle man between the population and its Health necessities.

The final expected result is than rather than the patient going to the Health Center after a problem occurs, the Physician or Nurse calls him before something bad happens.

The community is the ideal workplace. The Health professional has to know the socio-demographic characteristics of his population and manage all variables establishing focused objectives.

2.3.1 Family Health Team route in Portugal

After Munich conference, the "Ordem dos Enfermeiros" (institution that regulates nursing practice in Portugal) assumed the following compromises:

- Spread the word and thoughts throw all involved parties;
- Perform a national conference in order to establish action priorities;
- Encourage the experience share between professionals;
- Present the Health Ministry with a project to implement this methodology.

At a governmental level it was established the appointment of a work group in order to elaborate a monitoring guide.

The Institution that rules Physician practice in Portugal ("Ordem dos Médicos") shares many of the statements above stated. However it still has some reserves in Patients list being radically changed when they already have a health Problem assigned.

All this, and the use an information certified system plus the possibility of giving economic incentives to groups wanting to work with this design, gave a thrust in Primary Health Care towards its' effective implementation.

In 2007, the Ministry of Health of the Portuguese Government stated that the methodology to be adopted in Primary Care Nursing was: Family Nurse by Geographic Area (MS, 2007). In article 9 we can read: "*to each nurse should be entrusted with patients within 300 to 400 families by specific geographic area"*. In practice, this number corresponds in an average of 1500 to 1750 individual patients. List distributions to Physicians in terms of number of families and patients are exactly the same.

For the first time, Health professionals should reunite and distribute themselves among their influence area accordingly with the existing socio-demographics (ex. high residential areas have different necessities than rural environments).

All this taken in consideration, it urges a structured plan of implementing the Family Nursing Methodology and, in cases where this can be possible, to implement a Family Physician Methodology using geographic criteria. Its possible perceived advantages are (OE, 2002):

- Increases nursing visibility to the patient and health system;
- Improves the patient/nurse relation;
- Improves the health indicators;
- Allows a better involvement with other professionals;
- Gathering a geographic area working framework, it facilitates the link between family and the nurse;
- Diminishes the time spent to cure;
- Reconfigures the working framework to a more home centered;
- Allows a better measures implementation to risk groups.

2.3.2 Difficulties in implementing Geographic methodologies

The fundamental difficulties in implementing a family nurse oriented policy in Portugal are (OE, 2002):

- An organization centered in the family doctor image;
- Insufficient academic preparation
- A lack of structured policy by upper managers;
- Information support systems not adequate;
- The lack of perceived culture by the population in general about the figure and competences of a family nurse.

Studies from the Physicians point of view are not available, and opinion in general tends to be very variable.

Despite the clear orientation given by the Health Ministry, the methodology used by health units is not the desirable one. Many teams ignored the Geographic Area requirement and formed list by simple aggregating Doctors with Nurses with preexistent randomly formed lists.

The most difficult aspect of all the requirements is, in undoubtedly, the manual process that Health Professionals are obliged to do in order to divide and access the needs of their population.

There are still no studies in Portugal (and I didn't find any elsewhere too), regarding the methodology of Family Health Teams by Geographic Area.

2.4 Health and Information systems

One of the mandatory premises in order to successful implement a Family Nurse Methodology is the existence of a structured way of registering what nurses do (OE, 2007).

That existence includes legal-ethic issues and support decision in the continuity and quality of care, education, investigation and management (patient oriented or institution oriented).

Despite this requirement, studies demonstrate that visibility of nurse work is still low and shallow. The impact in the population in general is thus very difficult to measure.

The Portuguese government is trying to change this aspect, implementing a policy to help private companies and governmental institutions to develop information systems capable of showing rapidly and efficiently the intervention made and possible results.

In the Physician department, the work visibility is higher and much more systems are developed because they have to be more complex (Marc, 2001), however geographic issues are still not sufficiently addressed.

Population increased mobility and the necessity of following people through all life path, drives information systems into a computerized framework (WHO, 2000).

In 2007 a series of recommendations were written focusing the general principles Information Systems should follow in order to be accepted (OE, 2007):

- They should work in an integrated manner, guarantying interoperability with existing or to exist modules. Conformity with International Standards (ISO 18104, CEN, HL7, etc.) and with the state of the art in the moment of its creation must be accomplished;
- ICNP (International Classification for Nursing Practice) should be the classification used in order to develop a pattern of investigation, production of indicator and quality improvement;
- Systems should always be capable of document the nursing practice in all of its process;
- In conformity with the Minimum Data Set in Heath, systems should incorporate a classification of patients in term of dependence levels in Nursing Care.

Sharing information between institutions is also an important flag of the requirements imposed. It is recommended that the user accesses information in real time and different institutions can withdraw crucial data in order to create a web related way of manipulating patient information.

Protection, safety and data confidentiality must be guaranteed respecting legal and ethical requirements present in both Physician and Nurses deontological code.

2.4.1 Health Professionals and Informatics

The importance of informatics in Nursing is increasing. It is considered “an applied science”(McCormick et al., 2007). In the Physician department, informatics are seen as crucial to everything since Care Providing to investigating (McGlade et al., 2001).

Some considerate it as a Health specialty that integrates: Health care science, information science and computer science. It facilitates the integration of data, information and knowledge in order to support decision making to nurses and other providers (ANA, 2001).

More and more, Health Professionals who have academic preparation in informatics, play a vital role in health management. They can be researchers, educators, managers, consultants, etc. (McCormick et al., 2007).

2.5 Family Health Team

One of the main focus of the reform was to introduce a new paradigm on health, the methodology of allocate a “Family Health Team” to every patient (MS, 2007). A “Family Health Team” consists on a Physician and a Nurse, both in Primary Health context. Also, one of the main theories nowadays talks about clustering patients in geographic areas, meaning, dividing them accordingly to their home address. These theories come, mainly, from the nursing area but are extensible to both areas (Joel and Stallknecht, 2000; MS, 2007; OE, 2002, 2007).

The major difficulty on doing this implementation is to divide the patients geographically. Manual division is being use, thus consuming many human and time resources.

2.6 Geographic Information Systems

Geography takes a fundamental role in almost all decision we made. The choice of places, the appointing of market segments, the planning of distribution networks, response to emergencies scenarios, the redrawing of countries frontiers, all those problems address geographic issues(Nuckols et al., 2004). Geographic characteristics such as topography and geographic dispersion of population are fundamental factors in fair resources distribution.

It is defined by: "a computer-based system for integrating and analyzing spatially referenced data"(Cromley, 2003).

GIS crosses data and geography digitally with the purpose of building maps allowing us to visualize data with different degrees of complexity in a map. This gives us a useful way of reveling spatial and temporal relations between data.

A GIS integrates hardware, software, capture or recollection of data, management, analysis and the presentation of all ways of information geographically as seen in Figure 1.

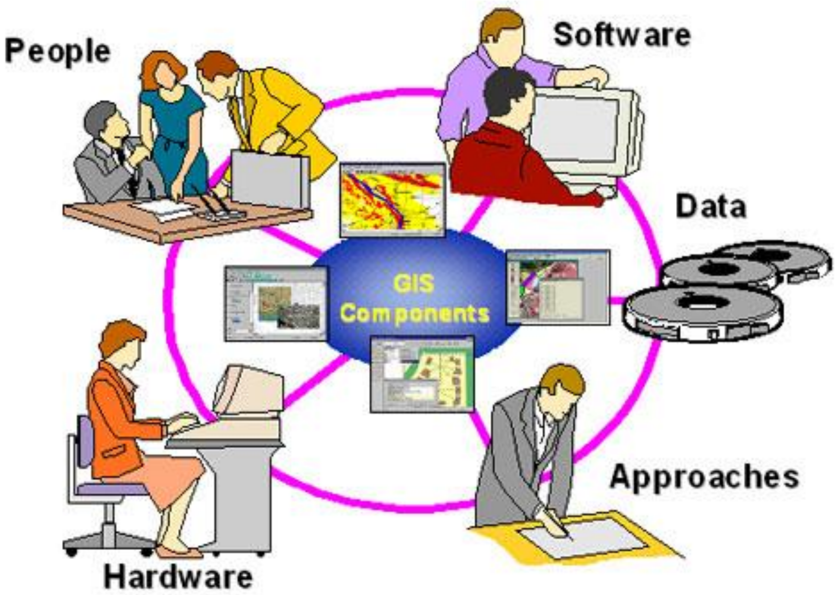

Figure 1: GIS components (educononline, 2009)

The allocation of references in a map can, frequently, represent a better perceptible graphic representation for the end user, than tables and text. That allows an improved access to information and a richer extraction of knowledge. For example, a doctor wants to visualize his families and their dependence levels crossing this information with the proximity of a day care center.

A GIS can be seen in 3 different perspectives:

- Database: A GIS is a unique form of database, a geographic database. It's an "Information System for Geography". Fundamentally it can be based in a structured database that describes the World in geographic terms.
- Map: It constitutes a set of intelligent maps that show characteristics and relations between many forms on earth surface. Information maps can be build and used as windows to a database in order to support queries, analysis and information editing.
- Model: an aggregate of tools that can transform information into a new set of data from existing one. These geoprocessing functions recollect information, apply analytical functions and show results in a new derivate set of data.

Combining data and applying some analytical rules, it is possible to create a pattern in order to help answer the question previously made (Esri, 2005-2011).

The GIS primary goals are(Maged and Kamel, 2004):

- Inform and educate health professionals and population;
- Support decision making in many levels;
- Prevent results before making any compromises;
- Select priorities in lower resources environments;
- Change bad practices and routines;
- Monitor and watch continuously changes implementations.

GIS allow us to:

- Understand, question, interpret and visualize data in many forms revealing relations, patterns and tendencies under the form of maps, globes, reports or graphs:

- Answer questions and resolve problems allowing looking at the data in a faster and easily shared way;
- Integration in almost any Information System within an organization;
- Resolve more problems than the simple use of a mapping program or the adding of data to an online mapping tool.

2.7 Geographic Information Systems and Health

Health related GIS have 2 main forms(Jean-Baptiste et al., 2005):

- Epidemiology – focus on the exploration, description and modeling of spatial-temporal incidence and prevalence of disease, the detection and analysis of disease clusters and patters;
- Health care – allowing analyzing and characterizing the distribution and the access of institutions (hospitals, health centers, blood centers, etc.)

Many of those systems possess simple functions such as measuring the distance between resources and the population as seen in Figure 2. These aspects, apparently basic, of a geo-politic-spatial infrastructure, can determine the population accessibility to health care.

Many, also have a Data mining tool associated with the data warehouse, allowing and interactive representation of knowledge to support the decision process.

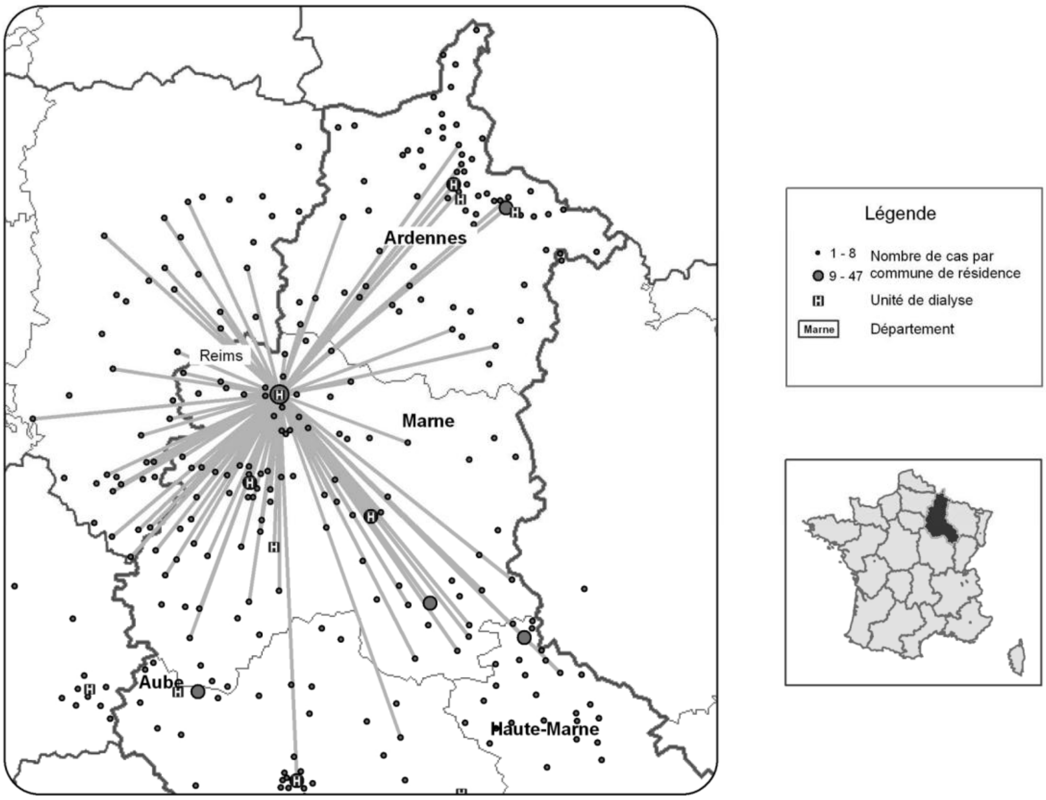

Figure 2: example of a GIS utility. It measures the distance between clinics and a central hospital (Jean-Baptiste et al., 2005)

2.7.1 How can GIS help

Investigators, Public Health professionals, policy making responsible and Health Professionals in particular can use GIS to better understand geographic relation that affect health results, risks, disease transmission, health care access and other public health concerns. They're being used more and more often to deal with problems in a local, regional, national or international overlay (NCHS, 2009).

The manual system Health Professionals are obliged to use in order to distribute their families is completely obsolete. It discourages the actual doing and takes a precious amount of time.

A GIS with a solid platform can not only help implementing but also test and validate this methodology. GIS are an example of tools that are being use in similar cases in order to present information on maps.

The simple visualization throw graphic tools can help rather the professionals in the effective care and also patients using the units itself.

2.7.2 The importance of using GIS in health environments

Health environments have a numeric-spatial scenario, meaning that they tend to need problem solving in these 2 aspects. As seen above in the type of systems mostly used by people, statistical try to take care of the first aspect and GIS the second. However, a robust GIS can also respond to these 2 aspects (Scotch et al., 2006). It can possess powerful analysis functions as well as the ability of showing things on a map.

It is proven that the use of GIS in health care is effective (Graves, 2008) and useful in discovering new patterns and optimize existing resources.

Studies still show that there is a slight underutilization of powerful GIS systems (Mowat et al., 2000.).

Specifically in Primary Care, the potentialities of GIS are tremendous. They could help tracking disease locations, incidences, surrounding environments, health care facilities, geographic boundaries of communities and other community infra-structures (Faruque et al., 2003).

However, increased efforts are needed to make health professionals and health organizations aware of the possibilities of these information products for empowering their decision making (Endacott et al., 2009).

Most databases tend to be data-rich, but information-poor (CWHPIN, 2000). Information extraction techniques must be developed in order to aid investigators perceive the real world.

Map representation can overcome these shortages. As quoted by Dredger S: “if a picture is worth a 1000 words, then a map is worth 1000 pictures” (Dredger et al., 2007).

Some characteristics must be present in any GIS:

- The presence of a common and easily perceived map (with familiar references such as roads, rivers, etc.);

- The possibility of creating graphs and tables to help contextualize the raw data;
- The possibility of allocate address-based client data;
- The inclusion of choropleth maps (where areas are colored or patterned in proportion to the measurement of the statistical variable being displayed, such as population density or per-capita income) (Slocum et al., 2009);
- The ability to cross multiple variables, including or excluding them according to preferences.

Despite the evident benefits of GIS use, its dissemination and utilization it’s not yet a generalized reality. Some possible explanation for this to happen can be (Rob, 2003):

- The lack of consideration towards user needs;
- Elevated cost of existing applications;
- The need to learn the way they function and operate.

At requirement level, we verify an almost total need of community involvement since the very beginning. Meaning, users and developers must work directly together in the project. Only in this way can projects be realistic, reasonable and sustainable (Weiner et al., 2002).

2.7.3 Geographic Information Systems Applied to Health Environments

“Application of Geographic Information Systems (GIS) in health sciences is relatively new, but it appears to be expanding faster than any other areas of GIS application” (Faruque et al., 2003).

Disease’s incidence location, health care facilities, community’s geographic boundaries, and other essential infrastructures have always been vital components of epidemiological and health care studies.

“The first real GIS, known as the Canada Geographic Information System, was operational for the Canada Land Inventory project in the sixties under the guidance of Roger Tomlinson who is regarded as the father of GIS” (Longley et al., 1999).

The use of GIS in health care research is concerned with the delivery of, and access to, health services (GATRELL and SENIOR, 1999).

Although GIS has been widely used in various epidemiological and health care related studies, few of them have been conducted in community health assessment, and even fewer are reported in nursing research (Faruque et al., 2003).

Community health is defined by as the identification of needs and the protection of and improvement of collective health within a geographically defined area (Allender and Spradley, 2001).

Geography and its aspects are very import in Community health because they define one another.

GIS can (Goodman and Wennberg, 1999):

- organize spatial data
- visualize spatial data
- analyze the data

"For policymakers, the concern is not how GIS works but what it actually achieves; that is why GIS has also been defined as a spatial decision-support system" (Cowen, 1988).

The primordial questions that these types of applications want to solve can be summarized in (Faruque et al., 2003):

- What exists at or near a particular location?
- What geographic areas meet certain criteria?
- What spatial patterns exist?
- What spatial association exists?
- What if the variables are changed?

2.7.4 Internet based GIS

There are many types of GIS platforms available, since close circuits, passing by wider range circuits such as Universities and/or institutions and lastly we can have web based ones (Leitner et al., 1998). These ones could use a server for geographic information (typically maps) and a platform to perform queries, presenting this way better accessibility. In this context, information and its manipulation capacity are made available to all authorized people.

In Figure 3 we can see an example of GIS architecture. It's composed by several layers: 1 the client; 2 dynamic web server and 3 a cluster of databases that are in charge of feeding the data Warehouse.

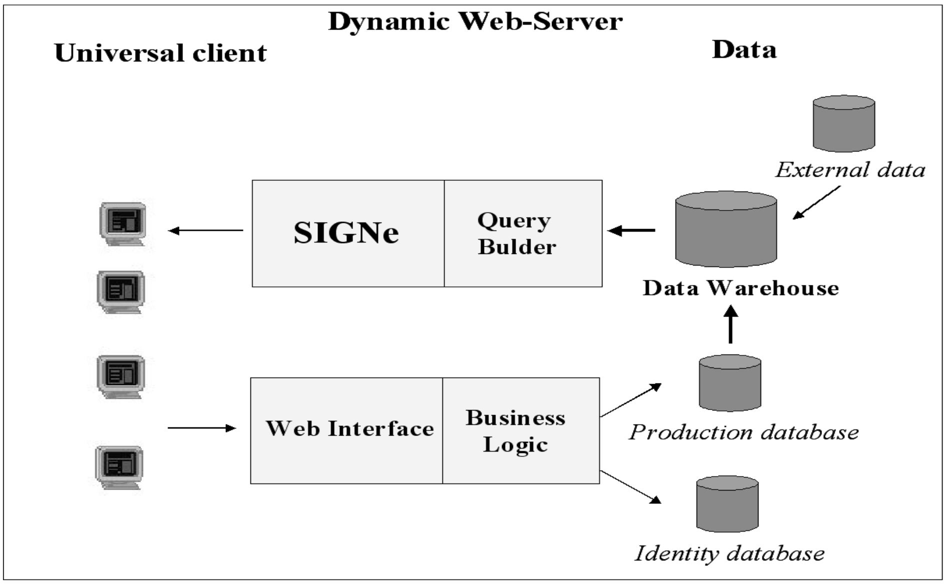

Figure 3: An example of a GIS architecture (Jean-Baptiste et al., 2005)

2.7.5 Google API and Google maps

One of the possibilities of having a free map server is using Google Maps.

Google maps, launched in 2005, is a survey and map visualization free service utilizing satellite images. It's provided by Google Inc. Corporation. The service, besides maps and satellite images, provides routes between pre-determinate spots, zoom, dragging the map, among others (Google, 2009).

In April 2006, version 2 of Google Maps API was launched, allowing users to become actual information producers. With a Google account it's now possible to create own routes, spots and areas, make remarks and share respective map access links. It's also possible to create a KML archive with coordinates and geometric forms in order to visualize in the map presented by Google Maps server, as shown in Figure 4.

Figure 4: coverage in France (GeoinWeb, 2009)

Google Maps API allows the use of JavaScript to incorporate Google Maps in Web pages. It provides a multiplicity of utilities in order to create maps and add content to it.

Google API it's a free service in testing beta phase, available to any free web site. All new innovator and important projects are being developed in a net based environment, thus the need of systems of this caliber.

The simplicity of using Google Maps is its biggest asset. The grabbing and dragging possibility, increase or decrease zoom without big delays in the web page are a few of the simple tasks that favor it.

Google Maps present an approach based on: "Asynchronous JavaScript + XML (Ajax). Ajax is defined by a new technology set, which one of them in its specific environment, coming together on a common purpose. It incorporates standard based presentations using XHTML and CSS; dynamic display and interaction using "Document Object Model"; exchange of data and its' manipulation using XML and XSLT; asynchronous data recovering using "XMLHttpRequest"; JavaScript unites all the above (Garret, 2005).

The functioning system of Ajax applications eliminates the interaction start-stop-start-stop on the internet, thru the introduction of a middleware (Ajax engine) between the user and the server. Its primary function is to assure that the page update is done only in pre-defined parts, reducing the interaction delay with the user.

2.7.6 GIS and statistical software

Although GIS are rapidly increasing in use, nowadays it's very common the use of another type of application to complement it: Statistical Software (Scotch et al., 2006).

Statistical applications are commonly used for analyzing and processing the data. Afterwards GIS present this information in a Geographic manner.

In Table 2 we can see that many of the process is being done by Statistical Software and only the spatial boundaries establishment is allocated to GIS.

Table 2: Types of software most used in different problem solving categories (adapted from (Scotch et al., 2006))

Numerical-Spatial Problem Solving Category	Most Popular Type of Information Technology
Data Management/Access	Statistical Software
Data Navigation	Statistical Software
Geographic Comparison	Statistical Software
Spatial Boundaries	GIS Software
Spatial Modeling	Statistical Software
Ranking Analysis	Statistical Software

This shows us that, although GIS application have a tremendous potential, people still trust in other software due to the insufficient power of existing GIS applications.

2.7.7 Usability in Web-bases GIS

Usability and its' criterions in these applications are part of the investigation process. Spatial information visualization must be tested using usability engineering principles and any application in this field must be tested and validated using conceptual usability methodology (Vanmeulebrouk et al., 2008).

2.7.8 GIS and Open Source

Directly applied to GIS execution, Open Source methodology has the following advantages (Vanmeulebrouk et al., 2008):

- It can read and write all data formats;
- It can run in multiple platforms;
- Free of charge, meaning no license fees;
- Foredoom to change the source code; meaning independence from vendor;
- Easy learning tools;
- Avoids commitment with proprietary software;
- Compatibility with existing GIS Framework;
- Can be adapted to local cultures and languages;
- It deduces dependence of developing countries towards others;
- Boosts the grow and sustainability of local companies;
- They have achieved a state of maturity, sophistication, robustness, stability that rivals with proprietary software.

2.7.9 Why use Open Source

In this domain, it's not necessary an expensive, full of licenses system neither a specialized education in computing science nor plenty of previous experience (Purvis et al., 2006).

Open Source is viewed as a model of promoting the quality and accessibility improvement. Its advantages in this area are:

- The source code is available;

- Its license allows unlimited product distribution;
- It also allows the creation of derivate products;
- It can be use when and wherever the user wants;
- A community of developers is forms in order to assure maintenance, support, continuous development and project success;
- Its reliable throw peer review.

2.8 Data in Health environment

Data in Health is a milestone because it directly interferes with health care providing. In fact, the process of decision making is much influenced by data of various sources (seen in Figure 5). We can see that a complex process is used to finally make health decision and easily infer that any disruption in a data source can result in disaster.

Decision making itself can be summarized in:

1. Recollect data;
2. Analyze it;
3. Formulate an hypothesis/diagnosis;
4. Prescribe and perform a treatment;
5. Access efficiency;
6. Recollect new data;
7. Analyze it;
8. Back to step 1.

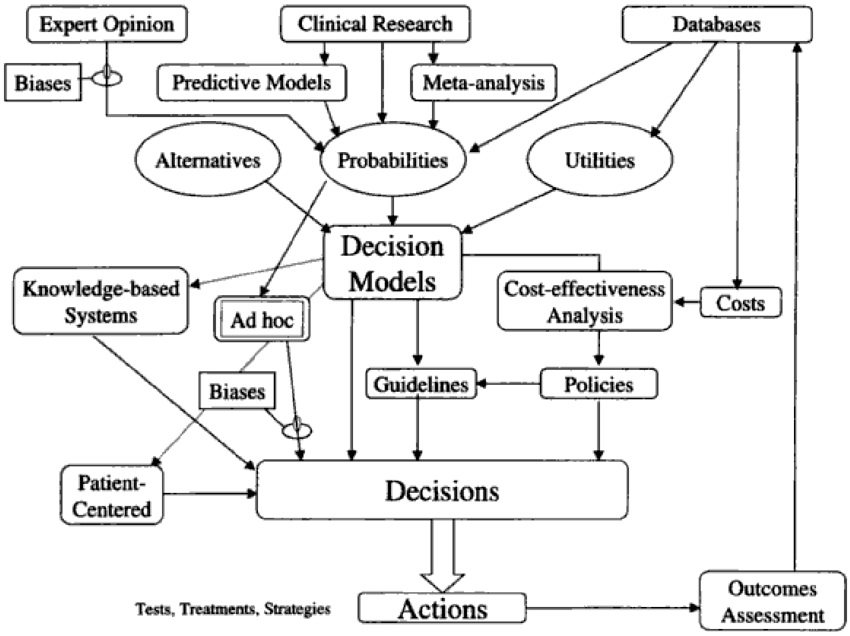

Figure 5: Overall schema of medical decision making *in (Chapman and Sonnenberg, 2003).*

The action to implement is dependent of the data recollected (Shortliffe and Cimino, 2006).

There are many types of different data that can exist in a health related environment. It could be a weight of a person, its vaccines and immunization status, medical exams results, diagnosis, etc. (Shortliffe and Cimino, 2006).

Because of this variety of data, there's also a multiplicity of ways to collect, store and process it. It depends mainly of the final person responsible for it.

Systematized and structured data in order to extract some knowledge can be defined as information (Palmisano and Rosini, 2003), this means that information can be an interpretation of data executed with a defined purpose. These interpretations can represent analysis, processing, comparison, classification and establish relationships in order to make decisions.

Final knowledge comes with the processing and interpretation of information.

2.8.1 Minimum Data Set

Many studies enhance the need to elaborate a structured minimum sub-set of data in order to correctly identify either patients or procedures (Hall et al., 2002). Some of the fields considered relevant are:

- Patient identifier;
- Practice identifier;
- National Health Index identifier:
- Sex;
- Date of birth;
- Ethnicity;
- Consultations file;
- Consultation date;
- Consultation identifier;
- Address:
    - Street name;
    - Door number;
    - City;
    - Country;
- Etc.

A quality and centralized minimum data set can identify without error patients and also minimize the manual error insertion of data (Landi et al., 2000).

2.8.2 Data quality

With the increase of medical and Health related information, the amount of errors in data represent a challenging aspect (Petrovskiy, 2003). In fact they can represent a rich field of information and inference exploitation because of their importance in the overall context of databases and information systems.

The identification of errors can incorporate (Gaspar, 2011):

- Anomalous cases – that not necessarily represent errors but rather can be deviant cases, cases that are far from average also called "*outliers*". These cases can be studied in order to be a point of reflection to future cases;
- Incorrect data – data that violates a determinate assumption and is, intrinsically incorrect. Examples of his can be: a person weight to be 87000 Kg, when is obvious that this in incorrect because it violates the weight gap acceptable to a human being;

and also false information fed to a system such as a street name that is not actually the patient address.

Due to the importance of these aspects, the need to investigate better ways to analyze and extract knowledge from large data sets is always present to the Health investigator (Freitas et al., 2010). The quality of data is an aspect that is present in almost every study recently made in order to rapidly identify and correct errors and also postulate rules in order that the errors are not repeated and prevented (Freitas et al., 2005).

Also, the quality of data is often seen as subjectively measured, because it depends greatly of the purpose data will be used. The user must assess its judgment in information pertinence and its use (Olson, 2003). In many scenarios, quantity of data accessible is less important than intrinsic quality. It's easily perceived that, for example, a simple typing error in a procedure in a single user can originate in serious malpractice.

2.8.3 Data processing

Real-world databases and specifically Medical related ones have a high probability of containing noisy (with data errors and outliers), missing (non-recorded instances), and inconsistent (incongruent and absurd values) data due to their typically huge size and their likely origin from, most of the times, human manual typing sources. Low-quality data will lead to low-quality mining results.

Process that can be used to process data (Han and Kamber, 2006):

1. Discrepancy detection poorly designed data entry forms that have many optional fields, human error in data entry, deliberate errors (e.g., respondents not wanting to divulge information about themselves), and data decay

    a. How to do it – use previous knowledge of the data (referred as metadata), such as type of data, boundaries, expected values, etc.." *Values that are more than two standard deviations away from the mean for a given attribute may be flagged as potential outliers" (Han and Kamber, 2006).* Data scrubbing tools use simple domain knowledge (e.g., knowledge of postal addresses, and spell-checking) to detect errors and make corrections in the data.

2. *Data transformation* – after finding discrepancies, data must be corrected. Both detection and transformation are iterative processes and one can influence the other.

Different types of data processing (Han and Kamber, 2006):

1. *Data Cleaning* – has the goal of identify, remove or correct data inconsistencies;
2. *Data Integration* – focused on assembling multiple sources of data in a coherent manner;
3. *Data Transformation* – has the same goal as data cleaning but generally normalizes data;
4. *Data reduction* – has the goal of reducing data sizes by eliminating redundancy, clustering, or aggregate similarities

These processes can be used separated or simultaneously.

Missing values processing:

- Ignore tuple;
- Manual filling;
- Fill with consistent value (unknown, none, etc.);
- Use central tendency measures (mean) to fill the value;
- Use central tendency measures (mean) within same classes;
- Use the most probable value to fill in (using with regression, inference-based tools using a Bayesian formalism, or decision tree induction)

It is important to be alert to the values that can predict missing values such as unknown, not applicable.

Data transformation available (Han and Kamber, 2006):

- Smoothing –in order to remove noise:
    - Binning – the replacement of values according to their neighborhood;

- Regression – smoothens values by trying to find a central linear tendency;
- Clustering – values are aggregated by groups or clusters. Within the cluster one can observe some homogeneity and outside it outliers may be found.

- Generalization – replacing data for a tighter interval (ex weight to under, correct, overweight, etc.)
- Normalization – modify attributes so they fall in a pre-determined range;
- Construction – build new attributes from the pre-existent ones.

Data reduction techniques available (Han and Kamber, 2006):

1. Data cube aggregation - aggregation operations are applied to the data in the construction of a data cube.
2. Attribute subset selection - irrelevant, weakly relevant or redundant attributes or dimensions may be detected and removed.
3. Dimensionality reduction - encoding mechanisms are used to reduce the data set size. Principal components analysis – the main attributes (the ones that better define the data) are selected and the other are ignored (Karhunen-Loeve, or K-L, method). This implies normalization of the attributes.
4. Numerosity reduction - the data are replaced or estimated by alternative, smaller data representations such as parametric models (which need store only the model parameters instead of the actual data) or nonparametric methods such as clustering, sampling, and the use of histograms.
5. Discretization and concept hierarchy generation (seen in Figure 6), - raw data values for attributes are replaced by ranges or higher conceptual levels. Data discretization is a form of numerosity reduction that is very useful for the automatic generation of concept hierarchies. Discretization and concept hierarchy generation are powerful tools for data mining, in that they allow the mining of data at multiple levels of abstraction.

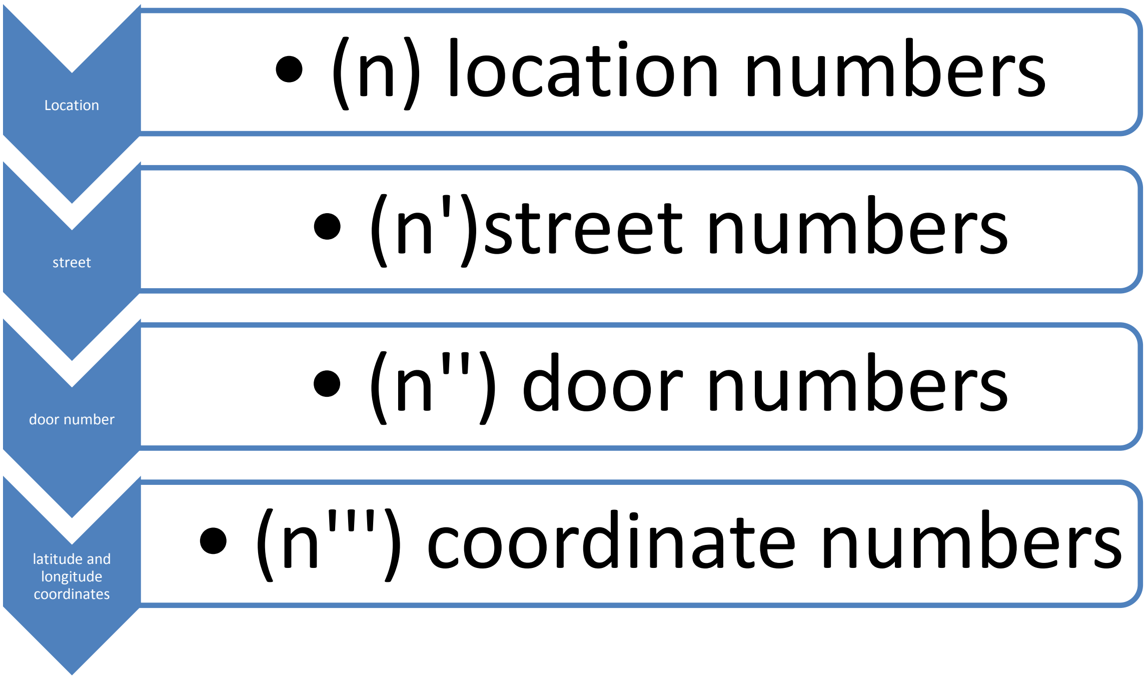

Figure 6: schema of concept hierarchy based on the number of distinct attribute values

3. Objectives

The aim of this study is to analyze and implement a distribution of patients of a health unit, and to develop a tool that can aid in this process. The distribution should be calculated throw various tested rules, using data from a health organization. The population studied has in common the fact that it does not have a definitive family physician and nurse (without a Family Health Team).

The main objective is to review the sate-of-the art in this particularly subject: the possibility of using GIS in a Family Health context.

3.1 Specific Objectives

- Investigate how is the actual status of Primary Care in Portugal;
- Verify what is being done in terms of distribute patients by geographic areas;
- Recollect data from a population without Family Physician and Nurse (without a family team);
- Verify the data integrity with the collaboration of Health Professionals. It is intended to identify record errors, missing values and incongruent data;
- Develop a method to allocate patients to different lists in order to give them a Family Health Team, using a theoretical method elaborated for this study (this method must include the requisites postulated by the theory of distributing patients by geographic area);
- Identify the best combination for Physician-Nurse pair;
- Test various methods of dividing and forming patient list using the requisites of equity, ethics and responsibility (always respecting patient rights and health improvement) and choose the better one;

- Perform the division manually;
- Elaborate a GIS tool that can help doing the above using Geographic Information Systems;
- Perform the division using the elaborated GIS;
- Present the patient division to the unit in order to proceed to the inclusion of patients in the lists, giving them a new Primary Care unit with both Physician and Nurse;
- Assess if distribution is accordingly to requirements;
- Identify future data that would be considered relevant in the GIS.

4. Methods

This study followed this methodology:

- Perform a bibliographic review of the subject;
- Select a health unit to collect data;
- Compile biographic and geographic data from the chosen unit;
- Process the data in order to filter errors;
- Execute a manual division of patient clusters;
- Test various methods to divide population;
- Develop an information system in order to help the division;
- Present the results to the health unit;
- Compare manual to automatic division;
- Verify the implications in health care of the above.

4.1 Bibliographic review

A search was made in PubMed, ISI web of knowledge and Scopus, including the following Mesh terms:

- GIS – geographic information systems;
- Public health;
- Public Health Nursing;
- Community Health Services;
- Community Health Nursing;
- Nursing;
- Family nursing;

- Primary Care Physicians.

Other terms where combined to search in the same sites and the web in general:

- Geographic area nursing;
- Geographic family area nursing.

A manual search was made in physical libraries. The following institutions were visited:

- Nurses Regional Institute (Ordem dos Enfermeiros, Secção regional do Norte);
- Porto School of Medicine, in the University (Faculdade de Medicina da Universidade do Porto)
- Porto Nursing School (Escola Superior de Enfermagem do Porto)

Thousands of papers were retrieved. The following inclusion criteria were used: suitability to the theme, scientific validity, age of the paper (no time limits were set but preference was set for studies over the year 1998). All abstracts were read and full paper of relevant ones was retrieved (when available).

4.2 Study Location

The decision of doing the survey at "ACES Porto Ocidental", future Family Unit "Bom Porto" was a conjunction of factors:

- The first location chose was the investigators working place. This selection was proven ineffective because of the interference with the investigation itself and because of the refuse by the local team to participate in this study and to engage the following work methodologies changes.
- An opportunity emerged when governmental plans were made to create a new "Family Health Unit" within the "Porto Ocidental" health center. It was to be made involving 3 Health Center units. Its main goal is to provide a Family medical team (including Physician, Nurse and Health secretary, 4 of them respectively) to approximately 7000 patients that don't have it nowadays.

- The Health professionals from the future unit requested help to the University build patient list, respecting the assumption that those list were to be made in a geographical Area methodology;
- The department of Medical Informatics of Medicine Faculty received this help request and reviewed if there was any post-graduate study in this area. Being my project in this area, the request was sent to me;
- The present investigator was selected after finding common objectives to the study request to the department.

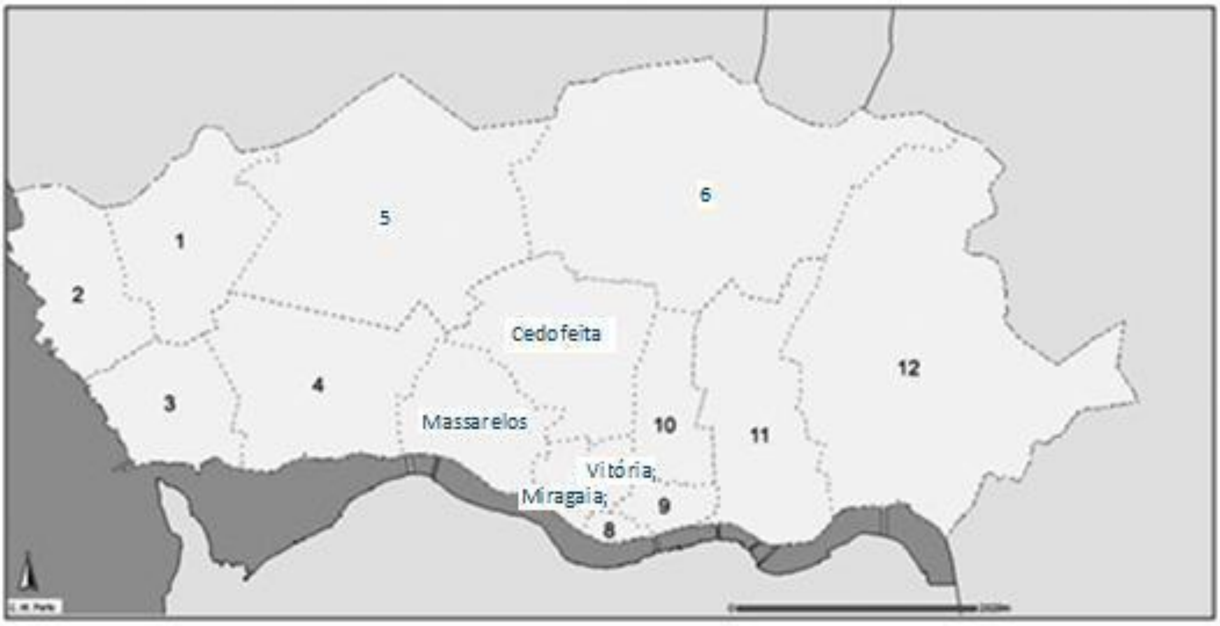

Figure 7: area of Future unit – 4 parishes in the center with name

This study was made in Portugal, in the city of Porto, in the western Health Center ("Agrupamento de Centros de Saúde Porto Ocidental") in order to help opening a new Family Health Unit - "Unidade de Saúde Familiar Bom Porto". For this unit, 5 lists of patients were needed because there were 5 Physicians and 5 Nurses.

In Figure 7 we can see all Porto City administrative divisions: 1 - Aldoar; 2 - Nevogilde; 3 - Foz do Douro; 4 - Lordelo do Ouro; 5 -Ramalde; Paranhos; 8 - S. Nicolau; 9 -Sé; 10 - Santo Ildefonso; 11 - Bonfim; 12 – Campanha.

4.2.1 Original data

In order to import data, a formal request was made to the Ethics Committee of ARS (Regional Health Administration) , the entity ruling over ACES Porto Ocidental where data was stored. This request was granted.

Data was contained in one information system called SINUS in 3 different servers corresponding to three different units. Although data in digital format was asked, that was not possible. Instead, the data was given in text format in many txt files containing the following variables:

- Code patient number;
- Patient name;
- Phone number;
- Original health center;
- Date of birth;
- Original physician;
- Type of inscription;
- Address city;
- Address street;
- Address door number;
- Zip code.

After obtaining the several txt documents, a series of manual procedures were made in order to transform them into digital format. Although in plain text, the data followed a structured format in which all attributes where separated by one of the following separators:

- Tab
- Paragraph
- :
- -
- ,

To separate different patients, 2 or more 3 paragraphs where used. An example of original data can be seen in Figure 8.

```
Code·number:······xxxxx· ··date·of·birth:·xx-xx-xxxx· ···Tel:··xxxxxxxxx¶
aaaaaaaaaaaaaaaaaaaaaaaaaaaaa¶
··adress:·xxxxxxxxxxxxxxxxxx,· xxxx· x··xxxx-xxxxxxxxxxxxx¶
¶
··Code·number:······xxxxx· ··date·of·birth:·xx-xx-xxxx· ···Tel:··xxxxxxxxx¶
·bbbbbbbbbbbbbbbbbbbbbbb¶
··adress:·xxxxxxxxxxxxxxx,· xxxx· x··xxxx-· xxxxxxxxxxx¶
¶
··Code·number:······xxxxx· ··date·of·birth:·xx-xx-xxxx· ···Tel:¶
··ccccccccccccccccccccccccccccccccccc¶
··adress:·xxxxxxxxxxxxxxxxxx,· xxx···xxxx· -·xxxxxxx¶
```

Figure 8: example of original data

The following steps were taken to ultimately transport data to excel format:

1. Transport txt document to docx (word) type document;
2. Systematic substitution of attribute delimiters (:, -, single paragraph, etc.) to tab;
3. Systematic substitution of individual delimiters (2 or more paragraphs) to single paragraph;
4. Transport docx document to xlsx (excel document);
5. Manual depuration of errors in all individuals.

After this, the data was exported to SPSS program in order to perform statistical analysis.

Because these manual methods were used and also because the original data was obtained in 18 different txt files, this task endured approximately 2 months to overcome.

All data was confidential, and for that, all fields that can identify patient (name, phone number and identity numbers) where removed from database or subjected to a randomized change. Attribute subset selection (Han and Kamber, 2006) was used in a primary task to remove phone number attribute and name. Only the health unit itself can identify the patients in order to effectively use the study results, the investigator cannot.

A *Stepwise backward elimination (Han and Kamber, 2006)* method determined that these attributes where irrelevant to the main feature evaluated. The procedure involved: considering all attributes and then eliminating one by one in order to determine which one had little or none interference with the outcome.

Only the relevant attributes where extracted from these units: original health center, birthdates, original physician code, address city, address street, address door number and zip code.

The data included 3 existing units as seen in Table 3.

Table 3: number of patients existing in original units

	Health Units		
	Unit X	Unit Y	Unit Z
Existing patients (n)	5439	7252	2021

Mixing all 3 databases resulted in a total of 14712 patients as seen in Table 3. From these, only the patients that lived in Porto city where selected n = 12848; after that a new selection was made excluding all patients that lived outside the Health Unit influence area. This area comprehended 4 parishes within Porto city, as seen in Figure 7. In this area we get a total of 7285 patients.

Data Cube aggregation technique (exemplified in Figure 9) was use in order to merge the 3 unit data that are in separate databases.

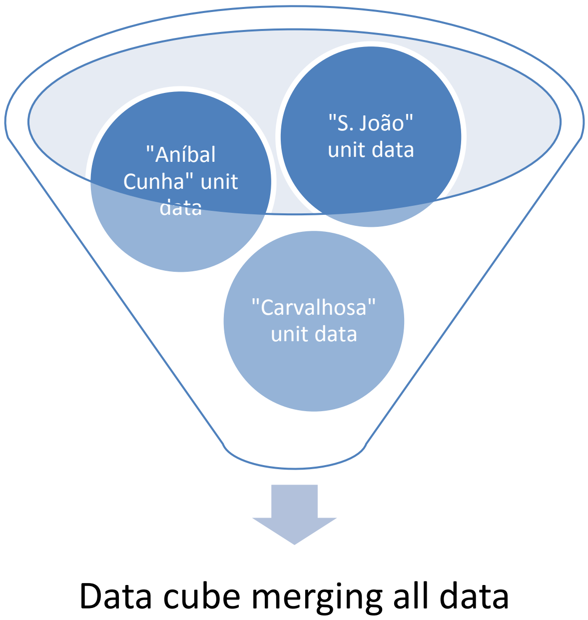

Figure 9: visualization of 3 units' data merging in one

This system is simple because all 3 data sources had the same variables in the lists and followed the same structure.

4.2.2 Data Processing

In order to detect inconsistence and errors in data 2 steps where made:

1. Using SPSS program, the records with missing values, and syntax errors (such as incorrect format of birth date or zip codes with incorrect values) were identified and in cases where attributes had to be filled, a manual validation in original information system was made. Inconsistences in non-relevant attributes were ignored.
2. Because the geographic placing of patients was one of the core aspects of this work, street names that presented errors were processed. This was accomplished by a cross validation between street names in obtained database and street codification in Portugal Postal Services – CTT. A free download of the all existing addresses in Portugal is available at CTT webpage and contains many fields. The ones important for this study were street names and cities. In order to fulfill the validation, only the records that represented the area of influence of the unit were cross validated. This was made manually, verifying if the street existing in database was actually a street existing in updated postal database.

4.2.3 Statistical analysis

For better understanding off the data and assess future results validity (Armitage et al., 2008), statistical methods were applied to database using SPSS program.

These methods included:

- Frequency tables;
- Central tendency measures (means, medians);
- Graphic analysis (histograms and bar charts);
- Hypothesis tests:
  - Parametric tests:
    - T-test;
    - Anova test;
  - Non parametric tests:
    - Kolmogorov Smirnov test;
    - Kruskal-Wallis Test.

4.2.4 Lists Formation

Driven by an existing law (MS, 2007), the patient number to distribute between physicians and nurses, has in consideration not only the number of people itself but also their calculated weight in workflow. So, these patients were divided by age intervals. A selection is made by age criteria. So, a child having 2 years counts as a 1.5 weighted patient. An elderly patient having 79 years old count as 2.5. This is explained because of health care search is higher in this ages. For now, diseases criteria are not being used.

For example: a 1350 people list may actually count as a 1570 list. This way, a list with multiple elder people having a great impact on a professionals' work is considerate as so. Criteria to build an "n" list that represented an expectable and desirable one are stated in Table 4.

The main methods that were used focused on dividing the population having in consideration their distance to the central location (the health unit), the distance between people themselves, the discrepancy between different lists, the minimum weighing number of people to exist within a list (1917 to 2412) and the real number of existing people acceptable by professionals (1550 to 1800).

Table 4: Existing patients in used database and their weighting.

Age interval	Average weight	n	Weighting formula	Weighted n
[0 to 6[	1,5	294	n * 1,5	441
[6 to 65[	1	5664	n	5664
[65 to 75[	2	608	n * 2	1216
[>= 75]	2,5	719	n * 2,5	1707,5
Total		7285		9028,5

Two divisions were made:

1. The first was manually and consisted in observing the map of the area required and to allocate street by street, patients residing in that street to a determinate group;
2. After the realization of the software, the GPS coordinates were retrieved from Google, both latitude and longitude. This was possible using the following junction of attributes: *"street name + door number + city"*. After doing this, these variables were added to database and mathematical distance between patient location and unit location was calculated.

Two distances were calculated: linear one and nonlinear. The linear represented the distance in a straight path between 2 points using the earth curvature, meaning the length of the ellipsoid line between 2 points (Figure 10). This was executed using the cosine rule from Spherical trigonometry (KryssTal, 2009) used in excel program. The formula is the following:

6371*ACOS(COS(PI()*(90-*unit latitude*)/180)*COS((90-*patient latitude*)*PI()/180)+SEN((90-*unit latitude*)*PI()/180)*SEN((90- *patient latitude*)*PI()/180)*COS((*unit longitude-patient longitude*)*PI()/180)) formula from (Romario, 2009), in which 6371 is the Earth ray in Km.

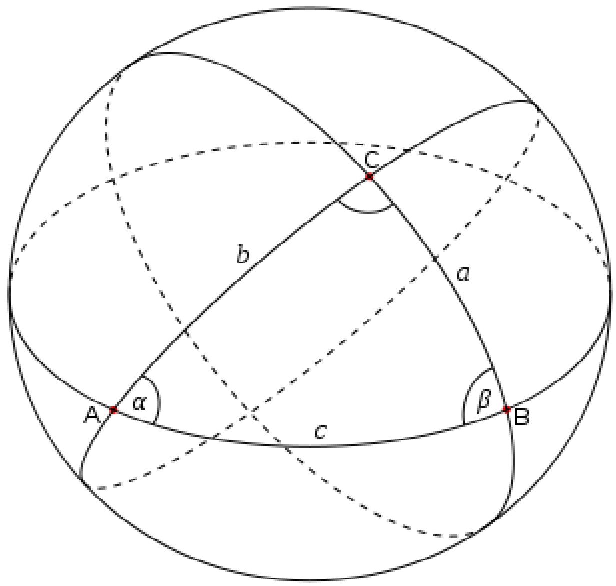

Figure 10: earth model and spherical trigonometry

The nonlinear distance was calculated using GoogleAPI throw intuitive path walk. The system calculates the shorter distance between point A and point B. However, it does not do this in a straight line but rather by the use of streets courses. For this purposes, walking path distances were used. Using the above formula: "*street name + door number + city*", the coordinate for the Health Unit was calculated and then a query requesting distance between point A (patient address) and point B (Health Unit address), Google was able to return a distance.

4.2.5 Physician-Nurse subgroup

To try helping the formation of groups of Physicians and Nurses, an attempt was made to identify which ones had more affinities. Although not requested, this task was considered relevant for the investigator.

A questionnaire was distributed to Health Professional (available in annex 8) in order to try to identify both personal and professional factors that could be a factor of proximity between the binomial groups to be formed. A total of 10 professionals constituted the universe. From these: 5 Physicians and 5 Nurses.

To all, a physical questionnaire was given and after a week, retrieved.

After data retrieval, the Hungarian algorithm (Kuhn, 2005) was applied to assess which pairs had more affinity. The algorithms consist in complex mathematical procedures involving a matrix display off data and the standardization of degrees assigned within.

4.2.6 Spatial distribution criteria

The division of patients was made by various methods (using trial and error forms):

- Manual Division;
- Matrix Grid division;
- Circular cluster division;
- Concentric circles;
- Triangular expansive out method.

Because these methods are intertwined with their results, they're better explained in the results chapter.

4.3 Information System Development

In order to correctly fulfill our purpose, new software was created. This was possible to the junction work of an engineering student from ISEP that joined the Medicine Faculty Department. His work was elaborating an internship report involving Health Information Systems. Because the previous work had already been done and rule assumptions were already pre-defined, a partnership was established between a Health Professional – myself- and computing sciences experts.

These experts were currently working in CIDES department. This department has as main activities:

- Teaching – to medical pre and post graduates;
- Investigation – in many areas linked to Health;
- Service providing in some areas such as Biostatistics, Evidence and Decision in Health and evaluation of technology and services in Health to investigators and Institutions.

Many work from this department resulted in Master and Doctoral dissertations and scientific articles published in national and international magazines or conferences.

The programming language used was MySQL, and 3 main tables where created: "Patients", "Address" and "Health Professional".

In the implementation of the prototype the following programming languages were used: PHP, HTML and JavaScript; the Eclipse developing tools for PHP and Toad for Oracle were also used.

The software development followed the iterative methodology where continuous versions where being made in a cyclic way. After each iteration, new features can be introduced and perfected. This made possible to deal with the final users expectations and to obtain feedback to make incremental changes. All this boosted the understanding of the business process.

If problems are identified earlier, risks are avoided. On the other hand, the architecture becomes more robust, because it is evaluated and improved at an early stage. A schema of the process is found in Figure 11.

Technologies used to elaborate this software were:

- Web technologies: PHP / HTML / CSS / JS / AJAX;
- Database Management System: MySQL;
- Web server: Apache;
- API Google Maps, API JQuery;
- Templates PHP.

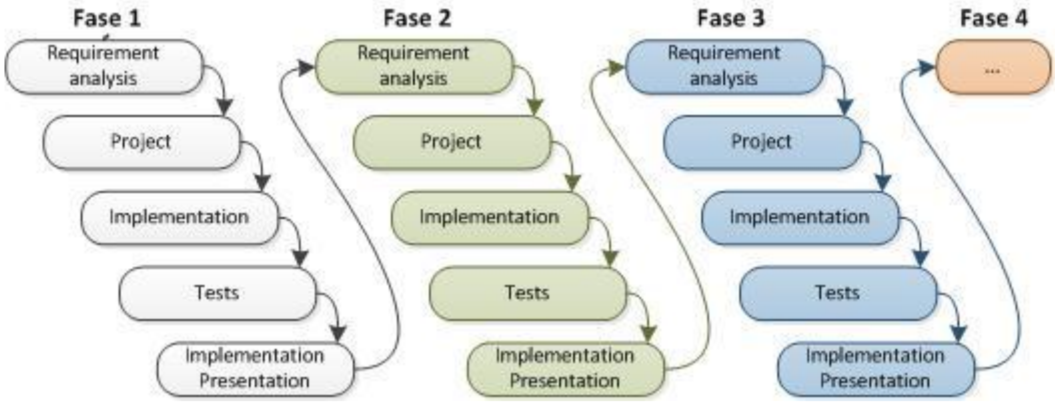

Figure 11: Iterative development model

The software had a series of functional requisites that were important. They can be seen in Table 5.

Table 5: Software functional requisites

number	Requisites
1	Patient visualization in Map
2	Patient visualization in Graphics
3	Patient visualization in Tables
4	Patient search
5	Search all patients from health unit
6	Search all patient from Health professional
7	Patient edition

4.3.1 Periodical meetings

In order to fulfill the requirements and having a full glance of what the final users wanted, meeting took place with a weekly periodicity. In this meetings, representatives of the final stakeholders and the system developers discussed iteratively the system layout, its goals, data to be fed, data to be extracted, etc.

These meeting started in March 2011 and ended in August with a total of 21. This intended to feed the iterative model described above and guarantee that final product was the desirable one.

4.4 Questionnaire

In the last part of this project a questionnaire was implemented (showed in annex 9) to Physicians and Nurses working in Family Health Units. This questionnaire had the purpose of evaluate if other Health indicators where considered relevant by end users to be posteriorly added to the software. Opinion surveys such as this are proven useful in many environments (Armstrong, 2001) to help cop and adapt people and workflows to changes.

These indicators where none other than the legal ones that Family Units are obliged to accept (ACSS, 2009) and can be seen in Table 6.

A simple questionnaire presenting this list of indicators was elaborated and asked people to access their relevance in being visualized in a map. It was diffused in a web group in which only Family Health Unit Professionals' had access. It also was distributed in some physic Health Center in which the researcher had more access. It used a simple Likert scale (Burns and Burns, 2008) from 0 to 4 in which the correspondence was:

0. Strongly disagree
1. Disagree
2. Neither agree nor disagree
3. Agree
4. Strongly agree

The sample choose to perform this questionnaire involved health Professionals (Physicians and Nurses) that were currently working in a Family Health Unit - N =4338. Only these were selected because they are the only clinical professionals that provide care in Family Health Units.

Table 6: List of legal health indicators currently in vigor to Family Health Units

	Indicator name
1	Percentage of consultations to the patient made by their own family doctor
2	Overall utilization rate of consultations
3	Rate of medical home visits (per 1000 subscribers)
4	Rate of nursing home visits (per 1000 subscribers)
5	Percentage of women aged 25 to 64 years with updated cytology (one in three years)
6	Percentage of diabetics with at least two HbA1C recorded in the last 12 months, covering two semesters
7	Percentage of hypertensive patients with blood pressure recording in the last 6 months
8	Percentage of children with updated vaccination at 2 years
9	Percentage of first appointments made to newborns until the 28th day
10	Percentage of first consultation for pregnancy in the first quarter
11	Percentage of women aged between 50 and 69 years with mammography performed in the last 2 years
12	Percentage of hypertensive patients with blood pressure in registration of each semester
13	Percentage of children with updated vaccination to 6 years
14	Average cost per user of prescription drugs
15	Utilization rate of nursing consultations in family planning
16	Percentage of pregnant women with six or more consultations by nurses in maternal health
17	Percentage of home care visits in the postpartum to watched pregnants
18	Percentage of pregnant women with postpartum review
19	Percentage of early diagnosis (THSPKU) made by the 7th day of life of the newborn
20	Percentage of home care visits for newborns up to 15 days old
21	Percentage of children with at least six consultations in child health surveillance from 0 to11 months
22	Percentage of children with at least three health consultations for children in 2nd year of life
23	Percentage of members with recorded height and weight in the last 12 months (2 years)
24	Percentage of children with updated vaccination at 2 years
25	Percentage of diabetics aged 18 to 75 years in nursing consultation
26	Percentage of diabetics with foot examination recorded in the year
27	Percentage of hypertensive patients with at least one record of Body Mass Index(BMI) in the last 12 months
28	Percentage of hypertensive patients with tetanus shot updated

The questionnaire was launched to a web group and distributed in physical Health Units.

The first method was an attempt of reaching to the highest number of Professionals, and due to the time constrains, seemed a fast way getting data. The second one was made because of the proximity and easy access for the investigator and can introduce some biases.

For validation purposes, and considering a 95% confidence interval with maximum error of 5 % a number of 385 filled questionnaires is desirable.

A personal survey was also made within the unit used to implement the division process in order to access which information they considered relevant to be presented geographically besides the one used.

The results were treated using statistical methods with the help of SPSS program: Frequency tables.

5. Results

5.1 Original data

In an initial phase, automatic retrieval to txt file documents of data was not possible. Internal database errors (inexistent or bad typed birth dates or addresses) prevented this migration as reported by the investigator and by the Health Professionals. The number of errors can be seen in Table 7. This fact had to be assessed verifying the fields with error one by one. These fields were target of correction, contacting each patient. This was made by the original Health Unit.

Table 7: Number of fields with integral errors

Unit	n	%
Unit X	7	0,05
Unit Y	58	0,39
Total	65	0,44

The final selected amount of patients was immediately divided accordingly to the city patients live. Because one of the requisites to list formation was to live in Porto city, the other patients were not considered to the following steps. The distribution of Patients between cities can be observed in Table 8.

Table 8: Distribution of initial patients among cities

		Frequency	Percent
	Others	1852	12,6
	Porto	12848	87,3
	Total	14700	99,9
	Missing	13	0,1
Total		**14713**	**100,0**

After this initial selection, only the patients that lived in Porto city where selected n = 12848; after that a new selection was made excluding all patients that lived outside the Health Unit influence area. This area comprehended 4 parishes within Porto city, as seen in Figure 7. In this area we get a total of 7285 patients.

5.2 Selected data

Data was processed and analyzed in two stages. The first prior to any distribution made and the second after the distribution was made.

5.2.1 Before distribution

Many errors were encountered within the database as shown in Table 9. The most important where the ones involving addresses (inexistent or incorrect ones). Address errors and misinformation where the most costly processes. Almost 25% of the addresses (street name) contained serious inconsistencies or major errors. These had to be addressed one by one in order to correctly input the coordinate in the map. This validation took months and proved to be exhausting. A cross validation with postal services database was executed in a primary instance. After the developing of the GIS system, we verified that incorrect street names had suggestions made by Google itself.

Table 9: Number and percentage of errors found in database.

Error description	n	%
Missing door number	511	7,01
Missing City	13	0,18
Missing street	25	0,34
Address incorrect (street name inexistent)	2137	29,33

We can see that errors in inexistent streets represent a quarter of all selected fields. Only after manual cross validation with postal database, these errors we corrected and were ready to be joined to the final database. All fields were successfully addressed.

The number of patients that came from the 3 different units can be seen in Table 10.

Table 10: Distribution of the number of patients from the original units

Original Unit	n	%
X	2792	38,3
Y	3741	51,4
Z	752	10,3
Total	7285	100

It is visible that Unit Z had less patient participation for this study, which is explainable because in that particular unit, many patients had already a Family Health Team allocated.

Units X and Y were had many problems with patient coverage because of human resource faults. In these units, different list already existed although not structured and definitive ones. To provide assistance to these patients, a rotating schedule of Health Professionals was in charge of minimum assistance. The distribution of patient throw initial lists is seen in Table 11. These different lists had, many times, overlapping teams allocated to them. It is evident that the number among them was not uniform in order to provide equal quality care. These differences in number are often the cause of Health Professionals frustration due to the fact that many are responsible for a much higher number of patients than others.

Table 11: number and percentage of patients in initial lists

Initial Lists	n	%
List 1	7	0,1
List 2	646	8,9
List 3	752	10,3
List 4	115	1,6
List 5	747	10,3
List 6	2342	32,1
List 7	674	9,3
List 8	2002	27,5
Total	7285	100

For a better understanding of the distribution among the initial lists and units, we can see Table 12. The contribution of the different units is varied and is easily perceived that list were unit oriented, meaning that they existed only in a single unit.

Table 12: Original Health Unit and Initial lists cross-tabulation

		Initial lists								
		List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	**Total**
Original Health unit	X	7	0	0	36	747	0	0	2002	**2792**
	Y	0	646	0	79	0	2342	674	0	**3741**
	Z	0	0	752	0	0	0	0	0	**752**
	Total	**7**	**646**	**752**	**115**	**747**	**2342**	**674**	**2002**	**7285**

Zip codes were one of the fields that were tried to use for distribution. In Porto city the following Zip codes exist: 4000; 4050; 4100; 4150; 4200; 4250; 4300 and 4350. The area of influence of the future Unit revolved over the 2 first ones. However, in the analysis made by cross validation, it was possible to detect that almost 50 % of the addressed had incorrect ZIP codes. In many times, even patients do not know their own ZIP code. The distribution can be seen in Table 13.

Table 13: Zip codes frequency in database

Zip Code	n	%
4000	1372	18,8
4050	4602	63,2
4100	407	5,6
4150	795	10,9
4200	106	1,5
4300	3	0

The age of patients among different list is an important variable. Its distribution was assessed and can be seen in Figure 12. The frequencies distribution did not followed a normal one, that was validated with Kolmogorov-Smirnov test - $p < 0{,}01$. For that reason, median values were considered for further analysis.

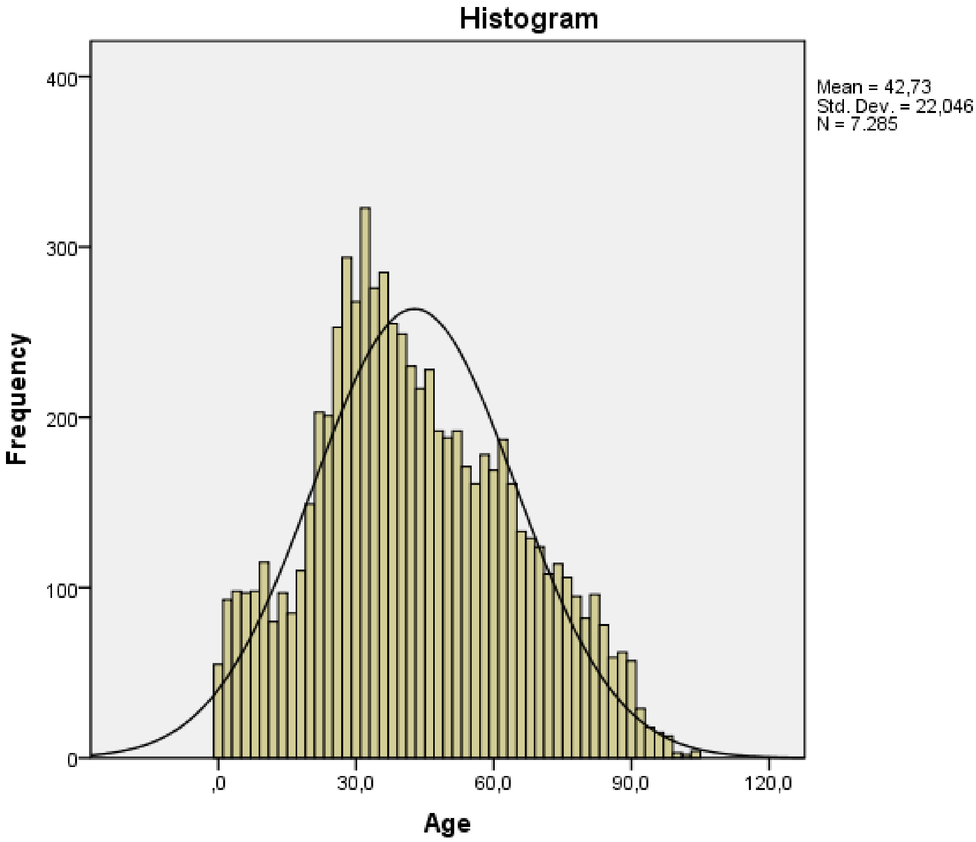

Figure 12: Graphic representation of variable age distribution with normal curve.

The distribution of patients' age among initial lists can be seen in Table 14.

Clear differences are observed between medians. List 1 for example, has median of 73 and on the other hand list 6 has 37. All other values are showed only to have a glance on the distribution. Minimum values of 0 correspond to babies that still don't have a full year completed.

As a curiosity, it's also visible the maximum age between lists, in this case a 104 years old patient present in list 3.

Table 14: Distribution of variable age between initial lists

		List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
Mean		69,429	48,755	47,715	49,417	45,221	39,371	43,184	41,288
95% Confidence Interval for Mean	Lower Bound	57,219	46,92	45,985	44,948	43,665	38,516	41,783	40,329
	Upper Bound	81,638	50,591	49,446	53,887	46,777	40,226	44,585	42,247
Median		**73**	**49**	**48**	**51**	**42**	**37**	**39**	**40**
Std. Deviation		13,2017	23,7568	24,1752	24,1958	21,6593	21,1092	18,5265	21,8761
Minimum		50	0	0	0	0	0	0	0
Maximum		83	103	104	99	101	103	100	98

This distribution can also be seen in Figure 13.

Executing the Kruskal-Wallis hypothesis test in order to assess if medians are not significantly different between the Initial lists we obtain a $p < 0,001$ meaning that they have significant differences.

To better understand the age distribution, patients were grouped according to existing rules (explained in chapter 4). The distribution of these groups among initial lists can be seen in Table 15. Many variations can be observed.

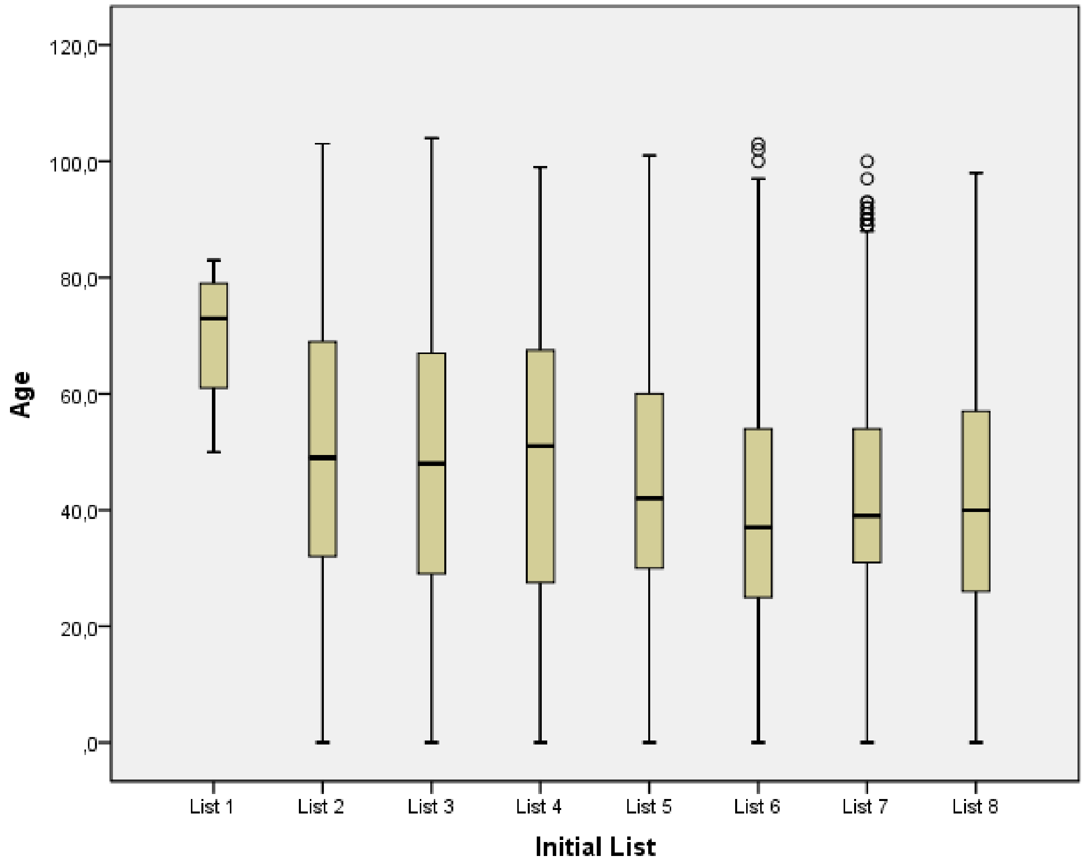

Figure 13: Distribution of age throw out Initial lists

Table 15: Distribution of grouped patients between initial lists.

		Patient ages groups				
		[0 to 6[	[6 to 65[	[65 to 75[	[>= 75]	Total
Initial lists	List 1	0	2	2	3	7
	List 2	19	437	81	109	646
	List 3	35	497	96	124	752
	List 4	2	76	17	20	115
	List 5	10	596	59	82	747
	List 6	127	1898	161	156	2342
	List 7	10	572	38	54	674
	List 8	91	1586	154	171	2002
Total		294	5664	608	719	7285

Distances were also object of analysis. Table 16 shows the main aspects of linear and nonlinear distances of patients' addresses to final Health Unit. Linear distances are expected to be inferior to nonlinear because of street urban geography and mathematic in general.

Table 16: Descriptive statistics of distances

	n=7285							
variable	**units**	**mean**	**std error**	**Median**	**std deviation**	**minimum**	**maximum**	**K-S p value***
linear distance	meters	2455,658	1272,962	781,68	108650,1	8,82	9263542,9	0,00
nonlinear distance	meters	1469,920	73,538	1000	6276,594	0	321000	0,00

*Kolmogorov Smirnov p value for a 0,05 significance level

Median values are also to be highlighted because they do not follow a normal distribution.

5.2.2 After distribution

The final display of how patients were divided among the 5 new structured lists is seen in Table 17. List c was the only one that was not filled to the maximum at Health professionals' request. The other lists present now a similar distribution.

Table 17: Number of patients in final lists.

Final lists	n	%
a	1747	24
b	1567	21,5
c	929	12,8
d	1443	19,8
e	1599	21,9
Total	7285	100

The contribution of initial to final lists is better seen in Table 18. Because no specification was earlier made regarding how many patients from each initial list had to be in final ones, no problem was raised. Also to note that initial lists had no geographic divisions and final one had. These tables (18 and 19) were a request of Health professionals and for that, were also included in this study.

Table 18: initial and final lists Cross tabulation

		final lists					
		a	b	c	d	e	**Total**
initial lists	List 1	1	0	2	1	3	**7**
	List 2	57	195	107	230	57	**646**
	List 3	195	74	86	96	301	**752**
	List 4	28	18	6	33	30	**115**
	List 5	206	192	83	94	172	**747**
	List 6	529	505	330	521	457	**2342**
	List 7	143	175	114	121	121	**674**
	List 8	588	408	201	347	458	**2002**
	Total	**1747**	**1567**	**929**	**1443**	**1599**	**7285**

In Table 19, the cross-tabulation between Original Unit, initial and final lists was made.

In order to compare with before lists, age in final lists was studied (as seen in

Table 20 and Figure 14). Median values are highlighted.

Executing the Kruskal-Wallis hypothesis test in order to assess if medians are not significantly different between the final list we obtain a $p < 0,01$ meaning that they have significant differences. However, a significant difference is observed between before and after dividing. Age differences were certainly smoothed among final lists.

Table 19: Original Health Unit, Initial list and Final list cross tabulation

original health unit			**Final list** a	b	c	d	e	Total
X	**Initial list**	**List 1**	1	0	2	1	3	**7**
		List 4	11	5	2	3	15	**36**
		List 5	206	192	83	94	172	**747**
		List 8	588	408	201	347	458	**2002**
	Total		**806**	**605**	**288**	**445**	**648**	**2792**
Y	**Initial list**	**List 2**	57	195	107	230	57	**646**
		List 4	17	13	4	30	15	**79**
		List 6	529	505	330	521	457	**2342**
		List 7	143	175	114	121	121	**674**
	Total		746	888	555	902	650	**3741**
Z	**Initial list**	**List 3**	195	74	86	96	301	**752**
	Total		**195**	**74**	**86**	**96**	**301**	**752**
Total			**1747**	**1567**	**929**	**1443**	**1599**	**7285**

Table 20: Age distribution among final lists

		Final Lists a	b	c	d	e
Mean		45,293	40,874	43,209	41,151	42,902
95% Confidence Interval for Mean	upper bound	44,244	39,806	41,835	40,048	41,777
	lower bound	46,341	41,943	44,583	42,254	44,026
Median		**44**	**38**	**41**	**39**	**41**
Std. Deviation		22,334	21,5675	21,3372	21,3563	22,9247
Minimum		0	0	0	0	0
Maximum		101	103	97	103	104
Range		101	103	97	103	104
Interquartile Range		34	30	30	29	33

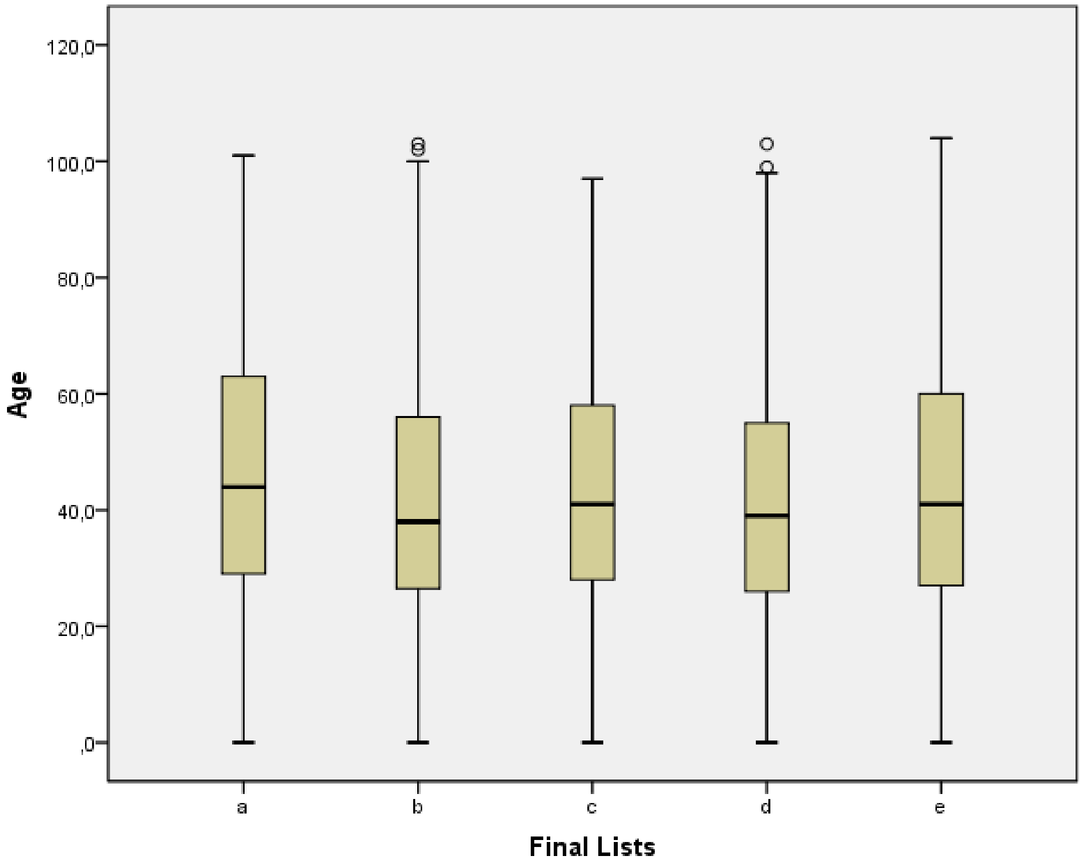

Figure 14: Age distribution throw out Final lists

Also for comparison purposes, the age grouping distribution among final lists was calculated. It can be seen in Table 21.

Table 21: Distribution of grouped patients between final lists

	Patient ages groups				
final lists	[0 to 6[	[6 to 65[	[65 to 75[	[>= 75]	Total
a	64	1296	193	194	1747
b	72	1261	103	131	1567
c	38	732	69	90	929
d	57	1153	115	118	1443
e	63	1222	128	186	1599
Total	294	5664	608	719	7285

Because distance to health unit was a variable worthy of interest, the values were calculate and presents by lists. The variable distribution also didn't present a normal distribution ($p < 0,01$ in K-S test), so median and interquartile values are presented. Values can be observed in Table 22 and Table 23.

Table 22: Distances of patient homes to Health Unit in initial lists

		nonlinear distance to unit		linear distance to unit	
initial lists	**N**	**Median**	**interquartile range**	**Median**	**interquartile range**
list1	7	700	800	1145,28	226,04
list2	644	800	700	627,255	598,09
list3	748	1000	900	780,26	715,43
list4	115	1100	700	814,97	605,22
list5	737	900	800	763,38	636,84
list6	2309	1000	800	801,87	647,32
list7	655	800	800	661,49	611,49
list8	1964	1000	900	802,26	690,92

Both distributions had significant differences in median values ($p < 0,01$ in Kruskal-Wallis test). This enforced another approach in dividing patients other than pure statistical analysis and served as base for map information assessment.

Table 23: Distances of patient homes to Health Unit in final lists

		nonlinear distance to unit		linear distance to unit	
final lists	**N**	**Median**	**interquartile range**	**Median**	**interquartile range**
a	1747	1100	800	829,83	723,65
b	1567	900	1000	784,35	881,46
c	929	700	700	496,95	524,22
d	1443	1000	600	787,35	445,34
e	1599	1100	800	830,32	614,98

Other results that were verified in dispersion tables that could represent an outlier were the existence of distances between Patient Homes and the Health Unit superior than 10 Km. This is a contradiction with the initial filter applied were selected Patients lived only in Porto city.

The total number of fields that met the condition were n = 63 (0,86% of total). These fields were then verified and the problem was in the coordinates retrieved from Google Api. These were in fact, patients that lived within the city but the Api returned a location in other cities or even countries.

5.3 Physician-Nurse subgroup

This particular part of the work consisted on retrieving all data from Health Professionals, and tries to structure it. Two subgroups had to be analyzed, Physicians and Nurses.

The initial questions had to be dichotomized in order to apply the Hungarian algorithm. So, for example, if someone responded they had special training in investigation a 1 value was given. If not, a 0 value. Using this method, a total of 30 different variables were retrieved and putted into a matrix display system.

Ex:	Physician A;	Physician B;	Physician C;	Physician D;
Nurse1;	0	1	0	1
Nurse 2;	1	1	1	0
Nurse 3;	0	0	0	1
Nurse 4;	1	1	0	1

All this 0 and ones corresponded to each question made in its dichotomized way (0= no; 1 = yes). After that, and in the same matrix structure, the sum of all existing features (the values 1).

The following steps of Hungarian algorithm were made:

- Multiply all values for -1;
- Subtract the lesser value found in each row for all rows;
- Assess results and try to identify which pairs had lesser values;
- Do the exact same for column values;

- Compare row and column results.

Four different pairs had values of 0 (the most affinity that can be present) and the fifth one had values of 4.

These results were given to the unit for reflection.

5.4 Patient Distribution Method

A variety of methods were used in order to divide the 7285 patients between 5 different groups. The principles of equity and costs reduction were the ones always present and subject to deep analysis.

5.4.1 Manual Division

The most common method used by Health Units to fulfill this endeavor is the manual one. In this scenario, Health professionals use printed maps and clipboards in order to distribute people within Health unit the influence area.

Manual lines are drawn and lists of patient are printed in order to view all data. Then, by a personal and subjective method, all patients living in a certain area are selected to professional "x" or "z" until reaching a certain amount of population (limit number). Proximity and equality issues are left to the professional decision. Because areas under the influence of a Health Unit are always of a considerable size, over-sized maps are needed. This is one of the complexities because detailed maps with street name and health units plotted are not always easy to obtain.

Also, one of the main problems here is that any manual division is always subjective and different from user to user. The distribution process is randomly and without any logic and can result in a rather not uniform distribution.

This method was used in order to primary divide people and lasted 9 months. Because maps in the size that was needed where not available, individual snapshots of portions of the city where printed and then merged using adhesive tape as seen in Figure 15.

Figure 15: Printed maps used to manually divide

In order to simulate and effectively do the patient division, various manual methods where used such as subdividing the geographic portions using blank sheets or measure tape (showed in Figure 16 and Figure 17).

There was still the question of geometrically how to divide the groups. Because this followed a trial and error method, this is better explained in the next chapters. When the question of how to divide was answered a return was made to manual division.

This process included the selection of a particular street (assuming that there was not going to be patients in the same street allocated to different lists – one of the Health group premises), and them selecting all patients living in that same street and the correspondent list.

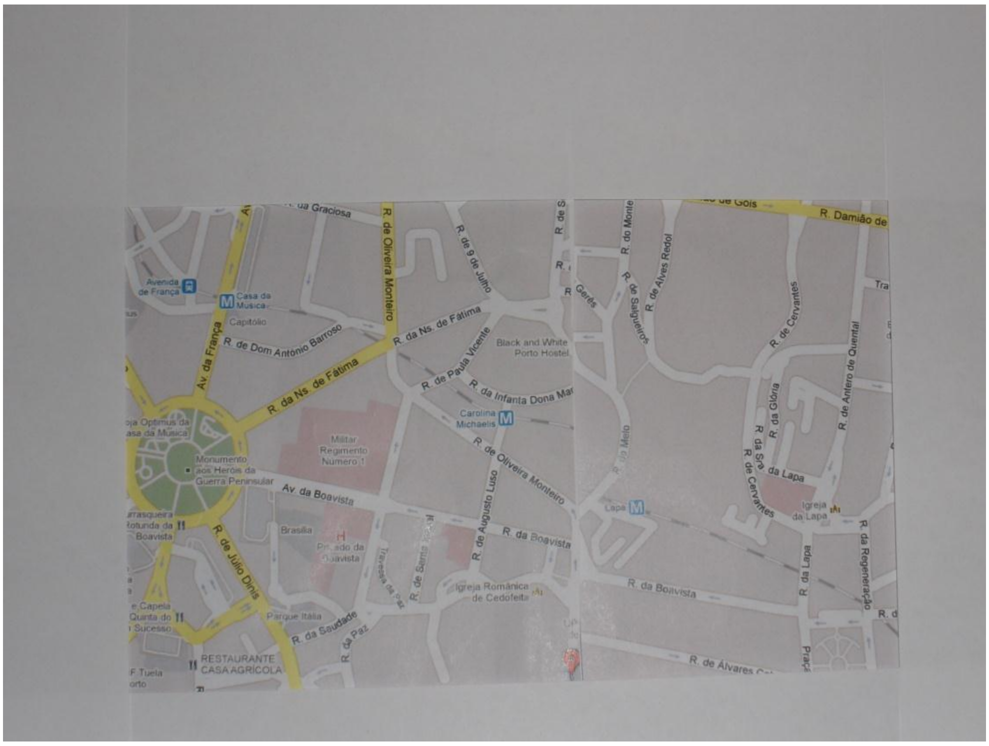

Figure 16: An example of a manual division method

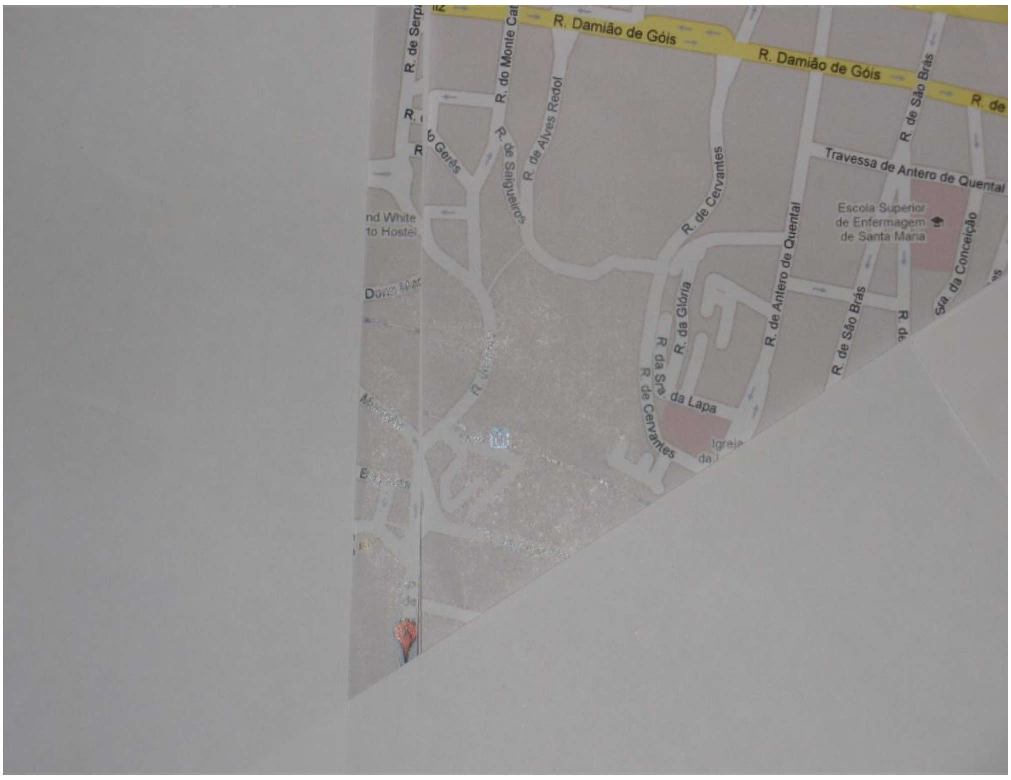

Figure 17: Another example of manual division method

5.4.2 Matrix Grid Division

A more systematized method implies a matrix division (grid display) of lists. The unit was placed in the geographic center and then a grid was applied (shown in Figure 18).

Portions of the matrix obtained where then allocated to different professionals.

The main difficulty here was the nonlinear design of city streets. Applying a matrix division in a non-matrix street design makes the distribution almost impossible and incoherent because a matrix portion may end when a street does not thus driving us to a decision in whether continue with the street and ignore the portion (destroying so the whole methodology) or break a street for two or more health professionals (this also being confusing to professionals themselves).

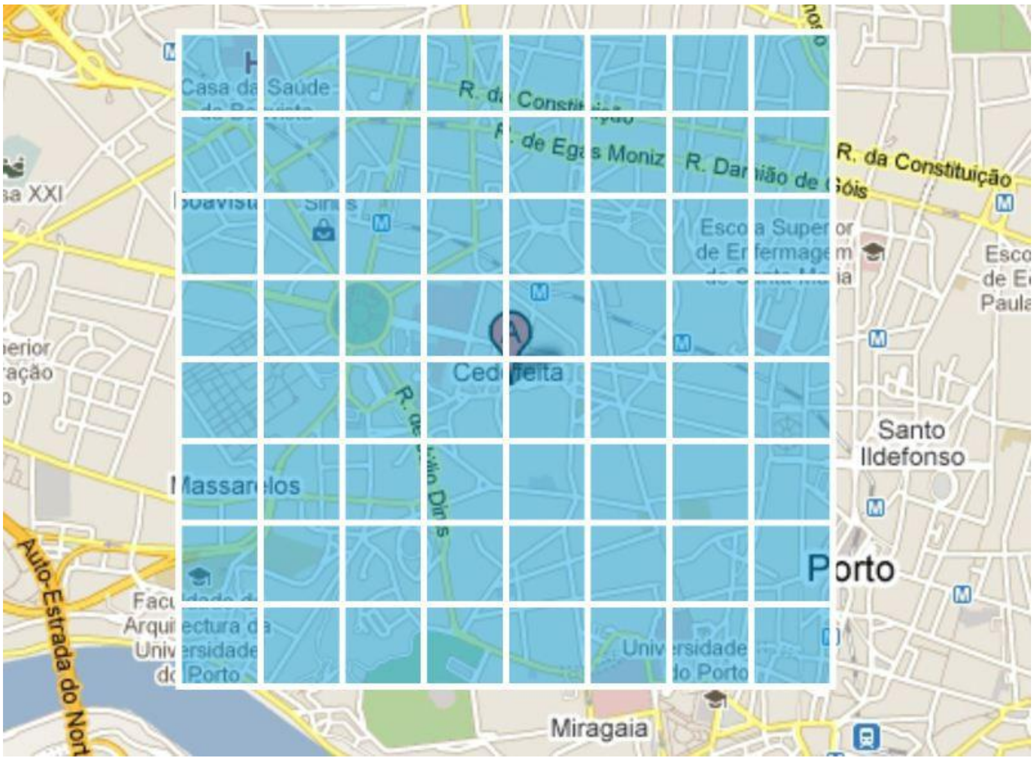

Figure 18: Matrix grid division method

5.4.3 Circular Cluster Division

With this distribution, circular clusters were drawn not having the Health unit as center but rather by intersecting themselves in a point – the Health Unit location. Then circle sizes are suited to weighing number of people within (seen in Figure 19).

The overlapping sections constitute the main problems. After suiting the street design in order to keep the whole street to a single health professional, the exact place where clusters converge drive us also to a decision in which list to put it.

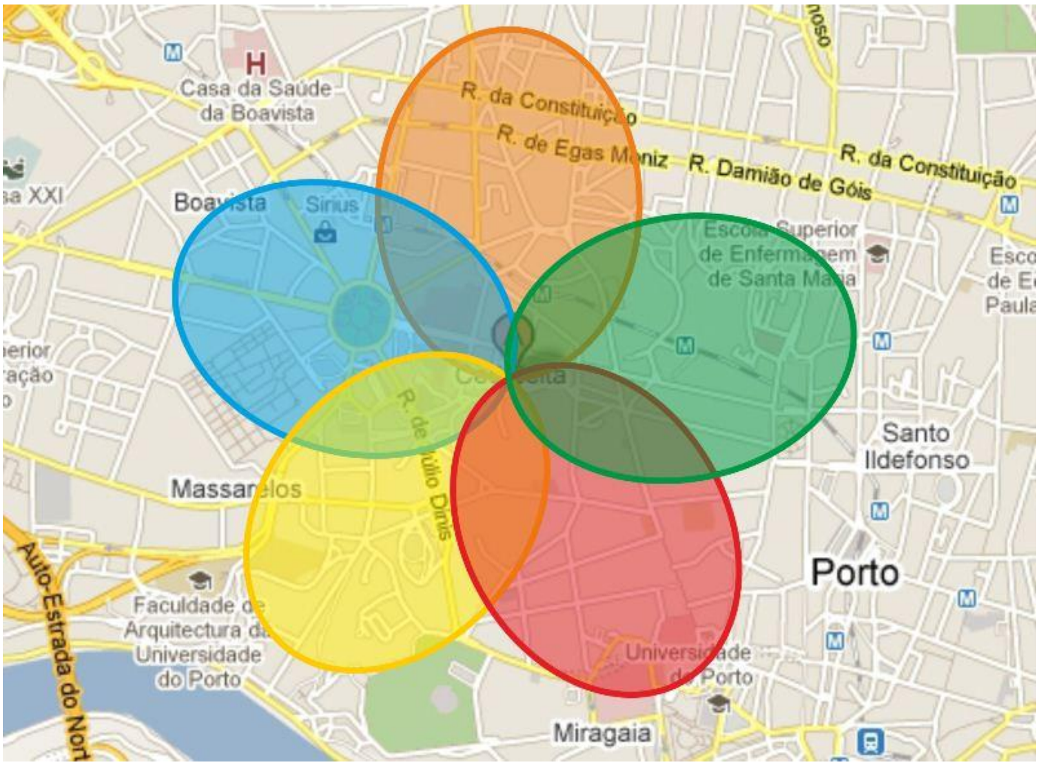

Figure 19: Circular Cluster Division method

5.4.4 Concentric Circles

In this method the unit was placed in the center of the map and concentric circles are drawn, from the inside to outside as seen in Figure 20.

Figure 20: Concentric Circles method

Right after visualizing this solution in the map, we can see its primary issues. The division it's not made having an equality of distances principle. Being that the professional that gets the closest circle cluster is favored in opposition with the one that gets the furthest. Having to perform house visits, the furthest circle will always take more time and money.

Also, the street design problem already discussed in this method is a complete nightmare because we can never allocate a full street to a single professional.

5.4.5 Triangular Expansive Out Method

Using this method, we place the Health unit in the center and then lines are drawn similar to a circle division in slices. The angles between lines may not be equal because people contained within, can also may not be (shown in Figure 21).

The street design here is also an issue but it can be relatively addressed. Only in cluster frontiers we can have such difficulty.

In this method, equality in distribution is fulfilled because all professionals have closest and furthest addresses and all cover a similar area (although not equal).

Figure 21: Triangular Expansive Out Method

This was the model chosen to both manual and later assisted distribution, but with some alterations (best explained in the next section).

5.5 GeoPrimaryHealth

The system obtained is a GIS destined to Health care area. It used HTML, CSS, Ajax, PHP and Javascript. The server has Apache and MySQL database manager. There is also use of a map server, Google Maps API 3.0.

The system sought to meet the following pre-established requirements:

- to know the distribution of users without a structured Family Team in the geographical areas of family health center;

- calculate the distribution of clients by health center for USF, a family physician and a family nurse;
- present the results on the map with markers in addition to allowing the viewing of them in graphical form or in tables, so helping distribute the patients by individual teams.

The distribution of patients by professional is represented by a marker (point on the map that represents the position of a user) with a color. This allows us to understand the geographic areas assigned to each professional.

Logical architecture can be seen in Figure 22. The web application is assessed throw a browser connected to server. To visualize the application, the browser interprets HTML and CSS in which the system was developed. The communication between client and server in reciprocal using HTTP protocols, with GET and POST methods. The server has Apache, which supports PHP programing and MySQL database manager. There’s also a map server Google Maps API allowing the use of systems integrated with PHP and MySQL.

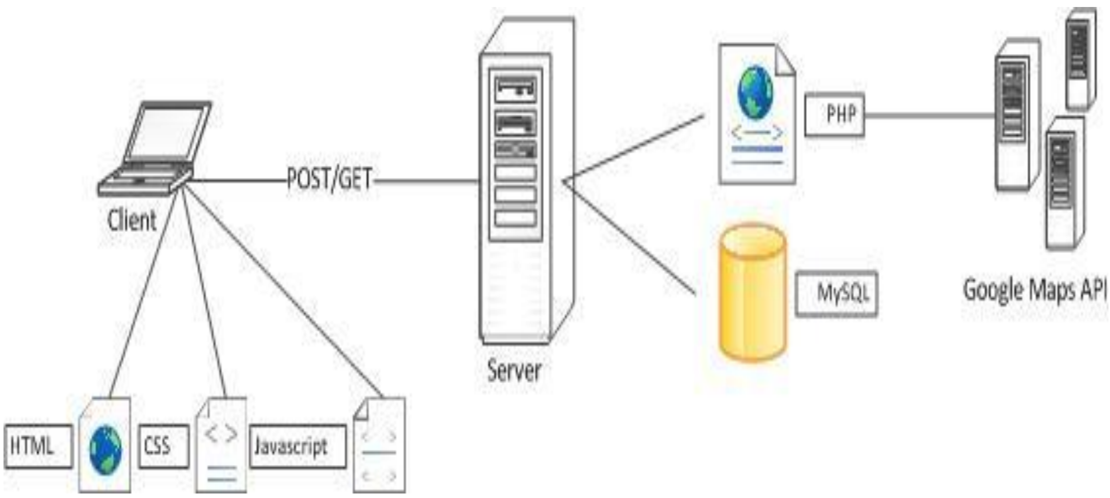

Figure 22: Logical Architecture of system

The system, fed with the filtered geographic information, was able to plot all patients in the map as seen in Figure 23.

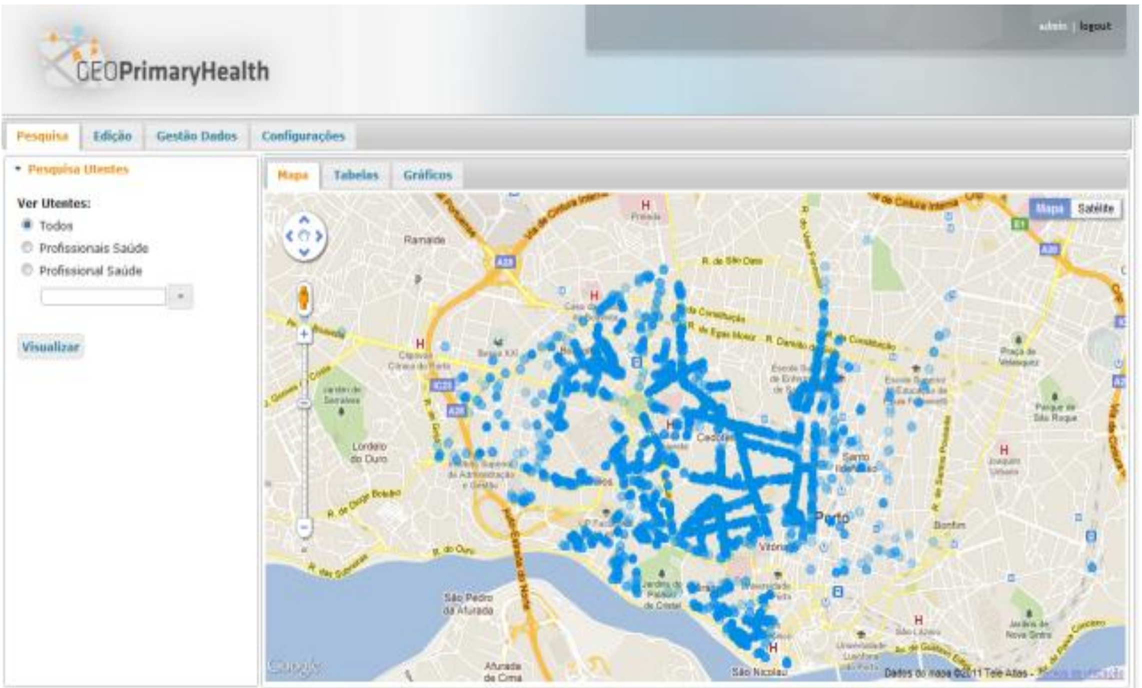

Figure 23: GeoPrimaryHealth - System obtained showing patients undistributed

This immediately represented a better comprehension of the problem that was in hand. Finally, the actual dispersion of patient addresses was visible and ready to be assessed in a more systematized way.

To enlighten higher processes and its stakeholders, a sequence diagram is showed in Figure 24. It shows how processes operate with one another and in what order (Odell et al., 2001).

A sequence diagram shows object interactions arranged in time sequence. Sequence diagrams typically are associated with use case realizations in the Logical View of the system under development. User, System, database and Google Api are displayed and also how they intervene in the patient allocation.

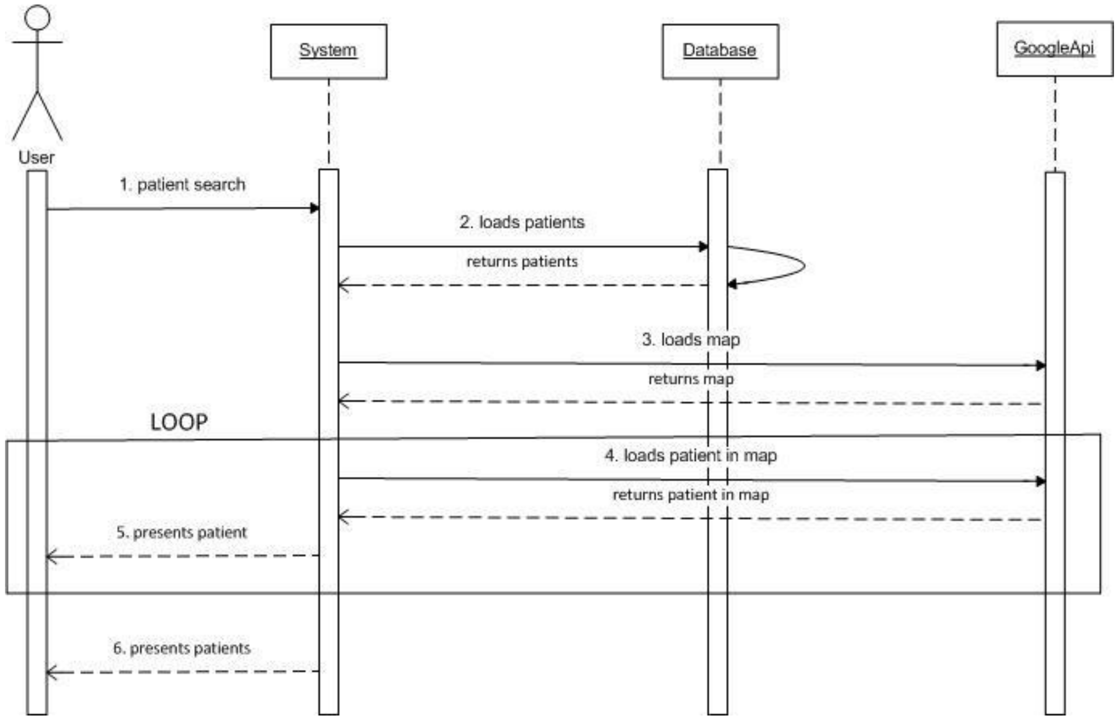

Figure 24: Sequence diagram of GeoPrimayHealth

Another representation in UML language is the Activities Diagram (Dumas and ter Hofstede, 2001). It constitutes a graphical representation of workflows of stepwise activities. It includes actions with support for choice, iteration and concurrency. They can be used to describe the business and operational step-by-step workflows of components in a system, showing its' overall flow of control

An activities diagram of this particular system can be seen in Figure 25. All the process of searching patients is visualized.

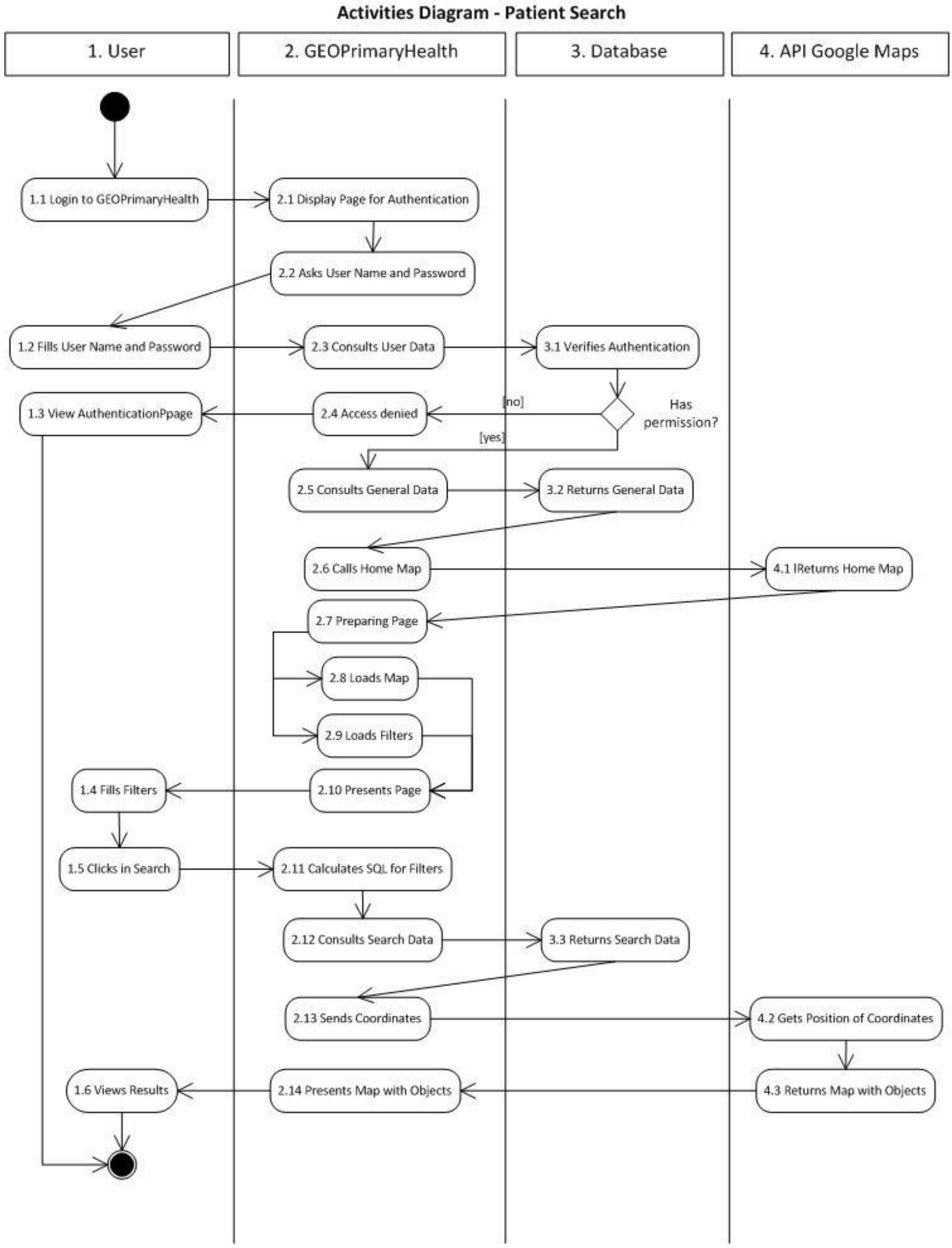

Figure 25: Activities diagram of GeoPrimaryHealth

The system can be accessed in an experimental version in: ***geoprimaryhealth.gim.med.up.pt.***

5.5.1 GeoPrimaryHealth database

The database that was used to feed the system was, as stated earlier, a subdivision of initial data. It comprehended patients that lived only in the area of influence of the unit n =7285 from 14714 representing 49,5%.

A new database was created by the computing sciences experts in order to optimize the system feeding. The database logical model is seen in Figure 26.

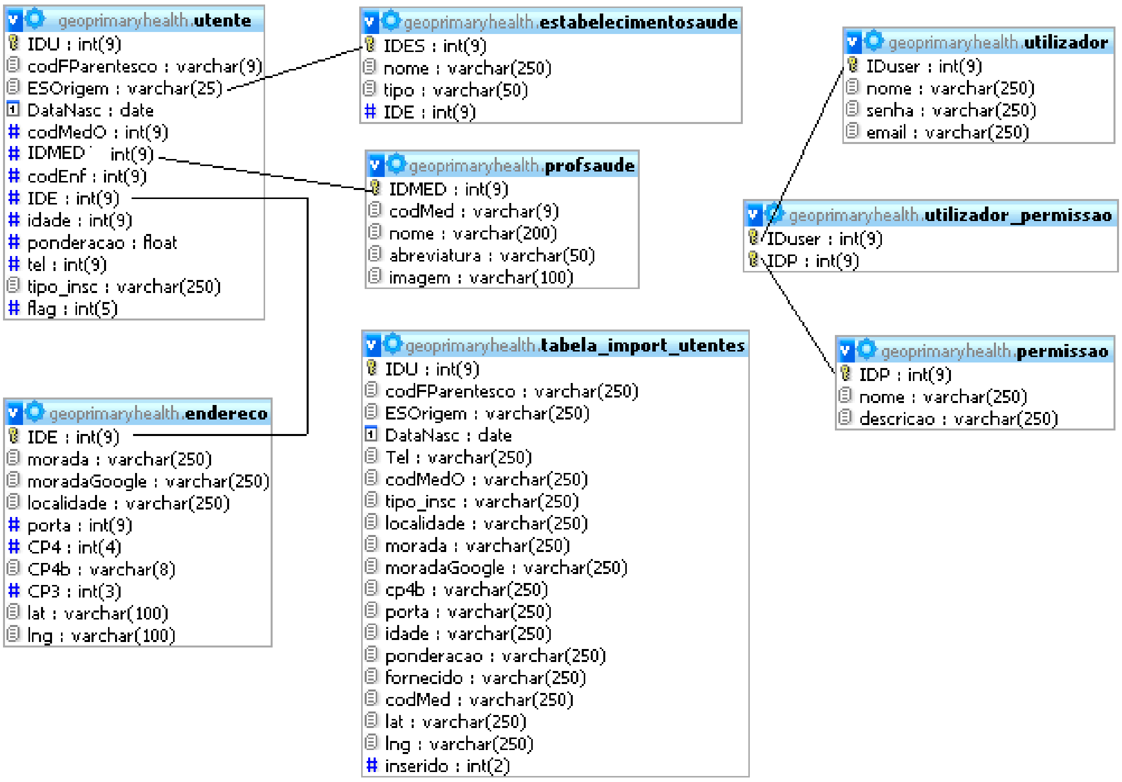

Figure 26: Final database logical model

In this representation, the different tables that final database had, are presented. The most important ones are:

- Patient;
- Health professional;
- Health Unit;

- Address.

Because many patients belong to the same address, being in the same family and also living in vertical construction and thus having the same street name and door number, the table address was created. Each pair “street” + “door number” had its own ID stores in this table.

5.5.2 Coordinates retrieval

Although a manual correction of data has already been made, there were still errors in importing the coordinates from Google Api. Total errors were n = 28 representing less than 0,39% of total data. If earlier correction hadn’t been made, we could expect percentages of errors superior to 26%.

The latitude and longitude values were retrieved and putted in new database in the format seen in Table 24.

Table 24: An example of GPS values retrieval from Google API

	GPS coordinates	
patient x	41,1536065	-8,6209005
patient y	41,153247	-8,6211238
patient z	41,1574126	-8,6208933
patient A	41,1641342	-8,6304844
patient B	41,1645227	-8,6271893

5.6 Patient distribution with GeoPrimaryHealth

Unsystematized division of the patients can be seen in Figure 27. This division is obviously going to difficult health care because of asymmetries between different lists. Also this scenario is the one existing in most heath units nowadays. Because of randomly attribution to Physicians and nurses, the map gets confused and, when house visits are necessary, costs are higher because transportation routes intersect themselves in a non-order way.

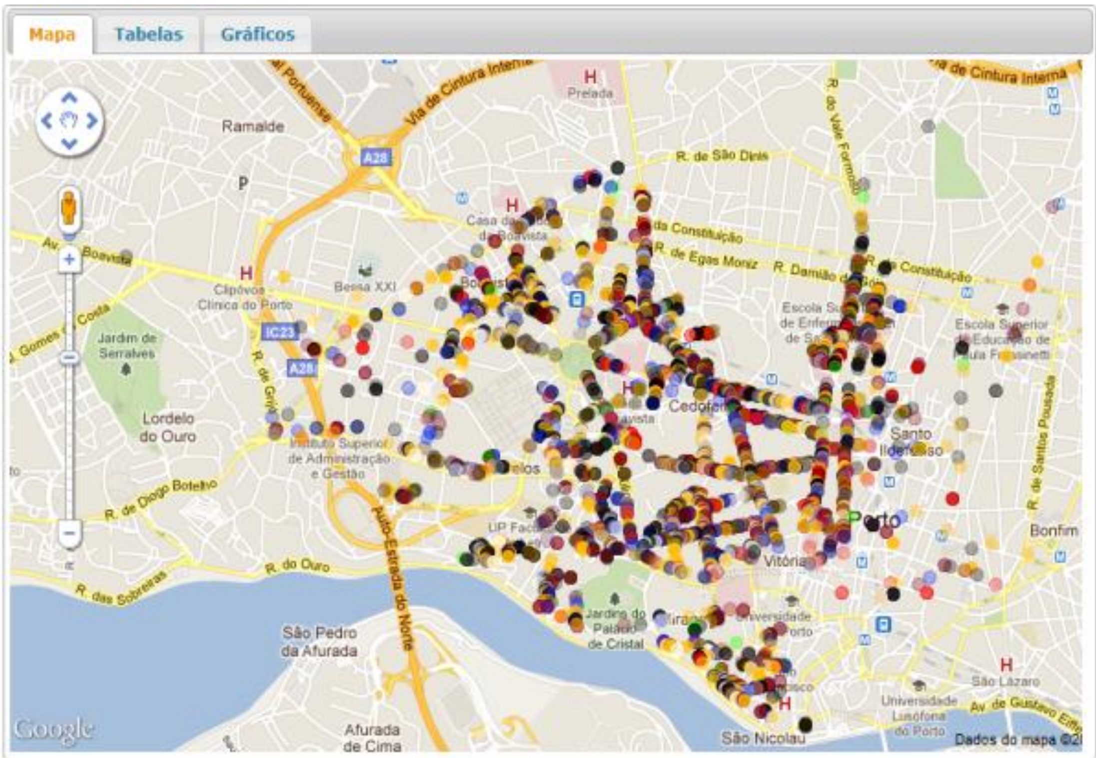

Figure 27: System showing the initial patient distribution

It is evident that an initial confusion and inability to interpret anything was one of the issues. Also important was the fact that despite the initial analysis using statistics was tried, this visualization was able to provide greater comprehension of distribution problems.

Using the already method called Triangular Expansive Out method, clusters of patients were divided. All the previous methods (with the exception of the manual one) were also tested in the system and the same results and conclusions were made. However, the exact method was not possible to execute. If we draw perfect lines on a city map, we are bound to fragment street to 2 or more lists.

So, for the above reasons, a new method was finally used based on the one described.

It was finally called: "*Semi-automatic Street based Triangular Expansive Out method*".

A better understanding of the used method can be seen in Figure 28 and Figure 29. The first one has patients plotted directly in the map, divided by different colors. Not only we can see the division but we can also see the density of patients among different city areas.

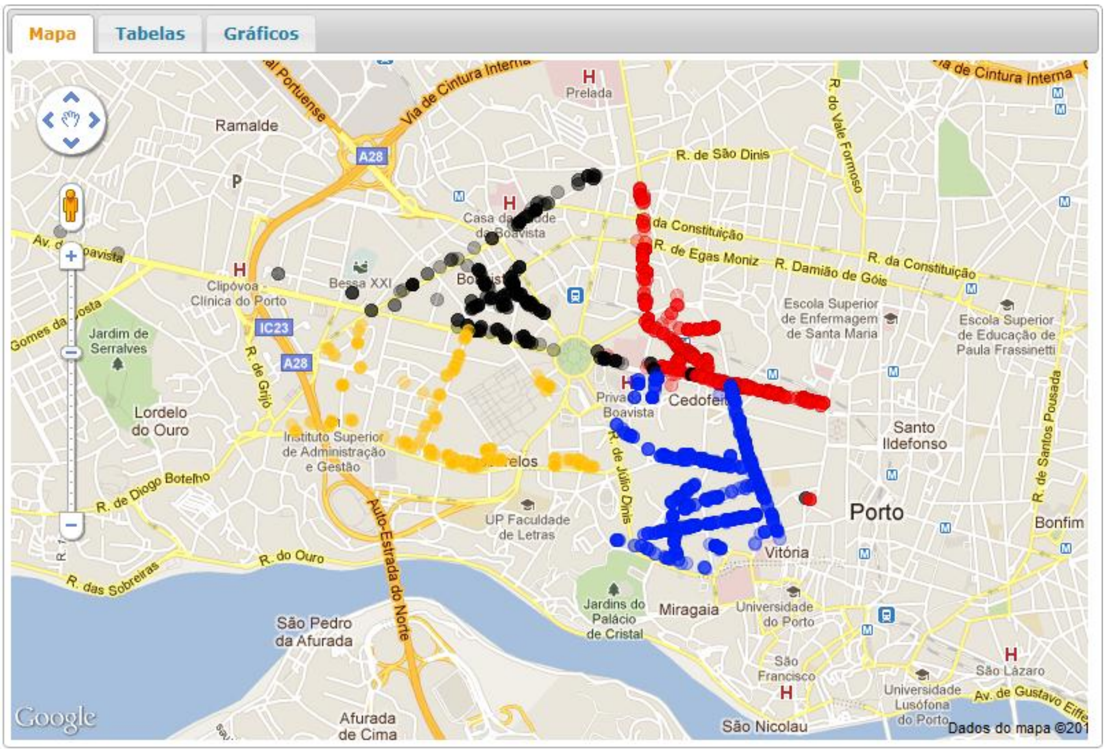

Figure 28: Representation of patient division methods used

All this combined with the zoom in, zoom out possibility of Goggle maps, represented a great advantage in patient visualization.

In Figure 29 we can see also one of the many possibilities of visualization, being the representation of the Health unit with a red marker.

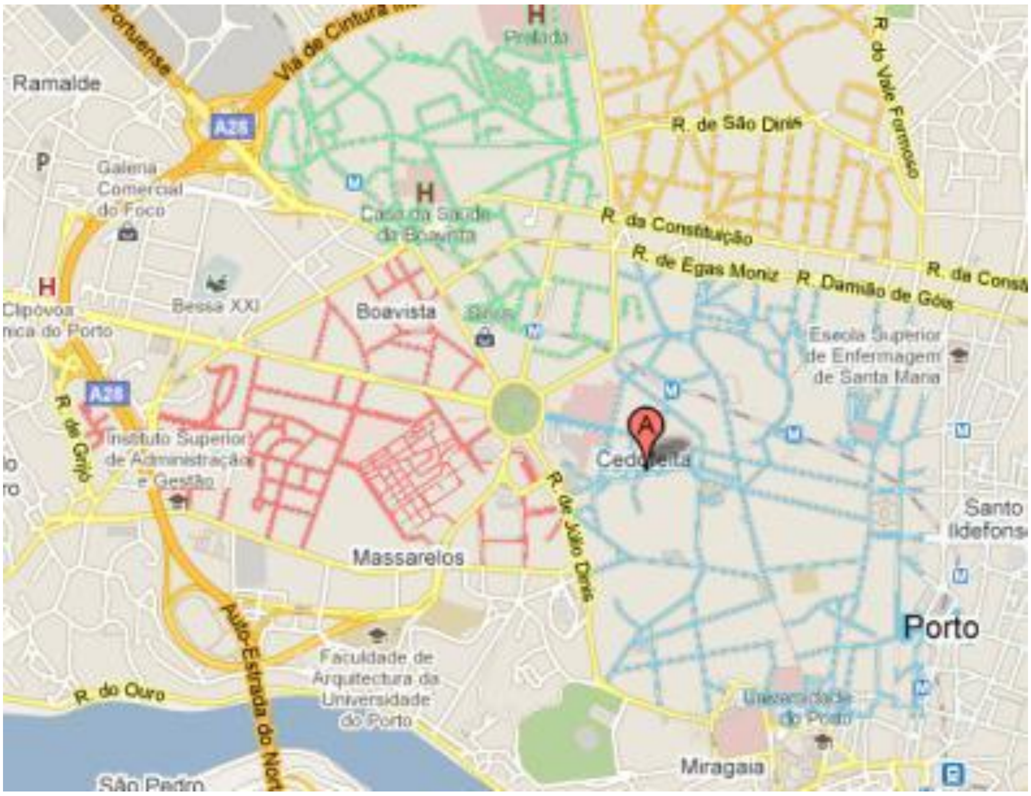

Figure 29: An example of a semi-automatic division of patients using the system

5.7 Health Professionals opinion

A total of 95 questionnaires were retrieved, thus failing the initial ones aimed for a representative sampling. In Table 25 we can see the all the units that are currently in activity and also their respectively number of patients and Health professionals.

Table 25: Health professionals existing in actual Family units

Units	Patients	Physicians	Nurses	Total Health Professionals
311	3.858.744	2.177	2.161	4.338

The scale used for scoring answers was posteriorly compacted into 3 values to assess negation, affirmation or doubt about whether the indicators are perceived to be important or not. So, all values 3 and 4 were considered as an affirmation. Results in Table 26.

Table 26: percentage of responses 3 and 4 of health Indicators

Health indicators	n	%
Percentage of children with updated vaccination at 2 years	78	83
Percentage of first appointments made to newborns until the 28th day	76	80,9
Percentage of children with updated vaccination to 6 years	74	78,7
% of hypertenses with blood pressure	72	76,6
Percentage of pregnant women with six or more consultations by nurses in maternal health	70	74,5
Percentage of children with updated vaccination at 2 years	69	73,4
Percentage of diabetics with foot examination recorded in the year	67	71,3
% of diabetics with 2 HbA1C	66	70,3
Percentage of women aged between 50 and 69 years with mammography performed in the last 2 years	66	70,2
Percentage of home care visits for newborns up to 15 days old	65	69,2
Percentage of early diagnosis (THSPKU) made by the 7th day of life of the newborn	63	67
Percentage of children with at least six consultations in child health surveillance from 0 to11 months	63	67
Percentage of children with at least three health consultations for children in 2nd year of life	63	67
Percentage of first consultation for pregnancy in the first quarter	62	65,9
Percentage of hypertensive patients with blood pressure in registration of each semester	62	65,9
rate of nursing home visits	59	65,5
% of women ages 25 to 64 with citology	60	63,9
consults by family doctor	56	59,6
overall rate of consultations	56	59,6
rate of medical home visits	53	56,4
Utilization rate of nursing consultations in family planning	53	56,3

Percentage of home care visits in the postpartum to watched pregnants	52	55,3
Percentage of hypertensive patients with at least one record of Body Mass Index(BMI) in the last 12 months	51	54,3
Percentage of hypertensive patients with tetanus shot updated	51	54,3
Percentage of pregnant women with postpartum review	48	51,1
Percentage of diabetics aged 18 to 75 years in nursing consultation	47	50
Average cost per user of prescription drugs	31	32,9
Percentage of members with recorded height and weight in the last 12 months (2 years)	17	18,1

We can see that only the 2 lower indicators didn't have a value > 50% of perceived importance to be added to the system. For all the rest, Health professionals would consider very important to visualize in a map. AS a curiosity, we can verify that the first 3 indicators are related with child welfare.

Other indicators that currently don't exist but that can be interesting for Health professionals (written in the suggestions) were the following:

- % of obese;
- % of smokers;
- % of hypertenses with BMI > 25;
- % of diabetics with BMI > 25;
- % of patients with BMI > 30.

5.7.1 Results of opinion of Family Unit

After gathering the opinion of Health Professionals from the unit chosen for the primary part of this study a few other fields where considered relevant to be showed in the GIS and can be an interesting beginning point of other studies:

- Displaying the different streets with different colors;

- To be able to show the different families existing within an area and to distinguish them by number (for example to put a yellow marker in 1 person families, a red in 2 element families, etc.
- To display the population divided by another grouping method, in this case – from 0 to 18; 19 to 64 and 65 or plus;
- Apply the above rule to weighting rules;
- Display the location of public transportation;
- To show the location of other Health Units including other family units, hospitals and private clinics;
- To be able to filter information displayed using the different Health programs used to aggregate patients;
- child health, maternal health, youth health, adult health, elderly health, group risk hypertension, diabetes risk group; etc.;
- To be able to calculate and show the best time and money saving course between the different addresses in which house visits had to be made in a particular day.

6. Discussion

6.1 Unit data

After the primary analysis without the help of any system, the difficulty of finding grounds to distribute patients started to arise. In fact, the initial theory that tried to assess distances of patients to Health Unit, revealed itself insufficient to fulfilling the task of distributing. However, it proved to be useful in identifying outliers and possible problems the system could have.

Age, used as criteria for forming lists revealed some particularities. It was attempted to form list having similar distribution of age by final lists. However, this proved to be impossible. If geographic assumptions are followed, we can never be sure that different lists have non-significant differences between age values (means or medians). Although there was an improvement in attenuate the differences in age distribution among initial to final lists, in final, statistical differences were still there.

For this situation, the Health Professionals seem to have the answer to validate why this happens. If the age intervals (Table 4) are used to put a weighing in total list formation we are sure to have age discrepancies and list different from one and other. However, this can be overcome by the simple fact that, for example, if one list has too many elderly patients, it will most certainly have less patient overall than one with many patients of an adult age.

6.2 Data errors

Health databases tend to have a serious problem in the quality of the data they possess. In this case we could verify that practically 30% of processed data has errors.

This was one of the most costly processes to attend because in order to correctly allocate patient clusters, addresses and identifying data must be accurate. The first attempt of dividing the group of patient using raw data proved itself frustrating. Because important errors occurred in original system that had the patient information, not even the data import was possible to do without data processing.

Many records had 0 has birth date or even didn't have name or address. This can only be possible because the patient first inscription is done manually in each system and has no cross-validation. The administrative staff can easily input incorrect data due to simple bad typing or to false information given away by patients in order to belong to a specific health unit.

False addresses are one of the most important difficulties in cluster dividing. The result that we obtain is polluted and full of errors. In this practical case, many cross-validations of addresses had to be made. The standardization of addresses by a single and reliable database is also required. This was attempted but with many errors. Lastly, a personal contact in dubious addresses and processes had to be made. This is very time-consuming and expensive, and delays health care providing processes. Also it raises problems within patients because they're confronted with the information they gave.

Also for the system developed to be used correctly, accuracy of data is in order. One of the recurrent problem that the system presented was the difficulty in calculating the precise place of a determinate address due to many existing ways of identifying a street name. An example of this can be: "Rua 5 de outubro", this being literally 5th October street. The postal and official database presents the street name as the one displayed above but many variations where encountered. Because Google API was being asked to plot geographically the addresses, many variations in street names resulted in rather a non-existing place or an incorrect location suggestion.

I'm certain that if we putted the raw data in the system, it would return geographic information with approximately 30 % placing error.

An interesting tool that could be designed would be a system that could accurately predict correct street names and locations fed with raw data.

After cluster dividing, a high percentage of people can easily be misplaced. One way to avoid this can be to centralize the raw data of all population for health database feeding.

Addresses, phone numbers, birthdates are a few of fields that have to be validated and centralized in order to avoid patient process duplication or misinformation. This could represent a minimum data set nationally synchronized.

6.3 Problems in the Division Process

In order to successfully allocate and divide cluster of people many problems emerged.

One of the most important one was the fact that: if we really allocate cluster geographically, what will we do with the following future migrations within or without the major primary cluster. We nowadays live in a world that urban masses are in constant shifting and moving. Urban population has a great percentage of house renting, so, it's expected that families that today live in one cluster may move to another or outside the major one. Moving outside all clusters can be easily addressed, but moving within different cluster can be difficult to address. Maintaining its family team, especially the physician is critical to patients. Also, moving from team to another is not always possible or desirable and brings all sort of ethical dilemmas.

We can divide cluster now, but after 1, 2, 5 or more years it's expected that the matrix of divisions changes in an unknown percentage.

Reality and legal environment of Portugal shows that moving people from one physician to another is almost impossible due to the lack of professionals in the country. Also, in the beginning we are geographically dividing clusters of people in the assumption that it's better for work practice and for patient health. Changing the rules in the future destroys all of the primary purposes.

Another problem that emerged was the initial process. In order to execute a perfect geographic allocation, all of the patients must not have family heath team. If not, the initial division forces people to change from one to another, or geographic cluster to be totally unshaped and ineffective to one of its purposes (to provide a more efficient health care distribution when providing health care in the community).

This perfect condition actually occurred in this experience, but it will be very difficult to encounter another in the same conditions.

Another very common problem can be the constant cycle of in and out of health professionals (rather by hiring new ones, or retiring, etc.). For example, we make a geographical cluster with all of its population, but in the same area there are other patients that have different family teams. In the beginning they're not considered to enter the cluster, but they actually live within its borders. If the family team of the second groups moves away (retiring, goes to work in another place, etc.), a group of people emerges within a already full formed cluster. What to do? Have they not the same right as the other to belong in an existing formed cluster with family team? Does the team have to enlarge its number limits and endanger healthcare quality? Have we the right to destroy a full formed and functional cluster? These are all questions rather difficult if not impossible to answer.

6.3.1 Physician-Nurse pairing

The suggestion for forming the 5 subgroups of Physician-Nurse was given to Health Unit in order for them to assess its importance and pertinence. Only one method was used to verify this, and, as stated before, the instrument to recollect the data was not sufficiently validated to ensure the absolute accuracy of final information.

Nevertheless, it served as an important subject of study and served as a complement for this dissertation.

It is easily perceived that choosing a subgroup can bring many questions. Personal ones are the most difficult to address. If someone doesn't like other and, for some reason, has many similarities with the same person, working together can be quite a challenge.

6.4 Division Method Chosen

The method that had the most benefits in this particular scenario was the triangular expansive out one. This is so because: with this distribution there's always a non-significant difference between the portions of geographic areal allocated to a list. This is a major issue because of 2 factors: the fact that in many cases Health Professionals must go to patients houses, and also because using such a methodology allows people within different list to have a similar accessibility to the Health Unit.

Due to the fact that there's no perfect geometric division in cities (and in this case in particular), the method choose to implement in the system was called: "Street based semi-automatic triangular expansive out". In this case, all people that live in the same street are considered to the same list, and the final divisor is still the Health Professional.

One question arises: can this method be used in other places? Well, if we take in consideration urban maps the answer can be yes. Cities in Portugal are similar in topography and using this methodology can easily fulfill the desired purpose needing, of course, further testing and research. However, if we focus our study in rural environments, topographic differences are significant and replication is not assured.

Future studies in other Units can help answer these questions and also assess the validity of this method.

6.5 GeoPrimaryHealth Suitability

Taking in consideration that the manual method used took months (approximately 9) and that with the help of the system developed time spent was only hours, its' suitability and advantage is significant.

Of course, this is only due to the fact that polluted data was no longer present and a correct coordinate could be extracted without a doubt from patient addresses. This means that is still much to do in order to clean databases and obtain a reliable source of information.

Other evaluations are still required to correctly assess suitability of this new tool. For this case in particular:

- There was an initial problem;
- The problem was addressed by various ways;
- A tool was designed;
- The problem was resolved.

For this particular unit, patient distribution is no longer an issue.

After the completion of the system, other plotting possibilities are being considered, such as the display of other type of information. However, this will require another and more profound study.

The type of data that is perceived to be important is easy to assess. And most likely, if a new and completely validated questionnaire was distributes, the answers would be similar. Almost all information is desired to be shown, but we cannot forget that plotting other variables in maps others than location alone, represents an exponential matrix growth of complexity.

6.6 Opinion of Health Professionals

It is always important to involve the final users of changes in the actual process of change itself. In this case, a simple opinion given can prove to be an asset to future implementations within information systems developed.

The opinion the professionals were asked to give comprehended mainly the Health indicators they are obliged to follow by existing law. They are considered to represent a basic framework to provide good quality Health care to Primary care and include many fields.

A constraint identified in this study was the lack of Health Professionals comparison variables. This was after the beginning of data processing. Immediately it could be seen that a simple addition of variables such as Professional category (distinguishing between Physician and Nurse), age (to assess important changes in opinion among youngest or elderly Professionals) or even gender. The constraints of time made a new distribution of questionnaires and a new evaluation impossible. This can be a starting point to future work.

But for the purpose in hand, the opinions identified, were satisfactory.

Almost all indicators currently existing are perceived to be important for being plotted in a map and showed to the Health team that is responsible for it. This can be related also with the importance people nowadays give to graphic visualization of information. It also serves as a confirmation that GIS are always good complements to existing tools.

7. Conclusion

This was definitely a satisfying project and all that was treated represented an immense growth both for me as an individual and as a Health professional. In fact, and despite all the tremendous work that was done that consumed all my energy in these last 2 years, when all the outcomes we retrieved and processed, this whole process made sense.

Being a Health professional myself, and working in Primary Care, the importance of execute studies in this environment are tremendous. After concluding this work, I realize that not only Primary care in Portugal has very good qualities and potentialities but also is completely permeable to changes and to new technologies adaptation.

The past attempts in geographically diving groups of patients for Primary Care Purposes in Portugal, have all been made in close environments. They did not project themselves to the science community in order to share information allowing all intervenient to learn and improve these processes.

Many manual attempts are not described to permit future users to learn with them.

The concept of Family Health Team is introduced by this study and I'm sure that this pairs of Health Professionals working together in community environment can bring significant changes to Care Providing.

The manual methods earlier used and the semi-automatic ones used later with the help of the system developed permitted a good comparison and allowed the awareness for future situations.

For final considerations, I would like to enhance the importance of a good preparation for the realization on any study that involves Health Data. Disregarding or enhancing some aspects and variables are always difficult and can cause unexpected results.

Data quality must be improved in order to make this all process faster and Primary Care databases must be a target of deep in field work. The time spent in initial data cleansing can be a discouragement to any investigator and contribute even more for the lacking of work in this area. Other major issues that have to be considered are that the results there selves can be easily distorted by insufficient data processing.

In this case, only with the alliance of using information tools and data analysis, could we successfully reach the goal of patient distribution.

After the completion of the information system, it can be concluded that new technologies in general and the system developed in particular, can help patient allocation and that they represent a breakthrough in time-saving.

As a matter of fact, doing this automatically after the system is fully developed, took 1 hour in opposition to months doing it manually.

Many studies are still lacking, investigating ways of dividing patients into similar clusters, and this one intends to be a starting point on this matter.

7.1 Implications to the future

The field of health geography could benefit from studies that provide greater understanding of patterns of geography, healthcare access, and health outcomes.

The literature cited collectively indicates the interrelated aspects of geography, accessibility, and health. What is not yet understood is the specific relationship of specific populations in their unique geographical contexts.

More research is needed to explore specific social and geographical variables of specific at-risk populations. Further research is also needed in the use of GIS technology to both visually identify and empirically measure spatial relationships of geographical, environmental, and social influences on disease and other health issues. GISs can provide the technology and methodology for the study of the web of causation of health disparities.

Information regarding access to healthcare services for specific populations could better describe the healthcare needs of those at risk, such as rural, elderly, low-income, and others, as well as specific geographical areas. Research of this nature could serve to assist healthcare planners and those who make decisions about the location of healthcare services.

The field of health geography is evolving through the use of evidence-based studies. Sufficient research is available to support the use of GIS as an effective technology for the study of healthcare access and health outcomes. Knowledge generated from empirical research can form a basis for the understanding of health access and health outcomes and for the development of intervention programs to resolve health disparities.

7.2 Future work

If GIS remain in closed environments, their potentiality is not going to be spread.

The best way of developing any prototype is to utilize a Participatory Design (Schuler and Namioka, 1993). With this method, users and developers work together providing automatic feedback all along the way.

A construction of a GIS system in order to help Nurses cope with new methodologies is desirable and can help answer many existing doubts that are holding back some important changes to be made.

One of the future work that can be developed is to follow the Health unit that adopted this distribution of Patients in order to identify what changes occurred. In order words, to validate this patient distribution methodology, identify its' actual improvements in Heath impact or not and the advantages or downsides that it brought.

Other projects can be to actually implement an address error correction, being by a centralized database or by a synchronization of RNU (nation patient process) with local Health Units Databases.

Much new functionality can be added to the developed system. Having in consideration the Health Professionals opinion, new characteristic could be target of further research.

Finally, we can also distribute this system between all Health Units that may need it in order to verify if this method of distribute patients is usable in all population scenarios (small villages, rural environments, islands, places with geographic significant difficulties such as mountains, rivers, etc.).

8. References

ACSS, A.C.d.S.d.S. (2009). UN. SAÚDE FAMILIAR E UN. CUIDADOS DE SAÚDE PERSONALIZADOS; CÁLCULO DE INDICADORES DE DESEMPENHO; CRITÉRIOS A OBSERVAR NA SUA IMPLEMENTAÇÃO (Portugal).

Allender, J.A., and Spradley, B.W. (2001). Community health nursing: concepts and practice, 5th edn (Philadelphia, Lippincott).

ANA, A.N.A. (2001). Scope and standards of nursing informatics practice. In American Nurses Publication (Washington, DC).

Arkowitz, H. (2002). Toward an integrative perspective on resistance to change. Journal of Clinical Psychology *58*, 219-227.

Armitage, P., Berry, G., and Matthews, J.N.S. (2008). Frontmatter. In Statistical Methods in Medical Research (Blackwell Science Ltd), pp. i-vii.

Armstrong, J.S. (2001). Principles of forecasting: a handbook for researchers and practitioners (Kluwer Academic).

Ash, J.S., Berg, M., and Coiera, E. (2003). Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. Journal of the American Medical Informatics Association *11*, 104-112.

Atun, R. (2004). What are the advantages and disadvantages of restructuring a health care system to be more focused on Primary Care services? . WHO Regional Office for Europe (Health Evidence Network report).

Burns, A., and Burns, R. (2008). Basic Marketing Research (New Jersey, Pearson Education).

Chapman, G.B., and Sonnenberg, F.A. (2003). Decision Making in Health Care: Theory, Psychology, and Applications (Cambridge University Press).

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S.C., and Shekelle, P.G. (2006). Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. Annals of Internal Medicine *144*, 742-752.

Cowen, D.J. (1988). GIS versus CAD versus DBMS: What are the differences? PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING *54*.

Cromley, E.K. (2003). GIS and disease. Annu Rev Public Health *24*, 7-24.

CWHPIN (2000). A framework for evaluating the utilization of health information products Hamilton. Central West Health Planning Information Network.

DGS, D.G.d.S. (2007). Centros de Saúde e Hospitais: recursos e produção do SNS. In Estatísticas da Direcção Geral da Saúde (Lisboa).

Dredger, S.M., Kothari, A., Morrison, J., Sawada, M., Crighton, E.J., and Graham, I.D. (2007). Using participatory design to develop (public) health decision support systems through GIS. Int J Health Geogr *6*, 53.

Dumas, M., and ter Hofstede, A. (2001). UML Activity Diagrams as a Workflow Specification Language

«UML» 2001 — The Unified Modeling Language. Modeling Languages, Concepts, and Tools. In, M. Gogolla, and C. Kobryn, eds. (Springer Berlin / Heidelberg), pp. 76-90.

educononline (2009). singapore educational consultant; GIS.

Endacott, R., Kamel Boulos, M.N., Manning, B.R., and Maramba, I. (2009). Geographic Information Systems for healthcare organizations: a primer for nursing professions. Comput Inform Nurs *27*, 50-56.

Esri (2005-2011). Gis.com; Geographic Information Systems

Faruque, F.S., Lofton, S.P., Doddato, T.M., and Mangum, C. (2003). Utilizing Geographic Information Systems in community assessment and nursing research. J Community Health Nurs *20*, 179-191.

Foy, R., Eccles, M., and Grimshaw, J. (2001). Why does primary care need more implementation research? Family Practice *18*, 353-355.

Franks, P., Clancy, C.M., and Nutting, P.A. (1992). Gatekeeping revisited--protecting patients from overtreatment. N Engl J Med *327*, 424-429.

Freitas, A., Brazdil, P., and Costa-Pereira, A. (2005). Mining Hospital Databases for Management Support. Paper presented at: IADIS Virtual Multi Conference on Computer Science and Information Systems.

Freitas, A., Silva-Costa, T., Marques, B., and Costa-Pereira, A. (2010). Implications of data quality problems within hospital administrative databases. Paper presented at: In 12Th Mediterranean Conference on Medical and Biological Engineering and Computing - MEDICON 2010 (Greece).

Garret, J.J. (2005). Ajax: A new Approach to Web Applications.

Gaspar, J. (2011). Sistematização dos Problemas de Qualidade de Dados em Base de Dados Hospitalares: Detecção de Inconsistências, Erros e Outliers (Porto, Faculdade de Medicina da Universidade do Porto), pp. 138.

GATRELL, A., and SENIOR, M. (1999). Health and health care applications. In Geographical information systems: Principle and technical issues Wiley, ed. (New York), pp. 925-938.

GeoinWeb (2009). Carte des associations et communautés sans fil en France.

Goodman, D., and Wennberg, J. (1999). Maps and health: The challenges of interpretation [editorial]. Public Health Management Practice, xii–xvi.

Google (2009). Google Maps: System requirements

Graves, B.A. (2008). Integrative literature review: a review of literature related to geographical information systems, healthcare access, and health outcomes. Perspect Health Inf Manag *5*, 11.

Gulliford, M. (2002). Availability of primary care doctors and population health in England: is there an association? Journal of Public Health Medicine, 14.

Hall, J., Tomlin, A., Martin, I., and Tilyard, M. (2002). A general practice minimum data set for New Zealand. Journal of the New Zealand Medical Association *115*.

Han, J., and Kamber, M. (2006). Data mining: concepts and techniques (Elsevier).

ICN (2002). Nurses always for you: caring for families (Geneva, International Council of Nurses).

Jean-Baptiste, R., Toubiana, L., Le Mignot, L., Ben Said, M., Mugnier, C., Le Bihan-Benjamin, C., Jais, J.P., and Landais, P. (2005). A Web-based GIS for health care decision-support. AMIA Annu Symp Proc, 365-369.

Joel, L., and Stallknecht, H. (2000). A global connection. The International Council of Nurses works to unify the profession. Am J Nurs *100*, 109-111.

KryssTal (2009). Spherical Trigonometry.

Kuhn, H.W. (2005). The Hungarian method for the assignment problem. Naval Research Logistics (NRL) *52*, 7-21.

Landi, F., Tua, E., Onder, G., Carrara, B., Sgadari, A., Rinaldi, C., Gambassi, G., Lattanzio, F., Bernabei, R., and Bergamo, f.t.S.-H.S.G.o. (2000). Minimum Data Set for Home Care: A Valid Instrument to Assess Frail Older People Living in the Community. Medical Care *38*, 1184-1190.

Leitner, H., McMaster, R., Elwood, S., McMaster, S., and Sheppard, E. (1998). Models for making GIS available to community organizations: dimensions of difference and appropriateness. In Empowerment, Marginalization and GIS (Santa Barbara CA).

Longley, P., Goodchild, M., Maguire, D., and Rhind, D., eds. (1999). Geographical Information Systems: Principle and technical issues, 2nd edn (New York).

Maged, N., and Kamel, B. (2004). Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom, Bath.

Marc, B. (2001). Implementing information systems in health care organizations: myths and challenges. International Journal of Medical Informatics *64*, 143-156.

McCormick, K.A., Delaney, C.J., Brennan, P.F., Effken, J.A., Kendrick, K., Murphy, J., Skiba, D.J., Warren, J.J., Weaver, C.A., Weiner, B.*, et al.* (2007). Guideposts to the future--an agenda for nursing informatics. J Am Med Inform Assoc *14*, 19-24.

McGlade, K.J., McKeveney, C.J., Crawford, V.L.S., and Brannigan, P. (2001). Preparing tomorrow’s doctors: the impact of a special study module in medical informatics. Medical Education *35*, 62-67.

MCSP, M.P.o.C.d.S.P. (2008). Sucessos e problemas das Unidades de Saúde Familiar Um estudo qualitativo (Lisboa).

Mowat, D., Gosselin, P., and Bedard, Y. (2000.). Improving health surveillance in Canada - What are the needs? Victoria, BC, Canada.

MS, M.d.S. (2005). Decreto lei nº 88/2005, M.d. Saúde, ed. (Diário da Republica).

MS, M.d.S. (2007). Decreto-Lei n.º 298/2007 de 22 de Agosto. In 161, M.d. Saúde, ed. (Diário da República).

NCHS, N.C.f.H.S. (2009). GIS Geographic Information Systems (Atlanta, National Center for Health Statistics).

Nuckols, J.R., Ward, M.H., and Jarup, L. (2004). Using geographic information systems for exposure assessment in environmental epidemiology studies. Environ Health Perspect *112*, 1007-1015.

Odell, J., Van Dyke Parunak, H., and Bauer, B. (2001). Representing Agent Interaction Protocols in UML

Agent-Oriented Software Engineering. In, P. Ciancarini, and M. Wooldridge, eds. (Springer Berlin / Heidelberg), pp. 201-218.

OE (2002). A cada família, o seu enfermeiro, O.d. Enfermeiros, ed.

OE (2007). Sistema de Informação em Enfermagem; Princípios Básicos da Arquitectura e principais requisitos técnico-funcionais, O.d. Enfermeiros, ed.

Olson, J.E. (2003). Data Quality - The Accuracy Dimension (Morgan Kaufmann Publishers).

Palmisano, A., and Rosini, A.M., eds. (2003). Administração de sistemas de informação ea gestão do conhecimento (Sao Paulo).

Petrovskiy, M.I. (2003). Outlier Detection Algorithms in Data Mining Systems. Programming and Computer Software *29*, 228-237.

Purvis, M., Sambells, J., and Turner, C. (2006). Beginning Google Maps Applications with PHP and Ajax, From Novice to Professional (New York).

Rob, M.A. (2003). Some challenges of integrating spatial and non-spatial datasets using a geographical information system. Information Technology for Development *10*, 171–178.

Romario, J. (2009). como-calcular-distancia-entre 2 pontos usando coordenadas GPS (Sao Paulo).

Sakellarides, C. (2006). De Alma-Ata a Harry Potter: um testemunho pessoal. In Revista Portuguesa de Saúde Pública, pp. 101-108.

Schuler, D., and Namioka, A., eds. (1993). Participatory Design: Principles and Practices (New Jersey, Lawrence Erlbaum).

Scotch, M., Parmanto, B., Gadd, C.S., and Sharma, R.K. (2006). Exploring the role of GIS during community health assessment problem solving: experiences of public health professionals. Int J Health Geogr *5*, 39.

Shi, L., Starfield, B., Kennedy, B., and Kawachi, I. (1999). Income inequality, primary care, and health indicators. J Fam Pract *48*, 275-284.

Shortliffe, E.H., and Cimino, J.J. (2006). Biomedical informatics: computer applications in health care and biomedicine (Springer).

Slocum, T.A., McMaster, R.B., Kessler, F.C., and Howard, H.H. (2009). Thematic Cartography and Geovisualization, 3 edn (New Jersey, Pearson Prentice Hall: Upper Saddle River).

Stewart, A.L., Grumbach, K., Osmond, D.H., Vranizan, K., Komaromy, M., and Bindman, A.B. (1997). Primary care and patient perceptions of access to care. J Fam Pract *44*, 177-185.

Vanmeulebrouk, B., Rivett, U., Ricketts, A., and Loudon, M. (2008). Open source GIS for HIV/AIDS management. Int J Health Geogr *7*, 53.

Weiner, D., Harris, T.M., and Craig, W.J. (2002). Community Participation and Geographic Information Systems. In Community Participation and Geographic Information Systems (London).

WHO (2000). Munich Declaration: nurses and midwives: a force for health. In WHO MINISTERIAL CONFERENCE ON NURSING AND MIDWIFERY IN EUROPE (Copenhagen).

WHO, W.H.O. (2006). Report on the Evaluation of the WHO Multi-country Family Health Nurse Pilot Study.