Centre for Geo-Information

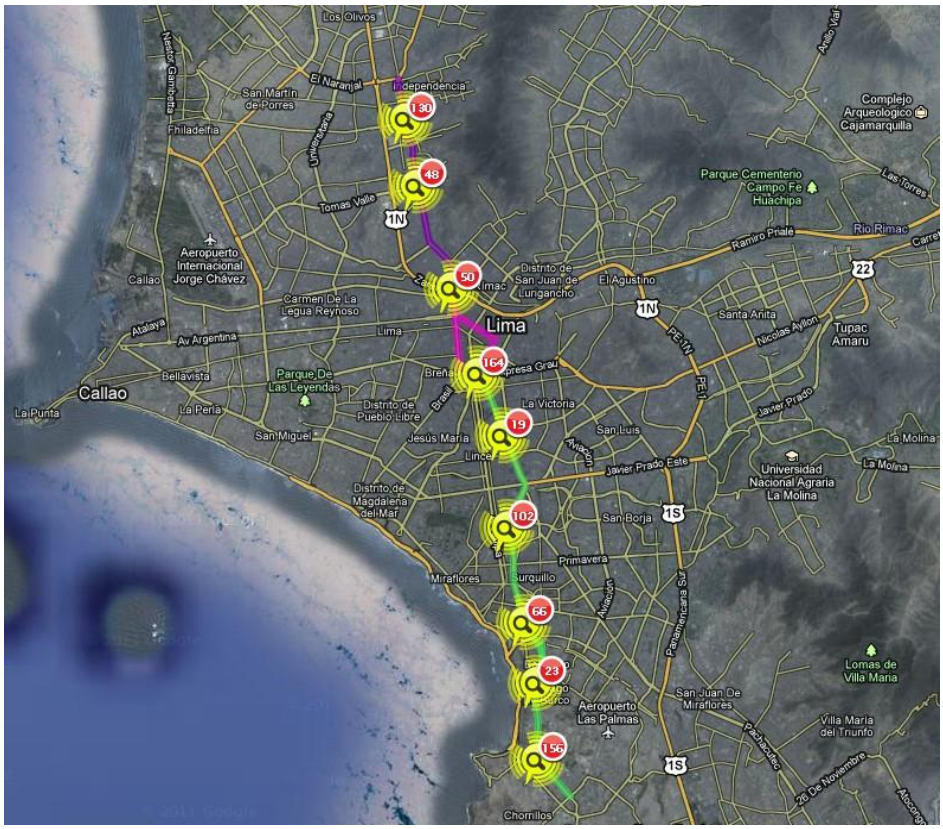
Thesis Report GIRS-2011-19

VOLUNTEERED GEOGRAPHIC INFORMATION

A Conceptual Model and Analysis of Volunteered Data

Silvia Carolina Loo Arancibia

August 2011



WAGENINGEN UNIVERSITY
WAGENINGEN UR

VOLUNTEERED GEOGRAPHIC INFORMATION

A Conceptual Model and Analysis of Volunteered Data

Silvia Carolina Loo Arancibia

Registration number 77 05 13 52 80 70

Supervisors:

Prof. Dr. Ir. Arnold Bregt
Dr. Ir. Lukasz Grus

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

August, 2011
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2011-19
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Foreword

This thesis report is the result of a long learning process that started in September 2009. It reflects the work I performed during the last period of the Master in Urban Environmental Management with major in Geo-Information Science at Wageningen University.

I would like to express my sincere gratitude to Arnold Bregt and Lukasz Grus for their supervision of my work during the period of the research. They guided me in the right direction with advices, suggestions and critics that made the improvement of the results possible. Their comments also encouraged and motivated me in the whole process of the study.

Thanks also to the National Police Department of Perú for their cooperation in providing the data I analyzed. Thanks to the teams of Dignidad Policial and La Factura for their time and support during the data collection period. I would like to extend my special thanks to The Netherlands Fellowship Programme (NFP), for the financial support they provided me during my studies.

I would also like to thank God for the two years of student experience in this country from which I have learned a lot. Thanks to my parents, family, and friends for their support and unconditional love. Thanks also to Thomas K. for his advices, patience and love.

Finally, thanks to my classmates and friends Samrawit Teclai and David Haberkorn for their support and good advices during this time.

Abstract

Volunteered geographic information (VGI) is a phenomenon which emerged in recent years due to the evolution of Internet technologies. Since its rising VGI attracted the attention of scholars and different studies have been conducted in order to explain its implications, but mainly focused in particular aspects of the phenomenon. This research attempts to bring a more holistic vision of the VGI. Therefore, the scope is focused in two main objectives.

The first objective is a proposal of a conceptual model for the VGI life cycle and the components participating in the process of the development of the VGI project. This proposal is achieved by revising diverse literature of VGI components, and previous models that are commonly used in software development. The model designed is describing phases, activities, actors, and driving forces in a VGI initiative. After that, it is tested in two Peruvian VGI projects in order to verify its functionality. When applied for characterizing the VGI projects it demonstrates its usability.

The second objective consists of an analysis of the content and quality of VGI data produced by the two Peruvian initiatives. The data content analysis is achieved by comparing VGI data with official sources and field observation collected data. The general results show coincidences between data. For data quality assessment three standards to analyze spatial information are used, which are attribute accuracy, positional accuracy and completeness.

The research contributes with an innovative conceptual model to the VGI field. At the same time, it provides a different perspective on the content and quality of VGI data.

TABLE OF CONTENTS

LIST OF FIGURES

.

1. INTRODUCTION

1.1 CONTEXT AND BACKGROUND

The Internet is defined as the worldwide interconnection of individual networks operated by government, industry, academia, and private parties (IWS 2011). Nowadays, due to the speedy development of digital network technologies, Internet has become an integral part of our society. It is a powerful platform to have access to, exchange and process information and connects hundreds of thousands of telecommunication networks (Peng and Tsou 2003).

These networks need a system which supports interactions between them. Web service technologies provide this functionality in the process of communicating data between organizations via Web-based applications. For the integration of such applications the following technologies provided by open standards are used: XML (to tag data), SOAP (to transfer data), WSDL (to describe the services available), and UDDI (to list what services are available) (Booth et al. 2004).

In recent years, Web services have experienced new improvements in terms of technology and role of users (De Longueville et al. 2009). The new version of the Web known as Web 2.0 is a platform that represents an evolution of traditional applications, which now are oriented to the final user (O'Reilly 2005). These new applications are generating collaboration, interaction, updates and services between users. In that way, they can voluntarily share their knowledge and experience with other internet users (De Longueville et al. 2009), and collect, access, and create information. Obviously, the success of the Web 2.0 world depends on a successful user experience (O'Reilly 2005).

As a result, many internet sites are based on the User Contribution System[1] (Cook 2008), which means that people from different organizations contribute with knowledge that will be useful for others. An important component of this System is the User Generated Content (UGC) that refers to any material created and uploaded to the Internet by individuals not necessarily professionals (IAB 2008).

UGC is one of the fastest growing forms of content on the internet (IAB 2008), and nowadays is generating a mass online collaboration. A good example of this is phenomenon is *Wikipedia,* the famous online free encyclopaedia written and updated frequently by amateurs (Cook 2008); another example is the site *Flickr*, the popular website where users upload and share personal photographs and videos.

Evidently, the scope of these advancements in information technology has reached the field of Geographic Information Systems[2] (GIS), as well. In the early 90's the Internet started to be used

[1] **Cook** (2008) confers exceptional value to user contributions. In such **User Contribution System** the *user* can be anyone: a customer, an employee, someone no linked to the organization. The *contribution* can be active (work, expertise or information) or passive (data automatically generated during the entry or use of the site). Finally, the *system* is the method, usually internet based, by which contributions are aggregated and made useful to others.

[2] **Geographic Information System** (GIS) is an organized structure that integrates a computer system (hardware, software and technic procedures), with spatially referenced (geographical) data management and analysis tasks (**Heywood** 2006). GIS allows users to access into large volumes of data for viewing, understanding, questioning, interpreting and visualizing.

for delivering geographic information and maps (Haklay et al. 2008). Since then, the Internet has changed how GIS data is accessed, processed, shared, and manipulated (Peng and Tsou 2003).

The Web 2.0 now offers a new and powerful opportunity to share and use geospatial information (Dangermond 2008). This evolution has motivated a shift in the paradigm of how we do GIS and also has influenced how we conceptualize public participation; the non-experts now are the prime generators of content on the Internet (Sieber and Rahemtulla 2010).

Expectably, GIS have expanded its limits into web-based applications both in digital mapping and spatial databases (Dangermond 2008). The information provision has changed from a linear model (centrally collection of data and information) to an inter-networked, participatory model; and the user-generated content is blurring the difference between producers and consumers of data (Haklay et al. 2008). Consequently, the role of GI users has changed from being data consumers to become data producers (Bishr and Janowicz 2008).

A special case of the user-generated content phenomenon is Volunteered Geographic Information (VGI)[3] (Goodchild 2007). VGI refers to the creation of new geographically-based initiatives and abilities in Web sites by large numbers of untrained private citizens with or without formal geographic expertise or qualification.

The information that these volunteer actors upload (such as geographical data, comments, photos or videos) is linked to a specific location through a spatial reference. For that reason, now vast amounts of place-based information are readily available (Flanagin and Metzger 2008). New products and projects have been created under the VGI strategy; a good example is OpenStreetMap, which is a community generated map of the world.

As a result, the Web unlocked GIS which now is available to everyone (Fu 2010) and obviously VGI Web sites assist the creation and dissemination of user-generated content (Agrios and Mann 2010). In such a way, GIS users can take advantage of the rapidly growing amount of VGI on the Web (Artz 2010).

An important value of VGI is that it can notice and monitor many local activities in various geographical locations (Goodchild 2007). Lately, with the proliferation of consumer GPS devices, smart phones and cameras, which help the users to easily upload the observed phenomena to the web, the public has turned into "6 billion walking sensors" (Goodchild 2007; Fu 2010).

Indeed, the usefulness of VGI and the capacity of response of volunteers have been confirmed with the 2010 earthquake in Haiti, just to highlight a recent example (Wolf 2010). After the natural disaster groups of volunteers updated detailed spatial information from the capital Port au Prince, using the platform of OpenStreetMap. In some days the city became one of the best mapped locations on earth (Wolf 2010). After this tragedy, VGI mapping has demonstrated its importance for taking action in crisis events such as earthquakes, floods, tsunamis, violence, and so on (Schade et al. 2010).

Contrarily to traditional and controlled creation of geographic data which consumes great amounts of money and time, VGI created data is essentially free or very low cost (Wolf 2010). For Goodchild (2010) these contributions are not yet formal (from official sources) but assertive

[3] Goodchild coined the term in 2007; however, before him scholars have given different names to this phenomenon, e.g. Turner in 2006 refers to Neogeography (Budhathoki 2010), Gouveia et al. in 2003 named it volunteer collected data.

(from volunteers), but they could include information collected by authoritative organizations and share it freely in the Web (Dangermond 2008).

VGI initiatives have a great potential because citizens are able to link their observations with their own experiences, being participants instead of just observers (De Longueville et al. 2009). If we create in the future useful VGI applications, they will be of enormous benefit to government, business, and individuals (Fu 2010).

1.2 PROBLEM DEFINITION

VGI as a relatively new research field has vast future potential applications in society (Harding et al. 2009) and offers the possibility to use and share information in different scenarios. Precisely, because VGI is a new phenomenon we still do not completely understand it (Budhathoki 2010). There is still a need to characterize the roles that VGI is expected to play as a knowledge and information production (Elwood 2008).

Research is required to determine VGI issues as if it is just a mere resource to improve the already existing spatial databases or is it possible to create complete new knowledge with it (Elwood 2008). There are tremendous amounts of generated geo-information; however they need to be standardized in order to be usable in combination with official data. For that reason standards should be studied and defined.

Of course, the VGI subject has been researched in different aspects. For instance, researchers already have given attention to volunteers and their motivations (Budhathoki 2010, Coleman et al. 2009, Parker et al. 2010). Nevertheless, it would be significant to determine who is originating VGI initiatives, and what motivations are encouraging these people to do it.

Likewise, there are studies available relating VGI with major established systems such as SDI (Spatial Data Infrastructure) (Budhathoki 2010, Castelein et al. 2010, Craglia et al. 2007, Goodchild 2007), PPGIS (Public Participation GIS) (Tulloch 2008), and GIS (Sui 2008). In such ways, VGI is described following these major systems structures but is not studied as a system itself, considering its own life cycle and functional activities.

Nowadays, any person with internet connection can create web content. Because the creators of that content are mostly anonymous, the quality of the information they provide cannot be guaranteed. In that way emerges one of the major shortcomings of VGI, which is the quality of the information (Bishr and Janowicz 2010). These contributions are not always following the quality criteria of formal GI (completeness, lineage, accuracy and consistency) (Bishr and Janowicz 2010). Consequently, that content is subject of criticism since it could be misrepresentative or not useful.

In a similar way, volunteer actors have different backgrounds and motivations and they are not necessarily experts in GI. This fact creates problems with the credibility of information because is very difficult to authenticate or validate the sources (Flanagin and Metzger 2008).

In this context emerges the interest of analysing two Peruvian VGI initiatives: criminality reports (www.quenoteroben.pe[4]) and transportation services reports (www.datea.pe[5]). These two recently created Web pages are using VGI in order to obtain information from citizens of Lima in the subject of safety. Quenoteroben,pe collects information related with crimes against property and datea.pe does the same with quality of public transportation.

1.3 RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

The main objective of this thesis is to propose a conceptual model for characterizing the life cycle of a VGI initiative and the elements that participates in such process. This thesis also aims at analysing the content and quality of the data provided by volunteers.

1. What are the phases and its components in the life cycle of a VGI initiative?
2. Which phases and its components of such process are present in the two Peruvian VGI initiatives?
3. What is the content and quality of information in the data produced by these initiatives?

1.4 OVERVIEW OF THE THESIS

- Chapter 1 of this thesis introduces the context and background of the research. Then, it describes the research objective and enunciates the research questions.

- Chapter 2 describes the methodology followed in order to achieve the research objective.

- Chapter 3 proposes a conceptual model for the life cycle of a VGI initiative.

- Chapter 4 evaluates the content of the two Peruvian VGI initiatives according to the VGI conceptual model proposed in chapter 3.

- Chapter 5 analyses the content and quality of data produced by the two Peruvian VGI initiatives.

- Finally, a discussion is stated and conclusions and recommendations are given.

[4] www.quenoteroben.pe

[5] www.datea.pe

2. METHODOLOGY

This chapter addresses the methodology followed for carrying out this study with the objective to answer the research questions. Hereafter the steps accomplished for each chapter included in this research are described.

Figure 1 shows a flowchart with the activities developed per chapter.

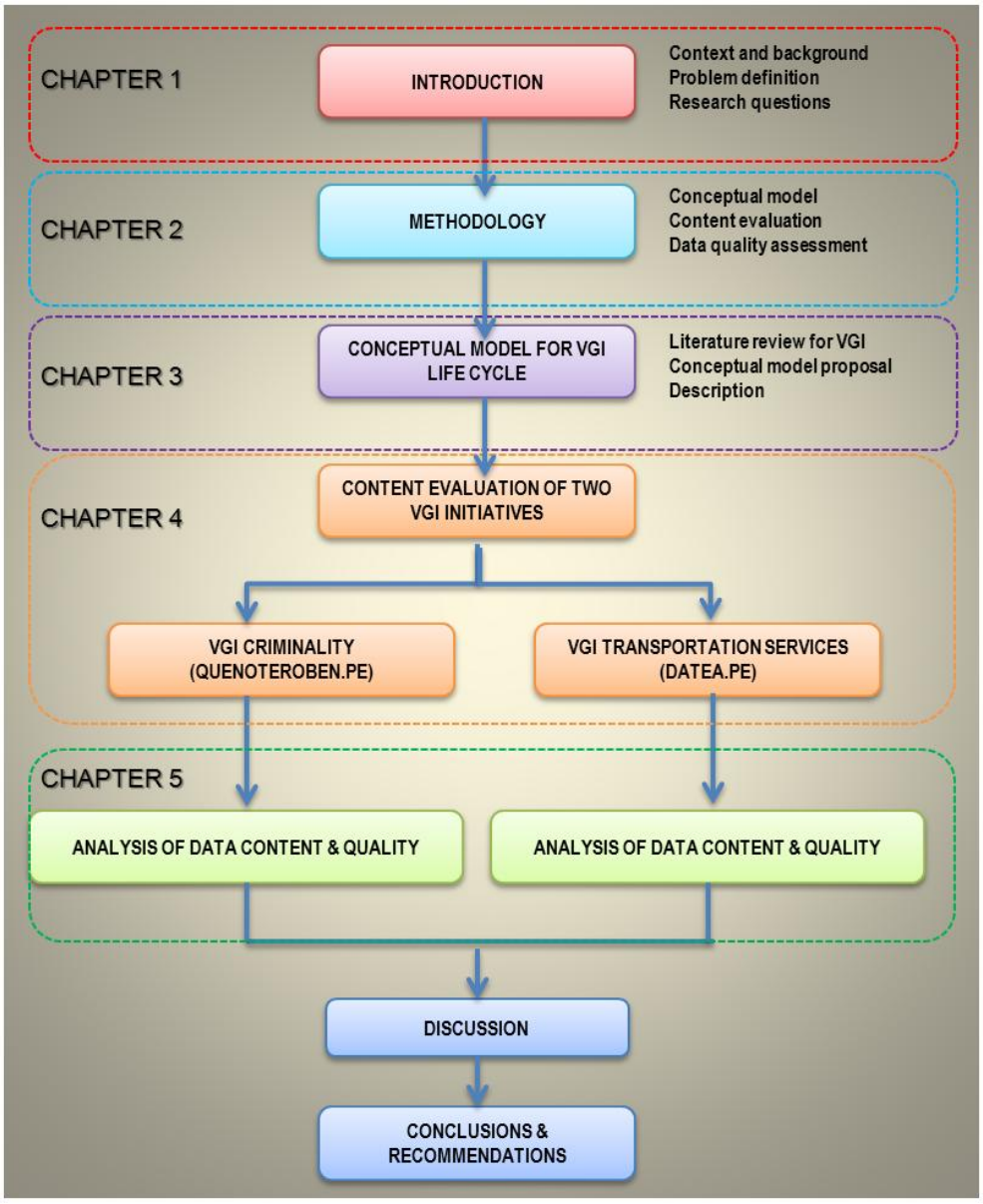


Figure 1: Research Methodology flowchart

In the next paragraphs the various research steps displayed in Figure 1 are explained in more detail. The stages of the study are presented from chapter 3 onwards.

2.1 CONCEPTUAL MODEL FOR VGI LIFE CYCLE

The aim of chapter 3 is to define a conceptual model to support and guide the characterization of the components in the life cycle process of a VGI initiative.

To attain this objective, a review of scientific literature is done, searching for previous studies in order to find relevant material and to identify useful aspects that could be included in defining such a conceptual model.

On the basis of this initial review, a model that could adjust the formulation of a conceptual model in a time line perspective is identified.

The result is a conceptual model whose construction implies the inference and description of phases involved in the life cycle of VGI and the components taking part in it.

Later on, the phases and the relation between the main components are described.

2.2 CONTENT EVALUATION OF TWO VGI INITIATIVES

The purpose of chapter 4 is to analyse, study and characterize two recently created Peruvian VGI initiatives: criminality (www.quenoteroben.pe) and transportation services (www.datea.pe), taking as a base the conceptual model proposed for this research.

To accomplish this objective an overview of Lima city's context is done, characterizing circumstances related with the two main topics of the VGI initiatives: public safety and crime, and public transportation services respectively.

Immediately after, an exploration of the activities required for developing a project is done. The main sources to obtain this information are interviews with the actors involved in the process of conceiving, designing and initiating of each project (initiators and programmers).

Later on, an identification of the main actors involved in the process is developed. Finally, the same process is carried out in order to determine the driving forces involved in each project. The aim is to find out the technological tools from the Internet network that are enabling the exchange of information along volunteers or contributors in these VGI initiatives.

2.3 DATA CONTENT AND DATA QUALITY ANALYSIS OF TWO VGI INITIATIVES

The content analysis is oriented to define which kind of information is possible to be extracted according to the contributions of volunteers in both initiatives. The quality analysis is oriented to determine if VGI is trustable after comparing it with official data. In this case the study is divided in each project.

2.3.1 VGI initiative for criminality

First of all, the sample for the Web project was chosen: Police stations of three districts from Lima: La Molina, San Borja, and San Isidro. The data corresponded to the months of May, June and July of 2010.

Second, information from the reports of the Web page was collected for the same months of 2010. As VGI data is presented as a small report in each crime point on the map, it was necessary to elaborate a table (tabular data) which included the fields from the report. For the later process of creating geospatial data each crime event was ascribed as a record with a unique identifier expressed in numbers. Likewise, it was annotated other variables such as address, date, modality of crime, gender of contributor, if he/she was a witness or a victim and if he/she reported the crime to the police department.

Third, the information was geo-referenced and the correspondent process of geospatial data creation started using ArcMap[6]. Point features were chosen for storing the data in a shapefile format. In total three shapefiles were created, one per district.

Fourth, besides a process of official and field data collection was made in the city of Lima. This data has been taken as a benchmark, in pursuance of having a basis from which it will be possible to compare the volunteered information. The three Police stations above mentioned were visited collecting data from the three correspondent months.

Fifth, the data was geo-referenced in ArcMap following the same procedure described above for the VGI data. The challenge in the procedure of geo-referencing data for both datasets is locating the exact position for some events. For instance, some records provided as address a block number in a certain avenue. However, a block normally has 100 meters of length and these reports did not specify exactly where the event happened. For these cases, it was decided to position the event in the middle of the block.

Sixth, the sample VGI collected and classified is analysed, in order to determine the content of the data. The content analysis is used to find out which modalities of criminality are prevailing in each district and the localization of offences. Then, the content of official data for the sample months is analysed.

Seventh, data quality analysis for analyzing spatial information is done. With this purpose elements defined and used by consensus in the identification of spatial data quality assessment are used (Coot 2008, Guptill & Morrison 1995). These elements are:

- Attribute accuracy: the closeness of attribute values to their true value. This element will be linked with the previous one. Values of datasets are compared at the same location. If there are coincidences in positions then the attribute comparison is carried out.
- Positional accuracy: position of geographic features. In this study this component is evaluated in relative accuracy. This refers to closeness of relative positions of objects in a dataset (VGI data) to those relative positions accepted as true (official data) (Coot and Rackham 2008). As mentioned before, in the geo-referencing process emerged the problem of the exact location for some records in blocks. For this reason a buffer of 100

[6] Is the main component of ESRI's ArcGIS suite of geospatial processing programs, and it is used primarily to view, edit, create, and analyze geospatial data

meters of ratio is created with the purpose to check if there are matches between both datasets.

- Completeness: presence and absence of objects in a dataset. For both datasets the error for omission in their information is evaluated

2.3.2 VGI initiative for transportation services

First, the sample for the Web project was chosen: three bus stations one per route from the Metropolitano bus service: Estación Caquetá, Estación Central Grau, Terminal Sur Matellini.

Second, the Web reports provided by volunteers in the bus stations were collected and tables were elaborated considering the topic, problem and suggested solution.

Third, the three main bus stops were visited for collecting data.Since there is not any possible official data to use for comparing with volunteer data, it was necessary to define a methodology for data collection. The methodology selected was field observation in situ. In such a way the real characteristics of each station were observed. For this task a matrix was elaborated following the same three topics used in the Web page for volunteers' reports, which are pedestrian access, especial access and service & security.

Fourth, the sample of VGI collected and classified is analysed, in order to determine the content of the data. The content analysis is used to find out which are the main problems reported in each station. Finally, the information collected in the field work is compared with the volunteered information uploaded in the VGI Web site.

For this initiative data quality analysis is not carried out. Due to the nature of information (numerous reports referred to the same point location or bus stop) it would be useless to geo-referencing the data.

3. CONCEPTUAL MODEL FOR VGI LIFE CYCLE

3.1 VGI PREVIOUS WORK

Since the term Volunteered Geographic Information was coined by Goodchild (2007), literature on its uses and implications has increased considerably. This attention received from GI researchers has been focusing mainly in studying the potential VGI applications in science, business and society (Elwood 2008, Budhathoki 2010).

Depending on the field of action and interests of researchers there are different approaches for characterizing and illustrating the VGI phenomenon. For means of this paper three main approaches were identified: VGI phenomenon and its relation with major systems, VGI explained through its components and VGI case study system architectures.

3.1.1 VGI and major systems

First of all, as it was mentioned above there are some works more oriented to emphasize VGI as part of other systems such as Spatial Data Infrastructure[7] - SDI (Goodchild 2007, Budhtahoki 2008, Craglia et al. 2007, Castelein et al. 2010), Public Participatory GIS[8] - PPGIS (Tulloch 2008), or even GIS (Sui 2008).

Accordingly, Goodchild (2007) suggests that VGI fits in the model of SDI in the sense of individuals participating and creating patchwork coverage instead of national mapping agencies producing and updating cartographic information. This response might be done according to the local needs of each community following standards and protocols provided for official agencies. After that, the various pieces of the patchwork can be fitted. This approach considers the functioning of VGI as an appropriate tool to provide information on the SDI system.

Besides other studies confer another perspective of the VGI phenomenon, which is its convergence with the SDI infrastructure. Craglia et al. (2007) states such convergence is already occurring. He centres the attention in three strands of convergence: through the developing of spatial data services for reaching broader audiences and delivering information; through sensor networks and sensor webs in real time for monitoring daily life motions; through some facets of VGI not that distant from SDI such as data production techniques (like GPS traces).

For their part, Castelein et al. (2010) characterizes VGI taking as a framework the model of SDI components proposed by Rajabifard et al. (2002) which are people, access network, policy, standards and data. The proposal defines thirteen characteristics of VGI grouped in categories within the five SDI components. After that, the study elaborates an analysis comparing both systems -based in the same five components- and determining relations and differences between them.

---

[7] According to the Executive Order of US (1994) **National Spatial Data Infrastructure (NSDI)** means the technology, policies, standards and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data.

[8] The term **Public Participation GIS** (**PPGIS**) was coined at the National Center for Geographic Information and Analysis (NCGIA), to cover a specific geographical context (North America), and for a particular purpose - how GIS technology could support public participation for variety of possible applications. (http://www.ppgis.net/ppgis.htm)

Following the same line with the SDI system there is another approach which proposes the hybridization of both systems VGI and SDI (Budhathoki 2010). In such hybrid system data collection, standards and quality control will be guided by authority represented by SDI. Thus, VGI would be a complement for SDI process, changing the roles of users to become *produsers*, having as a result richer GI infrastructure combining the long experience of SDI with the novel VGI.

These three approaches (Craglia et al. 2007, Budhathoki 2010, Castelein et al. 2010) have the vision of an evolutionary growth of both systems jointly. Probably because they are subsequent studies the VGI is considered not as a complement of SDI but a convergent system for SDI.

A similar vision but in a different direction was conducted by Tulloch (2008), comparing VGI with PPGIS and assuming that they share some ground. Notwithstanding it is unclear how much they overlap, there is a clear connectivity between them that could embrace these two areas as one. For the moment, the assumption is that VGI innovations would serve PPGIS, and the already research of the latter would benefit the less structured research and literature base in VGI.

Finally, there is a work (Sui 2008): which considers the phenomenon of VGI as the new development of GIS or the "wikification of GIS". Here, VGI is typified and its performance is described following the classical components of GIS:

- Hardware: owing to grid computing now networked computers can share computing power and work collaboratively to process large amount of information. This combined with human sensors via Web 2.0 might be the future scenario for computing infrastructure.
- Software: the open source and free software movement have growth rapidly due to the contribution of the user community. The open source of map severs in one of the driving forces for the next step which is the "wikification" of GIS software.
- Data: data production is one of the areas with dramatic changes. Since the transformation of volunteers from passive consumers to active producers of geospatial information, the amount and attribute information increased amazingly.
- People: nowadays we are witnesses of volunteers working for free and helping to collect geographical information, redefining traditional roles. This strategy has been followed even by official organizations such as the US Geological Survey.

3.1.2 VGI and components

In a different perspective there are some works oriented to understand or explain the VGI phenomenon by its components.

In his pioneer work Goodchild (2007) conferred a very important role to technology as the enabler component which facilitates the VGI development. He discusses and describes the following implications: Web 2.0, georeferencing, geotags, GPS, graphics, and broadband communication. In the same way he mentions participant populations as potential sources of data, the so-called human sensors. This is the innovative side of this kind of initiatives, because now the creation and dissemination of geographic information is a bottom-up process.

In a different direction Elwood (2008) ascribes interest to spatial data, demonstrating that research has been strongly oriented upon this component. In this work special attention is focused in three aspects of spatial data, which anyway are powerfully influenced by the role of technology:

- Production and sharing, these processes are influenced by hardware, software and Web services, for that reason now VGI services are more open to general public contributions.
- Content and characteristics, due to VGI tools such as simple interfaces and web accessibility the group of participants has increased. Similarly, the volume of digital spatial data and its expansion in topics has grown. Here is mentioned the role of SDI which research would contribute in functions as storage, manage, search and share of data in a digital environment. But mostly would be necessary to provide ground for interoperability.
- Purposes for use, obviously in means of the role of VGI in society.

More focused in VGI data Coot and Rackham (2008) describes some characteristics possible to be recognized that differentiate it from conventional data, such as:

- The creation of data is stimulated by lack of available data or limited access on existing conventional data sources.
- It is a collaborative process as long as involves capturing, processing and disseminating geographic information provided by individuals in a voluntary mode.
- The process of creating data is not necessarily done following standards and methodologies used by specialists. On the contrary, such process is more intuitive. Most people are not really conscious about its concept of GIS and attributes. Maybe they are exposed to GIS-products like Google Maps, but this does not mean that they recognize the technology behind them (Burtman 2009).
- The data is obtained using some open source approach, thus is possible to produce and consume data for free.

A group of studies are more focused in the volunteers of the data especially in their motivations for participating in such initiatives. This interest emerged forasmuch as for researchers resulted fascinating determine why people is contributing in this kind of projects if there is not reward for doing it.

Following this line of thought Budhathoki (2010) proposes a conceptual framework in which are described three sequential arenas, their interaction and their relationship within the VGI process:

- Motivational arena: divided in context of participation and motivations to contribute and manage a volunteered GI database. Again, as in other studies especial attention is given to technological developments to make this possible. The addition here is that the level of access to technological tools and the capacity to use them influence the ability to contribute to VGI.
- Action & Interaction arena: these are the mechanisms to interact and coordinate in order to address processes of decision making. The arena is integrated by structure, action and norms/ rules that affect people's decisions to contribute.
- Outcome arena: constituted by the contribution and the evaluative criteria.

This work is important because the author conceive the framework with the purpose of using it for deeper investigation in the motivation of participants contributing in GI.

In a similar line, Parker et al. (2010) investigate the range of stakeholders involved in the VGI process. Stakeholders are described as the net of users and those who have an influence on mashup design requirements. The originality consists in the introduction of a user-centered design perspective. Here, the authors try to describe the relationships and tensions or issues among these stakeholders (in terms of information flow). Since the net is quite wide they are categorized in: consumers, special interest –mapping- groups, local communities and professionals.

Finally, Coleman et al. (2009) conceptualize a characterization of the nature and motivations of users and producers. They support the use of the neologism “produsers” considering that participants in the production process are both users and producers. Here are proposed five overlapping categories of contributors: neophyte, interested amateur, expert amateur expert professional and expert authority. In this work are mentioned three different contexts within VGI operates:

- Market-driven context: when the contribution is directed to a commercial database or service. E.g. Tom Tom
- Social networking context: as its name indicates is done in social networks sites. E.g. OSM
- Civic/governmental context: when the contribution is done by a concerned citizen about his/her own town or city. E.g. public participation GIS.

3.1.3 VGI and case study systems architecture

For this topic we identified authors which are studying the VGI phenomenon in cases applied in real projects.

In an early work Gouveia et al. (2004) discussed and conceptualized the role of citizens contributing with their knowledge in data creation via Web platforms and the implications of this process. They proposed a system for environmental data collection by volunteers through information technologies as well as current VGI initiatives. This collaborative system describes four requirements for the monitoring process, which are:

- Data collection, via multimedia inputs using three main tools: annotation tools, geo-referencing tools and semantic tools.
- Data access and exploration, via searching and retrieval tools in order to facilitate spatial, thematic and temporal searches.
- Data validation, via tools and methodologies for guarantee the data quality and credibility.
- Monitoring communities, via communication, information exchange and teamwork tools, thus involving volunteers’ communities.

In this work is highlighted the decisive role of Information Technology as facilitator tool for carrying out such process. The design for the environmental collaborative monitoring system contemplates both front-end (user’s side) and blank-end (server’s side) solutions. Although, is not considering the inception step for the model, it provides clues for the components for such a system, which are data tier, logic tier, communication tier and presentation tier

The interesting proposal of this design is that it is introducing the volunteered concept before Goodchild (2007) coined the term. Likewise, is explaining and making visual the four tier architecture of this monitoring system. Moreover, in the model is inserted the requirements above mentioned and the functionalities that they cover.

De Longueville et al. (2009) just like Gouveia et al. (2004), provide an architectural model for an environmental case study. This work considers that there is some degree of vagueness in VGI – ignored in other systems- and to solve it a hybrid combination of open gazetteer approach and the concept of degree of truth is used. To carry out this proposal it is suggested a prototype web application based on a serve-oriented architecture called eVGI. The eVGI has three classical components: presentation tier, logic tier and data tier. According to them the innovative of the model is the way of components' implementation.

Clearly, the reviewed literature point out that VGI is an important component of the GIS evolution in the Web environment. Abundant research is motivated to analyze, describe and understand VGI phenomenon and its adjustment in the formal GIS landscape.

3.2 THE FRAMEWORK FOR THE VGI CONCEPTUAL MODEL

A conceptual model is defined as an abstract representation of concepts, their relationships, constraints, actions and interactions in a given domain (Tanrıöver and Bilgen 2010). It is used in different disciplines and contributes to represent, understand and communicate a situation occurring in a specific field.

Conceptual models are useful tools that give visual representation to abstract concepts turning them into something understandable and concrete. In this study a conceptual model helps to understand how various components of the VGI system interact in a time-line approach.

In the construction of the VGI life cycle conceptual model the main objective was first to determine different phases that are involved in the process. Well, the application of the life-cycle model is used in different domains and fields such as computer configuration, engineering systems, financial services, business managing, and strategic planning, among others (Ashafri, 1995). Likewise, this framework is very used in traditional software development as well as web sites development process.

According to this criterion and after critically evaluating various models it was decided to use and adapt some ideas from the System Development Life Cycle (SDLC) which encompass the conception of a life cycle. SDLC can be defined as the formal process of developing information systems through a number of successive stages or phases which comprises several sub-phases or steps (Stefanou 2003). The SDLC model prescribes a structured, phased, linear approach to system development (Agarwal, 1990).

It was considered that the logical adequacy of the SDLC approach could guide the initial stage of creation of the VGI model. Once determined the phases of the model, the development of the succeeding structure and the relationship between other components would be consistent.

The conceptual model of SDLC has evolved through the years. There are many different SDLC models and methodologies which generally consist of a series of defined steps or phases (NIST

2008). The key stages are growing in number and sophistication but they maintain the main logical process of development. In the following table (Table 1) it is possible to identify the different key stages proposed by different authors:

Table 1: Key phases of the SDLC conceptual framework according to different authors

KEY PHASES	AGARWAL (1990)	JAIN (1991)	ASHAFRI (1995)	STEFANOU (2003)	MODELL (2007)
1	Project definition	Investigation	Requirement specifications	Feasibility study	Project initiation
2	Requirements determination	Analysis and General Design	Design	System investigation	Analysis
3	Logical design	Detailed Design and Implementation	Coding	System analysis	Examination and Study
4	Physical design	Installation	Testing	System design	Design
5	Testing and implementation activities	Review	Maintenance	System implementation	Implementation of procedural solutions
6				Review and maintenance	Testing of the procedural solutions
7					Post-implementation review of results

All the previous key phases or activities are sequential. In a SDLC is possible to deal with different activities during the evolution of the model. However, there is no agreement concerned to the optimum number and names of the stages (Stefanou 2003). This important fact provides the allowance of choosing the phases that could be adjusted to the VGI model. It should be taken into account that this is a backtrack process. We are adapting an established model in which modifications will be done in order to suit the inquiry.

Dobson (2008) proposes an interesting adaptation of the SDLC waterfall model for geo-databases compilation which incorporates User Generated Content as one of the components of the system (Figure 2). This model attracted our attention because VGI is a type of UCG (Goodchild 2007). Indeed, the design is related with creation of databases for a navigating system. In this case UGC is considered an important method for augmenting the quality of updating information for commercial developers of navigation databases. The feedback content is provided by clients who purchased and are using a product. Therefore is not following the dynamics of a VGI process where there is not a market relation client-product.

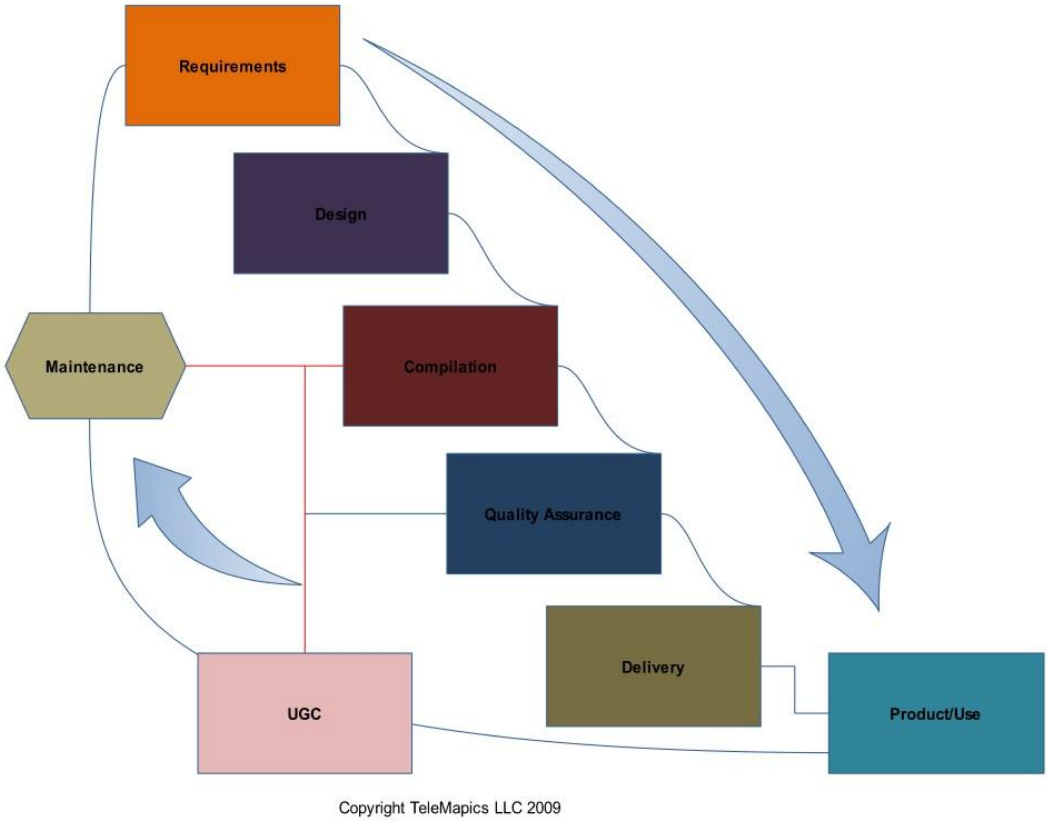


Figure 2: A model incorporating UGC in a SDLC for collecting navigating data (Dobson 2008).

As we can observe this model takes some phases from the SDLC framework such as requirements, design, maintenance, delivery and quality assurance. The innovative component and interesting for the purpose of this study is compilation, a phase that could be included as part of the VGI system as well. The UGC component is outside of the initial sequential process, when the product is being developed but is included in the feedback process, directly linked to maintenance and product use, and indirectly to compilation and quality assurance.

Nevertheless, the objective of this research is to study the VGI -a type of UGC- as an individual process and not as part of some other system like in this model. The following is the formulation of my position about VGI life cycle which is giving the direction to this study.

3.3 VGI LIFE CYCLE CONCEPTUAL MODEL

For this model is used a linear sequence of basic phases. Overlapping and iteration between them is usually inevitable (Stefanou 2003). However, this model attempts to delimitate the actions and elements within each phase in order to construct a coherent structure.

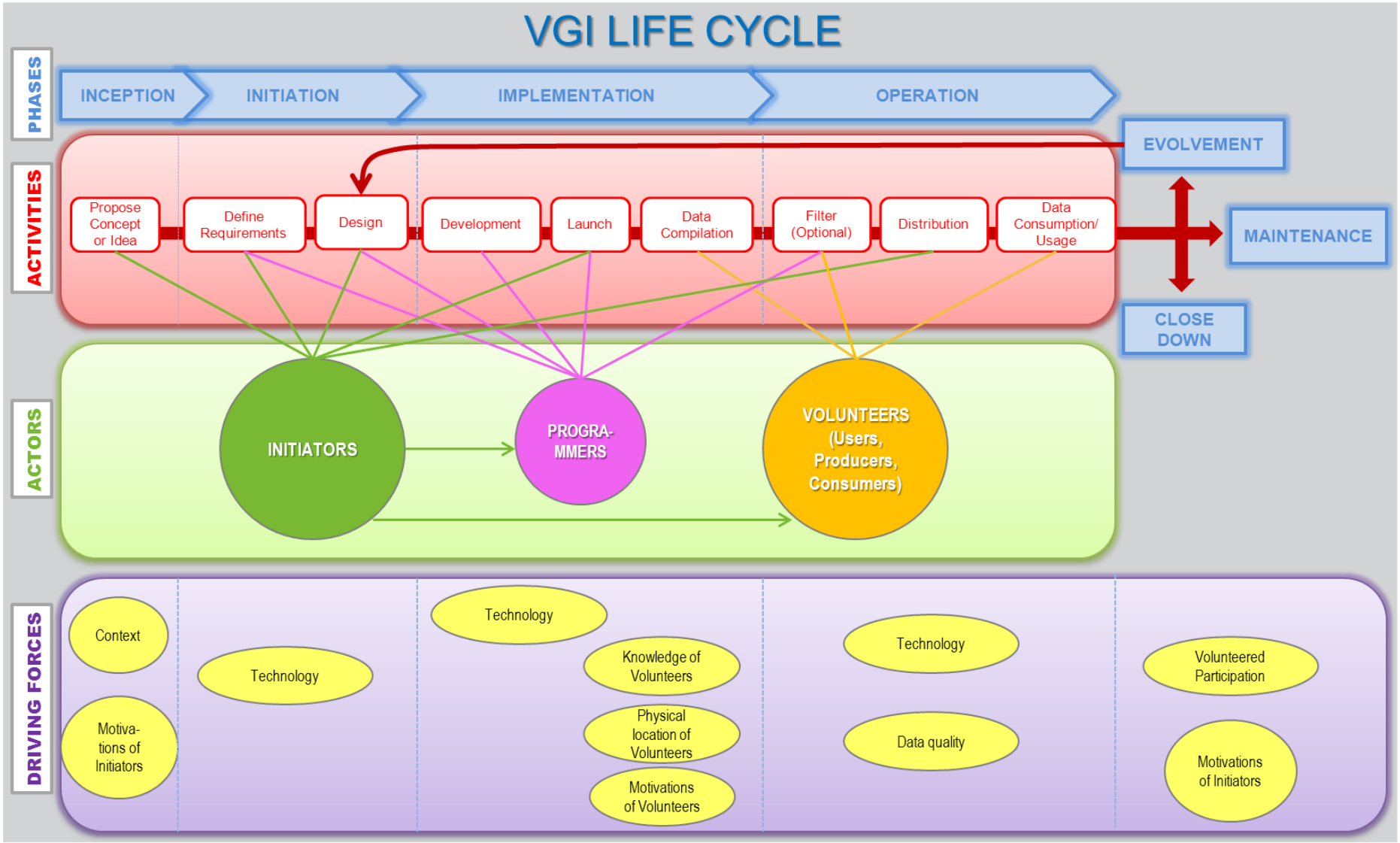


Figure 3: Conceptual Model for VGI life cycle.

As we can observe in Figure 3 this conceptual model is divided in four components: phases, activities, actors, and driving forces.

Phases: or stages for the VGI life cycle, composed by activities (or sub-phases). Each phase must be completed to a great extent before moving to the next one (Stefanou 2003).

Activities: the actions performed by actors at each stage of the developed project.

Actors: groups of people that intervene in the VGI life cycle and make possible the carrying out of the project. Each actor fulfils a role and accomplishes multiple activities. Initiators are the originators of the project and they are making it available. Programmers contribute with their technological knowledge to get the project going. Volunteers are the contributors of the geo-information; here are included users, producers, and consumers, since one single person could have the three functions at the same time.

Driving Forces: they are general trends that allow change to occur, having an effect on the other components of the system. Some of them, such context or technology influence the process in a more broadly and external way. Some others such knowledge or motivations are internal, and their influence is specific for this environment.

The logical sequence for reading the model is following the timeline five phases: inception, initiation, implementation, operation, and evolve/maintain/close down of the project. The latter will vary depending on the circumstances in which the project is executed. Specific activities are part of each phase, and one or more actors participate in them. The driving forces are influencing the performance or the other three components.

1. INCEPTION

In this very first phase is emerging the whole concept of the VGI initiative. The proposal of the idea comes from the initiators. They are influenced first by the context that somehow incentive and encourage them to start with the project. Also, their owninitiator motivations are important catalysts for carrying out the idea and later initializing the process.

2. INITIATION

The project initiation sets the boundaries and scope of the project (Modell, 2007). Here initiators contact programmers and together they define the requirements needed for the project. After this analysis, they will decide the logical (how the system fits the requirements) and physical (hardware, software, database and program specifications) design of the project (Stefanou 2003). In the design stage, decisions are made about how the project will be implemented.

Designing is a creative process and, in many cases, is mainly ad hoc (Ashrafi, 1995). Here are considered all the requirements of the project. Technology is the driving force that will influence and delimitate the two activities in how they will be developed.

3. IMPLEMENTATION

This stage is referred to the installing according to specifications and delivering the system into operation (Stefanou 2003). The development -or programming- is executed by programmers who construct the previously designed program. After that, the system is launched, that means installed and put into operation. Initiators use strategies to contact and attract volunteers and make them to participate in the project.

Immediately after starts the data compilation, contributed by volunteers. However for this activity we consider there is an overlapping between implementation and operation phases. The scope has not a clear delimitation and data compilation could be part of any of both phases.

The main driving force here is technology because the functionality of the three activities is directly related with it. Nevertheless knowledge, physical location and motivations of volunteers are inducing what, how and why volunteers are contributing data for compilation activity.

4. OPERATION

The first activity from this phase is filtering, which is an optional step because some projects execute it but some other simply release or publish the data just the same they receive it. There are cases where the system enables volunteers for filtering or editing data for correcting inaccuracies (e.g. OpenStreetMap).

Data is integrated into the system and then is distributed or published for its free use or consumption. Of course, technology plays again an important role as a driving force in the step of distribution. However, data consumption or usage will depend on the quality of the data, if consumers consider is inaccurate or misrepresentative they will not use it.

5. EVOLVEMENT/MAINTENANCE/CLOSE DOWN

Three different and possible destinies are expected for the project final phase. All of them are closely depending from the volunteered participation and the motivations of initiators.

The project will evolve or change according to the interest of the actors involved in the process. For instance some operations can be corrected or improved, some sub-products can emerge, or the topic can turn on into another one.

It is possible that the project stays with the same content and characteristics, so in this case it will be maintained as it is.

Finally, the project could be close down because completes its cycle and objectives. Another reason is because it failed and volunteers did not participate in it. Likewise, is possible that the interests of initiators go to a different direction and this will result in the close down of the project.

4. VGI CONCEPTUAL MODEL APPLIED IN TWO PERUVIAN INITIATIVES

Hereafter, the conceptual model of the VGI life cycle is applied for analysing the two study cases of VGI. Following the proposed structure the phases are characterized according with the other elements which are activities, actors, and driving forces. In such a way, the model is corroborated to determine its functionality.

4.1 VGI INITIATIVE FOR CRIMINALITY

This VGI initiative www.quenoteroben.pe is a project created by the group “Dignidad Policial”. It collects data associated with crimes that includes different modalities of robbery in the 49 districts circumscribed in the city of Lima and Callao.

In this Web site citizens via a Google map are able to upload information about crimes against property that they had suffered or witnessed, with geo-location, time and type of crime. The electronic map contains information that is constantly updated and “feed” by the citizens.

4.1.1 Inception

This phase is characterized by the context and the motivations of initiators.

4.1.1.1 Context: Public Safety and Crime

Public safety is an important factor of life quality in every city, and it is associated with the concept of confidence of people in their daily life. The capital city of Lima has the highest rate of homicides and criminality from the country (IOP, 2010). According to an official survey in public safety (IDL, 2010), during the year 2009, 6 out of 10 people in Lima have been victim of robbery or attempt to robbery. Likewise, from this survey it was possible to know that 61% of Lima’s citizens perceive that robbery on streets have augmented in the last year.

The Report on Public Safety and Human Rights from the Interamerican Commission of Human Rights (CIDH, 2009) confirms this tendency, not only in the Peruvian context but also in Latin America: “For the first time in decades, in Latin American countries, delinquency and crime have displaced unemployment as the main concern for population”.

The problem of delinquency in Lima affects daily all members of population no matter to which socioeconomic class do they belong (Blume, 2010). According to the Information System of Prosecution Office during 2008 the highest crimes incidence registered in Lima were crime against property with 29.65% of the total penal causes (Blume, 2010).

The citizens of Lima are concerned for the insecurity they should face as a consequence of street delinquency, reflected in the violent and threaten robberies on streets (IOP, 2010). As a consequence the population considers that most potentially insecure places are public spaces, precisely the places where they meet with other people doing quotidian activities. That means

they feel unsafe in bus stops, public transportation vehicles and on streets in general. Indeed, the main criminal scenarios are streets.

Besides surveys, another indicator used to measure the criminality is formal reports registered by the National Police Department. Between the years 2007 to 2009 these reports increased in 12,000 (IOP, 2010) which means that there is an aggravation in public safety problems. However, these numbers are not reflecting the real situation since a huge number of citizens do not denounce in a formal report when they are victims of a crime, especially if it is a matter of minor robberies. This is happening mainly because citizens have the feeling that presenting a report does not have results for them. Furthermore, according to the IOP survey (2010) 79% of people do not trust in the National Police Department.

This situation is really worrisome because the current situation is a previous stage for a more violent scenario, with higher levels of dangerousness and violence (Blume, 2010). Moreover, people have the feeling of defencelessness in confronting this circumstance. That is why they are always searching for measures to avoid crime and protect themselves.

4.1.1.2 Motivations of initiators

As we mentioned before people do not trust their local, regional and national authorities in the fighting against crime. For that reason, they are searching for new spaces in which they can express or manifest their perceptions or experiences facing crime in the city.

This VGI initiative was born when the director of NGO ‘Dignidad Policial’, the retired PNP General A. Jordán, had the idea to generate a Delinquency Map of Lima.

The main objective of initiators is to produce an alert site against crime for citizens where it is possible to report a crime. At the same time it prevents crime since the provided information might help other citizens to know about dangerous places and day period of crimes. As well, the final aim of the initiative is to become a tool for offering information not only to citizens but also to the National Police Department.

4.1.2 Initiation

Despite of his knowledge on public safety General A. Jordán did not know how to materialize and implement the idea of a crime map. The NGO made contact with a team of social communicators in order to carry out the initial idea and develop a Web page.

4.1.2.1 Technological requirements and design

This project is using Google services by the integration of Google maps into the Web page in a simple way and for free. The only requisite is to obtain an API key and specify in which domain will be used. The versatility offered by Google maps thanks to AJAX technology enables users to zoom in/out and move the map. The site quenoteroben.pe has opened an account in order to create 49 maps one per district of Lima city.

For the basic functionalities of the site PHP is used, which is a general-purpose scripting language that is especially suited for Web development and can be embedded into HTML[9]. This scripting language works for receiving the reports, publication of videos and for the head/top of the site.

Likewise, the site has been developed in a generic server which is Mediatemple[10]. Among others, the basic characteristics of this server are unlimited bandwidth, two gigabytes of space and possibility of receiving a quote of emails. Of course the web hosting has a .pe domain name and they pay annually approximately 30€ for the service.

4.1.3 Implementation

The inception and initiation phases started on January 2010. The launch of the site was on July 2010.

4.1.3.1 Data compilation

This site has a particular mechanism for managing the information. The reports are received via email (the hosting permits a maximum of 5 emails per day) and the programmer puts them on record on a database classified by type of crime, date of crime event, time of event and address of event. The data is stored in a MySQL database, which is an open source database and the reports are generated in an Excel .xls format.

Figure 4: Form filled by users with information about robberies.

[9] http://www.php.net/

[10] http://mediatemple.net/

This means that volunteers do not upload the information by themselves. They just should fill a form (Figure 4) without the requirement of registration. After that, a technician uploads the information on the map.

For technical and budget issues the initiators and programmers decided to use this process for obtaining the data. The procedure was chosen because generating a different system – volunteers uploading the information by themselves- turned out to be very expensive, difficult and time consuming.

4.1.4 Operation

4.1.4.1 Filtering

Initiators consider the modality of the technician uploading the information on the map more effective. The person who receives the reports makes the filter for incomplete or wrong reports that cannot be uploaded in the map. For them the filter is very important in order to verify the authenticity and usefulness of the reports, because in the virtual community and social networks it is not possible to expect always legitimacy of the contents.

4.1.4.2 Distribution

Besides the information available on the same Web site (Figure 5), daily, monthly and annually reports are generated and published via social networks such as Facebook and Twitter. This is a strategy to create continuous comments or debates between users and to demonstrate users that the information is always updated.

Likewise, the site promotes its own Youtube channel: Robotube. In this channel is possible to check videos related with crimes that the creators elaborate. These videos contain information about preventive tips of crimes modalities and potential robbery places as well as interviews with safety experts

Figure 5: Publication of the VGI in the Web site

4.1.5 Maintenance

After one year of operation this project remains the same. There are plans to work jointly with municipalities and gain money via advertisements. However, any of these plans did not take place yet.

4.2. VGI INITIATIVE FOR TRANSPORTATION SERVICES

The VGI initiative www.datea.pe is created by the NGO “Ciudad Nuestra” and “La Factura” communication group. It is related to the concept of tranquillity in transportation (IOP 2010). This site as well is collecting data from citizens related with the efficiency of the Metropolitan Bus Rapid Transit system, the very first experience of integrated public transportation service in the city of Lima.

4.2.1 Inception

4.2.1.1 Context: Public Transportation Services

The public transportation service is another main problem in the city of Lima. It is full of challenges because the system is deficient, informal and chaotic. And it is really far away from the meaning of a service. This problem is associated with the quality of life of the citizens.

According to the survey from IOP (2010), 86% of the population are not satisfied with the public transportation in Lima and they considered it as one of the main problems of the city. Taking into

account that more than 50% of citizens are getting around the city in mass public transportation the nonconformity with the service is very high. The problem is mainly associated with the over fleet and bad service offered by the vehicles called coasters, a 30 passenger conveyance that should be used for private transportation. And 69% of people are using these vehicles for transportation.

People also manifest that travelling in public transportation generates delays and stress. In a city with more than 70 kilometres of extension and more than 8 millions of people, this issue is a challenge that needs to solve problems in the short, mid and long term.

Nowadays, there are two big transportation infrastructure projects in the city. One of this is the Urban Electric Train –under construction-, and the other the Metropolitan Bus Rapid Transit system (Metropolitano). With these projects the regional government is planning to revert with order and control the informal growing of the public transportation sector from the last decades.

Metropolitano is the new integrated public transportation system for the city of Lima inaugurated in July of 2010. This system articulates a fleet of modern buses that circulates through exclusive corridors. It has been developed to interconnect the city from north to south, crossing 16 of 49 districts with benefits in transportation for near 1 million people. The current system is divided in three road corridors: North, Central and South.

4.2.1.2 Motivation of Initiators

This is a project promoted by the association of the NGO Ciudad Nuestra and the independent group of technology, design and participative strategies La Factura. The NGO Ciudad Nuestra is an organization expert in analysis and advices in urban public policies.

The initiative is created as a response to the context of changes and transformation that are happening in the public transportation system of Lima. The objective is to integrate the information and concerns from the users of the Metropolitano Bus Rapid Transit system.

Since the operation of the Metropolitano, some problems have been detected such as accessibility, pedestrian circulation, safety and service quality, among others. Whereas the Metropolitano project is in its initial stage, this initiative considers that users of the service should be part of the development and improvement of it. The involvement of users is not only oriented to report problems but also to propose solutions.

Users of the transport service are able to report problems that they are observing in each bus stop from the three corridors: North, Central and South, and after that they might propose solutions. The final aim of this initiative is to promote the active participation of citizens in order to make them contribute to the process of transformation of Lima into a modern city. For the creators public involvement is one of the main factors for the development of the city.

4.2.2 Initiation

4.2.2.1 Technological requirements and design

For the publication of this Web site –as well as the previous one- the map API of Google was used. Maps API are a free service, available for any web site that is free to consumers, with the possibility to overlay their own data on the top of them. This service is utilized in combination with Drupal[11] as a platform. Drupal is an open source content management (system and framework) platform and tries to balance flexibility and simplicity for effectiveness of Web designs. Its system is dynamic and it has been designed with the possibility of being developed and maintained by users' contributions.

The server for the Web hosting is hired from Hostgator[12], which is a company that offers different server packages, with the option of hosting almost 30,000 visits per day. The advantages of using this hosting are MySQL, transfer, storage, domain, and email account, all of them unlimited.

4.2.3 Implementation

The Web page was launched on July 16th 2010.

4.2.3.1 Data compilation

For uploading the information volunteers should first register and create a user account. After that they have the option of select between the main corridors for enter the report: North, Centre and South (Figure 4.2a). The three corridors specify the initial and final bus stop.

Figure 6: Three main corridors to upload the report for the Metropolitano service

Next, the corresponding map is displayed according to the chosen corridor. In this map are shown the different bus stops with the number of reports that have been uploaded (Figure 4.2b).

[4] http://drupal.org

[12] http://www.hostgator.com/

With this view is relatively easy to identify which bus stops are having more problems. Likewise, is also possible to read the comments that volunteers are uploading.

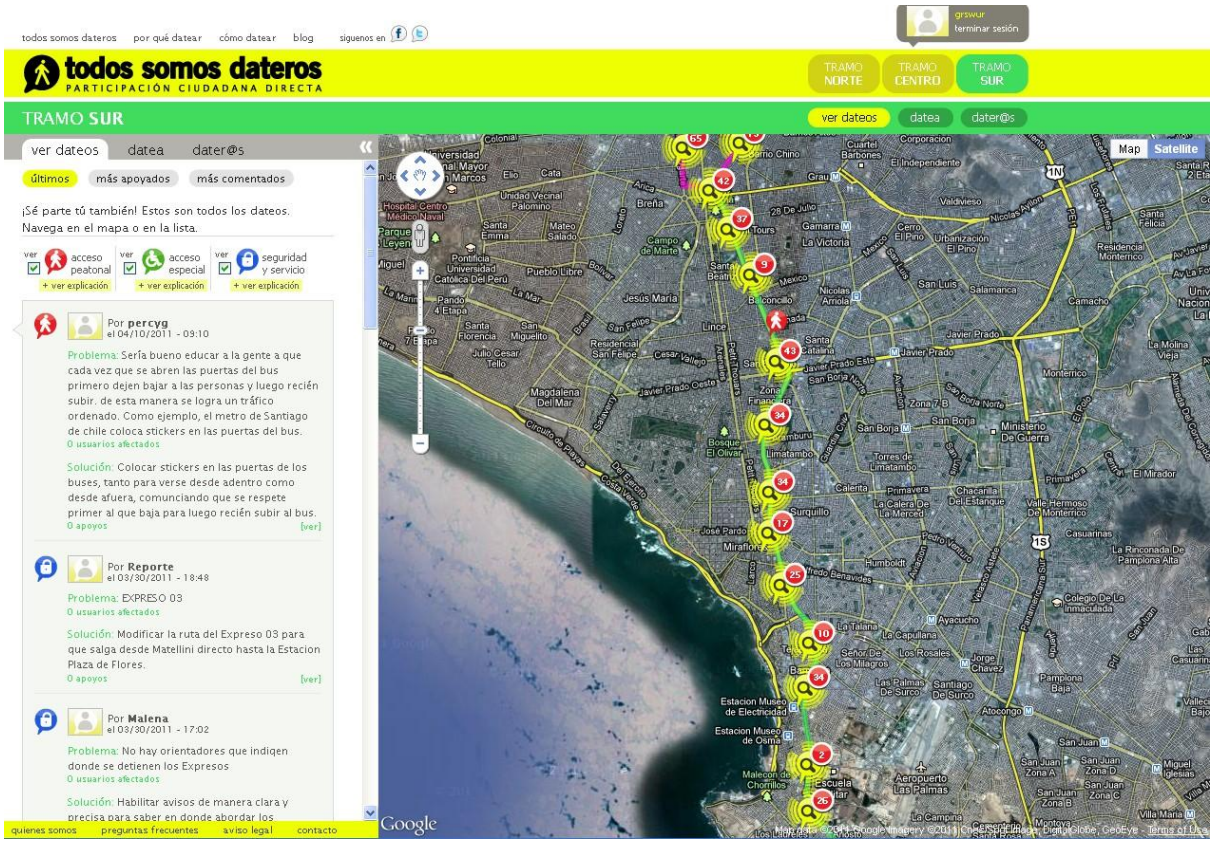


Figure 7: Map of the South Corridor showing the number of reports per each bus stop

If the user would like to make a report he or she should click in the option of 'datea' and a new window appears with options about three main topics: pedestrian access, especial access or safety & service (Figure 4.2c). Each of these options has an explanation about the topic in order to make clear the issues involved. Likewise each problem is classified in different options.

Figure 8: Table for choosing the problem volunteers want to report: Pedestrian access (red), Especial access (green), and Safety & service (blue)

After the input, volunteers should propose a solution about the problem or complaint they are reporting.

4.2.4 Operation

4.2.4.1 Distribution

The publication of the report is direct without any filter. Finally, the information is systematized and exported to Excel files.

After that, the NGO -Ciudad Nuestra- team in charge produces reports that are published in social networks and are sent to volunteers by email.

4.2.5 Evolvement

In this first period of the project, initiators noticed that there were numerous reports regarding to the critical situation of cyclists in the city. For that reason, they decided incorporate a new section oriented to this topic, collecting problems and proposing solutions in order to improve the few cycle paths within the city.

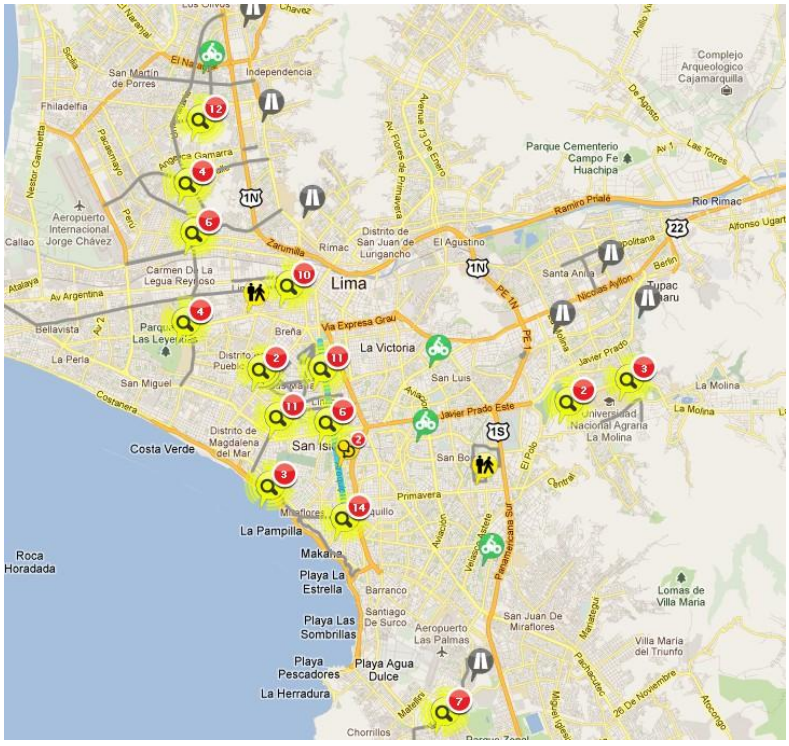


Figure 9: New section of datea.pe project: Cycle paths reports

The scope of this second period is going beyond the previous since is calling diverse authorities for being involved in the project. Currently, two important municipalities are participating.

The problems under consideration now are access, safety, infrastructure & signals, and behaviour.

Figure 10: Table for choosing the problem volunteers want to report in the new cycle paths project: Access (green), Safety (blue), Infrastructure & signals (grey), and Behaviour (yellow)

5. DATA CONTENT AND DATA QUALITY

VGI data in both case study initiatives is providing the location of different events such as crime offences and transportation service deficiencies. With this information it is possible to deduce characteristics of the problematic that each project involves and the distribution of these issues. However, because of criticism in VGI concerning the content and quality of the data it is necessary comparing it with official and trustable sources of information. In this chapter the aim is to accomplish such a comparison and analysis.

5.1 VGI INITIATIVE FOR CRIMINALITY

Regarding to the data content it is necessary to differentiate between thievery and robbery. In such a way it is easier to understand the situation of criminality. Thievery is referred to the taking in possession of any good illegally without threatening any person; there is not contact with the victim. On the contrary, robbery denotes the illegal taking in possession of any good using violence, threats and intimidation.

5.1.1 Data content analysis

As next step the characteristics of VGI data in the districts, which were chose as sample, are presented. Afterwards the VGI data is compared with official data. The analysis is focused on type and modality of robbery and the spatial distribution of the crime.

5.1.1.1 La Molina

As we can observe in Figure 11 there are few events registered by volunteers for the chosen sample months (May, June, and July 2010). Only two reports have been uploaded, one for June and one for July. One is a car theft offence and the other an alert of robbery in snatch modality. Very little could be concluded from this information.

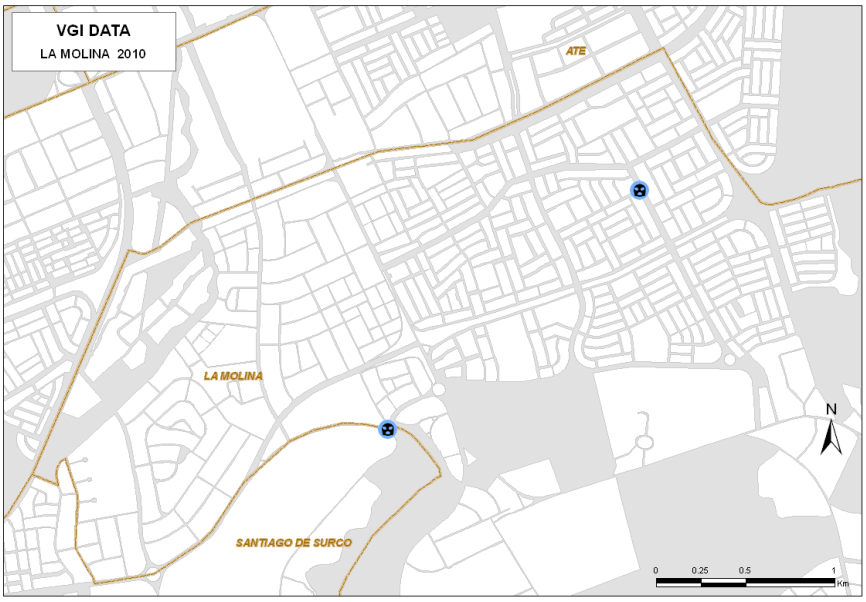


Figure 11: District of La Molina with volunteered reports.

The records registered for official data have the following characteristics for the district of La Molina. In total there are 112 offences registered for these three months. Most types of crimes reported are related with robbery from persons with firearms. It is difficult to establish comparisons with VGI data (Figure 11) because this includes only two reports.

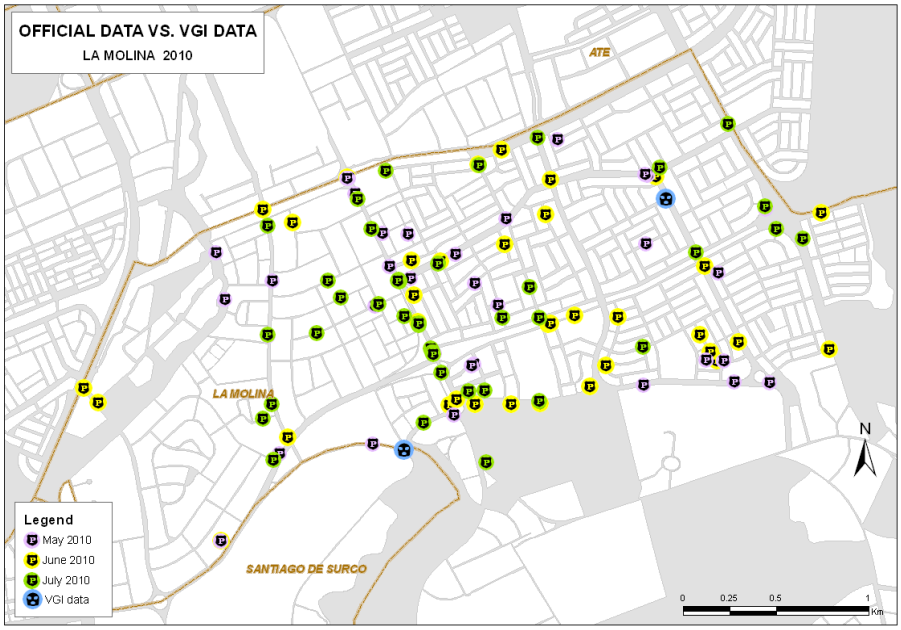


Figure 12: Official data vs. VGI data in La Molina district for the months of May, June and July 2010.

5.1.1.2 San Borja:

In San Borja VGI registered more events (15 in total) for the sample months (Figure 13).

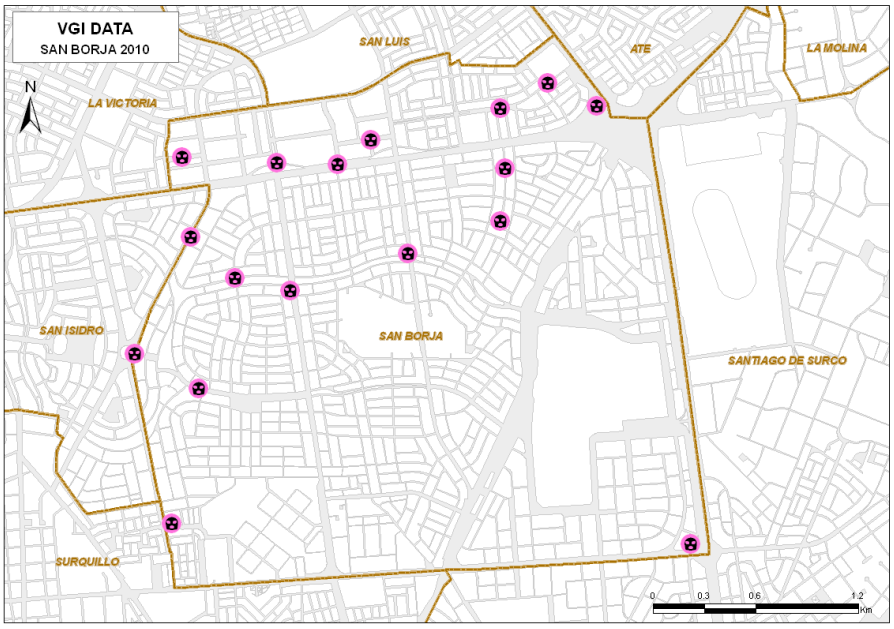


Figure 13: District of San Borja with volunteered data.

Crime: The most common crime is snatch robbery, which means the criminals have direct contact with the victims stealing from them their belongings. Likewise, there are assaults with firearms and as long as it is a dangerous situation, all of them were reported to Police. Most of the contributors indicated that they were victims in the incident. Moreover, practically all of them indicated they are men.

Spatial distribution: According to volunteer data, the spatial distribution of reports reflects that crimes are mainly happening in the north-east side of the district.

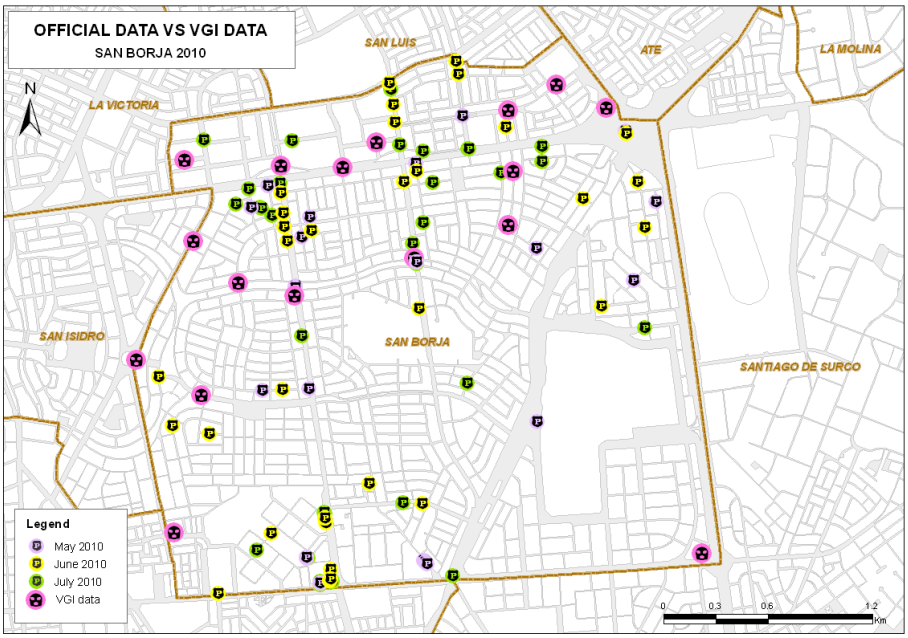


Figure 14: Official data vs VGI data in San Borja district for the months of May, June and July 2010.

On the other hand, official data (Figure 14) register 92 offences in total. For the three months the type of crime reported refers to robbery from person by snatch modality.

The official reports of crime are distributed among the whole district, differently than VGI reports that indicates north areas (Figure 13). However, the content of information is similar since in both datasets are indicated that robbery from a person in snatch modality is the major offence.

Numbers in official data evidence that in this district no matter people being are victim of minor offence, they are still reporting the crime in the Police Station. Likewise, there is a dangerous situation concerning to assaults with firearms.

The occurrence of crimes in the north area might be because it is a high frequented area by passengers. Two important and big avenues which interconnect the city (Javier Prado Av. and Aviación Av.) are confluence nearby. Moreover, many important institutions (National library, National Museum, educative institutes) are located in the surrounds; this implies large influx of people during the day.

Something similar occurs in the southwest area. There is not only a confluence of two important avenues (Aviación Av. and Angamos Av.) but also there are many shopping centers located in the area.

5.1.1.3 San Isidro

For San Isidro VGI report only 6 offences for the three sample months (Figure 15).

Crime: The most common modality of crime in this area is snatch robbery and since the volunteers were witness in most of the cases, they did not make a report to Police.

Spatial distribution: Reported crimes are located in the central-east side of the jurisdiction of this police station.

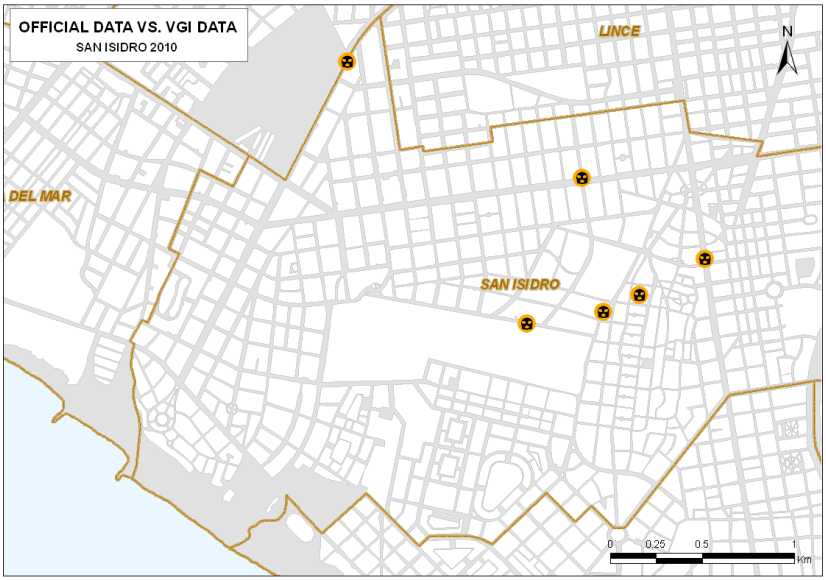


Figure 15: District of San Isidro with volunteered data.

Official data register 84 reports (Figure 16), and all of them referred robbery from a person in the modality of snatch as the main type of crime. The reports registered in the VGI initiative coincide with officials in the type and modality of offence (snatch). Looking at the data (Appendix 1 – Table 4) it is also possible to affirm that people are not reporting minor offences in Police stations.

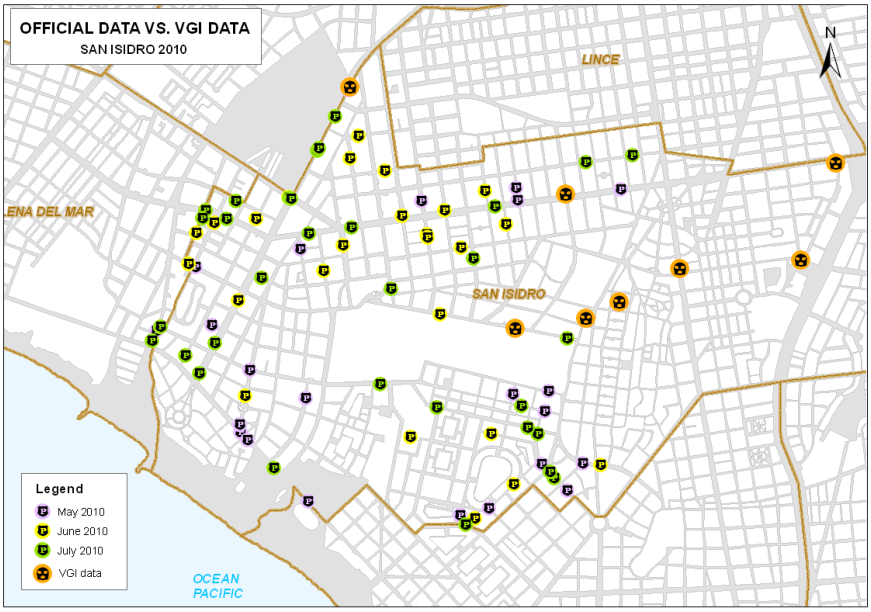


Figure 16: Official data vs VGI data in San Isidro district for the months of May, June and July 2010.

In summary, after analysing the content of both datasets it was found that for the districts of San Borja and San Isidro there are coincidences in the crime type, which is robbery in the modality of snatch. For the district of La Molina it is difficult to deduct any information because the VGI reports are too little.

5.1.2 Data quality analysis

For the analysis of data quality a buffer of 100 meters is applied[13] for each point of official data. This geo-processing tool for spatial analysis will help to determine proximity, overlapping or matches between events registered in both official and VGI datasets. Then, it is possible to analyse if there are attribute and positional accuracy for the matches.

5.1.2.1 Attribute accuracy

Attribute accuracy is referred to the carefulness and correctness on how the features in the data set are described.

After applying the buffer function of 100 meters around the records of official datasets, the only district which generated matches is San Borja in the month of July. The following map (Figure 17) was obtained and according to it, San Borja has three overlaps.

[13] This analysis tool is utilized to solve the problem of location to the true position of each crime event. It is impossible to locate offences in the exact place of occurrence. That is why a range of 100 meters is applied, which is the average length of street blocks in the city of Lima.

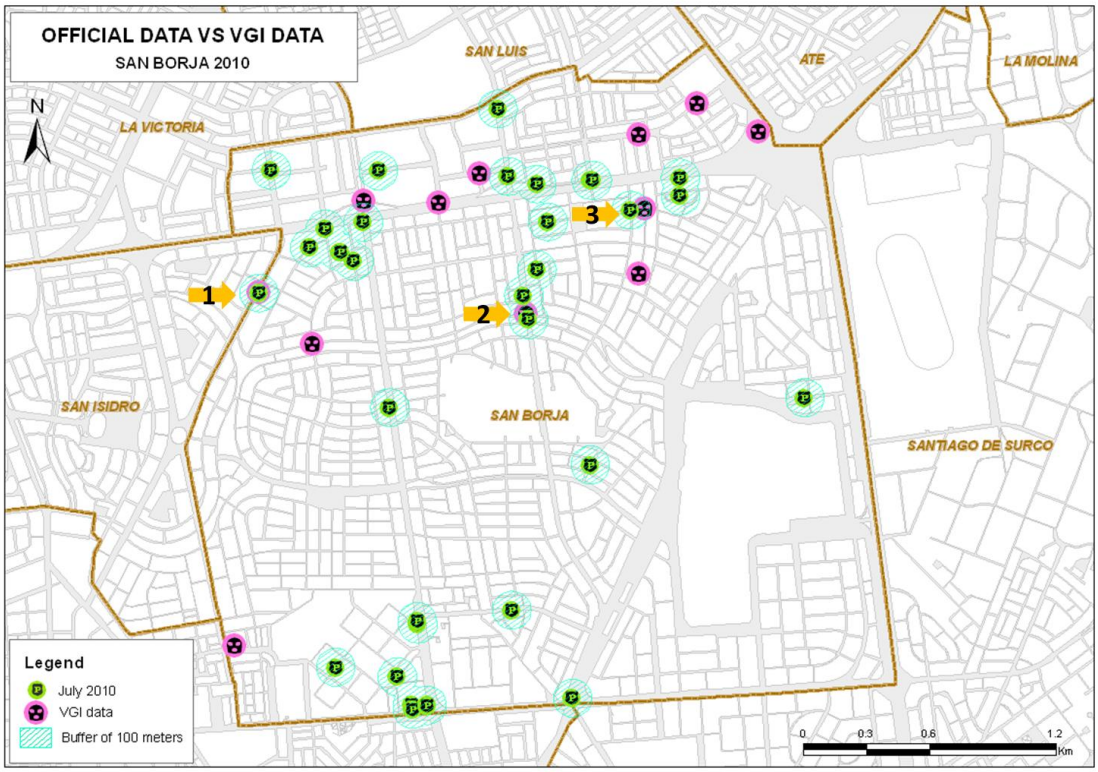


Figure 17: District of San Borja with buffer of 100 meters in official crime data July 2010.

Now, the content of such matches is analyzed with the purpose to determine attribute accuracy in the information that both datasets implies.

For match 1 there is address correspondence. However, while in the record of official data the type of crime is a robbery, the VGI report is not about a concrete event but is just an alert given by a witness in order to be careful in that area. Then, attribute accuracy is not given in this case.

For match 2 it is possible to determine positional correspondence, however just like in the previous match the type of crime is different: the VGI report is an alert of assault offence. On the other hand, official report indicates a theft. Definitely, the reports are not referred to the same event.

For match 3 the position of each event differs in one block. The contributor of the VGI report manifests that this event was reported in the Police Station. Both datasets register the same type of crime.

Summarizing the process, from all the data that has been collected the attribute accuracy is possible to be analyzed only in one district and just for three events. From two of them the reports are not containing the same information, which means that they are referred to different events. For the third one, it is not possible to assure that this is the same event, however coincidences lead to the assumption it could be.

5.1.2.2 Positional accuracy

The target here is to analyze the closeness of locational VGI features in correspondence to the position of official data attributes (assumed as the true position).

To carry out such a process data points for both datasets are needed, that are referred to the equal event, which means same crime offence, same date, and same place. Otherwise would be not possible to calculate positional accuracy, the comparison must be done with records that are referred to the same incident.

After corroborating that there is not correspondence or matches in information for this purpose because reports are describing different events it is not possible to calculate and evaluate positional accuracy.

5.1.2.3 Completeness

Completeness is a measure of the absence of data and the presence of excess data. Errors resulting from the first case are called errors of omission, and errors resulting from the latter are called errors of commission (van Oort 2005).

According to the information contained in the above displayed maps (Figures 12, 14, and 16) official data surpasses in great number the reports of crime offences in the three districts.

- La Molina: 112 (OD) vs. 2 (VGI)
- San Borja: 92 (OD) vs. 15 (VGI)
- San Isidro: 84 (OD) vs. 6 (VGI)

The comparison between number of records encompassed in official data and VGI data demonstrate that the latter is clearly incomplete. VGI data lacks great amount of information, for that reason is incurring in error of omission. It is interesting to observe that La Molina is the district with the major amount of official reports and at the same time the one with less VGI reports.

Although the target here is to analyze VGI data, as official data neither contains the VGI events in its database, it is incurring in error of omission as well.

5.2 VGI INITIATIVE FOR TRANSPORTATION SERVICES

5.2.1 Data content

By the time of data collection for this study, the project had 662 reports and 928 registered users in total. However, the volunteered reports analysed are not following necessarily the structure for data collection suggested in the Web site. Contributors included issues that are not contemplated in the form structure proposed by the initiators of the site. For instance, there are many reports about tariffs or number of bus stops, which are not part of the objective of data collection (the quality of the service offered by the Metropolitano).

For reporting the problems the information has been divided by:

(i) Pedestrian Access: all problems related with infrastructure and congestion (pedestrian and from traffic) that affects the access to the buses in sidewalks, contiguous streets and stations.

(ii) Especial Access: all problems related with infrastructure, signalling and congestion (pedestrian and from traffic) that affects the access of pregnant women, elder, and disabled people to the buses in sidewalks, contiguous streets and stations.

(iii) Safety and service: all problems of signalling, information, service quality, infrastructure and vigilance personnel in sidewalks, contiguous streets and stations.

5.2.1.1 Caquetá Station (Northern corridor)

In this station VGI reports are mostly related to safety and service and the major problem indicated is delinquency. For pedestrian access all the cases demand a pedestrian bridge because passengers cannot cross to the front side of the station.

According to the field observation, street vendors cause many problems outside Caquetá station because they obstruct the entrance way and the vision from other vehicles. Moreover, there is a bridge needed to facilitate the cross over from the front side to the station. However it is difficult to build such bridge since the station is located itself over a bridge.

Observation field: If we compared the information collected in field visit and VGI data there is a coincidence in the need of infrastructure to facilitate the cross of passengers to the station. Nevertheless, it seems that for volunteers street vendors do not represent a big issue, even when they evidently generate chaos and disorder (Figure 18).

Figure 18: Street vendors obstructing pedestrian pass.

In this station we recognized many problems regarding safety and service, especially associated with absence of infrastructure, such as emergency phones, toilets, cycle paths, bicycle parking, and guides (only three guides which is insufficient for the amount of people). However, any VGI contribution mentions these identified problems. Obviously, absence of cycle paths and bicycle parking areas is a constant along the city because there is not a bicycle culture. In such a way citizens are not used to them as an important part of any modern transportation infrastructure and they do not identify the situation as a problem.

In the topic of safety and services, VGI reports indicate that the main problem here is delinquency, of course outside the station. According to our observations this could be true. The surroundings are perceived as dangerous areas by citizens and reports of theft and robbery are common especially in peak hours.

5.2.1.2 Terminal Central Grau Station (Central corridor)

In this station VGI reports (83 in total) are mostly related with safety and service. The main complain reported is about crammed buses, especially in peak hours. In such way boarding is complicated and messy. This is a problem occurring in all main stations and despite it is not contemplated in the matrix for reporting issues, it is normally mentioned by users.

Regarding pedestrian access lack of signposting is the most commented issue beside the need of a pedestrian path to enter into the station. Apparently, boarding areas and buses schedules are not notorious in this station. Moreover, volunteers declare they feel confused in following directions inside the station.

For the especial access the contributions were focused in the absence of ramps for disabled people.

Observation field: we realized that this station is underground. All crosses for passengers occurred subsurface and the problems proposed for reports collection do not apply in this case. This is a big station with a small shopping center inside, and we did not identify major problems for accessing into the station. VGI complaints are mention lack of signposting as the major problem. However, we identified clear signals for entering into the station and inside of it. Also, there were electronic tables indicating buses schedule. (Figure 19)

Figure 19: Signals indicating entrance to Grau station and map showing location of places inside.

In this station the access of all entrances is through big ramps. Inside there are elevators with the correspondent signals. However, VGI reports indicate as a problem absence of ramps. This is not accurate because we identified that access to station is via a wide ramp (Figure 20).

Figure 20: Big ramp in entrance to Grau station and elevator for disabled people.

5.2.1.3 Matellini Station (Southern corridor)

VGI reports for Matellini station (101 in total, the most numerous records) are mainly focused on safety and service. The complaints are mostly oriented to the problems caused in the feeder route services. These are alternate service routes that transport passengers from diverse locations to the Metropolitano bus stations. The idea is to assemble an interconnected system of buses which mobilize passengers towards the Metropolitano. Basically, the problems with

feeders are associated with schedules and low number of routes. It is also mentioned the nonconformity with tariffs, passengers consider these are too high. It is pertinent to mention that these two problems are not contemplated in the form for collecting data. Another problem related with safety and service is the deterioration of infrastructure such as handrails, seats and doors inside the buses.

Regarding to pedestrian access the most remarked issue is the absence of adequate signals, specially the schedule of buses. It is also considered that access to buses for boarding is difficult in peak hours. They mention also the absence of a pedestrian bridge for accessing to the station.

Observation field: Matellini is the final station of the Metropolitano southern corridor. It is connected to a bus stop (of feeder routes) through a pedestrian bridge.The pedestrian bridge is connected with one of the entrances; however for the other entrance passengers should cross the street. The problem is that there are not traffic lights (these are located approximately at 20 meters away from the cross), just a transit police who controls the pass of people (Figure 21).

Figure 21: Pedestrian cross and transit police allowing the pass.

According to information from VGI reports the main problems are absence and wrong location of signals, and absence of a pedestrian bridge. This information is truthful, the traffic lights are wrong located, is difficult to cross to the other side of the street and a pedestrian bridge would be necessary.

In services topic we found many problems. As in the other stations there are neither cycle paths nor bicycle parking places. We noticed also absence of emergency phones and guides. Likewise, there was a presence of blind point for cars outside the station (Figure 22).

Figure 22: Blind point from a car, visibility is affected by the iron fence.

Well, for VGI reports the main problem is associated with feeder routes, but this issue is not contemplated in the form for data collection. However, it is a very common complaint that is why in the reports generated by the volunteers. Volunteers' feedback sometimes contributes with implementation or improvement of the VGI sites.

6. DISCUSSION

This chapter presents the discussion regarding the two main objectives proposed in this study, a conceptual model for VGI life cycle and an analysis of the data content and data quality of VGI.

6.1 VGI life cycle conceptual model

After proposing a conceptual model for a VGI life cycle, the findings of this research confirm its applicability and functionality in explaining the process of development of two Peruvian VGI initiatives.

The model encompasses the whole process of developing a VGI project, and the components which participate in it. Therefore, it is divided in phases, activities, actors and driving forces that may intervene in the process. This is a holistic approach, for that reason when the model was applied for describing two Peruvian VGI projects, it worked satisfactory.

The leading criteria used to define the VGI life cycle are based on the concept of System Development Life Cycle, which is widely used in software development. The idea of proposing sequential phases is based on this well recognized conceptual model, and it makes use of the knowledge and experience which already is incorporated in the System Development Life Cycle.

Likewise, other studies have similar efforts for characterizing and applying conceptual models to VGI phenomena and the relationships of components. However, the criteria of these are surely different from that one used for this research. In all these studies the attention is centred in a particular element of the VGI, such as motivations of volunteers (Budhathoki 2010), motivations of users and producers (Coleman et al. 2009), or characterization and relations between stakeholders (Parker et al 2010). Anyway, the formulation of these models helped to understand the VGI landscape and formulate the life cycle conceptual model.

Notwithstanding, the VGI conceptual model is not the unique procedure for studying VGI initiatives. Possibly, different approaches could be applied for analysing it as a process. Of course, all studies have limitations. In this case, since the model is an abstraction of reality it is subject of imperfections, for instance some activities like planning –between define requirements and design- were not considered as part of the phases. Moreover, it works for the studied initiatives which belong to a specific Peruvian context, but might not work in analysing other projects.

This conceptual model can be improved in structure, for instance adding more elements such as specific division of actors and their roles, and relations between them such as connection of driving forces with the other components. This has not been done in this study because the objective was to focus more into the main structure of the model rather than deepening in roles details.

6.2 Data content and data quality analysis for two VGI initiatives

Comparing VGI data with official sources of information seemed to be the best manner to determine the content and quality of the data. Indeed, VGI data content analysis suggests that relevant information could be obtained from VGI sources. Two projects, one for criminality reports and another for transportation services issues, were analysed in content and the results showed concordance with official and field observation data.

Some studies (Flanagin and Metzger 2008) consider that is very difficult to authenticate or validate the sources of VGI, for that reason emerged problems of credibility. However this study shows that for the reviewed projects there was coincidence between the VGI content and official sources, which at least is a plus for the credibility of the shown VGI projects. Of course it is not possible to make a statement considering the credibility of VGI projects in general, because this is dependent of many factors, but the approach shown in the thesis could be used to examine VGI projects concerning their data quality and so make statements about the usefulness of the data of a certain project in a certain case.

On the other hand, for the criminality project VGI data quality analysis could not be performed for its spatial mismatching with official data. The criteria chosen for this analysis were attribute accuracy, positional accuracy and completeness. However, the spatial inconsistency between datasets did not allow to carrying out the analysis. In previous studies (Bishr and Janowicz 2010) it has been mentioned that meeting the quality assessment of data with VGI is difficult.

Perhaps, if different samples (other districts, other months) were chosen the results would be different. For instance, one of the districts analysed presented only two records for the chosen period of time, so very little could be concluded from there.

Further research could be done in investigating how other methodologies of data quality assessment could influence the results which were achieved in this research and how using a larger dataset as basis could change the results.

7. CONCLUSIONS AND RECOMMENDATIONS

Referring to the research questions in chapter 1.3 and following the conceptual model evolved in this thesis, the phases in a life cycle of a VGI initiative are inception, initiation, implementation, operation, and evolvement-maintenance-closedown, with the last phase being one out of evolvement, maintenance or closedown depending on different factors concerning activities, actors and driving-forces of the project.

These phases and components could also be found in the two Peruvian VGI initiatives. In testing the model in the two VGI initiatives it seemed workable since it was possible to analyse the developing process of the projects following the structure of the model. Each project has its own characteristics and operates in a different way. For example, they proved to be in a different final phase, the criminality initiative is placed in maintenance of the project, and the transportation VGI project in evolvement.

The third research question is referred to the content and quality of the data produced by the VGI initiatives. For both projects the data content was analysed and for the criminality project an analysis about the data quality was performed.

Data content analysis demonstrated that VGI data could provide admissible content of information. For instance, at a district level in the criminality VGI project, two districts reported the same type and modality of offences than official data (robbery from a person in snatch modality); and in the case of the transportation services project, observation field corroborated information provided by volunteers about issues in bus stops (need of a pedestrian crossing in Caquetá Station).

However, in a spatial level for data quality analysis in the case of criminality VGI project, the results were unexpected. Since there were not matches between datasets it was not possible to evaluate attribute accuracy and positional accuracy in an appropriate way. Likewise, regarding completeness it was found that VGI data is incomplete making its comparison with official data, which is a more trustable source of information, difficult. Nevertheless official data is incomplete as well, since the official sources do not register in their database the same events than VGI reports. The main inference from this it is that VGI data might be used just as a complementary source of information for official data. By doing this, it is possible to achieve a more complete general database of crime offences.

Doing the data analysis for the criminality initiative the results lead to assume that in this case VGI is misrepresenting reality, especially for the small amount of volunteered contributions. On the contrary, for transportation initiative VGI contributes with new data, enhancing the information about transportation services problematic and consequently the awareness for adequate solutions.

For the side of the research, the VGI life cycle conceptual model is applicable in the case of the two Peruvian VGI initiatives, however might not work for describing other projects from different context. For that reason, more research is needed in order to improve the model for a wider and future application.

Regarding the analysis of data quality, more geo-spatial analysis tools could be used in order to make a more quantitative inquiry, since the results presented in this thesis are more qualitative oriented.

REFERENCES

Agarwal, R. and M. Tanniru (1990). "Systems development life-cycle for expert systems." Knowledge-Based Systems. 3.

Agrios, B. and K. Mann (2010) "Getting in touch with Volunteered Geographic Information. Use a JavaScript API live sample to build a Web editing application." News ArcUser.

Artz, M. (2010) "Structure: The Key to Volunteered Geographic Information: New Tools Enable the Public to Participate in GIS Database Development." ArcWatch magazine.

Ashafri, N., J. Kuilboer, et al. (1995). "Experts systems reliability: A life cycle approach." Information & Management 28(1995): 405-414.

Bishr, M. and K. Janowicz (2008). Can we trust information? The case of Volunteered Geographic Information. Towards Digital Earth: Search, Discover and Share Geospatial Data 2010″ at Future Internet Symposium, Berlin, Germany, September 20, 2010.

Bishr, M. and W. Kuhn (2007). "Geospatial Information Bottom-Up: A Matter of Trust and Semantics." The European Information Society. Lectures in GeoInformation and Cartography Part 9.

Blume, A. and N. Mejía (2010). Hoja de ruta para un plan local de seguridad ciudadana. Lima, Instituto de Defensa Legal.

Booth, D., H. Haas, et al. (2004) "Web Services Architecture." W3C Working Group Note.

Budhathoki, N. (2010). Participants' motivations to contribute geographic information in an online community. Doctor of Philosophy, University of Illinois at Urbana-Champaign.

Castelein, W., L. Grus, et al. (2010). A characterization of Volunteered Geographic Information. 13th AGILE International Conference on Geographic Information Science 2010. Guimarães, Portugal.

Chaure, J. (2006). "Lo básico para empezar a desarrollar con Google Maps I." Manuales Retrieved March 10th, 2011, from http://www.desarrolloweb.com/articulos/2434.php.

CIDH (2009). Informe sobre seguridad ciudadana y derechos humanos. OEA, Comisión Interamericana de Derechos Humanos. Doc. 57.

Coleman, D., Y. Georgiadou, et al. (2009). "Volunteered Geographic Information: the nature and motivation of produsers*." International Journal of Spatial Data Infrastructures 11(Special Issue GSDI).

Cook, S. (2008). "The Contribution Revolution. Letting volunteers build your business." Harvard business review October 2008: 1-11.

Coot, A. and L. Rackham (2008). Neogeography data quality - is it an issue? AGI Conference. C. Ltd. Chicago, ConsultingWhere Ltd.

Craglia, M., M. Goodchild, et al. (2008). "Next-Generation Digital Earth. A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science." International Journal of Spatial Data Infrastructures Research 3: 146-167.

Dangermond, J. (2008). "GIS and the GeoWeb." ArcNews Online, from http://www.esri.com/news/arcnews/summer08articles/gis-and-geoweb.html.

Elwood, S. (2008). "Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS." GeoJournal 72(2008): 173-183.

Elwood, S. and A. Leszczynski (2010). "Privacy, reconsidered: New representations, data practices, and the geoweb." Geoforum 42(2011): 6-15.

Flanagin, A. and M. Metzger (2008). "The credibility of volunteered geographic information." GeoJournal 72: 137-148.

Fu, P. (2010) "Web GIS: Principles and Applications." ArcWatch magazine June 2010.

Goodchild, M. (2007). "Citizens as sensors: the world of volunteered geography." GeoJournal 69: 211-221.

Guptill, S. and J. Morrison (1995). Elements of Spatial Data Quality.

Haklay, M., M. Singleton, et al. (2008). "Web Mapping 2.0. The Neography of the GeoWeb." Geography Compass 2(6): 2011-2039.

IAB (2008). User Generated Content, Social Media and Advertising - An Overview. Interactive Advertising Bureau.

IDL (2010). Informe Anual 2010 Sobre Seguridad Ciudadana

El legado del gobierno de Alan García: Un país más inseguro. Lima, Instituto de Defensa Legal. Primera edición.

IOP (2009). Estado de la Opinión Pública. Temas de Seguridad. Lima, Instituto de Opinión Pública. Año VI.

IWS. (2011). "Internet Growth Statistics. Today's road to e-Commerce and Global Trade Internet." Retrieved July 2011, 2011, from http://www.internetworldstats.com/.

Jovieau, T. (2011). "Neogeography: an extension of mainstream geography for everyone made by everyone?" Journal of Location Based Service 3(2): 75-81.

Longueville, B. D. (2010). "Community-based geoportals: The next generation? Concepts and methods for the geospatial Web 2.0. Computer." Environment and Urban Systems 34(2010): 299-308.

Longueville, B. D., N. Ostländer, et al. (2009). "Addressing vagueness in Volunteered Geographic Information (VGI) - A case study." International Journal of Spatial Data Infrastructures(Special Issue GSDI-11).

Maximilien, M., H. Wilkinson, et al. (2007). A Domain-Specific Language for Web APIs and Service Mashups. International Conference on Service Oriented Computing. Springer-Verlag. Berlin, ICSOC 2007: 13-26.

Modell, M. (2007). "A Professional's guide to System Analysis. Second Edition." from http://www.martymodell.com/pgsa2/pgsa03.html.

O'Connor, E. (2009). "The Web platform: what it is." Retrieved March 15th, from http://edward.oconnor.cx/2009/05/what-the-web-platform-is.

Oort, P. v. (2005). "Spatial data quality: from description to application." Geodesy 60.

O'Reilly, T. (2005). "What is Web 2.0." Web 2.0, from http://oreilly.com/web2/archive/what-is-web-20.html.

Orestein, D. (2000). "QuickStudy: Application Programming Interface (API)." App Development Topic Center Retrieved March 20th, 2011, from http://www.computerworld.com/s/article/43487/Application_Programming_Interface.

Parker, C., A. May, et al. (2010). An exploration of Volunteered Geographic Information stakeholders. GIS Research UK 18th Annual Conference, GISRUK 2010, University College London, 14th–16th April 2010., London : University College London.

Peng, Z. and M. Tsou (2003). Internet GIS. Geographic information services for the internet and wireless network. New Jersey, John Wiley & Sons Inc.

Schade, S., G. Luraschi, et al. (2010). Citizens as sensors for crisis events: Sensor Web enablement for Volunteered Geographic Information. 1st International Workshop on Pervasive Web Mapping, Geoprocessing and Services, Como, Italy.

Sieber, R. and H. Rahemtulla (2010). Model of Public Participation on the GeoWeb. Sixth international conference on Geographic Information Science, Zurich.

Stefanou, C. (2003). System Development Life Cycle.

Sui, D. (2008). "The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS." Computers, Environment and Urban Systems 32(2008): 1-5.

Tanrıöver, Ö. and S. Bilgen (2010). "A framework for reviewing domain specific conceptual models." Computer Standards & Interfaces 33((2011)): 448–464.

Tulloch, D. (2008). "Is VGI participation? From vernal pools to video games." GeoJournal 72: 161-171.

Warf, B. and D. Sui (2010). "From GIS to neogeography: ontological implications and theories of truth." Annuals of GIS 16(Taylor & Francis): 197-209.

Wolf, E. (2010) "Neogeography." Do it yourself GIScience.

Yu, J., B. Benatallah, et al. (2008) "Understanding Mashup Development." IEEE Internet Computing.

APPENDICES

APPENDIX 1- VGI crime reports per district 2010

Table 2: Reports for La Molina 2010

Id	ADDRESS	YEAR	MONTH	TYPE OF CRIME	MODALITY	STATUS	REPORT	GENDER
3	FLORA TRISTAN /M.BASTIDAS	2010	JUNE	CAR THEFT	PARKED	VICTIM	YES	M
6	DE LA RECOLETA	2010	JULY	ALERT	SNATCH	WITNESS	NO	-

Table 3: Reports for San Borja 2010

Id	ADDRESS	YEAR	MONTH	TYPE OF CRIME	MODALITY	STATUS	REPORT	GEN DER
1	PTE P. J. PRADO/ CIRCUNVALACION	2010	JULIO	CONCRETED ROBBERY	MUGGERY	VICTIM	NO	M
2	ENRIQUE OLIVERO	2010	JULIO	CAR ROBBERY	FIREARM ASSAULT	VICTIM	YES	F
4	JULIO BAYLETTI	2010	JULIO	ALERT	PARKED CAR	WITNESS	NO	
6	ALAMEDA D. ALOMIA ROBLES	2010	JULIO	ALERT	HOUSING THEFT	WITNESS	NO	M
7	JAVIER PRADO ESTE	2010	JULIO	CONCRETED ROBBERY	SNATCH	VICTIM	NO	M
9	J. PRADO/AVIACION	2010	JULIO	ALERT	SNATCH	WITNESS	NO	
11	BOULEVARD	2010	JUNIO	CONCRETED ROBBERY	FIREARM ASSAULT	VICTIM	YES	M
13	VERROCHIO/BRONSINO	2010	JULIO	CONCRETED ROBBERY	FIREARM ASSAULT	VICTIM	YES	M
15	CAVALLINI	2010	JULIO	CAR ROBBERY	FIREARM ASSAULT	WITNESS	NO	M
16	SAN LUIS	2010	JULIO	ALERT	ASSAULT	WITNESS	NO	
18	AVIACION	2010	JUNIO	CAR THEFT	PARKED CAR	VICTIM	NO	M
21	SAN BORJA NORTE/LAS ARTES	2010	JULIO	ALERT	NOT SPECIFIED	WITNESS	NO	
22	GUARDIA CIVIL	2010	JULIO	ALERT	NOT SPECIFIED	WITNESS	NO	
25	PUENTE PRIMAVERA	2010	JUNIO	CONCRETED ROBBERY	EXTRACTION	VICTIM	NO	M
31	SAN BORJA SUR	2010	JUNIO	CAR THEFT	PARKED CAR	VICTIM	NO	M
33	GUARDIA CIVIL	2010	MAYO	CONCRETED ROBBERY	SNATCH	VICTIM	NO	M
35	GALVEZ BARRENECHEA	2010	JULIO	CONCRETED ROBBERY	SNATCH	WITNESS	NO	M

Table 4: Reports for San Isidro 2010

Id	ADDRESS	YEAR	MONTH	TYPE OF CRIME	MODALITY	STATUS	GENDER	REPORT
4	PALMERAS/J. PRADO OESTE	2010	JUNE	ALERT	SNATCH	WITNESS		NO
8	JAVIER PRADO/LAS BEGONIAS	2010	JULY	ALERT	SNATCH	WITNESS		NO
9	J.PRADO (PTE PEATONAL VIA EXPRESA)	2010	JULY	CONCRETED ROBBERY	SNATCH	WITNESS	M	NO

12	J. PRADO/VIA EXPRESA (CE SAN AGUSTIN)	2010	JUNE	ALERT	ASSAULT	WITNESS		NO
14	GALVEZ BARRENECHEA	2010	JUNE	ALERT	CAR THEFT	WITNESS		NO
20	BEGONIAS/R. NAVARRETE	2010	JUNE	ALERT	SNATCH	WITNESS		NO
21	AREQUIPA	2010	JULY	ALERT	CAR THEFT	WITNESS		NO
23	CURA BEJAR	2010	MAY	CAR THEFT	PARKED	VICTIM	M	SI
24	CAMINO REAL	2010	JULY	ALERT	SNATCH	WITNESS		NO
25	EL GOLF	2010	JUNE	ALERT	SNATCH	WITNESS		NO

APPENDIX 2- VGI crime reports vs. official data per district 2010

Table 5: Number of reports and major modality of offences from the district La Molina

Month	Nr reports	Major Type crime	Modality
MAY	36	Car theft	Parked
JUNE	35	Robbery from person	Firearm
JULY	41	Robbery from person	Firearm

Table 6: Number of reports and major modality of offences from the district San Borja

Month	Nr reports	Major Type crime	Modality
MAY	25	Robbery from person	Snatch
JUNE	32	Robbery from person	Firearm/snatch
JULY	35	Robbery from person	Snatch

Table 7: Number of reports and major modality of offences from the district San Isidro

Month	Nr reports	Major Type crime	Modality
MAY	24	Robbery from person	Snatch
JUNE	27	Robbery from person	Snatch
JULY	33	Robbery from person	Snatch

APPENDIX 3- Overlapping between crime reports and official data in San Borja July 2010

Table 8: Match 1 between Official data and VGI data

DATA	ID	ADDRESS	NR	TYPE OF CRIME	MODALITY	STATUS	REPORT
OD	22	AV. GUARDIA CIVIL FARMACIA BTL	301	ROBBRY	ASSAULT	DOES NOT APPLY	DOES NOT APPLY
VGI	22	GUARDIA CIVIL	3	ALERT	ALERT	WITNESS	NO

Table 9: Match 2 between Official data and VGI data.

DATA	ID	ADDRESS	NR	TYPE OF CRIME	MODALITY	STATUS	REPORT
OD	1	AV. SAN BORJA NORTE/ AV. SAN LUIS		THEFT	THEFT	DOESN'T APPLY	DOESN'T APPLY
VGI	16	AV. SAN LUIS	22	ALERT	ASSAULT	WITNESS	NO

Table 10: Match 3 between Official data and VGI data.

DATA	ID	ADDRESS	NR	TYPE OF CRIME	MODALITY	STATUS	REPORT
OD	28	AV. BRONSINO/AV. ROUSSEAU		ROBBERY FROM PERSON	ASSAULT WITH FIREARM	DOES NOT APPLY	DOES NOT APPLY
VGI	13	AV. BRONSINO/ VERROCHIO		ROBBERY	ASSAULT	VICTIM	YES

APPENDIX 4- VGI transportation service reports per station

Table 11: Number of contributions for Caquetá Station

CAQUETA STATION	Nr.	MAIN PROBLEM
Safety and service	9	Delinquency
Pedestrian access	6	Absence of pedestrian bridge
Especial access	1	(Wrong information)
TOTAL CONTRIBUT	**16**	

Table 12: Number of contributions for Central Grau Station

GRAU STATION	Nr.	MAIN PROBLEM
Safety and service	55	Crammed buses
Pedestrian access	21	Signposting
Especial access	7	Absence of ramps
TOTAL CONTRIBUT	**83**	

Table 13: Number of contributions for Matellini Station

MATELLINI STATION	Nr.	MAIN PROBLEM
Safety and service	70	Feeder routes/Infrastructure deterioration
Pedestrian access	24	Lack of signposting
Especial access	7	(Wrong information)
TOTAL CONTRIBUT	**101**	

APPENDIX 5- Observation field matrix per station

Table 14: Field observation matrix for pedestrian access - Caquetá Station (Outside)

PROBLEM	OPTION	Yes*	No**	DESCRIPTION
Absence of	Sign		x	There is a sign for crossing through pedestrian cross
	Zebra crossing		x	There is zebra crossing
	Bridge	x		However, is not possible to build
	Sidewalk			There is sidewalk
Narrowness of	Bridge		x	Ok
	Sidewalk		x	Ok
Wrong location of	Infrastructure elements	x		Light post and electronic table next to pedestrian cross
	Objects		x	Does not apply
	Traffic lights		x	Right located
	Sign		x	Right located
Obstruct the way	Street vendors	x		At the end of the north cross and in both sides there are lots of street vendors.
	Vehicles		x	High traffic jam however they do not obstruct the way
	Infrastructure elements		x	Right located
Obstruct the vision	Advertising panels		x	
	Objects		x	
	Infrastructure elements		x	
	Street vendors	x		Obstruct the transit before crossing to the other side of the street

*Yes: a problem is identified

**No: any problem is identified

Table 15: Field observation matrix for pedestrian access - Caquetá Station (Inside)

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Sign		x	There is a sign for crossing through pedestrian cross
	Zebra crossing			Does not apply
	Bridge			Does not apply
	Sidewalk		x	There is sidewalk
Narrowness of	Bridge			Does not apply
	Sidewalk		x	Ok, however in peak hours queues are too long and people cannot circulate
Wrong location of	Infrastructure elements	x		The ticket machine and the ticket window are located one in front of the other.
	Objects		x	
	Traffic lights		x	Does not apply
	Sign		x	
Obstruct the way	Street vendors		x	They cannot enter into the station
	Vehicles		x	Does not apply
	Infrastructure elements		x	
Obstruct the vision	Advertising panels		x	There are panels but they are small and do not obstruct vision
	Objects		x	
	Infrastructure elements		x	

Table 16: Field observation matrix for especial access - Caquetá Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Wrong location of	Signal		x	
	Ramp		x	Deterioration of existent ramps and some are too steep
Narrowness of	Sidewalk		x	Ok
	Bridge			Does not apply
	Ramp		x	Wide
Obstruct the way	Infrastructure elements		x	
	Vehicles		x	
	Street vendors	x		They obstruct the way
Absence	Signal		x	Not signal outside the station, but inside there are signals for disabled people
	Recognizing places		x	
	Ramp		x	
Short time for crossing	Traffic light		x	

Table 17: Field observation matrix for safety and service - Caquetá Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Illumination		x	Inside the station is ok
	Emergency phones	x		No emergency phones, just public phones
	WC	x		No
	Cycle paths	x		No
	Bicycle parking	x		No
	Ticket machines		x	Ok
	Guides	x		Not enough personnel for such a big station
Presence of	Delinquents		x	Not inside, however outside is a very dangerous area
	Homeless		x	No
	Dangerous bend		x	No
	Blind point for cars	x		Outside the station other public transportation buses stop in forbidden areas, this could block visibility for car drivers
High speed	Vehicles		x	
Deterioration of	Signals		x	All news and in good state

Table 18: Field observation matrix for pedestrian access – Central Grau Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Sign		x	
	Zebra crossing			Does not apply
	Bridge			Does not apply
	Sidewalk		x	Underground entrance
Narrowness of	Bridge			Does not apply
	Sidewalk		x	Ok, narrow enough
Wrong location of	Infrastructure elements		x	
	Objects		x	
	Traffic lights			Does not apply

	Sign		x	
Obstruct the way	Street vendors		x	
	Vehicles		x	Does not apply
	Infrastructure elements		x	
Obstruct the vision	Advertising panels		x	
	Objects		x	
	Infrastructure elements		x	

Table 19: Field observation matrix for especial access – Central Grau Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Wrong location of	Signal		x	
	Ramp		x	
Narrowness of	Sidewalk		x	Wide enough
	Bridge			Does not apply
	Ramp		x	Wide enough
Obstruct the way	Infrastructure elements		x	
	Vehicles			Does not apply
	Street vendors		x	
Absence	Signal		x	The access areas and entrance have signals for disabled people
	Recognizing places		x	
	Ramp		x	
Short time for crossing	Traffic light		x	

Table 20: Field observation matrix for safety and service – Central Grau Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Illumination		x	Ok
	Emergency phones	x		No (however inside there is a medical center)
	WC		x	
	Cycle paths	x		No
	Bicycle parking	x		No
	Ticket machines		x	Ok
	Guides		x	Guides in the entrance and in the boarding platform
Presence of	Delinquents		x	No
	Homeless		x	No
	Dangerous bend	x		Dangerous bend for the buses before entering the station. However they transit this area in low speed
	Blind point for cars		x	Does not apply
High speed	Vehicles		x	Does not apply
Deterioration of	Signals		x	All news and in good state

Table 21: Field observation matrix for pedestrian access – Matellini Station (outside)

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Sign		X	
	Zebra crossing		X	Ok

	Bridge	x		No pedestrian bridges
	Sidewalk		X	Ok
Narrowness of	Bridge			Does not apply
	Sidewalk		X	Ok
Wrong location of	Infrastructure elements		X	
	Objects		X	
	Traffic lights		X	Ok
	Sign		X	
Obstruct the way	Street vendors		X	
	Vehicles		X	
	Infrastructure elements		X	
Obstruct the vision	Advertising panels		X	
	Objects		X	
	Infrastructure elements	x		Iron fence that obstructs the vision from the cars that are coming

Table 22: Field observation matrix for pedestrian access – Matellini Station (inside)

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Sign		x	
	Zebra crossing			Does not apply
	Bridge			Does not apply
	Sidewalk		x	
Narrowness of	Bridge		x	There is a bridge which connects the station with the bus stop of feeder routes buses. It is wide enough
	Sidewalk		x	
Wrong location of	Infrastructure elements		x	
	Objects		x	
	Traffic lights			Does not apply
	Sign		x	
Obstruct the way	Street vendors		x	
	Vehicles			Does not apply
	Infrastructure elements		x	
Obstruct the vision	Advertising panels		x	There are panels but well positioned.
	Objects		x	
	Infrastructure elements		x	

Table 23: Field observation matrix for especial access – Matellini Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Wrong location of	Signal		x	
	Ramp		x	
Narrowness of	Sidewalk		x	Wide enough
	Bridge		x	The pedestrian bridge is wide enough
	Ramp		x	The general entrance in through a wide ramp
Obstruct the way	Infrastructure elements		x	
	Vehicles		x	The transit police controls vehicles
	Street vendors		x	No
Absence	Signal		x	
	Recognizing places		x	

	Ramp		x	
Short time for crossing	Traffic light		x	The transit police controls traffic and crossing

Table 24: Field observation matrix for safety and service – Matellini Station

PROBLEM	OPTION	Yes	No	DESCRIPTION
Absence of	Illumination		x	Inside the station is ok
	Emergency phones	x		No
	WC		x	Ok
	Cycle paths	x		No
	Bicycle parking	x		No
	Ticket machines		x	Ok
	Guides	x		Guides in the entrance and in the boarding platform
Presence of	Delinquents		x	No
	Homeless		x	No
	Dangerous bend		x	
	Blind point for cars	x		Outside station there is an iron fence that obstructs vision for car drivers, especially when passengers go out of the station
High speed	Vehicles	x		Before entering station buses do U-turns, the area is very close to pedestrian entrances and it could be risky for pedestrians.
Deterioration of	Signals			All news and in good state