UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

An approach for assessing the quality of crowdsourced geographic information in the flood management domain

Lívia Castro Degrossi

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Livia Castro Degrossi

An approach for assessing the quality of crowdsourced geographic information in the flood management domain

Doctoral dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. João Porto de Albuquerque Pereira

Co-advisor: Profa. Dra. Renata Pontin de Mattos Fortes

USP – São Carlos
November 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

D321a Degrossi, Livia Castro
An approach for assessing the quality of
crowdsourced geographic information in the flood
management domain / Livia Castro Degrossi;
orientador João Porto de Albuquerque Pereira;
coorientadora Renata Pontin de Mattos Fortes. --
São Carlos, 2019.
143 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2019.

1. Crowdsourcing Geographic Information. 2.
Citizen Observatory. 3. Quality Assessment. 4.
Flood Management. I. Porto de Albuquerque Pereira,
João, orient. II. Pontin de Mattos Fortes, Renata,
coorient. III. Título.

Livia Castro Degrossi

**Uma abordagem para a avaliação da qualidade de
informações geográficas voluntárias no domínio de gestão
de inundação**

Tese apresentada ao Instituto de Ciências
Matemáticas e de Computação – ICMC-USP,
como parte dos requisitos para obtenção do título
de Doutora em Ciências – Ciências de Computação e
Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e
Matemática Computacional

Orientador: Prof. Dr. João Porto de
Albuquerque Pereira

Coorientador: Profa. Dra. Renata Pontin de
Mattos Fortes

USP – São Carlos
Novembro de 2019

To my parents, Suely C. de Castro Degrossi and Homero Luiz Degrossi,

To my sister, Marília Castro Degrossi,

To all people who know what a PhD is.

ACKNOWLEDGEMENTS

The PhD journey is arduous and I would never have been able to conclude it without the guidance of my advisors and support from my family and friends. Throughout my journey, many people have helped me to pursue my dream and I will be forever grateful to them.

I am truly grateful to my advisor, Professor João Porto de Albuquerque, for all your teachings in the past few years. Without your guidance, I would not have become the researcher I am today. During all those years, you have been my model of researcher and teacher. I hope to work with you in the near future.

I also would like to thank my co-advisor, Professor Renata Pontin de Mattos Fortes, for having welcomed me into your research group and all support you have given me in the most difficult moments of my journey. You have shown me another side of computer science, you have shown me the human behind the machines.

I would like to express my gratitude to my family for always believing that I was capable. You have always supported my love for science and my dream of becoming a researcher.

This work would not be possible without the help of my friends from the Laboratory of Software Engineering (LabES) who have shared with me their valuable knowledge and ideas. You have been my family for all these years. I am sincerely grateful to Alinne Corrêa, Francisco Carlos, Stevão Andrade, Faimison Porto, Flávio Horita, Misael Júnior, Brauner Oliveira, Ricardo Vilela, Rachel Reis, Kamila Lyra, Diego Damasceno, Sidgley Andrade, João Biazoto, João Choma, Danilo Reis and Claudinei Junior. For sure I will forget some names.

To my friends from the Laboratory of Bioinspired Computing (Biocom) who have welcomed me so well in the laboratory. Thank you Daniel Tozadore, Victor Hugo Barella, Victor Padilha, Rafael Mantovani, Adam Henrique, Jefferson Oliva, William Rosa, Saulo Mastelini, Edésio Alcobaça, Everlândio, Caetano Ranieri, Murilo Rehder and Daniel Cestari. You will be forever in my heart.

To my friends from the University of Heidelberg who have helped me during my stay abroad. Thank you Bettina Knorr, Carolin Klonner, Melanie Eckle, Amin Mobasher, Rui Nunes, Nikos Papapesios, Rene Westerholt, Adam, Dirk, Luisa, Chiao-Ling, Enrico, Sabrina, Amandus and Martin. Germany would not have been the same without you.

Special thanks to all professors who contributed to my education.

To all the employees of ICMC, your work is very important for us.

I would like to acknowledge funding provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Grant No. 88887.091744/2014-01).

“If I have seen further, it is by standing upon the shoulders of giants”.
(Isaac Newton)

RESUMO

DEGROSSI, L. C. **Uma abordagem para a avaliação da qualidade de informações geográficas voluntárias no domínio de gestão de inundação.** 2019. 143 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Informação geográfica de *crowdsourcing* (do inglês, CGI) consiste numa informação geográfica fornecida por não especialistas de maneira “ativa/consciente” e “passiva/inconsciente”. O uso de CGI no domínio da gestão de inundações é consideravelmente recente e tem sido motivado pelo seu potencial como fonte de informação geográfica em situações em que dados oficiais são escassos ou indisponíveis. Contudo, a qualidade desse tipo de informação é uma preocupação fundamental quando a utilizamos, visto que os cidadãos podem ter diferentes níveis de conhecimento e experiência. A usabilidade das plataformas de *crowdsourcing* é um ponto importante visto que pode impactar a qualidade de CGI, uma vez que o aumento da complexidade desses sistemas pode levar o cidadão ao fornecimento de informações errôneas ou imprecisas. Embora aspectos de usabilidade sejam cada vez mais discutidos entre projetistas e desenvolvedores de sistemas computadorizados, ainda há uma escassez de estudos que investiguem estratégias para o aprimoramento da usabilidade de plataformas de *crowdsourcing*. A avaliação da qualidade de CGI é outro ponto importante para determinar se a informação geográfica é adequada a um propósito específico. Na literatura, a avaliação da qualidade de CGI é realizada para cada CGI individualmente. Em situações de crise, contudo, há pouco tempo para analisar uma grande quantidade de dados e, portanto, minimizar a sobrecarga de informações é extremamente importante. Uma estratégia interessante e pouco explorada é a avaliação da qualidade dos elementos CGI agregados, ao invés de um único elemento. Esta tese de doutorado propõe uma abordagem para a melhoria e avaliação da qualidade de CGI no domínio de gestão de inundações. A abordagem consiste em uma taxonomia de métodos para a avaliação da qualidade de CGI na ausência de dados oficiais, um método para avaliar a qualidade do CGI e uma interface para o Observatório Cidadão de Enchentes. Os resultados obtidos na avaliação das principais contribuições revelam que o método proposto pode explicar a qualidade de CGI e a usabilidade da nova interface melhorou.

Palavras-chave: Informação Geográfica Voluntária, Observatório Cidadão, Avaliação da Qualidade, Gestão de Inundação.

ABSTRACT

DEGROSSI, L. C. **An approach for assessing the quality of crowdsourced geographic information in the flood management domain.** 2019. 143 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Crowdsourced Geographic Information (CGI) encompasses both “active/conscious” and “passive/unconscious” georeferenced information generated by non-experts. The use of CGI in the domain of flood management is considerably recent and has been motivated by its potential as source of geographic information in situations where authoritative data is scarce or unavailable. Given that citizens may vary greatly in knowledge and expertise, the quality of such information is a key concern when making use of CGI. Moreover, the usability of the crowdsourcing platforms is another critical point that impacts the quality of CGI, since increasing complexity of such systems can lead to the provision of erroneous or inaccurate information. Although usability aspects have been increasingly discussed among designers and developers of computerized systems, there is a lack of studies that investigate strategies for the enhancement of the usability of crowdsourcing platforms. In this perspective, the assessment of CGI quality is an important step to determine if the information fits a specific purpose. A common way of assessing the quality of CGI gathered by crowdsourcing platforms is the evaluation of each CGI item. However, in crisis situations, there is short time to scrutinize a great amount of data and, therefore, minimizing information overload is critically important. An interesting, but poorly explored, strategy is the assessment of the quality of aggregated CGI elements, instead of a single one. This doctoral thesis proposes an approach for the improvement and assessment of CGI quality in the domain of flood management. It describes a taxonomy of methods for the assessment of CGI quality in the absence of authoritative data, as well as proposes a method for evaluating the quality of CGI and a new interface for the Citizen Observatory of Floods. Results obtained in the evaluation of the main contributions reveal that the method can explain the quality of CGI and the usability of the new interface increased.

Keywords: Crowdsourced Geographic Information, Citizen Observatory, Quality Assessment, Flood Management.

LIST OF FIGURES

Figure 1 – Conceptual model for the assessment of CGI quality.	30
Figure 2 – Architecture of the AGORA approach.	32
Figure 3 – Levels of citizen engagement.	38
Figure 4 – Classification schema of CGI projects.	44
Figure 5 – Steps of the SLR.	46
Figure 6 – SLR process and the number of included and excluded studies in each step. .	49
Figure 7 – Development process of the taxonomy.	50
Figure 8 – Study area.	71
Figure 9 – Spatial units of analysis.	73
Figure 10 – Distribution of on-topic tweets.	79
Figure 11 – Distribution of flood events according to CGE.	80
Figure 12 – Extent of flood-prone area per spatial unit.	80
Figure 13 – Area of flood-control reservoirs per spatial unit.	81
Figure 14 – Old interface of the Citizen Observatory of Floods.	92
Figure 15 – Steps of the user-centered design process.	94
Figure 16 – Steps of the methodology.	94
Figure 17 – Most used Social Media.	98
Figure 18 – Systems employed by users to search rain- or flood-related information. . .	99
Figure 19 – Types of information users would share.	100
Figure 20 – Hierarchical Task Analysis of the old interface.	102
Figure 21 – Hierarchical Task Analysis of the new interface.	104
Figure 22 – New interface of the Citizen Observatory of Floods.	105
Figure 23 – New system interface after heuristic evaluation.	108

LIST OF CHARTS

Chart 1 – Data quality elements for the measurement of geographic data quality.	39
Chart 2 – Search string.	47
Chart 3 – Electronic Databases.	47
Chart 4 – The inclusion and exclusion criteria employed for the qualitative review. . . .	48
Chart 5 – Objective ending conditions.	51
Chart 6 – Subjective ending conditions.	51
Chart 7 – Dimensions and characteristics of the taxonomy.	53
Chart 8 – Examples of on-topic and off-topic tweets.	75
Chart 9 – Design-science research method.	93
Chart 10 – The proposed tasks to be performed by the evaluators.	96

LIST OF TABLES

Table 1 – Data quality elements for the measurement of CGI quality.	39
Table 2 – Summary of the quality assessment methods of the taxonomy.	54
Table 3 – Taxonomy of quality assessment methods.	55
Table 4 – Previous studies on the assessment of CGI quality in the flood management domain.	67
Table 5 – Hypotheses for the description of CGI quality.	69
Table 6 – Plausibility indicators for the description of CGI quality.	70
Table 7 – Degree of correlation among the plausibility indicators.	81
Table 8 – Coefficients and p-values of the plausibility indicators.	84
Table 9 – Requirements of the Citizen Observatory of Flood based on user needs. . . .	100
Table 10 – Problems identified by evaluators.	103
Table 11 – Frequency of violated usability heuristics.	107
Table 12 – Description of participant profiles.	109
Table 13 – Participants answers in the SUS questionnaire.	111

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
ATKIS	German Authority Topographic-Cartographic Information System
CGI	Crowdsourced Geographic Information
GIS	Geographic Information Science
GPS	Global Positioning System
HE	Heuristic Evaluation
HTA	Hierarchical Task Analysis
MAUP	Modifiable Areal Unit Problem
POI	Point-of-Interest
PPGIS	Public participation GIS
RQ	Research Question
SLR	Systematic Literature Review
SRI	Surface Rain Intensity
SSA	Secure Situation Awareness System
SUS	System Usability Scale
UCD	User-centered Design
UGC	User-generated Content
UTC	Coordinated Universal Time
VGI	Volunteered Geographic Information
VIF	Variance Inflation Factor
WoS	Web of Science

CONTENTS

1	INTRODUCTION	27
1.1	Contextualization	27
1.2	Problem Statement and Motivation for the Research	28
1.3	Objectives	29
1.4	An Approach for Quality Assessment	30
1.5	Research Context	31
1.6	Thesis Outline	32
2	BACKGROUND	35
2.1	Overview	35
2.2	Crowdsourced Geographic Information	35
2.3	Quality of Crowdsourced Geographic Information	37
2.4	Final Remarks	42
3	A TAXONOMY OF QUALITY ASSESSMENT METHODS	43
3.1	Overview	43
3.2	Related Work	43
3.3	Methodology	45
3.3.1	<i>Systematic Literature Review</i>	45
3.3.2	<i>Development of the Taxonomy</i>	48
3.4	Taxonomy of Quality Assessment Methods	53
3.4.1	<i>Geographic context analysis</i>	53
3.4.2	<i>Redundancy of volunteered contributions</i>	56
3.4.3	<i>Scoring volunteered contributions</i>	56
3.4.4	<i>Expert assessment</i>	57
3.4.5	<i>Automatic location checking</i>	57
3.4.6	<i>Spatiotemporal clustering</i>	58
3.4.7	<i>Volunteer profile; reputation</i>	58
3.4.8	<i>Error detection/correction by crowd</i>	59
3.4.9	<i>Extraction/learning of characteristics</i>	59
3.4.10	<i>Ranking/filtering by linguistic terms</i>	60
3.4.11	<i>Historical data analysis</i>	60
3.5	Discussion	61

3.6	Final Remarks	63
4	METHOD FOR QUALITY ASSESSMENT	65
4.1	Overview	65
4.2	Related Work	65
4.3	Quality Assessment Method	68
4.4	Case Study and Datasets	69
4.4.1	<i>Study area</i>	69
4.4.2	<i>Datasets</i>	71
4.5	Methodology	72
4.5.1	<i>Determination of the spatiotemporal resolution</i>	72
4.5.2	<i>Obtaining rainfall data from weather radar</i>	73
4.5.3	<i>Data preparation</i>	74
4.5.4	<i>Measurement of plausibility indicators</i>	75
4.5.5	<i>Statistical Analysis</i>	77
4.6	Results	78
4.6.1	<i>Data description</i>	78
4.6.2	<i>Statistical analysis</i>	81
4.7	Discussion	83
4.8	Final Remarks	86
5	CITIZEN OBSERVATORY OF FLOODS	89
5.1	Overview	89
5.2	Related Work	89
5.3	Citizen Observatory of Floods	90
5.4	Methodology	91
5.4.1	<i>User-centered Design</i>	93
5.4.1.1	<i>Requirement Analysis</i>	93
5.4.1.2	<i>Design and Prototype</i>	95
5.4.1.3	<i>Evaluation</i>	96
5.5	Results and Discussion	97
5.5.1	<i>Understanding users</i>	97
5.5.2	<i>Design and Prototyping</i>	100
5.5.3	<i>Heuristic Evaluation</i>	101
5.5.4	<i>User-based evaluation</i>	107
5.6	Final Remarks	110
6	CONCLUSIONS	113
6.1	Thesis Contributions	113
6.2	Limitations and Future Work	114

BIBLIOGRAPHY 117

**APPENDIX A DECLARATION OF ORIGINAL AUTHORSHIP AND
LIST OF PUBLICATIONS 127**

**APPENDIX B QUESTIONNAIRE FOR USER PROFILE CHARAC-
TERIZATION 131**

APPENDIX C QUESTIONS OF INTERVIEW 135

**APPENDIX D QUESTIONNAIRE FOR USER PROFILE CHARAC-
TERIZATION (USABILITY TEST) 137**

APPENDIX E SYSTEM USABILITY SCALE 141

INTRODUCTION

1.1 Contextualization

Over the past decades, significant advancements on Web 2.0 and mobile technologies have resulted in changes in the way data are created. Such advancements have promoted a growth in User-generated Content (UGC) (DAUGHERTY; EASTIN; BRIGHT, 2008), since common citizens (users) went from data consumers to data providers. Therefore, citizens are nowadays central to the process of creating, sharing, consuming and disseminating data (ANTONIOU; MORLEY; HAKLAY, 2010). The introduction of Global Positioning System (GPS) receivers in mobile devices has increased the amount of geographic information provided by citizens; they are providing data that were exclusively produced by authoritative agencies (SEE *et al.*, 2016).

Crowdsourced Geographic Information (CGI) is becoming an important source of information in many domains. The use of CGI in the domain of flood management is considerably recent and has been motivated by the potential of CGI to be a source of geographic information in situations where authoritative data is scarce or unavailable. CGI has gained special attention in this domain due to its potential for the detection of rain and flood events (ANDRADE *et al.*, 2017; LONGUEVILLE *et al.*, 2010), improvement of flood forecast and monitoring (MAZZOLENI *et al.*, 2017; RESTREPO-ESTRADA *et al.*, 2018; FAVA *et al.*, 2018), and complementing of authoritative data (LANFRANCHI *et al.*, 2014; DEGROSSI *et al.*, 2014). The use of CGI has grown owing to a number of key features, e.g., it is free, updated and abundant, and the absence and obsolescence of authoritative data in many parts of the world. Nevertheless, the quality of the information is a key concern when making use of CGI, since citizens may vary greatly in knowledge and expertise and have no formal training or cartographic skills. Concerns regarding its quality, for instance, have hampered its adoption by emergency agencies, e.g., Civil Defense.

Information quality is an important feature of any type of information, since it is crucial to its effective use (DEVILLERS; JEANSOULIN, 2006). Information is usually considered

to be of poor quality if it is incomplete, ambiguous, inaccurate or inconsistent, or has typos and misspellings. The quality of CGI has become a very popular topic amongst academics and researchers (ANTONIOU; MORLEY; HAKLAY, 2010) owing to its idiosyncrasies; CGI is “often regarded as insufficiently structured, documented, and validated to be a reliable source of information” (LONGUEVILLE *et al.*, 2010). The information that is supplied by non-experts does not have mostly to comply with any quality standards, and there is no control over the creation process. Quality assessment thus becomes an important step to understand if the information is fit-for-purpose with regard to the way it will be used (BALLATORE; ZIPF, 2015).

Quality assessment aims at verifying the level of concordance of CGI to a quality criterion¹. Nowadays, several methods have been proposed for the assessment of CGI quality (see Chapter in the literature 3). These methods differ with regard to the type of CGI evaluated (e.g., social media, crowd sensing and collaborative mapping), reference data types (e.g., methods can be classified as either extrinsic or intrinsic), among other factors.

1.2 Problem Statement and Motivation for the Research

The assessment of CGI quality is an important step that any initiative willing to use CGI must perform. Critical factors that influence the quality of CGI are the knowledge of citizens and their level of engagement, the crowdsourcing-based platforms used to provide the data, the variety of data structures, and the lack of control over the creation process (BORDOGNA *et al.*, 2016). Given that the provision of information occurs when citizens interact with the crowdsourcing platform, the characteristics of the system interface, e.g., its usability, may influence on the quality of CGI. Therefore, the development of strategies for the improvement of CGI quality is important by, for instance, controlling the creation of CGI, evaluating citizens’ reputation, assessing CGI quality, and improving the quality of crowdsourcing-based platforms.

The development of crowdsourcing-based platforms in the domain of disaster management has grown in the past few years (see Section 5.2). Nevertheless, little attention has been paid to the usability of those platforms, since most of them are designed in *ad hoc* manner. The enhancement of their usability, however, can be a way of improving the quality of CGI. The increase of the system interface complexity can lead to the provision of erroneous or inaccurate information by citizens and, as a consequence, affect the quality of the information. Although usability aspects have been increasingly discussed among designers and developers of computerized systems, there is a lack of studies that investigate strategies for the enhancement of the usability of crowdsourcing-based systems.

The assessment of the quality of CGI gathered by crowdsourcing platforms is an important step for the removal of low quality information. A critical point of this assessment is the context in which the information is applied, since information quality is determined by its

¹ An element to describe a certain aspect of CGI quality

“fitness for use” within this context (BORDOGNA *et al.*, 2016). In the literature, there are several strategies for the assessment of CGI quality in the flood management domain, which can be classified as either extrinsic or intrinsic, depending on the type of reference dataset used. A common characteristic of those strategies is the evaluation of the quality of each information. However, crowdsourcing platforms normally collect a large volume of CGI and the assessment of a single CGI item is often challenging. Another issue of this assessment is the access to the right data by practitioners in a feasible time. Particularly, in a disaster scenario, decision makers have short time to make decisions and, therefore, minimizing information overload in crisis situations is critically important (HUNG; KALANTARI; RAJABIFARD, 2016). An interesting, but poorly explored, strategy is the assessment of the quality of aggregated CGI elements; thus, instead of evaluating the quality of a single CGI element, all elements are evaluated as a whole. In this perspective, research focusing on the assessment of aggregated CGI can contribute to minimize information overload in the domain of flood management. Another characteristic of those strategies is the employment of only one method for the assessment of CGI quality. The quality of CGI, however, has many perspectives, as user reputation, number of contributions, analysis of the geographic context, among others. Nevertheless, as far as we know, the possible combination of these perspectives for the assessment of CGI quality has not been explored in the literature.

1.3 Objectives

According to the research gaps presented in the previous section, the main objective of this doctoral work is the development of an approach to the assessment of CGI quality in the flood management domain. Therefore, this thesis addresses the following research question: *How can the quality of CGI be improved and assessed in the flood management domain?* Based on this research question, four objectives are defined, which are described as follows:

- *Review of the existing methods for quality assessment of CGI:* we seek to investigate the quality elements used to describe a certain aspect of CGI quality, the methods for the assessment of those quality elements, types of information evaluated, and reference datasets employed. By carrying out a literature review, we aim at providing an overview of the characteristics of the existing methods.
- *Definition of a taxonomy of methods for quality assessment:* we aim at defining a taxonomy of methods for the assessment of CGI quality in the absence of authoritative data. By creating this taxonomy, we form a basis for researchers and designers of collaborative platforms.
- *Definition of a method for the assessment of CGI quality:* our main objective in this thesis is to propose a method for the assessment of the quality of CGI in the flood management

Figure 1 – Conceptual model for the assessment of CGI quality.

Source: Elaborated by the author.

domain. The method consists of a combination of the methods described by the taxonomy.

- *Design and development of an interface for a citizen observatory:* we aim at designing and developing an interface for the Citizen Observatory of Floods. By designing a new interface, we enhance the usability of the citizen observatory and, as a consequence, increase the quality of information provided by citizens.

The next section introduces a conceptual model for the assessment of CGI quality in the flood management domain. The model depicts the main components of our approach and illustrates the way they are connected.

1.4 An Approach for Quality Assessment

The outcomes of the objectives constitute a conceptual model for the assessment of CGI quality in the flood management domain, as shown in Figure 1. The model consists of three main components, i.e., (i) a taxonomy of methods for evaluating the quality of CGI, (ii) a method for the assessment of CGI quality in the flood management domain, and (iii) a crowdsourcing-based system for the gathering of flood-related information from citizens.

A Systematic Literature Review (SLR) discovered the methods and quality elements reported in the literature for the assessment of CGI quality (Chapter 3); however, owing to the large number of methods, selecting one is not a trivial task. The first component comprises a taxonomy of methods for the assessment of CGI quality when authoritative data is not available (Chapter 3). The purpose of the taxonomy is to form a basis for the researchers and designers of collaborative platforms based on CGI by discussing the idiosyncrasies of each method, so that they can select the best method for their purposes.

The second component consists of a method for the assessment of CGI quality in the flood management domain (Chapter 4). The method is a combination of the ones presented in the taxonomy and aims to assess the quality of CGI from crowdsourcing platforms by measuring a set of indicators. The obtaining of a large set of information from those platforms poses a challenge. Therefore, messages from Twitter, as known as tweets, are used as a proxy for the evaluation of the method.

Finally, the third component consists of a crowdsourcing platform, i.e., a Citizen Observatory of Floods (Chapter 5). The application aims at gathering information from citizens, as water level in a river bed, flood-affected areas, and rain intensity.

1.5 Research Context

This doctoral project is part of a research project, namely **A Geospatial Open Collaborative Approach for Building Resilience against Flooding (AGORA)** (ALBUQUERQUE; ZIPF, 2012), which has been developed by an interdisciplinary team formed by computer scientists and hydrological engineers. The acronym AGORA is inspired on the Greek word *Agorá* (literally, “gathering place”), which is considered the birthplace of democracy. The latter word lends itself as a metaphor for AGORA, a transdisciplinary approach for gathering organizations and individuals by enabling them to bring their information supply and demands to a common platform (Figure 2). Thus, AGORA combines traditional and alternative sources of information, as official data, static and mobile sensors and “people as sensors” (RESCH, 2013), which can be analyzed and interpreted by different experts, researchers and stakeholders.

AGORA comprises three pillars, namely Acquisition, Integration, and Application. The **Acquisition** pillar aims at gathering information from static and mobile sensors, citizens as sensors, social media, and collaborative mapping platforms, and make such information available to other systems. The **Integration** pillar aims at combining, managing and assessing the quality of information collected by the Acquisition pillar. The **Application** pillar aims at assisting the tasks carry out by decision makers and government agencies, e.g., Civil Defense and Emergency Services, and providing data for real-time flood monitoring, local early warning systems, and risk analysis.

The author developed a crowdsourcing-based approach for the obtaining of CGI in the flood management domain as part of her Master thesis (Crowd Sensing component - Acquisition pillar). The approach consists of a set of interpretation mechanisms for the determination of water level in a river bed and crowdsourcing platform named Citizen Observatory of Floods (DEGROSSI *et al.*, 2014). When making use of CGI, however, ensuring the quality of the information is a main concern, since information is provided by citizens that have little formal qualifications (GOODCHILD, 2007). Therefore, the evaluation of CGI quality should be undertaken before CGI is used as source of information in flood management activities (e.g. flood prediction).

Figure 2 – Architecture of the AGORA approach.

Source: [Albuquerque et al. \(2017\)](#).

This doctoral work aims at evaluating and improving the quality of CGI by developing a quality assessment method and proposing a new interface for the Citizen Observatory of Floods. The outcomes of this work contribute for the development of the component “Information Quality Assurance” of the Integration pillar.

1.6 Thesis Outline

The remainder of this doctoral thesis is structured as follows. Chapter 2 introduces the main concepts and terminology related to the topics investigated in this thesis, which are CGI and its quality. Initially, the definition of CGI is presented and its main characteristics are described. Later, the concepts of data quality, spatial data quality and CGI quality are discussed and the elements used for the measurement of quality are provided.

Chapter 3 describes a Systematic Literature Review (SLR) to discover the methods reported in the literature for the assessment of CGI quality and the development of a taxonomy of the methods identified. The literature review and taxonomy were published in a conference and journal, respectively ([DEGROSSI et al., 2017b](#); [DEGROSSI et al., 2018](#)).

Chapter 4 describes our method for the assessment of CGI quality, which encompasses a

set of the existing ones in the literature. Different quality indicators that were derived from the selected methods are also presented. The indicators were used to measure the quality of CGI in the flood management domain.

Chapter 5 presents the design and development of the new interface for the Citizen Observatory of Floods. An heuristic evaluation and usability test were carried out with usability experts and users, respectively. The results of the initial design and heuristic evaluation were published in a conference (DEGROSSI *et al.*, 2018).

Finally, Chapter 6 concludes this doctoral work, highlighting the main contributions, limitations, and presenting opportunities for future research. A list of publications resulting from this doctoral work is presented in Appendix A.

BACKGROUND

2.1 Overview

This chapter introduces the main concepts and definitions employed in this doctoral thesis. Section 2.2 is devoted to the definition of Crowdsourced Geographic Information and its main characteristics; Section 2.3 discusses the concepts of data quality, spatial data quality and CGI quality and provides a set of quality elements used in the measurement of quality. It also describes the main characteristics of the existing quality assessment methods.

2.2 Crowdsourced Geographic Information

The literature reports numerous terms that denote geographic information provided by citizens. Crowdsourced Geographic Information is an umbrella term that encompasses geographic information generated both actively (consciously) and passively (unconsciously) by citizens and has been used as a broader replacement for existing ones (SEE *et al.*, 2016) (e.g., UGC, Volunteered Geographic Information (VGI)), since some terms suggest information is voluntarily provided by citizens. Goodchild (2007), for instance, defined VGI as geographic information voluntarily provided by untrained citizens with few formal qualifications. This type of information, however, can range from purposefully shared to reflexively distributed and derived from users' actions without their conscious control (POORTHUIS *et al.*, 2016).

Researchers argue geographic information is actively contributed when citizens consciously or intentionally provide it (HAKLAY, 2013), e.g., geographic information is actively contributed to a crowdsourcing system/campaign (SEE *et al.*, 2016). OpenStreetMap¹ is a well known crowdsourcing project in which citizens actively create data on geographic features. On the other hand, geographic information is passively contributed when citizens act as an

¹ <https://www.openstreetmap.org>

observation platform, and data are collected without their active engagement (HAKLAY, 2013). This category comprises several forms of social media (e.g., Twitter², Facebook³), through which users share data with a network of friends, and these data can be later used for other purposes not envisaged by the user (POORTHUIS *et al.*, 2016). Active and passive CGI can be gathered by three types of collaborative activities, described as follows.

Social Media: comprises volunteers sharing geographic information on social media, which are Internet-based applications built on the ideological and technological foundations of Web 2.0 (KAPLAN; HAENLEIN, 2010). Volunteers use social media to share their experiences and/or opinion in “feeds” or “messages”, which may contain a geographic reference and be used as a source of “ambient geographic information” (STEFANIDIS; CROOKS; RADZIKOWSKI, 2013).

Crowd Sensing: involves the use of collaborative technologies for the gathering of “in situ” observations through specific platforms. Term “people as sensors” (RESCH, 2013) and related forms of “citizen science” are also associated with the activities performed and rely mostly on dedicated software platforms. They aim at collecting specific and structured information observed “on the ground”.

Collaborative Mapping: entails the generation of a particular type of digital data, i.e., data on geographic features, which can be understood as characteristics of the geographic space. It requires volunteers for the production of a very specific type of georeferenced data, e.g., geographic data on points-of-interest, streets, roads, buildings, land use, etc. Albuquerque, Herfort and Eckle (2016) classified collaborative mapping tasks into three types of analytical tasks. In classification tasks, volunteers analyze some geographic information and classify it into a category that best represents it. It might involve volunteers that interpret satellite imagery to classify land cover (SALK *et al.*, 2016). In digitization tasks, volunteers create geographic data (including a geometry and a location) of a real-world geographic object by, for instance, digitizing building footprints (MOONEY; CORCORAN; WINSTANLEY, 2010). Finally, in conflation tasks, volunteers analyze and interpret geographic information from multiple sources, and conflate them to find matching features/objects and produce new geographic information, e.g., detection of changes in geographic objects (ANHORN; HERFORT; ALBUQUERQUE, 2016).

The description of the collaborative activities enables the obtaining of different types of geographic information. Such a diversity results from the gathering of geographic information through geographically explicit and implicit applications (ANTONIOU; MORLEY; HAKLAY, 2010). In geographically explicit applications, citizens aim to create geographic information and, therefore, interact directly with geographic features. Collaborative Mapping applications and

² <https://twitter.com/>

³ <https://www.facebook.com/>

some Crowd Sensing and Social Media applications, such as Panoramio⁴, encourage users to provide information on spatial entities, e.g., capture spatial entities in their photos. On the other hand, in geographically implicit applications, citizens provide text- or image-based data with a geographical reference, i.e., information on location is implicitly available in metadata and/or descriptions, as coordinates, geotagges, etc. In most Social Media applications, as Twitter and Instagram⁵, geographic information is not the core feature.

Those different categories will impact on the level of citizen engagement and require training and knowledge (HAKLAY, 2013). According to Haklay (2013), citizen engagement can be classified into four different levels (Figure 3). The first comprises citizens providing only resources; therefore, the cognitive engagement is minimal. The second involves the use of citizens' cognitive ability as a resource, and citizens are asked to take some training. The third corresponds to participatory science, and citizens are responsible for both problem definition and data collection. Finally, in the fourth level, professional scientists and experts act as facilitators and citizens are part of the scientific problem definition, data collection, and analysis of results, depending on their level of engagement. In most cases, the first level, i.e., crowdsourcing, represents passive CGI and includes both spatially explicit and implicit data. The other three levels correspond to active CGI and also encompass spatially explicit and implicit data (KLONNER *et al.*, 2016).

This summary of the CGI characteristics highlights potential issues related to the quality of information. For instance, the different types of collaborative activities result in geographic data of varying degrees of accuracy, structure and format standardization. Furthermore, while passive CGI has implications on data quality due to the lack of control of the data collection, active CGI may require citizens to have a certain level of expertise and training. As a result, CGI is often considered of heterogeneous quality and uncertain credibility (FLANAGIN; METZGER, 2008), which might affect the usability of the crowdsourced information (BISHR; KUHN, 2013).

2.3 Quality of Crowdsourced Geographic Information

Data quality is a complex concept of many definitions. Quality can be understood as the “degree at which a set of inherent characteristics fulfills requirements” (ABNT, 2005) or the “totality of characteristics of a product that bear on its ability to satisfy stated or implied needs” (BATINI; SCANNAPIECO, 2016 apud ISO, 1994, p. 5). According to such definitions, data quality represents the degree at which the obtained data meet specific requirements.

Similarly to the definition of quality of ordinary data, the concept of spatial data quality refers to the degree at which the data represent the reality and precision of their location, i.e., precision of the spatial dimension (DEVILLERS; JEANSOULIN, 2006). It can be described

⁴ Panoramio is a discontinued social media platform for photo sharing with the world.

⁵ <https://www.instagram.com>

Figure 3 – Levels of citizen engagement.

Source: Adapted from [Haklay \(2013\)](#).

by different data quality elements, which represent a certain aspect of the quality of geographic data ([ISO, 2013](#)) and measure the difference between the data obtained and the reality they represent. A set of elements for the measurement of geographic data quality was proposed by [ISO \(2013\)](#) (Chart 1). The quality elements are grouped into six categories, namely (i) completeness, (ii) logical consistency, (iii) positional accuracy, (iv) thematic accuracy, (v) temporal quality and (vi) usability element, and designate two important aspects of data quality, i.e., internal quality and external quality

Internal quality represents the level of correspondence between the data obtained and ground truth data (i.e., “perfect” data). For instance, completeness, positional accuracy and thematic accuracy measure the internal quality of a dataset. On the other hand, external quality corresponds to the level of conformity between the data obtained and users’ needs or expectations in a given context. Therefore, in the perspective of external quality, data quality is not always absolute, since a dataset can have different levels of quality for different users. The usability element represents external quality ([DEVILLERS; JEANSOULIN, 2006](#)).

According to the literature, the measurement of CGI quality is often based on those quality elements, which are traditionally used for the assessment of the quality of geographic data ([ISO, 2013](#)). However, the particular features of this type of information make its quality assessment different from traditional evaluations of geographic data ([MOHAMMADI; MALEK,](#)

Chart 1 – Data quality elements for the measurement of geographic data quality.

Data quality element	Description
Completeness	Indicates the presence and absence of geographic features, their attributes and relationships.
Logical consistency	Indicates the degree of adherence to logical rules of data structure, attribution and relationships.
Positional accuracy	Indicates the accuracy of the position of geographic features within a spatial reference system.
Thematic accuracy	Indicates the accuracy of quantitative attributes and the correctness of both non-quantitative attributes and classifications of features and their relationships.
Temporal quality	Indicates the quality of temporal attributes and temporal relationships of geographic features.
Usability element	Indicates specific users' requirements that cannot be described by other quality elements.

Source: ISO (2013).

2015). For instance, the quality of CGI largely depends on the characteristics of the citizen, type of information, and way in which the information is produced (i.e., active or passive) (BORDOGNA *et al.*, 2016); therefore, researchers have added new elements towards assisting in its assessment (e.g., trust), or made new definitions for the existing quality elements, as shown in Table 1.

Table 1 – Data quality elements for the measurement of CGI quality.

Data quality element	Description
Completeness	<p>Indicates the absence of data and presence of excess data in the database (GIRRES; TOUYA, 2010).</p> <p>Measures the difference in the areas covered by OSM and German Authority Topographic-Cartographic Information System (ATKIS) buildings and the number of buildings in OSM that are recorded with attributes (attribute completeness) (FAN <i>et al.</i>, 2014).</p> <p>Determines the number of highly conserved areas identified by Public participation GIS (PPGIS) participants (BROWN; WEBER; BIE, 2015).</p> <p>Corresponds to the total area mapped out of the whole existing area (ARSANJANI; VAZ, 2015).</p>
Logical consistency	Assesses the degree of internal consistency as modeling rules and specifications (GIRRES; TOUYA, 2010).
Consistency	Compares the land cover classification contributed by experts and non-experts (SEE <i>et al.</i> , 2013).

Positional accuracy	<p>Evaluates the relation between the coordinate value of a building in OSM and the reality on the ground (FAN <i>et al.</i>, 2014).</p> <p>Measures the difference between the conservation score of PPGIS points and the mean conservation importance score derived from NaturePrint (BROWN; WEBER; BIE, 2015).</p> <p>Measures the difference between the encoded location of farmer markets and user-generated locations (CUI, 2013).</p> <p>Measures the distance between the published image position (latitude and longitude) and the estimated camera position based on image content (ZIELSTRA; HOCHMAIR, 2013).</p>
Geometric accuracy	<p>Assesses the positioning and geometries resolution from the ground reality (GIRRES; TOUYA, 2010).</p> <p>Measures the similarity of a building footprint in OSM to the shape of the building footprint in reality (FAN <i>et al.</i>, 2014).</p>
Location correctness	<p>Checks the visibility of the point of interest from the position of the visually generated VGI (observer point) (SENARATNE; BRÖRING; SCHRECK, 2013).</p>
Thematic accuracy	<p>Assesses the accuracy of quantitative attributes, correctness of non-quantitative attributes and classification of features (GIRRES; TOUYA, 2010).</p> <p>Assesses the correspondence between the semantics carried by the objects and the real world (GIRRES; TOUYA, 2010).</p> <p>Checks if real world buildings are recorded as building objects in OSM (FAN <i>et al.</i>, 2014).</p> <p>Indicates the degree of attribute matches between the land use/-cover dataset from OpenStreetMap and the reference dataset (ARSANJANI; VAZ, 2015).</p> <p>Compares the land cover classification contributed by experts and non-experts (FOODY <i>et al.</i>, 2013).</p> <p>Indicates the correctness of the value associated with the highway key in OSM street network data (JILANI; CORCORAN; BERTOLOTTO, 2014).</p>
Temporal accuracy	<p>Evaluates the actuality of the database in relation to changes in the real world (GIRRES; TOUYA, 2010).</p>
Usage	<p>Assesses the adequacy of the database to its use (GIRRES; TOUYA, 2010).</p>
Lineage	<p>Assesses the lineage of objects, their capture and evolution (GIRRES; TOUYA, 2010).</p>

Reliability	Compares volunteered land cover information to control data, which correspond to locations labeled by experts (COMBER <i>et al.</i> , 2013). Indicates the number of votes an opinion receives. A higher number of votes increases the opinion's score (LERTNATTEE; CHOMYA; SORNLERLAMVANICH, 2010).
Trustworthiness	Compares a user's classification of an event to its actual value (BODNAR <i>et al.</i> , 2014).
Credibility	Describes whether or not a piece of information can be believed (LONGUEVILLE <i>et al.</i> , 2010).

Source: Elaborated by the author.

Those elements can be measured by different methods, which differ according to the type of information evaluated, reference data types, and other factors. Depending on the reference dataset used, quality assessment methods can be classified as extrinsic and intrinsic. Extrinsic methods use external knowledge to measure the quality of CGI. Although authoritative data are commonly employed as external knowledge, their use can be constrained by financial costs, licensing restrictions (MOONEY; CORCORAN; WINSTANLEY, 2010), and currency (GOODCHILD; LI, 2012). On the other hand, intrinsic methods do not rely on external knowledge for assessing the quality of CGI and may, for instance, analyze historical metadata as a way of inferring the inherent quality of the data. Therefore, the quality of CGI can be evaluated despite a reference dataset being available or not. However, in most cases, intrinsic methods provide only rough estimates of CGI quality rather than absolute statements (BARRON; NEIS; ZIPF, 2014). Barron, Neis and Zipf (2014) proposed new intrinsic methods and indicators for assessing the quality of OpenStreetMap data.

Goodchild and Li (2012) designed three approaches to assess the quality of CGI when authoritative data are not available. The *Crowdsourcing approach* relies on the ability of a group of individuals (peers) to validate and correct erroneous information provided by another individual. In this sense, the term crowdsourcing might have two interpretations relevant to the assurance of CGI quality. According to the first, quality can be guaranteed on the basis of a number of independent and consistent observations, e.g., CGI reporting a flood event can be strengthened by additional information from the same point or point in the surrounding area. In the second interpretation, quality can be assured in terms of the ability of the crowd to converge on the truth. In OpenStreetMap, for instance, individuals (peers) are expected to edit and correct erroneous geographic information provided by others (MOONEY; MINGHINI, 2017).

The *social approach* can also be called hierarchical approach, since it relies on a hierar-

chy of individuals who act as moderators or gatekeepers of crowdsourcing platforms. Therefore, quality is assured by a group of individuals who maintain the platform integrity, prevent vandalism and infringement of copyright, and avoid the use of abusive content. In the Flood Citizen Observatory (DEGROSSI *et al.*, 2014), for instance, the platform administrator acts as a gatekeeper by assessing the veracity of CGI and classifying it as checked or unchecked. Finally, the *geographic approach* involves the comparison of an item of geographic information with the body of geographic knowledge. The approach adheres to rules, as the First Law of Geography, according to which “everything is related to everything else, but near things are more related than distant things” (TOBLER, 1970). Geographic information should, for instance, be consistent with what is known about the location and surrounding area. In other words, it should be related to the space in which the knowledge has been provided. Albuquerque *et al.* (2015) demonstrated a tendency for ‘relevant’ on-topic tweets to be closer to flood-affected catchments.

Quality assessment methods can also be employed in the light of two temporalities, namely (i) *ex ante* and (ii) *ex post* (BORDOGNA *et al.*, 2016), which differ according to the time at which the assessment is carried out in comparison with the creation time of CGI. The *Ex ante* strategy is employed prior to the creation of CGI and aims to avoid the production of low-quality CGI (BORDOGNA *et al.*, 2016). Apart from offering mechanisms that control the creation of data, the methods provide volunteers with resources for guiding the way information is produced. In contrast, the *Ex post* strategy is employed after a CGI item has been created, and aims at removing errors and improving the CGI quality, which involves the checking of its quality and filtering.

2.4 Final Remarks

This chapter has addressed the definition of term crowdsourced geographic information, which comprehends geographic information generated by citizens through geographically explicit and implicit applications, and its characteristics. Citizens create geographic information actively and passively through three types of collaborative activities, namely Social Media, Crowd Sensing and Collaborative Mapping. The level of citizen engagement may vary from the supply of (only) resources (Level 1) to the definition of the scientific problem and type of data collection and analysis (Level 4), which influence the CGI quality.

The measurement of CGI quality is based on different quality elements derived from those traditionally employed for the assessment of the quality of spatial data. However, researchers have made new definitions or added new elements to describe certain aspects of the CGI quality, since CGI has specific characteristics. Moreover, those elements are measured by different methods, which differ according to the type of reference dataset, approach used when authoritative data are unavailable, and temporality. The next chapter discusses the methods that assess the quality of CGI when authoritative data are not available.

A TAXONOMY OF QUALITY ASSESSMENT METHODS

3.1 Overview

This chapter introduces a Systematic Literature Review to discover the existing methods to assess the quality of CGI when authoritative data is not available and presents a taxonomy of the methods. The purpose of this taxonomy is to form a basis for the researchers and designers of collaborative platforms based on CGI, so that they can select the best method for their purposes, by discussing the idiosyncrasies of each method. In Section 3.2, there is an overview of related work. In Section 3.3, the methodology employed for the conduction of SLR and development of the taxonomy is presented. In Section 3.4, the proposed taxonomy is described in detail. In Section 3.5, our findings, the limitations of our taxonomy, and suggestions for future research are discussed. In Section 3.6, our conclusions are summarized.

3.2 Related Work

Several critical literature reviews (or surveys) involving the categorization of quality assessment methods have been conducted to provide an overview of this area (e.g., [Wiggins *et al.* \(2011\)](#), [Bordogna *et al.* \(2016\)](#), [Mirbabaie, Stieglitz and Volkeri \(2016\)](#), [Senaratne *et al.* \(2017\)](#)). An analysis of the quality assessment methods was carried out by [Wiggins *et al.* \(2011\)](#), where the authors analyzed the data validation policy and quality assessment in citizen science projects (i.e., crowd sensing). They found that the most common type of data validation is based on expert reviews conducted by trusted individuals or moderators. [Bordogna *et al.* \(2016\)](#) also analyzed CGI in citizen science projects. They initially reviewed and categorized CGI projects on the basis of a classification schema (Figure 4), which describes the characteristics of CGI projects, and analyzed the way each project deals with CGI quality. This work also provided a critical

Figure 4 – Classification schema of CGI projects.

Source: [Bordogna et al. \(2016\)](#).

description of the strategies currently adopted to assure and improve CGI quality. [Bordogna et al. \(2016\)](#) and [Wiggins et al. \(2011\)](#) conducted an important overview of quality assessment methods and made significant recommendations for improving CGI quality in research projects. However, these authors only analyzed studies proposing methods for quality assessment of CGI in citizen science projects and did not take into account other CGI sources such as collaborative mapping and social media.

[Senaratne et al. \(2017\)](#) conducted a critical literature review of the existing methods to assess the quality of the main types of CGI: text, image, and map. This review examined methods that are based on theories and discussions in the literature, and provided examples of the practical applicability of all the different approaches. In doing so, the authors provided a general description of the methods used in each paper analyzed; however, they did not create a taxonomy of methods for quality assessment. Moreover, this is a traditional literature review and, as many researchers have pointed out, traditional reviews are prone to bias (i.e., authors

may decide only to include studies with which they are familiar or which support their particular standpoint) (MULROW, 1994; BIOLCHINI *et al.*, 2005). In an attempt to minimize this kind of bias, Systematic Literature Review (SLR) has been proposed as a replicable, scientific, and transparent approach to locate the most significant literature about a given topic or discipline (BRERETON *et al.*, 2007; KITCHENHAM; CHARTERS, 2007).

Mirbabaie, Stieglitz and Volkeri (2016) conducted a systematic literature review on CGI in disaster management. The main goal of this review was to provide information about the quality elements that are used, as well as the methods that are employed to measure these elements. They found that attributes such as “accuracy” and “consistency” are mainly used as criteria for quality assessment, while other factors such as “trustworthiness” are not fully taken into account. However, they did not conduct an in-depth analysis of the existing methods with regard to their applications and limitations, and were only concerned with the existing methods for disaster management. Moreover, some key databases, such as Web of Science (WoS) and Scopus, were not used by Mirbabaie, Stieglitz and Volkeri (2016).

3.3 Methodology

3.3.1 *Systematic Literature Review*

Systematic Literature Review (SLR) was first applied to support evidence-based medicine. SLR is a kind of secondary study that aims at identifying, analyzing and interpreting all the available evidence related to a research topic (KITCHENHAM; CHARTERS, 2007). Differently from the usual process of literature review, an SLR is undertaken in a formal, rigorous and systematic way (BIOLCHINI *et al.*, 2005; OKOLI; SCHABRAM, 2010), i.e. in a way that is unbiased and (in a certain degree) repeatable (KITCHENHAM; CHARTERS, 2007). With this methodology is possible to summarize existing evidence and identify any gaps in a research topic, and provide a framework for position new research activities (KITCHENHAM; CHARTERS, 2007). Recently, SLR has been applied to the field of Geographic Information Science (GIS) for analyzing the current state of research, for instance, on the use of CGI for disaster management (HORITA *et al.*, 2013); concerning methodologies application and use cases of Twitter as a Location-based Social Network (STEIGER; ALBUQUERQUE; ZIPF, 2015); and on the use of CGI within natural hazard analysis (KLONNER *et al.*, 2016).

In this work, an SLR was carried out to discover current research on CGI within the scope of quality assessment. More specifically, each study was analyzed with regard to the method designed for assessing the quality of CGI. In conducting the SLR, this work complies with the guidelines recommended by Kitchenham and Charters (2007). The SLR follows a sequence of well-defined steps, which comprise (i) planning the review, (ii) conducting the review and (iii) reporting the review (Figure 5).

Figure 5 – Steps of the SLR.

Source: Elaborated by the author.

Review Planning

An important activity of the planning phase is to draft a clear and concise Research Question (RQ) (BRERETON *et al.*, 2007; OKOLI; SCHABRAM, 2010), since it will be used as a guide for the entire SLR process. As the main goal of this work is to discover studies proposing methods for the quality assessment of CGI, the following RQ has been raised:

RQ) What are the methods used to assess the quality of CGI?

In an SLR, existing evidence to answer a RQ can be obtained by carrying out a search string¹ on electronic databases. To build the search string, the main terms of the RQ were first selected, i.e., crowdsourced geographic information and quality assessment. The synonyms for each term were also identified in order to maximize the number of returned studies. Particularly, the synonyms for the former term were extracted from See *et al.* (2016). Finally, the Boolean operator OR was applied to join the synonyms of each term and the Boolean operator AND to join the main terms. The search string is shown in Chart 2.

The identification of primary studies began with the search string being applied to 5 (five)

¹ The search string is a logical expression that combines keywords and synonyms through boolean operators.

Chart 2 – Search string.

(VGI OR "volunteered geographic information" OR "volunteered geographic data" OR "crowdsourced geographic information" OR "crisis mapping" OR "collaborative mapping" OR OpenStreetMap OR ((crowdsourcing OR "crowd sourcing" OR "crowd-sourcing" OR "crowd-sourced data" OR "user-generated content" OR "social media" OR Twitter OR Flickr OR "collective intelligence" OR "collective knowledge" OR "citizen based")) AND (geographic OR spatial OR geotagged OR georeferenced))) AND ("quality assessment" OR "quality assurance" OR "quality evaluation" OR "quality control" OR reliability OR credibility OR trust OR accuracy)

Source: Elaborated by the author.

electronic databases (Chart 3). This set of electronic databases was particularly selected in order to maximize the number of studies, since a single database may not find all the existing evidence concerning a research topic (BRERETON *et al.*, 2007). Moreover, this set was selected because of their relevance to the research field, i.e., these electronic databases index the main journals and conferences of the area. However, owing to the idiosyncrasy of each electronic database, the search string had to be adjusted to them because there were few relevant studies in this field. Hence, some synonyms of the search string (e.g., social computing, quality analysis and quality enhancement) had to be removed.

Chart 3 – Electronic Databases.

Electronic Database	URL
IEEE Xplore	www.ieeeexplore.ieee.org
ACM Digital Library	www.portal.acm.org
Web of Science	www.webofknowledge.com
Science Direct	www.sciencedirect.com
SCOPUS	www.scopus.com

Source: Elaborated by the author.

The inclusion and exclusion criteria assisted the selection of key studies that could be used to answer the research question (BIOLCHINI *et al.*, 2005; PETERSEN *et al.*, 2008). Considering the main goal of this SLR and the aforementioned RQ, a set of inclusion and exclusion criteria (Chart 4) was defined, which were used as a basis for the selection of studies. Besides this set, a study was also excluded if (i) it is not published between 2004 and November 2015, (ii) it is not written in Portuguese or English, (iii) it is an SLR, a poster, a technical report or an ongoing research (iv) it is not available and (v) it is duplicated or incomplete.

Review Conduction

The search in the electronic databases resulted in a total of 551 primary studies included,

Chart 4 – The inclusion and exclusion criteria employed for the qualitative review.

IC1: the study sets out or adopts an approach for the quality assessment of crowdsourced geographic information

EC1: the study is related to quality assessment, but not to crowdsourced geographic information

EC2: the study is related to crowdsourced geographic information, but not to quality assessment

EC3: the study is not related to crowdsourced geographic information nor quality assessment

Source: Elaborated by the author.

after duplicate studies had been removed (Figure 6). After reading the title and abstract, 499 studies that are not directly related to the quality assessment of CGI were identified. However, if the objective of the work was not clear in the abstract, the study was included for a more in-depth analysis. After this stage, 52 studies were included for a complete reading.

While performing the complete reading, the included studies were analyzed to determine if the study was in fact a candidate to answer the RQ. If the study satisfies at least one exclusion criterion, it was excluded. However, if doubts emerged with regard to include the study; the opinion of another reviewer was taken into account to decide about it. After a close analysis, 18 studies were found that discuss methods for assessing CGI quality.

3.3.2 Development of the Taxonomy

According to Nickerson, Varshney and Muntermann (2013), a taxonomy is a set of n dimensions each consisting of mutually exclusive and collectively exhaustive characteristics. The mutually exclusive and collectively exhaustive restrictions mean that no object can have two different characteristics in a dimension and each object must have one of the characteristics in a dimension. For developing the taxonomy, a rigorous and systematic method was adopted (NICKERSON; VARSHNEY; MUNTERMANN, 2013), providing guidance during the development stage.

The definition of the characteristics of the taxonomy is based on a meta-characteristic (i.e., the most comprehensive characteristic), which serves as the basis for choosing them. In particular, the meta-characteristic is defined in terms of the purpose of the taxonomy which, in turn, depends on its expected use. This can be carried out by determining who the users are and what use they could make of the taxonomy. These are both researchers and designers of crowdsourcing-based platforms and are interested in the quality assessment methods that can be employed on those platforms when no authoritative data is available. Thus, the purpose of our taxonomy is to distinguish between quality assessment methods on the basis of their ability to assess the quality of CGI. With the aid of this taxonomy, the main goal is to assist the users in understanding the unique features of the current methods and suggest ways of devising new methods. Thus, the meta-characteristic for designing this taxonomy is the way in which the

Figure 6 – SLR process and the number of included and excluded studies in each step.

Source: Elaborated by the author.

Figure 7 – Development process of the taxonomy.

Source: Elaborated by the author.

method assesses the quality of CGI.

Since this method is iterative (Figure 7), a subset of the ending conditions (i.e., objective and subjective), which are required to terminate it, was selected. The main ending conditions that the taxonomy must satisfy are mutually exclusive and collective exhaustive characteristics. Additionally, the taxonomy must also satisfy three objective conditions (Chart 5) and five subjective ending conditions (Chart 6), which represent the qualitative attributes of a useful taxonomy.

After the definition of the meta-characteristic and the ending conditions, the identification of the dimensions and characteristics can begin. This entails adopting two approaches (i.e., empirical and conceptual approaches), both of which were used for the development of the taxonomy, with the iterations so as to have several perspectives. In the conceptual-to-empirical approach, the dimensions of the taxonomy are conceptualized without examining the objects.

Chart 5 – Objective ending conditions.

Objective ending condition	Comments
All objects or a representative sample of objects have been examined	If all objects have not been examined, then the additional objects need to be studied
At least one object is classified under every characteristic of every dimension	If at least one object is not found under a characteristic, then the taxonomy has a "null" characteristic. We must either identify an object with the characteristic or remove the characteristic from the taxonomy
No new dimensions or characteristics were added in the last iteration	If new dimensions were found, then more characteristics of the dimensions may be identified. If new characteristics were found, then more dimensions may be identified that include these characteristics

Source: Adapted from [Nickerson, Varshney and Muntermann \(2013\)](#).

Chart 6 – Subjective ending conditions.

Subjective ending condition	Comments
Concise	Does the number of dimensions allow the taxonomy to be meaningful without being unwieldy or overwhelming?
Robust	Do the dimensions and characteristics provide for differentiation among objects sufficient to be of interest?
Comprehensive	Can all objects or a (random) sample of objects within the domain of interest be classified?
Extendible	Can a new dimension or a new characteristic of an existing dimension be easily added?
Explanatory	What do the dimensions and characteristics explain about an object?

Source: Adapted from [Nickerson, Varshney and Muntermann \(2013\)](#).

On the other hand, in the empirical-to-conceptual approach, a subset of the objects are classified ([NICKERSON; VARSHNEY; MUNTERMANN, 2013](#)).

In Iteration 1, the conceptual-to-empirical approach was adopted, where the dimensions of the taxonomy are conceptualized without any examination of the objects ([NICKERSON; VARSHNEY; MUNTERMANN, 2013](#)), but rather based on conceptual distinctions drawn from the SLR. Some quality assessment methods may use external knowledge for assessing the quality of CGI (i.e., extrinsic), whereas others may not depend on this, but rather require internal knowledge to carry out this task (i.e., intrinsic). Both characteristics were grouped into the reference dimension. Since one dimension was added in this iteration, one more iteration was needed.

In Iteration 2, the empirical-to-conceptual approach, in which a subset of the objects is

classified (NICKERSON; VARSHNEY; MUNTERMANN, 2013), was selected. On the basis of the studies found in the SLR (Section 3.3.1), several quality assessment methods were identified and a set of these methods, i.e., *redundancy of volunteered contributions*, *volunteer profile; reputation*, and *spatiotemporal clustering*, was selected based on suitability. On the basis of my understanding of these methods, two new dimensions could be found, i.e., object and approach dimensions. In the object dimension, the object under evaluation could be the information or the volunteer. In the approach dimension, the method analyzes the geographical context (i.e., geographic approach), the social hierarchy (i.e., social approach) or uses a group of people for evaluating quality (i.e., crowdsourcing approach). Since new dimensions were added in this iteration, a further iteration was needed.

In Iteration 3, the conceptual-to-empirical approach was adopted once more in order to capture further conceptual distinctions made in the field. As noticed before, quality assessment methods can be distinguished by comparing the time of their application to the creation time of the CGI item (BORDOGNA *et al.*, 2016). Some methods can be employed after a CGI item has been collected (i.e., *ex-post*), with the aim of data cleansing and improving CGI quality, whereas other methods take place before a CGI item has been created (i.e., *ex-ante*), preventing the creation of low-quality CGI. Both characteristics were grouped into the temporal dimension. Since one dimension was added in this iteration, a further iteration was needed.

In Iteration 4, the empirical-to-conceptual approach was again adopted because there were still some other methods that needed to be examined. The methods *automatic location checking*, *extraction/learning of characteristics*, *ranking/filtering by linguistic terms*, *volunteer profile; reputation* and *scoring volunteered contributions* were selected from the SLR. On the basis of my understanding of these methods, several quality elements (i.e., positional accuracy, thematic accuracy, fitness-for-use, trust, reliability, and plausibility), which were used to measure CGI quality, were identified. Those were grouped into the criteria dimension. Since one dimension was added in this iteration, one more iteration was needed.

In Iteration 5, the empirical-to-conceptual approach was used once again since there were more methods that needed to be examined. The methods *geographic context analysis*, *expert assessment*, *error detection/correction by crowd*, *historical data analysis* were selected from the SLR. After analyzing them, any new characteristics and dimensions were identified and, thus, they were classified in accordance with the characteristics and dimensions outlined above. Since no new dimension was added in this iteration and all the methods from the SLR were examined, it can be concluded that the objective ending conditions were met. Furthermore, the taxonomy met the subjective ending conditions.

The final set of dimensions and characteristics is shown in Chart 7. They were used to build the taxonomy of quality assessment methods, which is discussed in the following section.

Chart 7 – Dimensions and characteristics of the taxonomy.

Dimension	Characteristics
Reference	<i>Extrinsic</i> : when the method uses external knowledge <i>Intrinsic</i> : when the method does not use external knowledge
Object	<i>Content</i> : the information is the object under evaluation <i>Volunteer</i> : the volunteer is the object under evaluation
Approach	<i>Crowdsourcing</i> : the method uses a group of individuals for evaluating quality <i>Social</i> : the method analyzes the social hierarchy <i>Geographical</i> : the method analyzes the geographical context
Temporality	<i>Ex-post</i> : the method is employed only after the creation of CGI <i>Ex-ante</i> : the method is employed only before the creation of CGI <i>All</i> : the method is employed before and after the creation of CGI
Criteria	Positional accuracy, thematic accuracy, fitness-for-use, trust, reliability, plausibility

Source: Elaborated by the author.

3.4 Taxonomy of Quality Assessment Methods

The taxonomy results from the methods identified in the SLR (Table 2) and comprises 11 methods that can be employed for assessing the quality of CGI when authoritative data are not available. Moreover, it consists of a set of dimensions (i.e., reference, object, approach, temporal, and criteria dimensions) (Table 7) that represent an abstraction of the main attributes of each method. Each dimension has a set of possible values, which are called the characteristics of a method. The dimension “approach”, for instance, comprises three possible characteristics (i.e., crowdsourcing, social, and geographical approaches). Each method was classified in accordance with these dimensions and characteristics (Table 3).

3.4.1 Geographic context analysis

Description: each place has its own distinguishing characteristics. The basic idea of the “geographic context analysis” method consists of investigating the area surrounding the location of CGI to determine its geographical features and employ them for assessing the likelihood that the citizen-generated data is consistent with those features. Both physical and human geographical features (e.g., hydrological, geomorphological, socio-economical, demographic, etc.) could potentially be used in this method.

Example: [Senaratne, Bröring and Schreck \(2013\)](#) examined the geographic context around Flickr photographs to determine which areas can be viewed from the CGI location, or more specifically, whether the Point-of-Interest (POI) lies within the line of sight departing from the coordinates given in the CGI geotag. Similarly, [Zielstra and Hochmair \(2013\)](#) carried out an investigation to evaluate the positional accuracy of geotagged photos from Flickr and Panoramio. Here, the

Table 2 – Summary of the quality assessment methods of the taxonomy.

ID	Method	Description
M1	Geographic context analysis	Investigating the area surrounding the location of CGI to determine its geographical features and employ them to assess the quality of CGI
M2	Redundancy of volunteered contributions	Requesting several volunteers to provide information about the same geographic feature to find out if or not there is a convergence of the information produced by different volunteers
M3	Scoring volunteered contributions	Asking volunteers to rate every piece of CGI that is contributed by other volunteers
M4	Expert assessment	Submitting CGI to experts who are responsible for checking the information content and correcting it if necessary
M5	Automatic location checking	Estimating the quality of CGI by the distance between geocoded coordinates, obtained from multiple geocoding services, and the location (i.e., an address) provided by the volunteer
M6	Spatiotemporal clustering	Creating spatiotemporal clusters of CGI elements using prior information about a phenomenon of interest and, later, evaluating the significance of the resulting clusters for a specific purpose
M7	Volunteer profile; reputation	Analyzing volunteer's profile or reputation and using it to estimate the quality of CGI
M8	Error detection/correction by crowd	Several volunteers acting as gatekeepers and, thus, correcting errors introduced by other volunteers
M9	Extracting/learning of characteristics	Extracting characteristics from each type of geographic feature, learning the information implicit in them and, later, using the information to estimate the quality of CGI
M10	Ranking/filtering by linguistic terms	Evaluating CGI items based on different criteria that are expressed linguistically, rank them in degrees of criteria satisfaction and, later, filter them based on the constraints of the application domain
M11	Historical data analysis	Deriving (intrinsic) indicators from the history of the data and, later, using them to make statements regarding the quality of CGI

Source: Elaborated by the author.

Table 3 – Taxonomy of quality assessment methods.

ID	Method	Object			Approach			Reference			Temporality			Criteria				
		INF	VOL	GEO	CWD	SOC	EXT	INT	EP	ALL	PA	TA	FFU	TRU	REL	PLA		
M1	Geographic context analysis	x		x			x		x		x							
M2	Redundancy of volunteered contributions	x			x		x		x			x						
M3	Scoring volunteered contributions	x			x		x		x						x			
M4	Expert assessment	x			x		x		x			x						
M5	Automatic location checking	x		x			x		x		x							
M6	Spatiotemporal clustering	x		x				x	x						x			
M7	Volunteer profile; reputation		x			x		x		x			x					
M8	Error detection/correction by crowd	x			x		x		x		x							
M9	Extracting/learning of characteristics	x		x				x			x							
M10	Ranking/filtering by linguistic terms	x		x				x					x					
M11	Historical data analysis	x			x			x	x									

Source: Elaborated by the author.

Note – INF=information; VOL=volunteer; GEO=geographic approach; CWD=crowdsourcing approach; SOC=social approach; EX=extrinsic; INT=intrinsic; EP=ex-post; PA=positional accuracy; TA=thematic accuracy; FFU=fitness for use; TRU=trust; REL=reliability; PLA=plausibility

positional accuracy was estimated by measuring the distance between the image position and the estimated camera position based on the imagery metadata.

Limitation: a key factor when applying this method is determining which geographical features are related to the CGI content, since they must be matched. Furthermore, establishing reliable geographic relations is not a trivial problem, since they may vary in accordance with the physical area and domain. Finally, this method may also be constrained by the unsatisfactory quality of the reference data (e.g., the background map) used for the measurements.

3.4.2 *Redundancy of volunteered contributions*

Description: this method involves requesting several volunteers to provide information about the same geographic feature independently. Later, CGI quality is determined by comparing the volunteered contributions in order to find out whether or not there is a convergence of the information independently produced by different volunteers. In other words, the CGI quality is assumed to be higher when there is agreement among volunteers with regard to the information content.

Example: [Comber et al. \(2013\)](#) estimated the degree of reliability of CGI by asking volunteers (i.e., experts, postgraduate students, scientists, and novices) to identify the type of land cover based on satellite imagery of a series of locations. Similarly, [See et al. \(2013\)](#) employed this method to evaluate the accuracy and consistency of volunteers when labeling land cover and determining the human impact on the environment. In contrast, [Foody \(2014\)](#) explored the redundancy of contributions to situations in which a large proportion of data is provided by poor sources and/or is incomplete.

Limitation: this type of analysis is applicable to situations where it is practical to obtain several versions of the volunteered information independently (i.e., in which different volunteers provide information about the same geographic feature(s) independently and the quality can be associated with the degree of agreement among them). However, it is worth noting that the ability to reach an agreement among volunteers will also depend on the difficulty of the task(s). In other words, there is a greater chance of reaching an agreement when the tasks are “easy” than when they are moderately hard to difficult ([ALBUQUERQUE; HERFORT; ECKLE, 2016](#); [SALK et al., 2017](#)). Moreover, since the information is based on the knowledge of volunteers, it could cause problems as some of it may be less reliable and of poorer quality ([COMBER et al., 2013](#)).

3.4.3 *Scoring volunteered contributions*

Description: in this method, volunteers are asked to rate every piece of CGI that is contributed by other volunteers. A score is thus calculated from the individual rates given and assumed to represent “content quality”.

Example: [Lertnattee, Chomya and Sornlertlamvanich \(2010\)](#) used a score calculated from the

number of clicks on the button “Vote” given by members of the community to herbal information, a type of information related to medicinal herbs.

Limitation: the success of the method depends on how successful a platform is in gathering rates from volunteers, since it might require a large number of raters to arrive at a score that truly represents the “content quality”.

3.4.4 Expert assessment

Description: feedback and contributions from individuals that are familiar with an area or domain may greatly assist in estimating the quality of CGI (KARAM; MELCHIORI, 2013). In this method, CGI is submitted to experts (i.e., individuals who have a greater knowledge of the geographic area or domain and are responsible for checking the information content and correcting it if necessary). Later, all the CGI are ranked on the basis of the corrections made by the individuals. When CGI is ranked in a lower position, it indicates the presence of malicious and/or wrong corrections. In other words, it is CGI of a poorer quality.

Example: Karam and Melchiori (2013) ranked CGI by employing four metrics: (a) the historical records of activities carried out by a volunteer; (b) the number of activities carried out by experts; (c) the feedback received after the change was made; and (d) the score of the user that submitted the information.

Limitation: a drawback of this method is that it requires (expert) individuals who are able to check the CGI provided by volunteers.

3.4.5 Automatic location checking

Description: the location of a volunteer can be ascertained in different ways, such as GPS, manual georeferencing, or an address. However, the latter could contain errors or typos. A way of automatically determining the correct location of a volunteer is to compare geocoded coordinates, from multiple geocoding services, with each other. Hence, the address should be submitted to several geocoding services, which will result in coordinates that are represented by latitude and longitude. To qualify as reference data, the different geocoding services should yield concordant results within a certain distance (CUI, 2013). The quality is, thus, estimated as the distance between the geocoded coordinate and the location provided by the volunteer.

Example: Cui (2013) employed automatic location checking to determine the spatial accuracy of the location of farmers’ markets.

Limitation: the reference data and geocoded methods used in the geocoding service might contain errors and, thus, lead to erroneous results (CUI, 2013). Hence, the use of multiple geocoding vendors’ services can improve the reliability of the resulting geocoded coordinates. Furthermore, the CGI location must be provided through an address or place name that will be submitted to the geocoding services.

3.4.6 Spatiotemporal clustering

Description: CGI quality can be addressed by aggregating information from several volunteers (MUMMIDI; KRUMM, 2008) and, later, by evaluating the significance of the resulting clusters for a specific purpose (LONGUEVILLE *et al.*, 2010). Thus, instead of checking the quality of a single CGI element, all elements are evaluated as a whole (i.e., the quality of the CGI cluster is assessed).

This method consists of creating spatiotemporal clusters of CGI elements using prior information about a phenomenon of interest. The clusters are created on the basis of the assumption that “CGI elements created at the same place and time refer to the same event” (LONGUEVILLE *et al.*, 2010).

The process starts by creating temporal clusters, which are the CGI elements clustered in several temporal classes. After this, the temporal classes are divided into sub-classes in compliance with spatial criteria. These steps convert raw CGI of unknown quality into spatiotemporal clusters, the importance of which can be quantified by means of a ranking score, which reflects the likelihood that an event took place in the time period and area that each cluster refers to.

Example: Longueville *et al.* (2010) employed spatiotemporal clustering to assess the likelihood that a flood event took place (i.e., to locate inundation areas of flood events that occurred in the United Kingdom between January 1st, 2007 and March 31st, 2009).

Limitation: as Longueville *et al.* (2010) point out, when employing this method, a large dataset must be made available, since a low amount of data prevents a robust statistical analysis from being conducted. Hence, this method is less suitable for sparsely populated areas (e.g., rural areas), where only a few VGI elements are available.

3.4.7 Volunteer profile; reputation

Description: the volunteer is an important factor in the quality of CGI, since his/her knowledge and background can have an influence on the data produced. By employing this method, the volunteer profile or reputation can be analyzed and used as a basis to estimate the quality of CGI. Bishr and Janowicz (2010) argue that if the volunteer has a reputation for being trustworthy, this means that his or her contribution is likely to be trustworthy too.

Example: Bodnar *et al.* (2014) analyzed volunteers profiles to establish the veracity of volunteered contributions with regard to four security-related events that took place in the USA. In contrast, Bishr and Kuhn (2013) employed this method to assess the trustworthiness of volunteers statements regarding the quality of water from a well.

Limitation: this method can be adversely affected if some of the metadata is missing or inaccurate (i.e., the metadata concerning a volunteer profile). Thus, before employing it, it is essential to check if reliable information about the volunteers is available. Another limitation is that this

method does not consider the extent to which users have the necessary skills for a particular task or context. Volunteers may be reliable at producing data about the surroundings in which they live, but when generating data about geographic areas for which they do not have any contextual knowledge, they may produce less reliable geographic information (KLONNER *et al.*, 2017).

3.4.8 Error detection/correction by crowd

Description: this method is based on the so-called Linus's law, according to which "given enough eyeballs, all bugs are shallow" (RAYMOND, 1999). In the case of CGI, this can be understood as "given enough volunteers, all errors can be identified and corrected".

The basic idea behind this method is that a single individual might unintentionally introduce an error in a crowdsourcing-based platform. Later, other people might notice this error and correct it, and hence the community of volunteers acts as gatekeepers. This method is different from the *redundancy of volunteered contributions*, since it is not based on an aggregation of information that is independently provided by volunteers, but rather relies on peer verification of the information produced by other volunteers as a self-correcting mechanism.

Example: Haklay *et al.* (2010) investigated whether Linus's law applies to the positional accuracy of OSM data (i.e., if the positional accuracy of a given geographic feature in OSM increases incrementally with the number of volunteers).

Limitation: as Linus's law states, there should be enough eyeballs (volunteers) to identify and correct the errors. However, this could be a drawback, since a large number of people may be needed to achieve good quality. This can be especially problematic in sparsely populated areas, such as rural areas, where the number of volunteers is small, as well as in virtual communities that have a small number of users.

3.4.9 Extraction/learning of characteristics

Description: each geographic feature has characteristics (i.e., shape, size, etc.) which can be used as a classification criterion. This method consists of extracting these characteristics from each type of geographic feature, learning the information implicit in them, and, later, using the information to estimate the quality of CGI. For instance, distinct characteristics can be extracted and learned from CGI with corresponding reference data and, later, used to estimate the quality of CGI with no corresponding reference data.

Example: Ali and Schmid (2014) designed a classifier that learns the correct class of existing entities (i.e., parks and gardens) on the basis of their characteristics (e.g., size) and used it to predict the correct class of a new entity. Similarly, Jilani, Corcoran and Bertolotto (2014) extracted geometrical and topological properties of OSM street network data that are representative of their semantic class, to infer the "road class" from the new data. Finally, Mohammadi and Malek

(2015) estimated the positional accuracy of OSM data without corresponding reference data by extracting patterns from OSM data that have corresponding reference data.

Limitation: this method does not depend on any external source when being employed. However, a large amount of data is required to properly learn the characteristics of geographic features. Moreover, it is context-dependent, since geographic features in the same region might have more similar characteristics to those in different regions.

3.4.10 *Ranking/filtering by linguistic terms*

Description: the underlying principle of this method is the need to express the quality criteria linguistically. The linguistic terms are used to specify the desired values of the quality indicators and, together, these comprise a schema for quality evaluation. Each CGI item is first evaluated on the basis of each criterion that is expressed linguistically and, later, ranked in degrees of criteria satisfaction. Finally, CGI items are filtered by being subject to the constraints of the application domain.

Example: Bordogna *et al.* (2014) employed this method to filter CGI items (i.e., pictures) for a citizen science project on glaciology.

Limitation: one critical factor in this method is the schema for quality evaluation, since this is changed by each intended use of the CGI items. In other words, the schema depends on the application domain.

3.4.11 *Historical data analysis*

Description: in special cases, CGI comes with historical data. In OSM, for instance, a new version of an object is created whenever its geometry is changed. From the history of the data, it is possible to derive (intrinsic) indicators that allow approximate statements to be made regarding data quality (BARRON; NEIS; ZIPF, 2014). An example of an indicator is the number of contributors in a given area, since it has been demonstrated that this has an influence on the quality (HAKLAY *et al.*, 2010). Moreover, the historical data can be analyzed to identify patterns and make predictions of quality elements.

Example: Keßler and Groot (2013) produced a set of indicators based on historical data (i.e., number of versions, contributors, confirmations, tag corrections, and rollbacks) to assess the quality of OSM data in Muenster, Germany. Positive indicators (e.g., a high version number) were shown to correlate with high-quality CGI.

Limitation: this is an alternative method when no ground-truth data is available. However, a certain amount of historical data must be available before it can be applied. Otherwise, the value of resulting statements may be limited.

3.5 Discussion

This chapter presents a taxonomy of quality assessment methods for CGI. In contrast with existing works, this taxonomy summarizes what methods can be employed to assess the quality of CGI when authoritative data is not available. In creating this taxonomy, different types of CGI source (i.e., social media, crowd sensing, and collaborative mapping) were taken into account. As a result, the taxonomy can be useful for quality assessment in a larger number of crowdsourcing-based platforms. Thus, this taxonomy can be regarded as an extension of previous work (WIGGINS *et al.*, 2011; BORDOGNA *et al.*, 2016), which only includes methods for assessing the quality of CGI in Citizen Science projects (i.e., crowd sensing).

Some of the methods presented in the taxonomy have already been identified in previous studies (WIGGINS *et al.*, 2011; BORDOGNA *et al.*, 2016; MIRBABAIE; STIEGLITZ; VOLKERI, 2016; SENARATNE *et al.*, 2017), such as *redundancy of volunteered contributions*, *scoring volunteered contributions*, *expert assessment*, *volunteer reputation*, *ranking/filtering by linguistic terms*, and *historical data analysis*. Other methods have not been identified as such in the literature so far (i.e., a contribution of the present work is to identify a set of new methods to assess the quality of different types of CGI when there is no authoritative data).

As well as in previous studies, the way each method works was briefly described, they were classified according to their temporality and the approach employed, and practical examples were presented. In addition, the methods were classified in accordance with the type of reference dataset and the object under evaluation, and their limitation(s) discussed, which may prevent their employment. Moreover, the CGI sources in which each method could be employed were analyzed. With the exception of four methods (i.e., *scoring volunteered contributions*, *spatiotemporal clustering*, *error detection/correction by crowd*, and *historical data analysis*), the majority of the methods can be employed with all CGI sources (i.e., social media, crowd sensing, and collaborative mapping).

The method *scoring volunteered contributions* (M3) can be employed for some crowd sensing platforms, since it is possible to measure a score and attach it to the information. On the other hand, in some collaborative mapping and social media platforms, it is not possible to directly measure a score but only to give indirect indicators of information quality; this is owing to the idiosyncrasies of these platforms.

The method *spatiotemporal clustering* (M6) can be employed for crowd sensing and social media platforms. In these platforms, volunteers are able to share information about a specific event. Moreover, this set of information is usually created in the same area and during a certain period of time. In contrast, in collaborative mapping platforms, the volunteers are not able to share information about an event, since these platforms usually collect data about less dynamic geographic features, such as roads and buildings.

The method *error detection/correction by crowd* (M8) cannot be employed for social

media platforms since no one is allowed to correct the contributions of anyone else. However, it can be applied to some collaborative mapping and crowd sensing platforms, if permission is granted to volunteers to edit the information.

Finally, the method *historical data analysis* (M11) can only be employed in crowdsourcing-based platforms that keep records of all the changes of a geographic feature, such as some collaborative mapping and crowd sensing platforms, owing to its main characteristic (i.e., the use of historical data). Social media platforms, however, do not keep any records, and this prevents them from using this method.

One advantage of the taxonomy is the fact that it is domain-independent (i.e., it can be employed in any field of study). This is different from the work of [Mirbabaie, Stieglitz and Volkeri \(2016\)](#), which is focused exclusively on the application domain of disaster management. Another advantage is the systematic process employed to develop the taxonomy. As a result of this, dimensions and characteristics that have not been identified in previous studies ([BORDOGNA et al., 2016](#); [SENARATNE et al., 2017](#)) were identified, such as object, reference, and criteria dimensions. Finally, the taxonomy provides a synthesis of the existing methods being employed to assess CGI quality, which adds a systematization of research in this area and thus complements previous works which have focused on summarizing existing studies in the literature related to quality assessment of CGI ([WIGGINS et al., 2011](#); [MIRBABAIE; STIEGLITZ; VOLKERI, 2016](#); [SENARATNE et al., 2017](#)).

An analysis of the methods presented in the taxonomy reveals opportunities for the development of new indicators and other methods to assess the quality of CGI. Furthermore, it allows researchers working with a particular type of CGI (e.g., social media) to learn about methods developed and employed in different types of CGI and application domains that could be transferred to their research focus.

As pointed out, quality assessment relies on the type and amount of data, the application domain, and the reference dataset available. As a consequence, only a few methods can be employed in different CGI sources and application domains. However, before their employment, there is a need to evaluate their appropriateness. If the methods available are not suitable for a specific purpose, then future investigation should be carried out in order to develop new method(s).

Unlike in existing works, the taxonomy was developed in a systematic way (i.e., the identification of existing methods through an SLR and development of the taxonomy). However, it still presents some limitations. SLR is a rigorous and systematic methodology, but there are some threats to its validity. These have been minimized by selecting several synonyms for both the main keywords, with the aim of discovering all the primary studies in the area. However, some synonyms were not employed in all electronic databases owing to their idiosyncrasies; in other words, some synonyms were excluded in some search engines because relevant studies could not be identified with them. The limited number of selected studies might be seen as a

consequence. In addition, the number of studies included might have been affected by language restrictions, as only studies written in English and Portuguese were taken into account. Thus, it is possible that some relevant studies were not included in this work.

3.6 Final Remarks

In this chapter, a taxonomy of methods for assessing CGI quality in the absence of authoritative data is proposed. Its development involved first looking at the state-of-the-art by conducting an SLR, where several works were found that employ quality assessment methods for CGI. After this, each method was scrutinized to determine its characteristics and, thus, be in a position to create the taxonomy. In addition, each method was described in detail, and its limitation(s) and potential application were discussed.

The taxonomy aims at assisting quality assessment in new and existing crowdsourcing-based platforms. Quality assessment is an important stage in all CGI, since the information comes from unknown sources and is of unknown quality. As well as this, the scientific community can also benefit from the results of the taxonomy, because it provides an overview of existing methods, but also offers scope for future research projects.

METHOD FOR QUALITY ASSESSMENT

4.1 Overview

This chapter introduces our method for the assessment of CGI quality in the flood management domain. Section 4.2 discusses the methods reported in the literature for the same purpose; Section 4.3 describes the hypotheses tested in our method and existing methods employed to assess CGI quality; Section 4.4 describes the case study and datasets used in this work; Section 4.5 discusses the materials and methods employed for the development of the work; Section 4.6 presents the results obtained; Section 4.7 discusses the main findings and contributions.

4.2 Related Work

Several methods for the assessment of CGI quality have been proposed (WIGGINS *et al.*, 2011; BORDOGNA *et al.*, 2016; MIRBABAIE; STIEGLITZ; VOLKERI, 2016; SENARATNE *et al.*, 2017; DEGROSSI *et al.*, 2018). An important aspect for such an assessment is the determination of the intended use of information (i.e., external quality), since quality depends on the context in which it is applied (BORDOGNA *et al.*, 2016). Researchers have investigated methods to assess the quality of CGI for flood management (Table 4). Degrossi *et al.* (2014) and Moreira, Degrossi and Albuquerque (2015), for instance, estimated the quality of water level observations by comparing them with authoritative data (i.e., sensor data). Ogie and Forehead (2017) also used authoritative data to evaluate the positional accuracy of flood-related Twitter messages. However, the lack of authoritative data in many parts of the world hampers traditional quality analyses that compare CGI to reference data sources. Consequently, alternative methods that assess the quality of CGI when authoritative data are unavailable have been proposed. Intrinsic methods aim to evaluate CGI quality without a reference dataset (i.e., ground-truth data), and researchers have explored features of the target dataset to be used as proxies for quality

measurements. Such intrinsic quality indicators provide approximate statements on the data quality (BARRON; NEIS; ZIPF, 2014).

Poser and Dransch (2010), for instance, evaluated the accuracy of crowdsourced observations regarding flood inundation depth by considering the characteristics of the geographic context. More specifically, the observations were compared against the inundation depth measured from water marks in 409 buildings in Eilenburg (Germany) obtained by means of a digital elevation model, risk model, land use, buildings' characteristics, etc. (SCHWARZ; MAIWALD; GERSTBERGER, 2005). Hung, Kalantari and Rajabifard (2016) developed a probability method for assessing the credibility of CGI and a set of geographic features was assumed to influence credibility and used as model predictors. They correspond to (i) distance to flood risk zones, (ii) DEM value of CGI point, (iii) distance to nearest CGI, and (iv) distance to water resource areas.

Feng and Sester (2018) extracted and learned characteristics of Twitter messages to determine high-quality relevant messages. Messages containing pluvial flood-related keywords were filtered and labelled according to historical weather data for the identification of relevant text messages on rainfall events. subsequently, natural language processing and deep learning algorithms were trained to learn the characteristics. A similar approach was employed for the extraction and learning of characteristics of rain- and flood-related Twitter photos. Arthur *et al.* (2018) also extracted and learned characteristics of tweets to identify messages describing floods "currently" occurring in the UK.

A common characteristic of the aforementioned studies is the evaluation of the quality of each CGI item. However, the increasing amount of CGI hampers the evaluation of individual data, especially in dynamic environments, as flood management. An advantage of the quality assessment through the aggregation of CGI elements is it can indicate areas of higher flooding probability and, thus, provide a "big picture" of the situation before decision makers scrutinize data, which facilitates a big data analysis. Longueville *et al.* (2010), for instance, derived the likelihood of a flood event taking place by converting raw CGI into spatiotemporal clusters, which were later evaluated. Nevertheless, this type of analysis has still been poorly explored.

Another characteristic is the employment of a single method for the assessment of CGI quality. Such use, however, explores only an one-sided perspective of the quality of CGI and does not examine a combination of distinct perspectives. Thus, there is still a lack of studies that investigate the possible combination of different methods for the same purpose. In addition, the aforementioned studies have employed a few methods reported in the literature and have not explored other methods so far.

In contrast to previous studies, in this work, the quality of CGI is assessed by aggregating CGI items into spatial units. The aggregation of CGI aims at assisting decision makers by providing a "big picture" of the situation and, hence, facilitating big data analyses. Moreover, CGI quality is evaluated through a combination of several methods reported in the literature. The goal is to investigate the relation between rainfall phenomena and social media messages and the

Table 4 – Previous studies on the assessment of CGI quality in the flood management domain.

Study	CGI Source	Method	Quality element(s)
Degrossi <i>et al.</i> (2014)	Crowd sensing	Comparison with authoritative data	Fitness for use
Moreira, Degrossi and Albuquerque (2015)	Crowd sensing	Comparison with authoritative data	Fitness for use
Ogie and Forehead (2017)	Crowd sensing	Comparison with authoritative data	Positional accuracy
Poser and Dransch (2010)	Crowd sensing	Geographic context	Attribute accuracy
Hung, Kalantari and Rajabifard (2016)	Crowd sensing	Geographic context analysis	Fitness for use
Feng and Sester (2018)	Social Media	Extraction/learning of characteristics	Credibility
Arthur <i>et al.</i> (2018)	Social Media	Extraction/learning of characteristics	Relevance
Longueville <i>et al.</i> (2010)	Social Media	Spatiotemporal clustering	Relevance
			Credibility

Source: Elaborated by the author.

way knowledge on geographic context and historical events strengthens this relation.

4.3 Quality Assessment Method

In Brazil, floods occur when either excessive rainfalls in a short period of time lead to overflowing rivers (i.e., flash flood), or heavy rainfalls saturate an urban drainage system (i.e., pluvial flood). The occurrence of rain and/or flood events influences Social Media activities. According to [Andrade *et al.* \(2017\)](#), social media users often provide rainfall-related messages 10 min before the actual phenomenon and the correspondence between rainfall data and rainfall-related messages is higher when the temporal scale increases. They also upload georeferenced flood-related photos at the same time traditional media talk about floods ([LONGUEVILLE *et al.*, 2010](#)). In addition, on-topic social media messages (i.e., flood-related georeferenced messages) are also more likely to occur near flood-affected areas ([ALBUQUERQUE *et al.*, 2015](#)). The location of the event is important for the creation of a context and use of what is known about the location ([CRAGLIA; OSTERMANN; SPINSANTI, 2012](#)). In light of such findings, the spatiotemporal characteristics of rain- or flood-related CGI correspond to the spatiotemporal characteristics of rain and/or flood events; therefore, spatial units with a large number of on-topic tweets (i.e., rain- or flood-related georeferenced messages) are more likely to be affected. Nonetheless, the spatial distribution of social media messages only represents population distribution, i.e., “tweets are more like to occur in places where people live” ([POORTHUIS *et al.*, 2016](#)). Social media activities, thus, are not homogeneous throughout the spatial units due to the difference in social media users density ([FENG; SESTER, 2018](#)), and each spatial unit has an average number of daily collected on-topic tweets. In addition, rain and/or flood events do not occur in isolation; in general, several spatial units can be affected. Consequently, a spatial unit has a higher probability of being affected if nearby units also have a large number of on-topic tweets, i.e., spatial units exhibit similar behavior.

A failure in the drainage system is not the only cause of pluvial flooding; its occurrence is also influenced by the topography of the city, gutters status and other factors. Therefore, analyses of the number and frequency of historical (pluvial) flood events provide additional information on flood-prone locations, i.e., locations of higher flooding probability. Such locations may also correspond to areas nearby rivers. According to the National Water Agency¹ (ANA, in Portuguese), an area nearby a river is considered more vulnerable to flooding based on the frequency of occurrence and impact degree of flood events ([Agência Nacional de Águas, 2018](#)). Since flood-prone rivers may have different extents and be situated in different parts of the watershed, spatial units with a larger extent of flood-prone locations are more likely to register a flood event.

The probability of floods decreases according to the presence of flood-control reservoirs,

¹ www.ana.gov.br

Table 5 – Hypotheses for the description of CGI quality.

Hypothesis	Description
H1	Spatial units with a large number of on-topic tweets are more likely to be affected.
H2	Spatial units with a large number of social media users have more on-topic tweets.
H3	Spatial units with larger extents of flood-prone areas are more likely to register a flood event.
H4	Spatial units with larger areas of flood-control reservoir have lower flooding probability.
H5	Spatial units with a large number of flood events registered have higher flooding probability.
H6	Spatial units with higher frequency of flood events have higher flooding probability.
H7	Spatial units have higher probability to be affected if nearby units also have a large number of on-topic tweets.

Source: Elaborated by the author.

which are man-made structures that store rainwater temporarily. Their construction along or parallel to rivers aims at storing the excess water and, thus, reduces the number of flood events. In general, each reservoir has a different storage capacity, since their size depends on the available area nearby the river. Therefore, spatial units with larger areas of flood-control reservoir have lower flooding probability.

Our method thus explores the relationship between a set of hypotheses (Table 5) and rainfall phenomena for determining the likelihood of occurrence of an event, i.e., its plausibility. Moreover, a set of indicators, which were derived from a set of the quality assessment methods of the taxonomy, is set upon which the hypotheses are tested (Table 6).

4.4 Case Study and Datasets

4.4.1 Study area

São Paulo was selected as the study area (Figure 8). This large city, with approximately 12,176,866 inhabitants (IBGE, 2018), is the capital of São Paulo state. Its tropical weather has dry winters and rainy summers and, therefore, flood events occur more frequently from December to March. Floods in São Paulo are mainly characterized by flash floods caused by excessive rainfalls in a short period of time, i.e., usually less than 6 (six) hours (National Weather Service, 2018).

Table 6 – Plausibility indicators for the description of CGI quality.

Hypothesis	Method	Indicator	Description
H1	Redundancy of volunteers' contribution	(x_1) number of on-topic tweets	Number of rain- or flood-related tweets within a spatial unit.
H2	-	(x_2) ratio	Daily average of on-topic tweets.
H3	Geographic context	(x_3) flood-prone areas	Extent of flood-prone rivers within a spatial unit.
H4	Geographic context	(x_4) flood-control reservoirs	Total area of flood-control reservoirs within a spatial unit.
H5	Historical data analysis	(x_5) number of historical flood events	Number of previous flood events within a spatial unit.
H6	Historical data analysis	(x_6) frequency of flood events	Frequency of occurrence of flood events, i.e., time interval between two events within a spatial unit.
H7	-	(x_7) autocorrelation between spatial units	Degree of similarity of observations in spatial units.

Source: Elaborated by the author.

Figure 8 – Study area.

Source: Elaborated by the author.

4.4.2 Datasets

Twitter data

Twitter² is a Social Media (KAPLAN; HAENLEIN, 2010) for users to post messages, known as tweets, which can be up to 140 characters long. Together with the message, users can also add *hashtags*, which are keywords or phrases preceded by the symbol #, which describes an interesting topic or a theme, so that other users can follow it more easily. Each tweet contains several metadata fields, e.g., a unique ID that identifies every tweet, a timestamp indicating when the user posted the message (timestamp values are always in Coordinated Universal Time (UTC)), and coordinates representing the exact location (latitude/longitude) from which the tweet was sent; however, only a small percentage of tweets is georeferenced (BURTON *et al.*, 2012). Different metadata fields describe the tweet's user, e.g., a unique identifier, a name, which does not necessarily represent a person's name, and a screen name denoting how the user identifies himself/herself.

Tweets can be automatically retrieved through an Application Programming Interface (API). Twitter's streaming API was used in the present study, since it provides access to 1% of the total tweets and offers a varying number of filters, such as language, user ID, location, etc. Two bounding boxes were defined for the retrieval of tweets; one covers the northern region of São Paulo city (-46.96, -23.62, -46.26, -23.21), whereas the other covers its southern region (-46.96,

² <https://twitter.com>

-24.11, -46.26, -23.62). The initial dataset contained 11,848,923 million tweets posted from 07 November, 2016 to 28 February, 2017. However, only 891,367 tweets had been geotagged (7,52%).

Authoritative data

Authoritative data were obtained from 4 (four) information sources. First, the map of flood-prone areas in Brazil (i.e., rivers prone to flooding) was obtained from the National Water Agency. The map was defined by practitioners from several national agencies that have analyzed the frequency of occurrence and degree of impact of flood events for each area. An area is classified as highly, moderately or slightly vulnerable if the time interval between two flood events is shorter than 5 (five) years, between 5 (five) and 10 (ten) years, and longer than 10 (ten) years, respectively. Flood-control reservoir data were acquired through the online system³ of the municipal administration of the city of São Paulo (GeoSampa, in Portuguese). Information on flood events was gathered from the Emergency Management Center⁴ (CGE, in Portuguese). The events are identified and reported to CGE by the Traffic Engineering Company⁵ of São Paulo, which manages and supervises the road system.

Finally, rainfall data derived from the weather radar data of the Department of Water and Power (DAEE, in Portuguese) and Polytechnic School of the University of São Paulo were used. The radar, located 60/70 km away from São Paulo city, provides rainfall rates in a spatial resolution of 250 meters per 1 degree azimuthal width. The previous datasets, i.e., CGE data and weather radar data, were obtained for the period from 07 November, 2016 to 28 February, 2017.

4.5 Methodology

This section describes the methodology employed in this study. Section 4.5.1 addresses the definition of the spatiotemporal resolution; Section 4.5.2 explains the obtaining of rainfall rates; Section 4.5.3 describes Twitter data preparation; Section 4.5.4 discusses the measurement of the plausibility indicators; finally, Section 4.5.5 reports on the statistical analysis conducted.

4.5.1 Determination of the spatiotemporal resolution

The analysis of social media data activity are dependent on the spatial and temporal resolutions (POORTHUIS *et al.*, 2016). For the aggregation of tweets and avoidance of the Modifiable Areal Unit Problem (MAUP), a 90 km² spatial resolution and a 12-hour temporal resolution for each day were adopted. Such spatiotemporal resolution (i.e., 90 km²/12 hour) provided the highest mean and lowest coefficient of variation throughout the Twitter data analysis for rainfall monitoring. The spatial units are represented as hexagonal cells, because of their

³ <http://geosampa.prefeitura.sp.gov.br>

⁴ <https://www.cgesp.org/v3/>

⁵ <http://www.cetsp.com.br>

Figure 9 – Spatial units of analysis.

Source: Elaborated by the author.

higher representational accuracy (POORTHUIS *et al.*, 2016) and more homogeneous spatial division, since the distance between the hexagons' centroids is always the same (LEE *et al.*, 2016). A 2-km buffer was created around the administrative limit of São Paulo city for the generation of the hexagonal cells, which covered the entire city (Figure 9). The vector layer of such cells was then generated upon the previous one and 35 hexagonal cells were obtained.

4.5.2 Obtaining rainfall data from weather radar

The weather radar of the Department of Water and Power (DAEE, in Portuguese) and Polytechnic School of the University of São Paulo obtained rainfall rates on a 5-min temporal resolution and 250-meter spatial resolution per 1 degree azimuthal width (SELEX, 2015). However, instead of working with polar rain yields, a radar product known as Surface Rain Intensity (SRI), projected in Cartesian coordinates with 500x500 meter resolution, was used to derive the rainfall rates (SELEX, 2015).

Such rates were combined in a 10-min temporal resolution for each location of the radar's

coverage area. In case of missing data, rainfall rates were generated through linear interpolation and aggregated into 90 km^2 spatial units. A spatial unit rainfall rate corresponds to the average of all rainfall rates within the unit. Finally, the rainfall rates were again combined for obtaining rates in a 12-hour temporal resolution.

4.5.3 Data preparation

The data preparation comprised three steps, namely tweets retrieval, keyword-based filtering, and content analysis and classification, explained in what follows.

Retrieval of georeferenced tweets

Georeferenced tweets were retrieved from the initial dataset (Section 4.4) according to location and publication date. Only tweets published between 07 November, 2016 and 28 February, 2017 and within the hexagonal cells were considered.

Keyword-based filtering

The tweets were filtered according to keywords for the identification of rain and flood-related information. Those containing Brazilian Portuguese keywords ‘chuva’ (rain), ‘garoa’ (drizzle), ‘tempestade’ (storm), ‘alagamento’, or ‘enchente’, or ‘inundação’ (flood) and ‘enxurrada’ (running rainwater) were included for content analysis and classification, regardless of capitalization. Those containing variations in the main keywords, as ‘chuva’ (chuvisco, chuvarada, etc.), ‘garoa’ (garoando), ‘tempestade’ (temporal, tempo ruim, etc.), ‘alagamento’ (alagado), and ‘inundação’ (inundado) were also included. The selection of this set of keywords was based on the most common words used by Brazilian citizens to talk about rain or flood events.

Content analysis and classification

After filtering, each tweet was analyzed independently by three expert researchers and classified as on-topic or off-topic (Chart 8). A tweet was classified as on-topic if its content contained information on a rain or flood event that occurred, was occurring or would occur, according to its publication date. On the other hand, a tweet was classified as off-topic if it did not fit the previous definition. In case of disagreement, the researchers classified it in the majority class. For instance, if two researchers classified a tweet as on-topic and one researcher classified it as off-topic, it was considered on-topic.

The researchers also analyzed the content of available hyperlinks (e.g., photos), since many keywords have other meanings. Keyword ‘garoa’ (drizzle), for instance, is used as a popular reference to the city of São Paulo, which is known as the land of drizzle.

The initial dataset contained 7,099 georeferenced tweets; only 3,089 tweets were classified as on-topic, while 4,010 tweets employed the keywords with another meaning, or were outside spatial units. According to [Andrade et al. \(2018\)](#), some keywords are noisy and, therefore,

Chart 8 – Examples of on-topic and off-topic tweets.

Date/time	Tweet (Portuguese)	Tweet (English)	Status
11.18.2016 11:24	Atrasada! Porém aqui! Chuva e trânsito. (@Bio Ritmo in Barueri, SP) https://t.co/kBEOcI4Fh3	I am late! But I am here! Rain and traffic (@Bio Ritmo in Barueri, SP) https://t.co/kBEOcI4Fh3	On-topic
11.10.2016 20:53	Depois da chuva #Sun- set @Rua Artur Azevedo https://t.co/ahi0nsW8SR	After rain #Sunset @ Rua Artur Azevedo https://t.co/ahi0nsW8SR	On-topic
01.25.2017 22:25	Fim de tarde (antes da chuva) #Mairiporã #saopaulo #grandeSP #cantareira #sp #clouds #day https://t.co/dnvzJ9MFCN	Late afternoon (before raining) #Mairiporã #saopaulo #grandeSP #cantareira #sp #clouds #day https://t.co/dnvzJ9MFCN	On-topic
01.09.2017 12:37	Mais chuva e neve de- vem atingir a Califór- nia após inundações https://t.co/jx4b1H9NZB #SDV Via @exame https://t.co/dEGtvWwJ7F	More rain and snow will reach California after floods https://t.co/jx4b1H9NZB #SDV Via @exame https://t.co/dEGtvWwJ7F	Off-topic

Source: Elaborated by the author.

may introduce false positive messages. The keyword ‘chuva’ (rain), for instance, was used to refer to a rain event, Brazilian food (“bolinho de chuva”), and song lyrics (e.g., “Chuva de arroz” and “Ela só quer paz”).

Determination of degree of agreement

The degree of agreement between the raters (researchers) was measured for the assessment of the reliability of tweets classification. The most elementary approach is the counting of number of cases agreed by raters; however, a certain amount of agreement is expected by chance (COHEN, 1960). Kappa coefficient is a quantitative measure of raters agreement and represents the proportion of agreement after the removal of the agreement by chance (COHEN, 1960). The coefficient determined the classification consistency.

4.5.4 Measurement of plausibility indicators

The quality of CGI was measured according to a set of indicators derived from the methods present in the taxonomy (Chapter 3). Next sections address the calculation of each indicator.

Calculation of number of on-topic tweets

The calculation of the number of on-topic tweets within a spatial unit was preceded by the definition of the spatial unit to which each tweet belonged. Afterwards, for each temporal

resolution (i.e., 12 hours), the number of on-topic tweets within the unit was counted. The calculation considered all tweets within a spatial unit, including multiple messages send by the same user.

Calculation of the extent of flood-prone areas

The city of São Paulo has several flood-prone areas heterogeneously distributed among the spatial units, i.e., the flood-prone river segments in some units are longer than in others. The length of flood-prone rivers within a spatial unit was computed after the determination of the ones intersecting it. Since the total length did not change during the period of analyses, the value obtained for each spatial unit was the same for each temporal resolution.

Calculation of the area of flood reservoirs

The total flood-control reservoirs area in a given spatial unit was measured after the proportion of reservoir contained within the unit⁶ had been determined. If the reservoir was within the spatial unit, its total area was considered. Conversely, if only part of the reservoir was within the unit, the part of this feature intersecting the unit⁷ was determined and its area was later computed. Finally, the areas of the reservoirs in each cell were summed. The total area of flood-control reservoirs did not change during the period of analyses and, therefore, the total area in each unit was the same for each temporal resolution.

Determination of number of historical flood events

The computation of the indicator was preceded by the definition of the spatial unit to which the flood event belonged. Subsequently, the number of events in each unit was counted and aggregated along the periods of time.

Determination of frequency of flood events

Prior to the establishment of the frequency of flood events, the spatial unit to which an event belonged was determined. On the one hand, if no event had been registered in a unit, the value ‘No event’ was assigned to it. On the other hand, the time interval between events was computed. According to the National Water Agency, the frequency of occurrence is high, moderate or low if the time interval between two events is shorter than 5 years, between 5 and 10 years, and longer than 10 years, respectively. According to the classification proposed and the time intervals computed, categories ‘High’, ‘Moderate’ and ‘Low’ were assigned to each spatial unit.

Measurement of autocorrelation between spatial units

After the number of on-topic tweets had been calculated, the Local Moran’s I coefficient was computed for identifying locations with a local spatial autocorrelation. The measurement

⁶ A feature is said to contain another if the interior of the former contains the boundary and interior of the latter.

⁷ Points of intersection of the features with the unit.

of the coefficient comprised three steps, namely neighbours definition, spatial weights matrix definition and Local Moran's I calculation. The conceptualization of spatial relationship, i.e., definition of neighbours, was performed by polygon contiguity. The contiguity condition was set to Queen's Contiguity (first order), since it considers polygons that share edges and corners as neighbours.

Subsequently, a spatial weights matrix was created to quantify the spatial relationship among the polygons. The weights were established according to row-standardized weighting, since it creates proportional weights when polygons have an unequal number of neighbours. Therefore, the row sum was one in each case. Given the spatial weights matrix, the value of Local Moran's I was computed for the identification of spatial clusters and spatial outliers (ANSELIN, 1995). A positive value of the coefficient indicates the polygon is surrounded by other polygons of similar values (i.e., spatial cluster) and a negative value indicates the polygon is surrounded by other polygons of dissimilar values (i.e., spatial outliers). Since the number of on-topic tweets in each spatial unit is different due to social media users density (POORTHUIS *et al.*, 2016), the measurement of the coefficient was based on the ratio.

A particular aspect of the coefficient significance assessment is individual locations will tend to be correlated, since spatial units might contain common elements (ANSELIN, 1995). Owing to this correlation, the coefficient significance of a spatial unit was assessed according to the Bonferroni criterion, which sets the significance cut-off at α/n . For example, the number of units is 35 and $\alpha = 0.05$; therefore, the significance at an individual level would be $\alpha_i = 0.0014$.

4.5.5 Statistical Analysis

The statistical analysis aimed at testing the hypotheses (Table 5), identifying significant indicators for the determination of CGI plausibility and assessment of the model's goodness-of-fit. Such objectives were achieved by regression analysis, which is a technique for the modelling of relationships between a set of independent variables and a dependent variable (MILES; SHEVLIN, 2001). The relationships are depicted by a regression equation, where the coefficients represent the average change in the dependent variable given a one-unit change in the independent variable, while the other independent variables are kept constant. In the study, the regression equation expresses the relationship between the rainfall rates (dependent variable) and plausibility indicators (independent variables).

Among several types of regression models, linear regression was selected because it is simple and often provides an adequate and interpretable description of the way independent variables affect the dependent variable (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). However, it is sensitive to outliers and multicollinearity.

Both independent and dependent variables were first transformed for the avoidance of outliers. The literature reports several transformation techniques, as square, cubic and fourth

root, min-max normalization, and standardization. The square root technique was employed for the transformation of the variables *rainfall rate*, *number of on-topic tweets*, *ratio* and *number of historical events*, whereas the cubic and fourth root techniques were applied to the variables *flood-prone area* and *flood-control reservoir*, respectively, since they had greater measurement ranges.

Multicollinearity occurs when independent variables are correlated⁸. As aforementioned, a key goal of regression analysis is the understanding of the relationship between dependent and independent variables. However, it cannot be achieved when multicollinearity occurs, since correlation indicates changes in an independent variable entail changes in another one. Multicollinearity was tested between the independent variables by the Variance Inflation Factor (VIF) (HAIR *et al.*, 2014). A 1.00 VIF value indicates variables are not correlated. On the other hand, higher VIF values denote higher multicollinearity degrees.

The value of the adjusted R-squared of the model was analyzed for the assessment of the goodness-of-fit of the model. The measure corresponds to the total amount of variance of the dependent variable explained by the independent variables (MILES; SHEVLIN, 2001). The model fits the data if it yields a high adjusted R-squared value.

4.6 Results

The results of the study are described in the following sections. The first provides a description of the data, and the second discusses the outcomes of the regression analysis.

4.6.1 Data description

The distribution of on-topic tweets throughout the spatial units was not homogeneous and concentrated mostly in the city center, where social media users density was higher (Figure 10). The spatial units with the greatest number of on-topic tweets correspond to Sé, Mooca, Jabaquara and Penha districts. On average, Twitter users posted 23 on-topic tweets per day during the study period. Conversely, the spatial units in the southern region did not register any on-topic tweet. A possible explanation is the southern region has mostly low income population.

The distribution of flood events was also heterogeneous. CGE registered 546 flood events from 07 November, 2016 to 28 February, 2017 mostly concentrated in the central and eastern regions of the city (Figure 11). Lapa, Pinheiros, Sé and Vila Mariana districts were the most affected, with 177 flood events, followed by Pirituba-Jaraguá, Freguesia-Brasilândia, Mooca, Vila Prudente and others, which registered 159 flood events in total. Those and other districts were constantly affected by floods, i.e., the frequency of occurrence was high. Conversely, either

⁸ “Correlation is a measure of the extent to which the variables are linearly related” (MILES; SHEVLIN, 2001)

Figure 10 – Distribution of on-topic tweets.

Source: Elaborated by the author.

the districts in the southern region have not been affected, or the Traffic Engineering Company has not registered any event.

The spatial distribution of flood-affected spatial units and on-topic tweets showed a correspondence. As several researchers have highlighted, the spatiotemporal characteristics of rain and/or flood events impact on the spatiotemporal characteristics of rain- or flood-related social media messages (LONGUEVILLE *et al.*, 2010; ALBUQUERQUE *et al.*, 2015; ANDRADE *et al.*, 2017). Nonetheless, on-topic tweets were also found in areas not affected by floods. An analysis of their content revealed they reported mostly rainfall events. In contrast, no on-topic tweets were registered in a spatial unit affected by floods, since it is located in a sparsely populated area.

The flood-affected spatial units also correspond to those with areas prone to flood (Figure 12). However, the unit mostly affected by floods is not the one with higher extents of flood-prone rivers. In fact, it has the fourth lowest extent of rivers prone to flood. For minimizing the occurrence of flood events, several reservoirs have been built in flood-affected spatial units. In general, the units with higher reservoir areas have recorded a lower number of flood events, as shown in Figure 13.

Figure 11 – Distribution of flood events according to CGE.

Source: Elaborated by the author.

Figure 12 – Extent of flood-prone area per spatial unit.

Source: Elaborated by the author.

Figure 13 – Area of flood-control reservoirs per spatial unit.

Source: Elaborated by the author.

Table 7 – Degree of correlation among the plausibility indicators.

Indicator	VIF value
(x_1) number of on-topic tweets	1.179071
(x_2) ratio	1.073001
(x_3) flood-prone areas	1.064529
(x_4) flood-control reservoirs	1.023091
(x_5) number of historical flood events	1.211491
(x_6) frequency of flood events	1.193397
(x_7) autocorrelation between spatial units	1.023737
(x_8) period of the day	1.013329

Source: Elaborated by the author.

4.6.2 Statistical analysis

The multicollinearity analysis revealed the plausibility indicators (Table 6) were slightly correlated, i.e., their VIF values were higher than 1.0 (Table 7); therefore, all indicators were considered in the regression analysis, since a change in an independent variable does not entail change in another variable. In particular, the *number of historical flood events* indicator showed the highest VIF value (i.e., 1.21).

Before the regression, an exploratory analysis revealed rainfall data were missing for

some temporal resolutions, although an interpolation technique had been applied; therefore, those observations were removed from further analyses for the avoidance of bias. It also showed eight spatial units in the southern region comprehended a forest region and, for almost all units, the independent variables had no values. Hence, those units were removed from the regression analysis and, as a consequence, a new α value (i.e., $\alpha = 0.05 / 27 = 0.0019$) was employed for the identification of spatial units with significant Local Moran's I.

The analysis also revealed a relation among rainfall rate, number of on-topic tweets and period of the day. A low number of on-topic tweets was registered when heavy rain had occurred after midnight; in contrast, a large number of on-topic tweets was posted when moderate rain occurred after midday. In light of this, the hypothesis raised is that “*on-topic tweets are more likely to be posted when rainfall events occur between midday and midnight*” (H8). Therefore, each day was divided in two periods, i.e., from midnight to midday and from midday to midnight. The period of the day was included in our regression model through a categorical independent variable.

As shown in Table 8, the existence of on-topic tweets within a spatial unit indicates the occurrence of an event, i.e., there is a relation between rainfall events and on-topic tweets indeed. In particular, the number of on-topic tweets within a unit increases along with the rainfall rate, since the results showed a positive relationship between them.

The number of on-topic tweets, however, varies across the spatial units owing to a difference of social media users density. The central region of São Paulo, for instance, has a higher users density, especially on working hours. The results revealed this difference in social media activity among spatial units is significant for the explanation of the relationship and supported our hypothesis. The *ratio* variable has a positive relation with rainfall rates and contributes to the percentage of explanation.

Additionally, as the exploratory analysis showed, the period of the day is an important indicator of the relationship between rainfall rates and number of on-topic tweets, i.e., there is a relation between rainfall events, on-topic tweets and period of the day, which confirms our hypothesis (H8). Moreover, the R^2 value increased 11.5% when the variable was added to the regression equation. This result reinforces an assumption highlighted by Longueville *et al.* (2010) at which a combination of (extreme) events and individuals to witness them is required.

The autocorrelation between spatial units is also a significant indicator of the likelihood of occurrence of an event. However, it showed a negative relationship with rainfall rates. A possible explanation is that events reported by citizens within a spatial unit may not be registered by its neighbours owing to the size of spatial units (spatial resolution). This result refutes our initial hypothesis (H7) at which a spatial unit has higher probability to be affected if its neighbours also register an event.

The analysis of historical flood events slightly strengthened the relationship between

rainfall rates and number of on-topic tweets. Although the frequency of occurrence of flood events suggests spatial units with higher flooding probability, the results of the regression analysis revealed this variable was not statistically significant and, therefore, our hypothesis (H6) was rejected. On the other hand, the number of historical flood events was a significant indicator and has a negative relationship with rainfall rates.

The knowledge on geographic context also slightly strengthened the relation. The regression analysis confirmed the extent of flood-prone areas within a spatial unit as a significant indicator; therefore, units with greater extent of flood-prone areas are more susceptible to flood. On the other hand, flood-control reservoirs aim at improving the resilience against flood by storing the excess water. The results confirmed our hypothesis (H4) and showed the flood-control reservoir variable was statistically significant. A change in the variable values, however, does not alter the dependent variable significantly, since their coefficient were lower in comparison to the others.

The plausibility indicators explained 14% of the variability in rainfall rates. The number of on-topic tweets, ratio and flood-prone areas had a positive relation with rainfall rates, whereas the flood-control reservoirs, autocorrelation between spatial units and period of the day had a negative one. On this basis, the initial hypothesis that there is a relation between rainfall phenomenon and social media messages, strengthened by knowledge on geographic context and historical events, was accepted.

4.7 Discussion

This chapter introduced a method for the assessment of CGI quality in the flood management domain. Our method explored a combination of existing methods that has not been used so far. In particular, the methods redundancy of volunteers' contribution and historical data analysis had not been employed for the assessment of CGI quality in the flood management. The main goal was the investigation of the relation among rainfall rates, on-topic social media messages, historical flood events and geographic context.

A set of hypotheses was established based on the characteristics of the domain and findings reported in the literature. For the measurement of each hypothesis, an indicator, which was derived from the methods selected, was developed. A regression analysis aimed at verifying the indicators that were statistically significant and, therefore, confirming or rejecting our hypotheses.

According to the analysis, the indicators were statistically significant, except for frequency of flood events, and explained 14% of the dependent variable. As reported in the literature, there is a relation between rainfall events and on-topic social media messages and the number of on-topic tweets within a spatial unit denotes the degree of plausibility of an event. Such result confirmed our hypothesis and reinforced the findings by [Andrade et al. \(2017\)](#); rainfall events

Table 8 – Coefficients and p-values of the plausibility indicators.

Indicators	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
(x ₁) number of on-topic tweets	0.26 (2.16e-15)	0.40 (<2e-16)	0.17 (0.000136)	0.17 (9.85e-05)	0.21 (9.04e-06)	0.21 (9.28e-06)	0.21 (5.77e-06)	0.21 (4.87e-06)
(x ₂) ratio			0.25 (1.25e-14)	0.26 (1.78e-15)	0.25 (1.12e-13)	0.25 (1.14e-13)	0.24 (2.46e-13)	0.24 (1.75e-13)
(x ₃) flood-prone areas							0.004 (0.0392)	0.01 (0.00788)
(x ₄) flood-control reservoirs							-0.01 (0.04607)	-0.01 (0.04607)
(x ₅) number of historical flood events					-0.02 (0.0281)	-0.02 (0.0704)	-0.02 (0.0124)	-0.02 (0.05907)
(x ₆) frequency of flood events						(0.8503)		
(x ₇) autocorrelation between spatial units				0.28 (0.0159)	-0.27 (0.0166)	-0.27 (0.0164)	-0.28 (0.0162)	-0.27 (0.01756)
(x ₈) period of the day		-1.11 (<2e-16)	-1.10 (<2e-16)	-1.10 (<2e-16)	-1.10 (<2e-16)	-1.10 (<2e-16)	-1.10 (<2e-16)	-1.10 (<2e-16)
Intercept	1.13 (<2e-16)	1.65 (<2e-16)	1.62 (<2e-16)	1.62 (<2e-16)	1.65 (<2e-16)	1.65 (<2e-16)	1.58 (<2e-16)	1.59 (<2e-16)
R-squared	0.01152	0.1263	0.1358	0.1366	0.1372	0.137	0.1377	0.1382

Source: Research data.

impact on social media activity and users post on-topic messages when they are affected by such events.

This relation, however, is influenced by the period of the day. Our analyses showed users normally post on-topic tweets when rainfall events occur after midday, whereas any or a few tweets are found for rainfall events occurring after midnight. Such finding was not reported by any related studies and is a contribution of our study. Longueville *et al.* (2010), for instance, created daily clusters, instead of 12-hour clusters and, therefore, their analysis was not influenced by the period of the day.

Additionally, our analyses confirmed social media users density is an important indicator of the relation between rainfall events and number of on-topic tweets, since the latter increases within spatial units with a large number of users. According to Poorthuis *et al.* (2016), the spatial distribution of tweets is strongly related to floating population density, since tweets are more likely to occur in places where people live or work. This aspect has not been explored by previous studies (Section 4.2) for the assessment of CGI quality and is another contribution of our study.

The autocorrelation indicator was statistically significant in our model, however it has a negative relationship with rainfall rates, and, in theory, this indicator should have a positive one. Although the number of on-topic tweets may vary across spatial units, adjacent units should, in a certain degree, behave similarly. The size of spatial units (i.e., 90 km^2) might have influenced this result. Further investigation should analyze the behaviour of this indicator for different spatial resolutions.

Flood-prone areas was a significant indicator for the explanation of CGI quality, as the result obtained by Hung, Kalantari and Rajabifard (2016). These areas are important features of the geographic context, since they show a higher likelihood of being flooded; thus, spatial units with greater extents of them are more likely to register an event. In addition, according to Albuquerque *et al.* (2015), on-topic tweets are more likely to be found near those areas. Another important feature of the geographic context, flood-control reservoirs were also statistically significant in our model. As previously mentioned, those reservoirs are a mechanism for storing the excess water in areas constantly affected by floods. However, as far as we know, previous studies have not explored this feature as an indicator for the assessment of CGI quality.

The number of historical flood events was also a significant indicator for the explanation of CGI quality, whereas the frequency of flood events was not. However, as the previous indicator, the former has not been used for such purpose yet. In Brazil, flood events also occur owing to failures of the drainage system and, as our result showed, spatial units with a large number of historical flood events have a higher flooding probability.

Another contribution of our study is the development of alternative models for the assessment of CGI quality (Section 4.6.2). The stepwise technique was used for the addition and removal of plausibility indicators, consequently alternative models were established throughout

the iterations. A common problem in disaster management is the lack of information sources, therefore, those models can be employed in such a case.

Our results, however, should be considered within the limitations of our study. Tweets content analysis revealed the messages contained hyperlinks with relevant information for flood management and the manual analysis enabled the use of this additional information for the classification of tweets. Therefore, in case of an automated approach, those content cannot be disconsidered. On the other hand, our analyses did not scrutinize tweet metadata, as publication date and coordinates, and, therefore, they were considered as a reference of the time and location of the occurrence of the phenomenon.

A linear regression model was selected for modelling the relationship between dependent and independent variables. Such model, however, does not capture the variation of the data structure over space and allow the modelling of different relationships at distinct point in space. To minimize this limitation, we included in our model a local indicator of spatial association, i.e., Local Moran's I or LISA ([ANSELIN, 1995](#)), which may be interpreted as indicator of local pockets of nonstationarity.

The aggregation of tweets was based on a 90- km^2 /12-hour spatiotemporal resolution. In particular, our analysis on the period of the day considered only two periods, i.e., from 12AM to 12PM and from 12PM to 12AM. We did not explore the day-night contrast of social media activity. In addition, further investigation should be performed for the analysis of plausibility indicators' significance in different resolutions and gathering more evidence on the results. Finally, our analyses should be replicated in different study areas towards a wider generalization of the results. Although the characteristics of floods in São Paulo city are similar to those in other cities, social media users may behave differently and the relationship between rainfall rate and on-topic tweets might not hold.

4.8 Final Remarks

This chapter introduced a method for the assessment of CGI quality in the domain of flood management when authoritative data is not available. A set of hypotheses was proposed for the explanation of CGI quality and a set of indicators were established for their measurement. The indicators were derived from methods reported in the taxonomy (Chapter 3). A linear regression was carried out for investigating the relationship between dependent variable (i.e., rainfall rates) and independent variables (i.e., indicators) and, thus, confirm or reject our hypotheses.

Except for frequency of flood events, all hypotheses were confirmed by the regression analysis, since the results showed the indicators were statistically significant. Linear regression, however, is a non-spatial model and does not take location into account in the analysis of the relationships between variables. Therefore, further investigation should be carried out for the modelling of different relationships over space. In addition, our analyses should be replicated in

different spatiotemporal resolutions and study areas towards a wider generalization of the results.

CITIZEN OBSERVATORY OF FLOODS

5.1 Overview

This chapter presents the design and development of the new interface of the Citizen Observatory of Floods. Section 5.2 discusses the Citizen Observatories reported in the literature for the gathering of information from citizens and usability aspects of emergency information systems; Section 5.3 is devoted to the description of the Citizen Observatory of Floods; Section 5.4 reports the methodology used for the development of the new interface; Section 5.5 describes and discusses the results of our work.

5.2 Related Work

Several crowdsourcing-based information systems have been developed towards gathering information from citizens in the natural disaster domain and used as an interaction mechanism between citizens and governmental agencies throughout emergency situations. Term “Citizen Observatory” refers to dedicated information systems for the gathering of specific and structured information from citizens through different devices (MIORANDI *et al.*, 2013).

The United States Geological Survey developed the system “*Did you feel it?*”¹ for collecting data from citizens who felt an earthquake and creating a map of damage extents (WALD *et al.*, 2011). A Citizen Observatory was also developed for gathering information from disaster-affected population after the Haiti Earthquake in 2010 (ZOOK *et al.*, 2012), in which individuals could provide information on infrastructure damage, services available, security threat, and public health. In the same year, an earthquake struck Chile and a Citizen Observatory aimed at collecting information from citizens on threats, response, vital lines, and emergency, which was used by emergency agencies and governmental institutions throughout

¹ <https://earthquake.usgs.gov/data/dyfi/>

disaster response (IACUCCI *et al.*, 2010). The Australian Broadcasting Corporation launched a crowdsourcing-based system to map flood-affected areas in Queensland during flood events in 2011 (MCDOUGALL, 2011). Similarly, the WeSenseIt project has developed a Citizen Observatory for collecting observations on water environment “with a focus on minimizing the effects of pluvial flooding and poor water quality” in Delft, the Netherlands (WESENSEIT, 2019). In the platform, citizens can share flood-related information or provide information on dangerous situations (LANFRANCHI *et al.*, 2014). Other *ad hoc* initiatives have developed Citizen Observatories for gathering information from citizens after floods in Pakistan² (2010) and Boulder³, USA (2011).

The development of citizen observatories, however, is not a trivial task, since those systems involve a wide diversity of users, as youngsters, the elderly and people with disabilities, who are not domain experts and have little formal qualifications. Furthermore, users often find themselves in critical situations with little time to provide and search for information. Therefore, usability constitutes a crucial aspect of any emergency information system, e.g., Citizen Observatories.

The aforementioned observatories have mostly been developed *ad hoc* after a disaster event and, therefore, do not comply with usability principles. According to Moreira, Degrossi and Albuquerque (2015), for instance, the Citizen Observatory of Floods was not considered an usable system by users, since they showed difficulty related to the submission of information. The literature reports a few studies that have investigated the usability and/or accessibility of emergency information systems. Wentz *et al.* (2014) evaluated the accessibility of sign-up pages for emergency alerts of counties in Massachusetts, New York and Maryland. The authors identified 21 emergency notification sign-up forms with one or more accessibility violations and the most problematic violations were the indication of required form fields and inaccessible progress indicators. Nurse *et al.* (2012) carried out a user-based evaluation of the usability of a Secure Situation Awareness System (SSA). According to the initial evaluation of SSA, further research should focus on facilitating customization of SSA interfaces, i.e., what should be displayed and how it should be shown, improving data input, and communicating information to users.

5.3 Citizen Observatory of Floods

Degrossi *et al.* (2014) proposed a crowdsourcing-based approach, which comprises a set of interpretation mechanisms and a crowdsourcing-based platform, for the obtaining of useful CGI in the domain of flood risk management in Brazil. The mechanisms aims at assisting citizens in interpreting environmental variables, and its development was based on three scenarios,

² <http://pakreport.org/ushahidi/>

³ <https://boulderflood2013b.crowdmap/>

namely controlled, semi-controlled and uncontrolled points, which differ according to the type of interpretation resource available in a riverbed.

The Citizen Observatory of Floods is a crowdsourcing-based platform⁴ for the collection of CGI on flood-affected areas and water level in riverbeds. Citizens must provide the following data for information provision: (i) *title*, which represents information main content; (ii) *description*, which describes citizen's observation; (iii) *category*, which represents an interpretation mechanism; and (iv) *place name*, which corresponds to the location from where citizen is sending the information (Figure 14).

Citizens can use both a Web site and a mobile application for sharing flood-related information. Particularly, the latter facilitates the provision of information, since citizens can provide information in near real time and “*in situ*” and there are almost 5 billion mobile phone users worldwide (STATISTA, 2019). However, as highlighted by Moreira, Degrossi and Albuquerque (2015), the interface of the citizen observatory presented usability issues; users had problems in interpreting the system interface at the time to submit information. In light of this, we developed a new interface for the Citizens Observatory of Floods by adopting usability and accessibility principles.

5.4 Methodology

Design science is a problem solving process that aims at creating and evaluating new and innovative IT artifacts to solve organizational problems (HEVNER *et al.*, 2004), as algorithms and implemented or prototype systems. The difference between routine design and design-science research is that the latter “addresses important unsolved problems in a unique or innovative ways or solved problems in more effective or efficient ways” (HEVNER *et al.*, 2004). In this chapter, a new interface and implemented system of the Citizen Observatory of Floods is the artifact designed. Hevner *et al.* (2004) proposed seven guidelines for design-science process to be complete. The six guidelines and their corresponding descriptions are summarized in Table 9. The seventh guideline corresponds to the communication of the research, which was already achieved by the publication by Degrossi *et al.* (2018).

As previously mentioned, usability aspects have been increasingly discussed among designers and developers of computerized systems. However, there is still a lack of studies that investigate how usability and accessibility principles should be employed to enhance the usability of crowdsourcing platforms. To address this problem, we developed an artifact that corresponds to a new interface and implemented system of the Citizen Observatory of Floods. For the development and evaluation of the artifact, user-centered design techniques were adopted, thus achieving the research rigor. The steps and technique adopted are detailed in the next section.

⁴ <http://www.agora.icmc.usp.br/enchente/>

Figure 14 – Old interface of the Citizen Observatory of Floods.

Add Report

Title

Enter a title for the report.

Description

Enter a description for the report.

Category

Select Category

Date and Time

May 11, 2018 10:51 AM

Location

Enter location

Map showing Africa, South America, and parts of Europe and Asia. Labels include: Oceano Atlântico Norte, Turquia, Iraque, Arábia Saudita, Argélia, Líbia, Níger, Sudão, Etiópia, Quênia, Tanzânia, Angola, Botsuana, África do Sul, Brasil, Oceano Atlântico Sul, Google.

Latitude Longitude

Photos

Add Photo

News

Enter news URL

Source: Elaborated by the author.

Chart 9 – Design-science research method.

Guideline	Description
Guideline 1: design as an artifact	The development of a new interface and implemented system of the Citizen Observatory of Floods.
Guideline 2: problem relevance	This research addresses the lack of studies that investigate strategies for the enhancement of the usability of crowdsourcing-based systems.
Guideline 3: design evaluation	The artifact is evaluated through a heuristic evaluation with usability experts and an usability test with actual users.
Guideline 4: research contributions	The contributions are represented by (i) a set of requirements for the development of citizen observatories, (ii) a new interface for the Citizen Observatory of Floods and (iii) an implemented system of the Citizen Observatory of Floods.
Guideline 5: research rigor	User-centered design techniques were adopted for both the construction and evaluation of the artifact.
Guideline 6: design as a search process	This research extends our previous work in developing an crowdsourcing-based approach for the obtaining of flood-related information.

Source: Elaborated by the author.

5.4.1 User-centered Design

User-centered Design (UCD) is a software design methodology that helps developers and designers to meet the needs of their users (NORMAN, 2013; LOWDERMILK, 2013). It aims at understanding the users, their needs and requirements for delineating the requirements for interactive computer-based artifacts to be usable. Thus, users are the main focus of the interface design, however, they are not always active participants (KATZEFF *et al.*, 2012). By placing users at the center of the design and development process, we remove ambiguity and, as a consequence, usability and product quality increases (ISO, 2010). Therefore, this methodology ensures that computerized systems maintains usability (LOWDERMILK, 2013). The methodology comprises four steps, namely (i) requirement analysis, (ii) (re)design, (iii) prototyping, and (iv) evaluation (Figure 15) (USABILITY.GOV, 2019).

UCD was employed for the design of the Citizen Observatory of Floods interface. Figure 16 depicts the steps performed throughout the design and development of the interface, showing in what years they had occurred. In the following, there is a discussion on the activities performed in each stage.

5.4.1.1 Requirement Analysis

Two data-gathering techniques, i.e., (i) questionnaire⁵ and (ii) interview⁶, were employed for the establishment of system's requirements. The questionnaire, which consisted of 27 open-

⁵ A list of questions to elicit specific information from a large group of people

⁶ An interviewer asks respondents a list of questions and records their answers

Figure 15 – Steps of the user-centered design process.

Source: Adapted from [usability.gov](https://www.usability.gov) (2019).

Figure 16 – Steps of the methodology.

Source: Elaborated by the author.

and closed-ended questions (Appendix B), characterized the users' profiles, identified the frequency at which they witnessed flood events, the way they responded to them, and the type of information shared and searched. Both types of question were combined, since closed-ended questions might limit answers or bias users towards to a certain response, whereas open-ended questions would provide more information (FARRELL, 2016).

The questionnaire was applied in two cities constantly affected by flood events, namely São Carlos/SP and Piracicaba/SP, from 10 November 2017 to 21 November 2017. Participants were randomly chosen, i.e., a student approached pedestrians walking in the city center and invited them to participate of the research and answer the questionnaire, voluntarily. Thirty individuals (16 from São Carlos and 14 from Piracicaba) answered the questionnaire; no participant mentioned any kind of disabilities.

The director of the Civil Defense, Mr. Pedro Caballero, responsible for the activities of the emergency agency, was also interviewed. The goal was to understand the actions taken by the Civil Defense before, during and after a flood event, the sources of information used to support its actions, type(s) of information gathered and made available to citizens, and type(s) of information it would like to receive. Ten questions were elaborated beforehand by two researchers (Appendix C). The 2-hour interview was held at the Institute of Mathematics and Computer Science of the University of São Paulo (Brazil) on 18 October 2017, recorded, and transcribed verbatim for the purpose of the analysis.

The answers from both questionnaire and interview enabled the identification of users' needs and establishment of the system's requirements, which were employed for the design of the new interface.

5.4.1.2 Design and Prototype

On the basis of the requirements, an investigation of the tasks users would perform to achieve a specific goal was carried out (BARBOSA; SILVA, 2010). Hierarchical Task Analysis (HTA) is a method that examines the tasks users perform and decomposes them into a hierarchy of sub-tasks (PREECE; SHARP; ROGERS, 2015). This method was employed for analyzing the tasks users should carry out to provide flood-related information in the former Citizen Observatory of Floods and, as an outcome, a diagram of those tasks was developed. Subsequently, a new HTA was proposed in order to optimize / minimize the number of tasks previously identified.

The new HTA was then used for the design of the new interface mockup⁷, which was developed by a student and presented to two researchers who are specialists in flood management and human-computer interaction. They evaluated it and suggested improvements regarding usability and flood management issues. When the prototype had reached a maturity level, a

⁷ A paper prototype.

prototyping tool, called Justinmind⁸, was adopted for the design of the interface final versions, since it enables the creation of high-fidelity prototypes and reproduction of the dynamic behavior of a system. The design of the interface emerged interactively through seven (7) “design-evaluation-redesign” cycles among the student and researchers.

All design decisions were made according to characteristics of flood management, users’ needs and four prioritized usability criteria, namely Task Conformance, Dialog Initiative, Familiarity and Consistency, which specify (i) the degree at which the system supports users’ tasks, (ii) the dialogue flow controller, (iii) prior knowledge on the task(s) and other systems, and (iv) system’s behaviour in similar situations, respectively (DIX *et al.*, 2004).

5.4.1.3 Evaluation

The prototype was evaluated according to two methods. The first verified the system compliance with users’ requirements, analyzing its functionalities and identifying potential problems, whereas the second validated the effectiveness of the users’ interaction. The methods are described in what follows.

Inspection-based Evaluation

Heuristic Evaluation (HE) (NIELSEN; MOLICH, 1990) was chosen for the inspection-based evaluation because it is one of the most popular methods used for the purpose of usability evaluation (SALGADO; FREIRE, 2014). Since it should be employed by a group of evaluators for the obtaining of the best outcome, four (4) experienced evaluators in usability analyzed the interface independently for detecting possible violations of the heuristics. Prior to the inspection, the main purpose of the Citizen Observatory of Floods and users’ profiles were briefly explained to the evaluators. It was proposed to each evaluator to perform, individually, a set of four typical tasks (Chart 10) for the interface inspection and detection of Nielsen’s usability heuristics’ violations. The evaluation was held at the Institute of Mathematics and Computer Science of the University of São Paulo (Brazil) from 27 February 2018 to 02 March 2018.

Chart 10 – The proposed tasks to be performed by the evaluators.

ID	Task
T1	Send a report on the water level in the river bed.
T2	Send a photo of a flood event.
T3	Search reports on rain intensity.
T4	Change the local(s) of interest to obtain information.

Source: Elaborated by the author.

Each violation identified was registered in a form, which consisted of a description of the Citizen Observatory of Floods platform and users’ profile, a list of Nielsen’s 10 heuristics (NIELSEN, 1994), and a list of violations. A violation comprehends a unique identifier (ID),

⁸ <https://www.justinmind.com/>

description of a problem, violated heuristic(s), and suggestion(s) for improvements. After the evaluation, all violations were summarized into a single document and labelled with the name of the evaluator who identified it. Similar violations were grouped as a single one.

User-based Evaluation

The Think Aloud technique (DIX *et al.*, 2004) was applied for the usability test. Users were encouraged to say out loud their thoughts, feelings, frustrations and progress while interacting with the interface. Their comments are important for improvements of the interface, especially when a comment is repeated multiple times.

Individuals were selected according to their ages (youngers and elderly), and knowledge of computers and flood management domain; most participants had little or no knowledge of computer and flood management. They were asked to perform two tasks, i.e., (T1) send a report on the water level in the river bed and (T2) send a photo of a flood event, similarly to those performed by the evaluators. Each participant interacted individually with the interface in an isolated room without external interference and the entire interaction was recorded for further analyses. The test occurred from 31 March 2019 to 13 April 2019.

Participants answered two questionnaires, one before and another after the test. The first aimed at obtaining information on participants' profiles and use of mobile devices by them (Appendix D), and the second, i.e., System Usability Scale (SUS) (BROOKE, 1996), aimed at measuring the usability of the system quantitatively (Appendix E).

SUS is a simple questionnaire in which ten statements are made and users indicate his or her degree of agreement or disagreement on a Likert scale, which ranged from strongly disagree to strongly agree (i.e., from 1 to 5). The main goal is the provision of a global view of system usability (BROOKE, 1996), which is computed according to some rules. For odd values, a user's answer is subtracted by one (1), and for even values, it is subtracted from five (5). The values are then summed and multiplied by 2.5.

5.5 Results and Discussion

5.5.1 Understanding users

According to the answers, all users have mobile devices, as mobile phones and tablets; however, only 87% use them everyday. In general, they use mobile devices to send and receive emails (20%) and/or messages on WhatsApp⁹ (83%), and access Social Media (97%), among which the most common ones are Facebook¹⁰ and WhatsApp (87%), followed by Instagram¹¹

⁹ <https://www.whatsapp.com/>

¹⁰ <https://www.facebook.com/>

¹¹ <https://www.instagram.com>

Figure 17 – Most used Social Media.

Source: Elaborated by the author.

(30%), Twitter¹² (10%), and YouTube¹³ (3%) (Figure 17). The elements of the new interface were chosen according to their similarity to those present on the most popular social media interfaces, since the criterion “Familiarity” suggests the use of user’s prior knowledge on existing systems for the development of new systems. For example, the menu icons are similar to the icons on Instagram and other interface elements were based on those of Facebook.

Only 50% answered positively when questioned about living or working in flood-prone areas, and only 40% of those have witnessed flood events more than 4 times (7%), 3 times (10%), twice (10%), and once (17%). Additionally, most users had never conveyed information on rain or flood events (67%) and some had shared warnings and photos with friends and relatives (33%) by WhatsApp (23%), SMS (10%) or phone call (7%). The low number of users who shared flood-related information might be related to the number of users who had witnessed flood events.

When asked about searching for information on rain and flood events, 57% answered negatively, whereas 43% reported searching for rain- or flood-related information on Facebook (20%), WhatsApp (3%), Internet (3%), traditional media (3%), email (3%), others (10%), and no answer (63%) (Figure 18). This result suggests users were more familiar with the elements of Facebook interface and, therefore, elements similar to those present on Facebook were incorporated in the interface, e.g., the timeline element.

Finally, users were questioned if they would use a specific system (e.g., Citizen Observatory of Floods) to share and search for rain- and flood-related information. 70% answered

¹² <https://www.twitter.com/>

¹³ <https://www.youtube.com>

Figure 18 – Systems employed by users to search rain- or flood-related information.

Source: Elaborated by the author.

positively, whereas 27% reported they would not, and 3% did not answer. The users, who have answered negatively, were asked about their willingness to use it and 63% answered yes if it would not be long. This result suggests the number of tasks users would perform to share and search for information should be as small as possible and reinforces the main goal in the elaboration of the HTA of the new interface. The participants were also inquired on the type of information they would share with others, and they answered flood risk (67%), increasing level of water on streets (53%), rain intensity (37%), and water level in a river bed (20%); and 10% did not answer (Figure 19). Such information was included in the system, so that users could share it with others.

The answers of Mr. Pedro Caballero, director of the Civil Defense, were analyzed after the interview. The agency uses different sources of information, as weather stations, pluviometers, water level sensors and "human as sensors" (GOODCHILD, 2007), to assist in their actions before, during and after a natural disaster. The Civil Defense has trained a group of volunteers to provide flood-related information and help other citizens to evacuate an area in case of a flood event. These volunteers communicate with the Civil Defense through phone calls and radio. The agency also monitors social media messages, however, it does not use them, because the users' perceptions of an event influence the information content.

"For example, a street might be flooded according to a social media message, but water does not block the passage of cars, and the river is not overflowing".

Pedro Caballero

The Civil Defense also provides information to citizens who have registered their phone

Figure 19 – Types of information users would share.

Source: Elaborated by the author.

Table 9 – Requirements of the Citizen Observatory of Flood based on user needs.

ID	Requirement
R1	The system should enable users to send photos
R2	The system should inform individuals on rain- or flood-related events
R3	The system should enable users to search for rain or flood-related information
R4	The system should enable users to share rain- or flood-related information
R5	The system should enable users to provide information on flood likelihood, water level on streets and rain intensity
R6	The system should enable the Civil Defense to sent warnings.

Source: Elaborated by the author.

numbers. Warnings are manually sent by the director through WhatsApp messages, which is a time-consuming task. Not all citizens receive the message, therefore, the development of a system for the sharing of rain- and flood-related information is fundamental also for sending warnings and receiving information from citizens.

The answers from the questionnaire and interview enabled the identification of users' needs and establishment of the requirements of the system. They were also employed for the design of the new system interface, and all decisions made in the design process were based on them. The requirements are shown in Table 9.

5.5.2 Design and Prototyping

Understanding of user needs and their requirements provided the foundation for the development of the new interface. According to the former interface, users had to perform 10

tasks to share flood-related information (Figure 20). However, in a natural disaster scenario (e.g., flood event), users find themselves in a critical situation and, therefore, must respond fast to adverse situations. Since their time to understand the interface's elements and share information is limited, the design aimed at reducing users' cognitive load by minimizing the number of tasks and making the interface more intuitive. Therefore, the number of tasks is lower in the new HTA (Figure 21), e.g., users perform 5 tasks (tasks 1, 2, 2.1, 3 and 5) for the provision of flood-related information.

According to the requirements, users must be able to send photos and share rain- and flood-related information, as flood likelihood, water level on streets and rain intensity (R1, R4 and R5). To facilitate their provision, such information was portrayed as buttons in the new interface (Figure 22b). In addition, since system's users comprehend all citizens, the needs of people with disabilities and the elderly were also considered for the interface design. Individuals with low vision have difficulty in reading texts, therefore, a bigger text font size and contrast between text and background were adopted. Moreover, there is a space between buttons and their sizes correspond to the devices widths for preventing users from clicking on more than one button concomitantly. Those design decisions were based on the literature guidelines on usability and accessibility principles (W3C Working Group, 2009; W3C Working Group, 2010; CHI, 2013).

Users must also be able to search for rain- or flood-related information (R3). Since Facebook is the most used social media, all information posted is displayed on a timeline; and users can filter it according to their type or location (Figure 22e), e.g., they can filter information within a given radius. The filtering options are exhibited on a menu bar (Figure 22e). The filtering functionality can be accessed by users through a superior menu. Another functionality is user's profile, at which users can turn on GPS, set a region of interest and check all his/her information posts (Figure 22f).

After the design stage, inspection- and user-based evaluations checked if all design decisions were in accordance with usability principles and validated the user interaction effectiveness.

5.5.3 Heuristic Evaluation

The heuristic evaluation revealed 36 violations mainly related to the header of the app, login page, user's profile page, main and inferior menu, and functionalities of the sharing and searching for information. Some problems violated more than one heuristic and/or were identified by more than one evaluator. All problems are shown in Table 10. Particularly, the second column presents the Problem ID, which was codified by a letter and number. The first corresponds to the first letter of evaluators name (i.e., A, F, O and S), whereas the latter represents the number of the problem found by an evaluator.

Analyzing the common usability problems, we observe that a first problem identified

Figure 20 – Hierarchical Task Analysis of the old interface.

Source: Elaborated by the author.

Table 10 – Problems identified by evaluators.

Category (Interaction type)	Problem ID	Description
Login	A1 and S1	The system does not provide any login feedback for users.
Help and documentation	O3 and S3	The system does not provide any help or documentation.
User's profile	A2, A7, O12 and S10	The profile icon does not correspond to the one used by other systems. Moreover, it should be placed together with the functionalities.
User's profile	A12, A15, F7, O10 and O11	The radius of interest of the profile is complex for lay person and does not provide a numerical feedback
User's profile	A16	The checkbox for the GPS activation is too small
User's profile	O11	The local of interest (position on the map) does not change.
User's profile	A15	The profile page is too complex.
Sharing information	F6, O3 and S5	The "Other" button is not intuitive, especially for elderly, since its main functionality is to send a photo.
Sharing information	A4, A5, A6, F4, O1 and O2	The "Send" button seems to be selected and users may send information unintentionally. Moreover, the message "Thank you for sharing", showed after information provision, rapidly disappears.
Sharing information	A8, A13 and F3	The "Sharing Information" option suggests there is an option for sending a photo and the icons used to represent the flood phenomenon are not familiar
Filtering functionality	A9, A13, A14, O5, O6, S7 and S8	The filtering menu style is based on the desktop style, and the font size and contrast hinder the text reading.
Filtering functionality	F5	The filtering functionality is not showing only information of a category.
Inferior menu	S2	The inferior menu icons is not be intuitive for elderly.
Inferior menu	A11, O9 and S9	The position of the icons change from a page to another.
Superior menu	O8, S6	The superior menu icons is not be intuitive for elderly, since it does not correspond to the one used by other systems.
Other	A10	The homepage is inefficient.
Other	A3	There is too much text in the header.

Source: Elaborated by the author.

Figure 21 – Hierarchical Task Analysis of the new interface.

Source: Elaborated by the author.

by evaluators was the lack of help and documentation in the system; however, all systems should provide users with information on their task(s) and steps to be carried out (Heuristic 10). Therefore, the evaluators suggested the addition of a documentation explaining all functionalities, which has not been included yet.

The icon to access user profile did not correspond to the one used in other mobile applications and might be unfamiliar to users (Heuristic 2), and they might want to know its actual meaning (Heuristic 4). In addition, people with disabilities might have difficulty in understanding figurative language or specialized usage (W3C Working Group, 2008). Therefore, the evaluators suggested changing it to a common user's profile icon (e.g., Google Material Design¹⁴) and to the inferior menu for keeping consistency with the other system functionalities (Figure 23a).

¹⁴ <https://material.io>

Figure 22 – New interface of the Citizen Observatory of Floods.

Source: Elaborated by the author.

The position of the icons in the inferior menu changes from a screen to another, which might affect users' understanding of the meaning of different elements and their actions (Heuristics 1 and 4), and the system might not prevent a problem from occurring (Heuristic 5). Such unexpected changes (e.g., a change in the icon's position) may cause difficulties for users, especially those with disabilities (W3C Working Group, 2008). Therefore, the maintenance of system behavior might avoid users from getting confused. The evaluators also suggested highlighting icons when they were selected and adding a caption under them.

Users could share photos by selecting option "Others", which may not be intuitive for some of them, i.e., the system did not speak users' language or use concepts familiar to them (Heuristic 2). Particularly, people with certain disabilities and the elderly may have difficulties in understanding non-literal and specialized words or figurative language (W3C Working Group, 2008), therefore, the system should always use words or language of same meaning to all users. Since the main functionality of this option is to send a photo, evaluators proposed changing its name to "Send photo".

Flood-related information can also be shared by users. However, according to one of the evaluators, this option suggests users can attach a photo; thus, the evaluator proposed then the addition of this functionality. Such suggestion was not accepted, because the provision of this information is assisted by a set of mechanisms that help citizens to better interpret environmental variables (DEGROSSI *et al.*, 2014).

When a user shared rain- or flood-related information, the system showed the message "Thank you for sharing", which rapidly disappeared and hindered the interaction of the people with disabilities, e.g., blind people, or people with low vision, cognitive limitations, or motor impairments, with the content. Therefore, increasing the time of interaction would enable them to react, type, locate things on screen and operate controls. The evaluators proposed the confirmation message should be similar to the one provided by Google Material Design.

The filtering information functionality followed a desktop system style (dropdown menu), which is not used in mobile applications, and its contrast hindered the reading of users. According to the evaluators, the system should not provide information that is irrelevant or rarely necessary to users (Heuristic 8) and should always keep consistency. They suggested adopting elements used in Social Media, where this functionality can be found in the application header (Figure 23e), and improving the contrast between the text and its background, so that people with low vision could read. They also highlighted the lack of accelerators that enable expert users to tailor frequent actions (Heuristic 7), e.g., after filtering information, a user must undo an action by clicking on the "Back" button to exhibit all information available. An evaluator proposed the inclusion of another element with similar functionality, which was not added yet.

For each problem identified, the evaluators suggested improvements, and most of them were implemented in the new version of the prototype (Figure 22). Thus, by adopting evaluators suggestions, we enhanced the usability of the new interface. However, some suggestions were

Table 11 – Frequency of violated usability heuristics.

N.	Heuristic	Frequency
1	Visibility of system status	26%
2	Match between system and the real world	28%
3	User control and freedom	0%
4	Consistency and standards	11%
5	Error prevention	9%
6	Recognition rather than recall	2%
7	Flexibility and efficiency of use	13%
8	Aesthetic and minimalist design	9%
9	Help users recognize, diagnose and recover from errors	0%
10	Help and documentation	2%

Source: Research data.

not taken into account, because the design decisions were also based on previous studies on the flood management domain. For instance, an evaluator reported “*the icons used to share information about a flood event are unfamiliar*”, however they were proposed by Degrossi *et al.* (2014) for the obtaining of useful information for flood risk management in Brazil.

The problems violated several heuristics, as shown in Table 11, except Heuristics 3 (User control and freedom) and 9 (Help users recognize, diagnose and recover from errors). The system supports undo and redo actions if users choose a system functionality by mistake, and the interface was designed to be as intuitive as possible for avoiding error messages.

5.5.4 User-based evaluation

As shown in Table 12, twelve individuals participated in the test; their age ranged from 22 to 77 years old (7 younger and 5 elderly) and most of them were women (8 women and 4 men). The educational level of participants also varied; eight individuals were graduated, one was an undergraduate student, one had finished high school, and two had not finished elementary school.

The majority of participants had mobile devices for more than 5 years (10 participants); and 2 participants for approximately 6 months. Ten participants often used mobile devices more than 4 times a day, 1 participant less than once a day and another participant more than once a week. Most individuals had no difficulty in using mobile devices, however 3 individuals (P10, P11 and P12) had problems in remembering the steps for the use of such devices or popular mobile applications, e.g., WhatsApp¹⁵.

No individual was afraid of using of mobile devices, however, two reported feeling insecure about the way of using some applications, i.e., P10 and P11. Eight individuals rarely or never felt lost while using mobile devices, differently to four individuals, i.e., P3, P8, P11

¹⁵ <https://www.whatsapp.com/>

Figure 23 – New system interface after heuristic evaluation.

Source: Elaborated by the author.

Table 12 – Description of participant profiles.

ID	Gender	Age	Educational Level
P1	Male	22	College (Incomplete)
P2	Female	27	College (Complete)
P3	Female	27	College (Complete)
P4	Female	30	College (Complete)
P5	Female	31	College (Complete)
P6	Male	32	College (Complete)
P7	Male	37	College (Complete)
P8	Male	61	High school (Complete)
P9	Female	64	College (Complete)
P10	Female	65	College (Complete)
P11	Female	76	Elementary school (Incomplete)
P12	Female	77	Elementary school (Incomplete)

Source: Research data.

and P12 (1 younger and 3 elderly), who always or sometimes did not know what to do. Two of them, P11 and P12, were taking classes on the use of mobile phones and tablets at the Institute of Mathematics and Computer Science of the University of São Paulo, Brazil, and one individual (P3) reported difficulties with the Operating System, i.e., iOS¹⁶. Additionally, those three participants sometimes felt without knowing what actions performing while using mobile devices and the other participants rarely or never had this feeling. Four participants, i.e., P3, P10, P11 and P12 (1 younger and 3 elderly), always or sometimes asked for help to use their mobile devices, while the others rarely or never asked for help. Finally, eight participants reported understanding error messages when they occurred and four elderly had never understood them (P9, P10, P11 and P12).

The evaluation with users revealed many problems in the interface. The analyses of the saved videos (of testing users) allowed to identify the main matters/issues. The first task (T1 in Chart 10) consisted in “sending a report on the water level in a river bed”. One young (P6) and the elderly participants did not recognize the inferior menu and, therefore, started analyzing other interface elements. Two elderly participants clicked on the Option button, which provides information filtering options, on the interface right top corner, for searching for a way to share the information. The other participants read all information available in the timeline and clicked on the Like button of the information that best represented the scenario. When the young participant (P6) was asked about the reason of such action, he said “*I liked the report that best represented the scenario described and did not occur to me to click on the Inform button*”. According to the participants, they did not recognize the menu, since the icons and their labels did not represent their corresponding actions, the label under each icon was too small, and the menu position in the interface did not correspond to the desktop applications, in which the menu is in the header.

¹⁶ <https://www.apple.com/ios/home/>

The analyses of the videos also revealed the icons contrast hampered their identification by users. For instance, the icons did not pop when mobile device's brightness was low.

The second task (**T2** in Chart 10) consisted in the “sending of a photo of a flood event”, which involved taking a picture, writing a comment and locating the flood event on a map. However, the task was not clear for two young participants (P2 and P3); they asked if they should send a photo and also report flood level on streets. Such questioning suggests the task was not intuitive. The participants decided to send only the photo, which was required by the task. Regarding location, three participants, i.e., P2, P7 and P10 (2 youngers and 1 elderly), asked if they should provide the location of the flood event on the map (Figure 23d) or if would be automatically determined. This question was raised because no additional information was provided to the users on the map, and the location was automatically captured by GPS in the first task. A young participant (P2) also highlighted the location of the event was not published with the photo in the timeline and, therefore, she suggested its inclusion, since it may facilitate the identification of event's location by other users.

In both tasks, confirmation message was not exhibited to users after information provision and an elderly (P8) was discontent because his comment was submitted with typos. Therefore, users must confirm their messages before they are actually sent. Another elderly participant (P10) also expressed discontentment with the undo functionality, because the system did not provide an undo button and the one available in the device always closed the mobile application.

In general terms, the evaluation has revealed **T1** was the most problematic since six participants considered it difficult, and **T2** was not intuitive to two participants (Figure 21). A participant (P4) said the provision of information is quick; however, users must first think before taking action. The first comment reinforces our main goal of reducing cognitive load by minimizing the number of tasks; however, as the participant highlighted, the design was not completely successful in making the interface more intuitive to users.

According to participant answers (Table 13), the interface scored 49 points, on average, in the SUS questionnaire and this result reinforces the issues identified by the Think Aloud technique. This average is lower among young participants (i.e., 46 points) and higher among the elderly (i.e., 52 points), although the latter participants faced more issues in the provision of information. Additionally, the average remains (almost) the same among female and male participants, i.e., 49 and 48 points, respectively.

5.6 Final Remarks

This chapter presented the design and development of the new Citizen Observatory of Floods interface. Initially, a questionnaire was applied for the identification of users' needs and thirty individuals answered it. The director of the Civil Defense was interviewed for the determination of the types of information sources the agency uses. An analysis of the answers

Table 13 – Participants answers in the SUS questionnaire.

ID	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7	Q.8	Q.9	Q.10	Total	Score
P1	1	3	1	3	3	3	1	3	1	3	22	55
P2	1	0	4	0	3	0	4	0	4	0	16	40
P3	1	3	4	0	0	0	4	0	1	3	16	40
P4	3	3	1	3	3	3	1	3	1	3	24	60
P5	4	0	4	0	0	0	4	0	4	0	16	40
P6	2	3	1	3	3	3	1	3	1	3	23	57.5
P7	1	3	1	0	3	0	1	0	4	0	13	32.5
P8	1	3	4	0	3	0	4	0	4	0	19	47.5
P9	1	3	1	3	1	0	4	0	1	3	17	42.5
P10	1	3	2	1	2	3	3	1	1	4	21	52.5
P11	3	4	3	4	2	0	1	1	4	4	26	65
P12	0	4	0	4	2	2	1	4	1	4	22	55

Source: Research data.

(questionnaire and interview) enabled the establishment of system's requirements.

User needs and their requirements were employed for the design of the new interface. An important distinction between old and new interfaces is the number of tasks users must perform for the provision of rain- and flood-related information; users must carry out 10 tasks in the former and 5 tasks in the latter.

For the design of the new interface, different usability principles were applied. Many interface elements, for instance, are similar to Social Media elements due to user familiarity with them. An heuristic evaluation and usability test were carried out to evaluate the usability of the new interface. Several problems were reported by evaluators and users, which will be included in a new version of the system.

By designing the new interface, we aimed at enhancing the usability of the Citizen Observatory of Floods, because users have faced problems when providing flood-related information in a previous study ([MOREIRA; DEGROSSI; ALBUQUERQUE, 2015](#)). The improvement of the system's usability contributes towards the improvement of information quality, since users would have a better understanding of interface elements and, as a consequence, the probability of providing inaccurate information would be lower.

CONCLUSIONS

Crowdsourced Geographic Information has emerged as an alternative source of geographic information in many domains owing to a number of key features; it is free, updated and abundant. CGI has been increasingly adopted in the domain of flood management for the detection of rain and flood events, improvement of flood forecast and complementing of authoritative data. The use of CGI, however, is always accompanied by concerns regarding its quality; the information that is supplied by non-experts mostly does not have to comply with any quality standards, and there is no centralized control over the creation process. Therefore, research focusing on the development of methods for quality assessment is important to consolidate the use of CGI in this domain.

This doctoral thesis contributes towards the assessment of CGI quality in the flood management domain by proposing an approach for this purpose. Achievements of this doctoral work include the identification of methods for quality assessment of CGI, the establishment of a taxonomy of methods for the assessment of CGI quality in the absence of ground-truth data, the development of a method for the evaluation of CGI quality in the domain of flood management, the establishment of requirements for the development of crowdsourcing-based platforms for flood management, and the design and development of an interface for the Citizen Observatory of Floods.

The contributions described along the doctoral thesis are summarized in Section 6.1. Section 6.2 discusses the limitations of the work and presents directions for future research.

6.1 Thesis Contributions

This section describes the main contributions of this doctoral work.

- **Definition of a taxonomy of methods for quality assessment:** a taxonomy of methods for the assessment of CGI quality when no reference data are available was defined (see

Chapter 3). The definition first involved looking at the state-of-the-art by conducting an SLR for the identification of studies on CGI within the scope of quality assessment. The taxonomy was established according to 18 studies. The eleven methods of the taxonomy were described in detail and its limitation(s) and potential application were discussed. The taxonomy aims at assisting quality assessment in new and existing crowdsourcing-based platforms. The scientific community can also benefit from the results of our taxonomy, because it provides an overview of existing methods and offers scope for future research projects.

- **Definition of a method for the assessment of CGI quality:** a method for the assessment of CGI quality in the domain of flood management was developed (see Chapter 4). A set of hypotheses was proposed for the explanation of CGI quality. The novelty of our method is the combination of existing methods, proposed in the taxonomy, for the measurement of such hypotheses. By developing the method, we aimed at assisting decision makers, emergency agencies and citizens in determining the likelihood of the occurrence of an event.
- **Requirements for crowdsourcing-based platforms:** a questionnaire was applied for the identification of user needs and a set of requirements was obtained from the analysis of the answers.
- **Design and development of an interface for citizen observatory:** we designed and developed a new interface for the Citizen Observatory of Floods (see Chapter 5). The interface was designed according to user needs, their requirements and usability principles. The improvements on the system usability aimed to overcome the limitations of the existing citizen observatory¹.

This doctoral work comprises an interdisciplinary research and contributes to the areas of Computer Science and Flood Management by the establishment of the taxonomy of quality assessment methods, the method for the assessment of CGI quality, and the design and development of a system for the gathering of CGI.

6.2 Limitations and Future Work

This section summarizes the main limitations of this doctoral work and highlights future directions of research in the areas associated with the assessment of CGI quality and design and development of crowdsourcing-based platforms.

Limitations of the systematic review: we did not include all synonyms of quality assessment and CGI in the literature review owing to the idiosyncrasies of the electronic databases. In

¹ <http://www.agora.icmc.usp.br/enchente/>

addition, we restricted our search to studies written in Portuguese and English, and published between 2004 and 2015. Therefore, the review should be updated to include studies published after the final year. The search was only carried out in 6 electronic databases and we did not perform manual, backward or forward search. Such restrictions may have hampered the identification of relevant results.

Limitations of the taxonomy: the taxonomy might have not included all methods for the assessment of CGI quality when authoritative data is not available due to the limitations of the literature review. As a consequence, the number of dimensions and characteristics might have also be affected, since the empirical-to-conceptual approach analyzes objects (methods) for deriving them. Although the taxonomy was developed according to a systematic and rigorous process, we did not evaluate it with experts. The evaluation is important for the obtaining of more evidence on its correctness. Moreover, an analysis of the methods presented in the taxonomy reveals opportunities for the development of new indicators and other methods to assess the quality of CGI.

Limitations of the quality assessment method: georeferenced tweets were retrieved based the available location (i.e., coordinates). Some tweets, on the other hand, refer to phenomenon location by hashtags or in the message content. Therefore, strategies should be developed to extract those references towards geolocating tweets. Moreover, the aggregation of tweets was performed based on a 90-km²/12-hour spatiotemporal resolution. Further research should investigate the aggregation of tweets for other spatiotemporal resolutions in order to obtain a wider generalization of the results. A simple linear regression was used for modelling the relationship between dependent and independent variables and did not capture the variation of the data structure over space. Thus, other methods that explores spatial nonstationarity should be investigated. In addition, a few indicators were established for the explanation of CGI quality and future research should investigate other indicators in other to increase the percentage of explanation.

Limitations of the requirement analysis: a questionnaire was applied for the identification of user needs; however, most participants had never witnessed flood events, or shared and/or searched for information on flood events. Therefore, the answers gathered during the requirement analysis were mostly based on what users would do rather than what they actually do.

Evaluation of the interface: an user-based inspection was conducted to evaluate the new interface. Although the results provided preliminary evidence of the system usability, the interface still presents many problems, especially for the elderly. Therefore, we plan to redesign the interface and perform another evaluation in order to gather more confidence on the system usability. In addition, we will include accessibility principles in the redesign process.

BIBLIOGRAPHY

Agência Nacional de Águas. Vulnerabilidade a inundações. Accessed on: 09.13.2018. 2018. Available: <<http://metadados.ana.gov.br/geonetwork/>>. Citation on page 68.

ALBUQUERQUE, J. P.; HERFORT, B.; BRENNING, A.; ZIPF, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. **International Journal of Geographical Information Science**, v. 29, n. 4, p. 667–689, 2015. Citations on pages 42, 68, 79, and 85.

ALBUQUERQUE, J. P.; HERFORT, B.; ECKLE, M. The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. **Remote Sensing**, v. 8, n. 10, p. 1–22, 2016. Citations on pages 36 and 56.

ALBUQUERQUE, J. P.; HORITA, F. E. A.; DEGROSSI, L. C.; ROCHA, R. dos S.; ANDRADE, S. C. de; RESTREPO-ESTRADA, C.; LEYH, W. Volunteered geographic information and the future of geospatial data. In: _____. [S.l.]: IGI Global, 2017. chap. Leveraging Volunteered Geographic Information for improving disaster resilience, p. 158–184. Citations on pages 32 and 128.

ALBUQUERQUE, J. P.; ZIPF, A. Collaborative information systems for disaster management: building resilience against disasters by combining participatory environmental monitoring and vulnerability communication. In: **Alumi seminar natural hazards - research on natural disasters, civil defense, disaster prevention and aid**. Teresópolis, Rio de Janeiro, Brazil: [s.n.], 2012. p. 71–74. Citation on page 31.

ALI, A. L.; SCHMID, F. Geographic information science. In: _____. [S.l.]: Springer, Cham, 2014. chap. Data Quality Assurance for Volunteered Geographic Information, p. 126–141. Citation on page 59.

ANDRADE, S. C.; DEGROSSI, L. C.; RESTREPO-ESTRADA, C. E.; DELBEM, A. C. B.; ALBUQUERQUE, J. P. de. Does keyword noise change over space and time? a case study of social media messages. In: **Proceedings of the 19th Brazilian Symposium on GeoInformatics**. Campina Grande, Brazil: [s.n.], 2018. p. 116–121. Citations on pages 74 and 129.

ANDRADE, S. C.; RESTREPO-ESTRADA, C.; DELBEM, A. C. B.; MENDIONDO, E. M.; ALBUQUERQUE, J. P. de. Societal geoinnovation. In: _____. [S.l.]: Springer, 2017. chap. Mining Rainfall Spatio-Temporal Patterns in Twitter: A Temporal Approach, p. 19–37. Citations on pages 27, 68, 79, and 83.

ANHORN, J.; HERFORT, B.; ALBUQUERQUE, J. P. de. Crowdsourced validation and updating of dynamic features in openstreetmap: An analysis of shelter mapping after the 2015 nepal earthquake. In: **Proceedings of the 13th International Conference on Information Systems for Crisis Response and Management**. [S.l.: s.n.], 2016. Citation on page 36.

ANSELIN, L. Local indicators of spatial association - lisa. **Geographical Analysis**, v. 27, n. 2, p. 93–115, 1995. Citations on pages 77 and 86.

ANTONIOU, V.; MORLEY, J.; HAKLAY, M. Web 2.0 geotagged photos: assessing the spatial dimension of the phenomenon. **Geomatica**, v. 64, n. 1, p. 99–110, 2010. Citations on pages 27, 28, and 36.

ARSANJANI, J. J.; VAZ, E. An assessment of a collaborative mapping approach for exploring land use patterns for several european metropolises. **International Journal of Applied Earth Observation and Geoinformation**, v. 35, p. 329–337, 2015. Citations on pages 39 and 40.

ARTHUR, R.; BOULTON, C. A.; SHOTTON, H.; WILLIAMS, H. T. P. Social sensing of floods in the uk. **PLoS ONE**, v. 13, n. 1, p. e0189327, 2018. Citations on pages 66 and 67.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO 9000:2005 Quality management systems – Fundamentals and vocabulary**. 2005. Citation on page 37.

BALLATORE, A.; ZIPF, A. A conceptual quality framework for volunteered geographic information. In: FABRIKANT, S. I.; RAUBAL, M.; BERTOLOTTO, M.; DAVIES, C.; FREUND-SCHUH, S.; BELL, S. (Ed.). **Proceedings of the 12th International Conference Spatial Information Theory (COSIT)**. Santa Fe, NM, USA: Cham, Switzerland: Springer, 2015. v. 9368, p. 89–107. Citation on page 28.

BARBOSA, S. D. J.; SILVA, B. S. **Interação Humano-Computador**. Rio de Janeiro: Elsevier, 2010. ISBN 978-8-53-521120-7. Citation on page 95.

BARRON, C.; NEIS, P.; ZIPF, A. A comprehensive framework for intrinsic openstreetmap. **Transactions in GIS**, v. 18, n. 6, p. 877–895, 2014. Citations on pages 41, 60, and 66.

BATINI, C.; SCANNAPIECO, M. **Data and Information Quality**. [S.l.]: Springer, 2016. ISBN 978-3-319-24104-3. Citation on page 37.

BIOLCHINI, J.; MIAN, P. G.; NATALI, A. C. C.; TRAVASSOS, G. H. **Systematic Review in Software Engineering**. [S.l.], 2005. Available: <<https://www.cos.ufrj.br/uploadfile/es67905.pdf>>. Citations on pages 45 and 47.

BISHR, M.; JANOWICZ, K. Can we trust information? the case of volunteered geographic information. In: **Proceedings of the Towards Digital Earth: Search, Discover and Share Geospatial Data Workshop**. [S.l.]: Berlin, Germany, 2010. Citation on page 58.

BISHR, M.; KUHN, W. Spatial information theory. cosit 2013. lecture notes in computer science. In: _____. [S.l.]: Springer, Cham, 2013. v. 8116, chap. Trust and Reputation Models for Quality Assessment of Human Sensor Observations, p. 53–73. Citations on pages 37 and 58.

BODNAR, T.; TUCKER, C.; HOPKINSON, K.; BILÉN, S. G. Increasing the veracity of event detection on social media networks through user trust modeling. In: **IEEE International Conference on Big Data**. Washington, DC, USA: [s.n.], 2014. p. 636–643. Citations on pages 41 and 58.

BORDOGNA, G.; CARRARA, P.; CRISCUOLO, L.; PEPE, M.; RAMPINI, A. A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. **Information Sciences**, v. 258, p. 312–327, 2014. Citation on page 60.

_____. On predicting and improving the quality of volunteer geographic information projects. **International Journal of Digital Earth**, v. 9, n. 2, p. 134–155, 2016. Citations on pages 28, 29, 39, 42, 43, 44, 52, 61, 62, and 65.

BRERETON, P.; KITCHENHAM, B. A.; BUDGEN, D.; TURNER, M.; KHALIL, M. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of Systems and Software**, v. 80, n. 4, p. 571–583, 2007. Citations on pages 45, 46, and 47.

BROOKE, J. Usability evaluation in industry. In: _____. [S.l.]: London: Taylor and Francis, 1996. chap. SUS-A quick and dirty usability scale, p. 189–194. Citation on page 97.

BROWN, G.; WEBER, D.; BIE, K. de. Is ppgis good enough? an empirical evaluation of the quality of ppgis crowd-sourced spatial data for conservation planning. **Land Use Policy**, v. 43, p. 228–238, 2015. Citations on pages 39 and 40.

BURTON, S. H.; TANNER, K. W.; GIRAUD-CARRIER, C. G.; WEST, J. H.; BARNES, M. D. “right time, right place” health communication on twitter: Value and accuracy of location information. **Journal of Medical Internet Research**, v. 14, n. 6, 2012. ISSN 1438-8871. Citation on page 71.

CHI, E. H. The false dichotomy between accessibility and usability. In: **Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility**. Rio de Janeiro, Brazil: [s.n.], 2013. Citation on page 101.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960. Citation on page 75.

COMBER, A.; SEE, L.; FRITZ, S.; VELDE, M. V. der; PERGER, C.; FOODY, G. Using control data to determine the reliability of volunteered geographic information about land cover. **International Journal of Applied Earth Observation and Geoinformation**, v. 23, p. 37–48, 2013. Citations on pages 41 and 56.

CRAGLIA, M.; OSTERMANN, F.; SPINSANTI, L. Digital earth from vision to practice: making sense of citizen-generated content. **International Journal of Digital Earth**, v. 5, n. 5, p. 398–416, 2012. Citation on page 68.

CUI, Y. A systematic approach to evaluate and validate the spatial accuracy of farmers market locations using multi-geocoding services. **Applied Geography**, v. 41, p. 87–95, 2013. Citations on pages 40 and 57.

DAUGHERTY, T.; EASTIN, M. S.; BRIGHT, L. Exploring consumer motivations for creating user-generated content. **Journal of Interactive Advertising**, v. 8, n. 2, p. 16–25, 2008. Citation on page 27.

DEGROSSI, L. C.; ABE, B. B.; ALBUQUERQUE, J. P. de; FORTES, R. P. de M. Enhancing usability of a citizen observatory based on user-centered design. In: **Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion**. Thessaloniki, Greece: [s.n.], 2018. p. 294–301. Citations on pages 33, 91, and 127.

DEGROSSI, L. C.; ALBUQUERQUE, J. P. de; FAVA, M. C.; MENDIONDO, E. M. Flood citizen observatory: a crowdsourcing-based approach for flood risk management in brazil. In: **Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering**. Vancouver, Canada: [s.n.], 2014. p. 570–575. Citations on pages 27, 31, 42, 65, 67, 90, 106, and 107.

DEGROSSI, L. C.; ALBUQUERQUE, J. P. de; RESTREPO-ESTRADA, C. E.; MOBASHERI, A.; ZIPF, A. Exploring the geographical context for quality assessment of vgi in flood management domain. In: **Proceedings of the 23rd Americas Conference on Information Systems**. Boston, Massachusetts, USA: [s.n.], 2017. p. 1–10. Citation on page 128.

DEGROSSI, L. C.; ALBUQUERQUE, J. P. de; ROCHA, R. dos S.; ZIPF, A. A framework of quality assessment methods for crowdsourced geographic information: a systematic literature review. In: COMES, T.; BÉNABEN, F.; HANACHI, C.; LAURAS, M.; MONTARNAL, A. (Ed.). **Proceedings of the 14th ISCRAM Conference**. Abi, France: [s.n.], 2017. p. 532–545. Citations on pages 32 and 128.

_____. A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. **Transactions in GIS**, v. 22, p. 542–560, 2018. Citations on pages 32, 65, and 127.

DEVILLERS, R.; JEANSOULIN, R. **Fundamentals of Spatial Data Quality**. [S.l.]: ISTE, 2006. ISBN 978-1-905209-56-9. Citations on pages 27, 37, and 38.

DIX, A.; FINLAY, J.; ABOARD, G. D.; BEALE, R. **Human Computer Interaction**. 3rd. ed. [S.l.]: Pearson Education Limited, 2004. ISBN 978-0-13-046109-4. Citations on pages 96 and 97.

FAN, H.; ZIPF, A.; FU, Q.; NEIS, P. Quality assessment for building footprints data on openstreetmap. **International Journal of Geographical Information Science**, v. 28, n. 4, 2014. Citations on pages 39 and 40.

FARRELL, S. **Open-Ended vs. Closed-Ended Questions in User Research**. 2016. Online. Accessed on May 10th, 2018. Available: <<https://www.nngroup.com/articles/open-ended-questions/>>. Citation on page 95.

FAVA, M. C.; ABE, N.; RESTREPO-ESTRADA, C.; KIMURA, B. Y. L.; MENDIONDO, E. M. Flood modelling using synthesised citizen science urban streamflow observations. **Journal of Flood Risk Management**, 2018. Citation on page 27.

FENG, Y.; SESTER, M. Extraction of pluvial flood relevant volunteered geographic information (vgi) by deep learning from user generated texts and photos. **International Journal of Geo-Information**, v. 7, n. 39, p. 1–25, 2018. Citations on pages 66, 67, and 68.

FLANAGIN, A. J.; METZGER, M. J. The credibility of volunteered geographic information. **GeoJournal**, v. 72, n. 3-4, 2008. Citation on page 37.

FOODY, G. M. Rating crowdsourced annotations: Evaluating contributions of variable quality and completeness. **International Journal of Digital Earth**, v. 7, n. 8, p. 650–670, 2014. Citation on page 56.

FOODY, G. M.; SEE, L.; FRITZ, S.; VELDE, M. V. der; PERGER, C.; SCHILL, C.; BOYD, D. S. Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. **Transactions in GIS**, v. 17, n. 6, p. 847–860, 2013. Citation on page 40.

GIRRES, J.-F.; TOUYA, G. Quality assessment of the french openstreetmap dataset. **Transactions in GIS**, v. 14, n. 4, 2010. Citations on pages 39 and 40.

GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, n. 4, p. 211–221, 2007. Citations on pages 31, 35, and 99.

GOODCHILD, M. F.; LI, L. Assuring the quality of volunteered geographic information. **Spatial Statistics**, v. 1, p. 110–120, 2012. Citation on page 41.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate Data Analysis**. 7. ed. [S.l.]: Pearson, 2014. ISBN 1-292-02190-X. Citation on page 78.

HAKLAY, M. Crowdsourcing geographic knowledge. In: _____. [S.l.]: Springer, 2013. chap. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation, p. 105–122. Citations on pages 35, 36, 37, and 38.

HAKLAY, M. M.; BASIOUKA, S.; ANTONIOU, V.; ATHER, A. How many volunteers does it take to map an area well? the validity of linus' law to volunteered geographic information. **The Cartographic Journal**, v. 47, n. 4, p. 315–322, 2010. Citations on pages 59 and 60.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference and prediction**. [S.l.]: Springer, 2009. ISBN 978-0-387-84857-0. Citation on page 77.

HEVNER, A. R.; MARCH, S. T.; PARK, J.; RAM, S. Design science in information system research. **MIS Quartely**, v. 28, n. 1, p. 75 – 105, 2004. Citation on page 91.

HORITA, F. E. A.; ALBUQUERQUE, J. P. de; DEGROSSI, L. C.; MENDIONDO, E. M.; UEYAMA, J. Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. **Computers and Geosciences**, v. 80, p. 84–94, 2015. Citation on page 128.

HORITA, F. E. A.; DEGROSSI, L. C.; ASSIS, L. F. F. G.; ZIPF, A.; ALBUQUERQUE, J. P. de. The use of volunteered geographic information and crowdsourcing in disaster management: a systematic literature review. In: **Proceedings of the Nineteenth Americas Conference on Information Systems**. Chicago, Illinois: [s.n.], 2013. p. 1–10. Citation on page 45.

HUNG, K.-C.; KALANTARI, M.; RAJABIFARD, A. Methods for assessing the credibility of volunteered geographic information in flood response: A case study in brisbane, australia. **Applied Geography**, v. 68, p. 37–47, 2016. Citations on pages 29, 66, 67, and 85.

IACUCCI, A. A.; SONOYAMA, S.; STAUFFER, C.; CARLSEN, J.; RODRIGUEZ, D.; DOUGHERTY, J. **Ushahidi Chile Grant Report**. 2010. Accessed on: October 20th, 2015. Available: <<https://www.mozilla.org/en-US/grants/reports/ushahidi-chile-report/>>. Citation on page 90.

IBGE. **São Paulo**. 2018. Accessed on 03 December 2018. Available: <<https://www.ibge.gov.br/estatisticas-novoportal/por-cidade-estado-estatisticas.html?t=destaques&c=3550308>>. Citation on page 69.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 8402:1994 Quality Management and Quality Assurance –vocabulary**. 1994. Citation on page 37.

_____. **ISO 9241:210 Human-centered design for interactive systems**. 2010. Citation on page 93.

_____. **ISO 19157:2013 Geographic information – Data quality**. 2013. Citations on pages 38 and 39.

JILANI, M.; CORCORAN, P.; BERTOLOTTO, M. Automated highway tag assessment of openstreetmap road networks. In: **Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems**. [S.l.: s.n.], 2014. p. 449–452. Citations on pages [40](#) and [59](#).

KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! the challenges and opportunities of social media. **Business Horizons**, v. 53, n. 1, p. 59–68, 2010. Citations on pages [36](#) and [71](#).

KARAM, R.; MELCHIORI, M. A crowdsourcing-based framework for improving geo-spatial open data. In: **Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics**. Manchester, UK: [s.n.], 2013. p. 468–473. Citation on page [57](#).

KATZEFF, C.; NYBLOM, A.; TUNHEDEN, S.; TORSTENSSON, C. User-centred design and evaluation of energycoach – an interactive energy service for households. **Behaviour and Information Technology**, v. 31, n. 3, p. 305–324, 2012. Citation on page [93](#).

KEßLER, C.; GROOT, R. T. A. de. Geographic information science at the heart of europe. In: _____. [S.l.]: Springer, Cham, 2013. chap. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap, p. 21–37. Citation on page [60](#).

KITCHENHAM, B. A.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007. Citation on page [45](#).

KLONNER, C.; ECKLE, M.; USÓN, T.; HÖFLE, B. Quality improvement of remotely volunteered geographic information via country-specific mapping instructions. In: **Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management**. Albi, France: [s.n.], 2017. Citation on page [59](#).

KLONNER, C.; MARX, S.; USÓN, T.; ALBUQUERQUE, J. P. de; HÖFLE, B. Volunteered geographic information in natural hazard analysis: A systematic literature review of current approaches with a focus on preparedness and mitigation. **International Journal of Geo-Information**, v. 5, n. 7, 2016. Citations on pages [37](#) and [45](#).

LANFRANCHI, V.; WRIGLEY, S. N.; IRESON, N.; CIRAVEGNA, F.; WEHN, U. Citizens' observatories for situation awareness in flooding. In: HILTZ, S.; PFAFF, M.; PLOTNICK, L.; SHIH, P. (Ed.). **Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management**. Pennsylvania, USA: [s.n.], 2014. p. 145–154. Citations on pages [27](#) and [90](#).

LEE, Y.; KWON, P.; YU, K.; PARK, W. Method for determining appropriate clustering criteria of location-sensing data. **ISPRS International Journal of Geo-Information**, v. 5, n. 9, p. 151, 2016. Citation on page [73](#).

LERTNATTEE, V.; CHOMYA, S.; SORNLERTLAMVANICH, V. Advances in intelligent information and database systems. In: _____. [S.l.]: Springer, Berlin, Heidelberg, 2010. v. 283, chap. Reliable Improvement for Collective Intelligence on Thai Herbal Information, p. 99–110. Citations on pages [41](#) and [56](#).

LONGUEVILLE, B. D.; LURASCHI, G.; SMITS, P.; PEEDELL, S.; EUROPEAN, T. D. G. Citizens as sensors for natural hazards: a vgi integration workflow. **Geomatica**, v. 64, n. 1, p. 41–59, 2010. Citations on pages [27](#), [28](#), [41](#), [58](#), [66](#), [67](#), [68](#), [79](#), [82](#), and [85](#).

LOWDERMILK, T. **User-Centered Design**. [S.l.]: O'Reilly, 2013. ISBN 978-1-449-35980-5. Citation on page 93.

MAZZOLENI, M.; VERLAAN, M.; AFONSO, L.; MONEGO, M.; NORBIATO, D.; FERRI, M.; SOLOMATINE, D. P. Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? **Hydrology and Earth System Sciences**, v. 21, p. 839–861, 2017. Citation on page 27.

MCDOUGALL, K. Using volunteered information to map the queensland floods. In: **Proceedings of the Surveying & Spatial Sciences Biennial Conference**. Wellington, New Zealand: [s.n.], 2011. p. 13–23. Citation on page 90.

MILES, J.; SHEVLIN, M. **Applying regression and correlation: a guide for students and researchers**. [S.l.]: SAGE, 2001. ISBN 978-0-7619-6230-4. Citations on pages 77 and 78.

MIORANDI, D.; CARRERAS, I.; GREGORI, E.; GRAHAM, I.; STEWART, J. Measuring net neutrality in mobile internet: Towards a crowdsensing-based citizen observatory. In: **Proceedings of the 2013 IEEE International Conference on Communications Workshops**. Budapest, Hungary: [s.n.], 2013. p. 199–203. Citation on page 89.

MIRBABAIE, M.; STIEGLITZ, S.; VOLKERI, S. Volunteered geographic information and its implications for disaster management. In: **Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)**. Koloa, HI, USA: [s.n.], 2016. p. 207–216. Citations on pages 43, 45, 61, 62, and 65.

MOBASHERI, A.; HUANG, H.; DEGROSSI, L. C.; ZIPF, A. Enrichment of openstreetmap data completeness with sidewalk geometries using data mining techniques. **Sensors**, v. 18, n. 2, p. 1–16, 2018. Citation on page 128.

MOHAMMADI, N.; MALEK, M. Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data. **Journal of Spatial Science**, v. 60, n. 1, p. 119–135, 2015. Citations on pages 39 and 60.

MOONEY, P.; CORCORAN, P.; WINSTANLEY, A. C. Towards quality metrics for openstreetmap. In: **Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems**. San Jose, California, USA: [s.n.], 2010. p. 514–517. Citations on pages 36 and 41.

MOONEY, P.; MINGHINI, M. Mapping and the citizen sensor. In: _____. [S.l.]: Ubiquity Press, 2017. chap. A Review of OpenStreetMap Data, p. 37–59. Citation on page 41.

MOREIRA, R. B.; DEGROSSI, L. C.; ALBUQUERQUE, J. P. de. An experimental evaluation of a crowdsourcing-based approach for flood risk management. In: **Proceeding of the 18th Ibero-American Conference on Software Engineering**. Lima, Peru: [s.n.], 2015. p. 393–403. Citations on pages 65, 67, 90, 91, 111, and 129.

MULROW, C. D. Rationale for systematic reviews. **BMJ**, v. 309, n. 597, 1994. Citation on page 45.

MUMMIDI, L. N.; KRUMM, J. Discovering points of interest from users' map annotations. **GeoJournal**, v. 72, n. 3, 2008. Citation on page 58.

National Weather Service. **Floods and Flash Floods**. 2018. Accessed on 03 December 2018. Available: <<https://www.weather.gov/ffc/floods>>. Citation on page 69.

NICKERSON, R. C.; VARSHNEY, U.; MUNTERMANN, J. A method for taxonomy development and its application in information systems. **European Journal of Information Systems**, v. 22, n. 3, p. 336–359, May 2013. ISSN 1476-9344. Citations on pages 48, 51, and 52.

NIELSEN, J. **10 Usability Heuristics for User Interface Design**. 1994. Online. Available: <<https://www.nngroup.com/articles/ten-usability-heuristics/>>. Citation on page 96.

NIELSEN, J.; MOLICH, R. Heuristic evaluation of user interfaces. In: ACM. **Proceedings of the SIGCHI conference on Human factors in computing systems**. [S.l.], 1990. p. 249–256. Citation on page 96.

NORMAN, D. A. **The design of everyday things: Revised and expanded edition**. [S.l.]: Basic books, 2013. ISBN 0465072992. Citation on page 93.

NURSE, J. R.; CREESE, S.; GOLDSMITH, M.; CRADDOCK, R.; JONES, G. An initial usability evaluation of the secure situation awareness system. In: ROTHKRANTZ, L.; RISTVEJ, J.; FRANCO, Z. (Ed.). **Proceedings of the 9th ISCRAM Conference**. Vancouver, Canada: [s.n.], 2012. p. 1–5. Citation on page 90.

OGIE, R. I.; FOREHEAD, H. Investigating the accuracy of georeferenced social media data for flood mapping. In: **Proceedings of the 4th International Conference on Information and Communication Technologies for Disaster Management**. Münster, Germany: [s.n.], 2017. Citations on pages 65 and 67.

OKOLI, C.; SCHABRAM, K. A guide to conducting a systematic literature review of information systems research. **Sprouts: Working Papers on Information Systems**, v. 10, n. 26, p. 1–49, 2010. Available: <<http://sprouts.aisnet.org/10-26>>. Citations on pages 45 and 46.

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: **Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering**. [S.l.: s.n.], 2008. p. 68–77. Citation on page 47.

POIANI, T. H.; ROCHA, R. dos S.; DEGROSSI, L. C.; ALBUQUERQUE, J. P. de. Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the nepal earthquake 2015. In: **Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)**. [S.l.: s.n.], 2016. p. 188–197. Citation on page 129.

POORTHUIS, A.; ZOOK, M.; SHELTON, T.; GRAHAM, M.; STEPHENS, M. Key methods in geography. In: _____. [S.l.]: SAGE, 2016. chap. Using Geotagged Digital Social Data in Geographic Research, p. 248–269. ISBN 978-1-4462-9858-9. Citations on pages 35, 36, 68, 72, 73, 77, and 85.

POSER, K.; DRANSCH, D. Volunteered geographic information for disaster management with application to rapid flood damage estimation. **Geomatica**, v. 64, n. 1, p. 89–98, 2010. Citations on pages 66 and 67.

PREECE, J.; SHARP, H.; ROGERS, Y. **Interaction design: beyond human-computer interaction**. 4. ed. [S.l.]: John Wiley & Sons, 2015. ISBN 978-1-119-08879-0. Citation on page 95.

RAYMOND, E. S. **The cathedral and the bazaar**. Sebastopol, CA: O'Reilly Media, 1999. Citation on page 59.

RESCH, B. Progress in location-based services. lecture notes in geoinformation and cartography. In: _____. [S.l.]: Springer, Berlin, Heidelberg, 2013. chap. People as Sensors and Collective Sensing-Contextual Observations Complementing Geo-Sensor Network Measurements, p. 391–406. Citations on pages [31](#) and [36](#).

RESTREPO-ESTRADA, C.; ANDRADE, S. C. de; ABE, N.; FAVA, M. C.; MENDIONDO, E. M.; ALBUQUERQUE, J. P. de. Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. **Computers and Geosciences**, v. 111, p. 148–158, 2018. Citation on page [27](#).

ROCHA, R. dos S.; DEGROSSI, L. C.; ALBUQUERQUE, J. P. de. A systematic literature review of geospatial web service composition. In: **Proceedings of the 18th Ibero-american Conference on Software Engineering**. Lima, Peru: [s.n.], 2015. p. 502–515. Citation on page [129](#).

SALGADO, A. de L.; FREIRE, A. P. Human-computer interaction: Applications and services. In: _____. [S.l.]: Springer, Cham, 2014. v. 8512, chap. Heuristic Evaluation of Mobile Usability: A Mapping Study, p. 178–188. Citation on page [96](#).

SALK, C. F.; STURN, T.; SEE, L.; FRITZ, S.; PERGER, C. Assessing quality of volunteer crowdsourcing contributions: lessons from the cropland capture game. **International Journal of Digital Earth**, v. 9, n. 4, p. 410–426, 2016. Citation on page [36](#).

SALK, C. F.; STURN, T.; SEE, L.; FRITZ, S. Limitations of majority agreement in crowdsourced image interpretation. **Transactions in GIS**, v. 21, n. 2, p. 207–223, 2017. Citation on page [56](#).

SCHWARZ, J.; MAIWALD, H.; GERSTBERGER, A. Quantifizierung der schäden infolge hochwassereinwirkung: Fallstudie eilenburg. **Bautechnik: Zeitschrift für den gesamten Ingenieurbau**, v. 82, n. 12, p. 845–856, 2005. Citation on page [66](#).

SEE, L.; COMBER, A.; SALK, C.; FRITZ, S.; VELDE, M. van der; PERGER, C.; SCHILL, C.; MCCALLUM, I.; KRAXNER, F.; OBERSTEINER, M. Comparing the quality of crowdsourced data contributed by expert and non-experts. **PLoS ONE**, v. 8, n. 7, p. e69958, 2013. Citations on pages [39](#) and [56](#).

SEE, L.; MOONEY, P.; FOODY, G.; BASTIN, L.; COMBER, A.; ESTIMA, J.; FRITZ, S.; KERLE, N.; JIANG, B.; LAAKSO, M.; LIU, H.-Y.; MILCINSKI, G.; NIKSIC, M.; PAINHO, M.; PODÖR, A.; OLTEANU-RAIMOND, A.-M.; RUTZINGER, M. Crowdsourcing, citizen science or volunteered geographic information? the current state of crowdsourced geographic information. **International Journal of Geo-Information**, v. 5, n. 55, p. 1–23, 2016. Citations on pages [27](#), [35](#), and [46](#).

SELEX. **Selex meteor manual**. [S.l.], 2015. Citation on page [73](#).

SENARATNE, H.; BRÖRING, A.; SCHRECK, T. Using reverse viewshed analysis to assess the location correctness of visually generated vgi. **Transactions in GIS**, v. 17, n. 3, p. 369–386, 2013. Citations on pages [40](#) and [53](#).

SENARATNE, H.; MOBASHERI, A.; ALI, A. L.; CAPINERI, C.; HAKLAY, M. M. A review of volunteered geographic information quality assessment methods. **International Journal of Geographical Information Science**, v. 31, n. 1, p. 139–167, 2017. Citations on pages [43](#), [44](#), [61](#), [62](#), and [65](#).

- STATISTA. **Number of mobile phone users worldwide from 2015 to 2020 (in billions)**. 2019. Online. Accessed on: April 21st, 2019. Available: <<https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/>>. Citation on page 91.
- STEFANIDIS, A.; CROOKS, A.; RADZIKOWSKI, J. Harvesting ambient geospatial information from social media feeds. **GeoJournal**, v. 78, n. 2, p. 319–338, 2013. Citation on page 36.
- STEIGER, E.; ALBUQUERQUE, J. P. de; ZIPF, A. An advanced systematic literature review on spatiotemporal analyses of twitter data. **Transactions in GIS**, v. 19, n. 6, p. 809–834, 2015. Citation on page 45.
- TOBLER, W. R. A computer movie simulating urban growth in the detroit region. **Economic Geography**, v. 46, p. 234–240, 1970. Citation on page 42.
- USABILITY.GOV. <https://www.usability.gov/what-and-why/user-centered-design.html>. 2019. Online. Accessed on: April 16th, 2019. Available: <<https://www.usability.gov/what-and-why/user-centered-design.html>>. Citations on pages 93 and 94.
- W3C Working Group. **Web Content Accessibility Guidelines (WCAG) 2.0**. 2008. Online. Accessed on: May 3rd, 2018. Available: <<https://www.w3.org/TR/WCAG20/>>. Citations on pages 104 and 106.
- _____. **Relationship between Mobile Web Best Practices (MWBP) and Web Content Accessibility Guidelines (WCAG)**. 2009. Online. Accessed on: May 3rd, 2018. Available: <<https://www.w3.org/TR/mwbp-wcag/>>. Citation on page 101.
- _____. **Mobile Web Application Best Practices**. 2010. Online. Accessed on: May 3rd, 2018. Available: <<https://www.w3.org/TR/mwabp/>>. Citation on page 101.
- WALD, D. J.; QUITORIANO, V.; WORDEN, C. B.; HOPPER, M.; DEWEY, J. W. Usgs "did you feel it?" internet-based macroseismic intensity maps. **Annals of Geophysics**, v. 54, n. 6, p. 688–707, 2011. Citation on page 89.
- WENTZ, B.; LAZAR, J.; STEIN, M.; GBENRO, O.; HOLANDEZ, E.; RAMSEY, A. Danger, danger! evaluating the accessibility of web-based emergency alert sign-ups in the northeastern united states. **Government Information Quarterly**, v. 31, p. 488–497, 2014. Citation on page 90.
- WESENSEIT. **WeSenseIt: The Delfland Case Study**. 2019. Accessed on: April 26th, 2019. Available: <<http://www.wesenseit.com/organisatie/>>. Citation on page 90.
- WIGGINS, A.; NEWMAN, G.; STEVENSON, R. D.; CROWSTON, K. Mechanisms for data quality and validation in citizen science. In: **IEEE Seventh International Conference on e-Science Workshops**. Stockholm, Sweden: [s.n.], 2011. p. 14–19. Citations on pages 43, 44, 61, 62, and 65.
- ZIELSTRA, D.; HOCHMAIR, H. H. Positional accuracy analysis of flickr and panoramio images for selected world regions. **Journal of Spatial Science**, v. 58, n. 2, p. 251–273, 2013. Citations on pages 40 and 53.
- ZOOK, M.; GRAHAM, M.; SHELTON, T.; GORMAN, S. Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake. **World Medical & Health Policy**, v. 2, n. 2, p. 7–33, 2012. Citation on page 89.

DECLARATION OF ORIGINAL AUTHORSHIP AND LIST OF PUBLICATIONS

I confirm that this doctoral dissertation has not been submitted in support of an application for another degree at this or any other teaching or research institution. It is the result of my own work and the use of all material from other sources has been properly and fully acknowledged. Research done in collaboration is also clearly indicated.

Excerpts of this dissertation have been either published or submitted for the appreciation of editorial boards of journals, conferences, and workshops, according to the list of publications presented below. My contributions to each publication are listed as well.

Publications Resulting from this Thesis

- **Degrossi, L. C.;** Albuquerque, J. P.; Rocha, R. S.; Zipf, A.; “*A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information*” (DEGROSSI *et al.*, 2018).

Journal: Transactions in GIS

Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with her contributors.

- **Degrossi, L. C.;** Abe, B. B.; Albuquerque, J. P.; Fortes, R. P. M.; “*Enhancing usability of a Citizen Observatory based on User-centered Design*” (DEGROSSI *et al.*, 2018).

Conference: 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion

Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with her contributors.

- **Degrossi, L. C.;** Albuquerque, J. P.; Rocha, R. S.; Zipf, A.; “*A Framework of Quality*

Assessment Methods for Crowdsourced Geographic Information: a Systematic Literature Review” (DEGROSSI *et al.*, 2017b).

Conference: 14th International Conference on Information Systems for Crisis Response and Management

Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with her contributors.

- Degrossi, L. C.; Albuquerque, J. P.; Restrepo-Estrada, C.; Mobasheri, A.; Zipf, A.; “Exploring the geographical context for quality assessment of VGI in flood management domain” (DEGROSSI *et al.*, 2017a).

Conference: 23rd Americas Conference on Information Systems

Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with her contributors.

Book Chapter

- Albuquerque, J. P.; Horita, F. E. A.; Degrossi, L. C.; Rocha, R. S.; Andrade, S. C.; Restrepo-Estrada, C.; Leyh, W.; “Leveraging Volunteered Geographic Information for Improving Disaster Resilience: Lessons learned with AGORA and future research directions” (ALBUQUERQUE *et al.*, 2017).

Book: Campelo, C. E. C.; Bertolotto, M.; Corcoran, P. (eds.) Volunteered Geographic Information and the Future of Geospatial Data

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the chapter writing.

Other Related Publications

- Mobasheri, A.; Huang, H.; Degrossi, L. C.; Zipf, A.; “Enrichment of OpenStreetMap Data Completeness with Sidewalk Geometries Using Data Mining Techniques” (MOBASHERI *et al.*, 2018).

Journal: Sensors

Level of contribution: Low - the PhD candidate helped in the development of the work.

- Horita, F. E. A.; Albuquerque, J. P.; Degrossi, L. C.; Mendiondo, E. M.; Ueyama, J.; “Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks” (HORITA *et al.*, 2015).

Journal: Computers & Geosciences

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the paper writing.

- Poiani, T. H.; Rocha, R. S.; **Degrossi, L. C.**; Albuquerque, J. P.; “*Potential of Collaborative Mapping for Disaster Relief: A Case Study of OpenStreetMap in the Nepal Earthquake*” ([POIANI et al., 2016](#)).

Conference: XLIX Hawaii International Conference on System Sciences

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the paper writing.
- Moreira, R. B.; **Degrossi, L. C.**; Albuquerque, J. P.; “*An experimental evaluation of a crowdsourcing-based approach for flood risk management*” ([MOREIRA; DEGROSSI; ALBUQUERQUE, 2015](#)).

Conference: 12th Workshop on Experimental Software Engineering

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the paper writing.
- Rocha, R. S.; **Degrossi, L. C.**; Albuquerque, J. P.; “*A Systematic Review of Geospatial Web Service Composition*” ([ROCHA; DEGROSSI; ALBUQUERQUE, 2015](#)).

Conference: 12th Workshop on Experimental Software Engineering

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the paper writing.
- Andrade, S. C.; **Degrossi, L. C.**; Restrepo-Estrada, C.; Delbem, A. C. B.; Albuquerque, J. P.; “*Does keyword noise change over space and time? A case study of social media messages*” ([ANDRADE et al., 2018](#)).

Conference: XIX Brazilian Symposium on GeoInformatics

Level of contribution: Low - the PhD candidate helped in the paper writing.
- Rocha, R. S.; **Degrossi, L. C.**; Albuquerque, J. P.; Hellingrath, B.; “*Collaborative Mapping for Crisis Management*”.

Conference: 22nd ICE/IEEE International Technology Management Conference (NiTiM International Graduate Summer School)

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the paper writing.
- Rocha, R. S.; **Degrossi, L. C.**; Horita, F. E. A.; Albuquerque, J. P.; “*AGORA-PL: Uma Proposta para Desenvolvimento de Famílias de Sistemas Colaborativos baseados em VGI para a Gestão do Risco de Inundação*”.

Conference: XI Simpósio Brasileiro de Sistemas Colaborativos

Level of contribution: Medium - the PhD candidate helped in the development of the work, as well as in the paper writing.

QUESTIONNAIRE FOR USER PROFILE CHARACTERIZATION

The questionnaire aims at collecting information from citizens for the design and development of the new Citizen Observatory of Floods interface. It characterized the users' profile, identified the frequency at which they witnessed flood events, the way they responded to them, and the type of information shared and searched.

1. Name: _____

2. City: _____

3. State: _____

4. Date of birth (DD/MM/AAAA):

5. Gender:

☐ Male

☐ Female

6. Civil status:

☐ Single

☐ Married

☐ Divorced

☐ Widower

7. Educational level:

☐ Elementary school (incomplete)

☐ Elementary school (complete)

- ☐ High school (incomplete)
- ☐ High school (complete)
- ☐ College (incomplete)
- ☐ College (complete)
- ☐ Other:

8. Do you have any mobile device (e.g., cell phone and/or tablet)?

- ☐ Yes
- ☐ No

9. If answered positively, how often do you use your mobile device?

- ☐ More than once a week.
- ☐ 2 or 3 time a week.
- ☐ 4 or 5 time a week.
- ☐ Every day.

10. Do you use any digital communication system (e.g., WhatsApp, email, etc.)?

- ☐ Yes
- ☐ No

11. If answered positively, which is the most used digital communication system?

12. Do you use any Social Media?

- ☐ Yes
- ☐ No

13. If answered positively, which Social Media do you use?

- ☐ Facebook
- ☐ Twitter
- ☐ WhatsApp
- ☐ Instagram
- ☐ Other: _____

14. Do you live or work in a flood-prone area?

- ☐ Yes
- ☐ No

15. Have you seen any flood event?

☐ Yes

☐ No

☐ Comment: _____

16. If answered positively, how many times?

☐ 1

☐ 2

☐ 3

☐ 4 or more times

17. If answered positively, do you take any preventive action against flood?

☐ Yes

☐ No

☐ Comment: _____

18. Have you provided any information about flood events?

☐ Yes

☐ No

☐ Comment: _____

19. Do you usually share information about flood events on Social Media?

☐ Yes

☐ No

20. If answered positively, which information do you share?

21. If answered positively, how do you share information?

☐ SMS

☐ Phone call

☐ WhatsApp

☐ Other: _____

22. Do you usually share or search for information about flood events?

☐ Yes

☐ No

23. If answered positively, where and which information do you share or search for?

24. Would you use a platform to share information about flood events?

25. If answered negatively, would you be willing to provide?

26. If answered positively, which information type would you be willing to share?

☐ Flood likelihood

☐ Water level in a river bed

☐ Rain intensity

☐ Water level on streets

☐ Other: _____

27. For you, what is the best way for receiving information?

☐ SMS

☐ WhatsApp

☐ Facebook

☐ Other: _____

QUESTIONS OF INTERVIEW

The interview with the director of the Civil Defense aims to understand the actions taken by the Civil Defense before, during and after a flood event, the sources of information used to support their actions, type(s) of information gathered and made available to citizens, and type(s) of information the Civil Defense would like to receive. The questions are described as follows.

1. What is the role of Civil Defense before, during and after a flood event?
2. What preventive actions are carried out by the Civil Defense before and after a flood event?
3. What sources of information support the actions?
4. How is the dataset collected?
5. Do the population contribute with information regarding flood events?
6. In case of a positive answer, how do they contribute? For example, what technologies are used for this purpose?
7. Does the Civil Defense use information from Social Media regarding rain or flood events?
8. What information is made available to the population? How is it made available?
9. Is there a Website for making the information available? Is it accessible to the population?
If not, would the Civil Defense use a Website to share and receive information?
10. What information has the Civil Defense not received, but would like to receive?

QUESTIONNAIRE FOR USER PROFILE CHARACTERIZATION (USABILITY TEST)

The questionnaire was answered by all participants before the usability test and aims at gathering information on the participants and use of mobile applications by them.

1. Name: _____
2. Age: _____
3. Gender:
 - ☐ Male
 - ☐ Female
4. Civil status:
 - ☐ Single
 - ☐ Married
 - ☐ Divorced
 - ☐ Widower
5. Occupation: _____
6. Educational level:
 - ☐ Elementary school (incomplete)
 - ☐ Elementary school (complete)
 - ☐ High school (incomplete)
 - ☐ High school (complete)
 - ☐ College (incomplete)

☐ College (complete)

☐ Other: _____

7. Do you have mobile device?

☐ No

☐ Yes

8. How long do you use mobile device?

☐ more than 5 years

☐ more than 1 year

☐ more than 6 months

☐ I do not use mobile device

9. If you do not use mobile device, what is the main reason not to use it?

☐ I do not have a mobile device.

☐ I do not have Internet connection.

☐ I do not know how to use it.

☐ It is too expensive.

☐ Other: _____

10. How often do you use your mobile device?

☐ More than 4 times a day

☐ More than once a day

☐ Less than once a day

☐ More than 3 times a week

☐ More than once a week

☐ Less than once a week

☐ Rarely

☐ Other: _____

11. Do you have any difficulty on the use of mobile devices? Please, describe briefly your(s) main(s) difficulty(ies).

12. Do you feel afraid while using mobile devices?

☐ No

☐ Yes

13. Why do you feel afraid while using mobile devices?

14. Do you feel insecure while using mobile devices?

☐ No

☐ Yes

15. Why do you feel insecure while using mobile devices?

16. How often do you usually feel lost while using mobile devices?

☐ Always

☐ Sometimes

☐ Rarely

☐ Never

17. How often do you usually feel without knowing what to do while using mobile devices?

☐ Always

☐ Sometimes

☐ Rarely

☐ Never

18. How often do you usually ask for help to use mobile devices?

☐ Always

☐ Sometimes

☐ Rarely

☐ Never

19. How often do you understand error messages while using mobile applications/devices?

☐ Always

☐ Sometimes

☐ Rarely

☐ Never

20. Do you have any suggestion to improve mobile applications and, thus, have a better experience on their use?

21. Do you have any problem that affects your memory or motor coordination?

☐ No

☐ Yes

22. Which problems do you have that affects your memory or motor coordination?

SYSTEM USABILITY SCALE

1. I think that I would like to use this system frequently.
 - a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
2. I found the system unnecessarily complex.
 - a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
3. I thought the system was easy to use.
 - a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
4. I think that I would need the support of a technical person to be able to use this system.
 - a. Strongly disagree.
 - b. Disagree.

- c. Neutral.
 - d. Agree.
 - e. Strongly agree.
5. I found the various functions in this system were well integrated.
- a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
6. I thought there was too much inconsistency in this system.
- a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
7. I would imagine that most people would learn to use this system very quickly.
- a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
8. I found the system very cumbersome to use.
- a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.
 - d. Agree.
 - e. Strongly agree.
9. I felt very confident using the system.
- a. Strongly disagree.
 - b. Disagree.
 - c. Neutral.

- d. Agree.
- e. Strongly agree.

10. I needed to learn a lot of things before.

- a. Strongly disagree.
- b. Disagree.
- c. Neutral.
- d. Agree.
- e. Strongly agree.

11. I could get going with this system.

- a. Strongly disagree.
- b. Disagree.
- c. Neutral.
- d. Agree.
- e. Strongly agree.