Computers, Environment and Urban Systems 59 (2016) 231–243

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management

Carlos Granell [a,*], Frank O. Ostermann [b]

[a] Institute of New Imaging Technologies, Universitat Jaume I of Castellón, Spain
[b] Department of Geo-Information Processing, Faculty of Geo-Information Science and Earth Observation, University of Twente, The Netherlands

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form 16 December 2015
Accepted 28 January 2016
Available online 16 February 2016

Keywords:
VGI
Crowdsourcing
Geo-social media, data analysis methods
Disaster management
Systematic mapping

ABSTRACT

This paper investigates research using VGI and geo-social media in the disaster management context. Relying on the method of systematic mapping, it develops a classification schema that captures three levels of main category, focus, and intended use, and analyzes the relationships with the employed data sources and analysis methods. It focuses the scope to the pioneering field of disaster management, but the described approach and the developed classification schema are easily adaptable to different application domains or future developments. The results show that a hypothesized consolidation of research, characterized through the building of canonical bodies of knowledge and advanced application cases with refined methodology, has not yet happened. The majority of the studies investigate the challenges and potential solutions of data handling, with fewer studies focusing on socio-technological issues or advanced applications. This trend is currently showing no sign of change, highlighting that VGI research is still very much technology-driven as opposed to theory- or application-driven. From the results of the systematic mapping study, the authors formulate and discuss several research objectives for future work, which could lead to a stronger, more theory-driven treatment of the topic VGI in GIScience.

1. Introduction

In 1941, the BBC called for people's collaboration to obtain holiday pictures and postcards of European beaches. Over 10 million of items were received.[1] This early precedent resembles in scale and quality of information modern responses to disasters: people generate and collect dozens of millions of media content for supporting diverse activities during all stages of disaster management (Meier, 2015). What makes the current utilization of user-generated content different is that many of the original technological limitations on the production, collection, and processing of data no longer exist. People now produce data in many ways, and many do not realize that they are constantly generating data through their own handheld devices (e.g. automated GPS traces).

The term Volunteered Geographic Information (VGI) was coined in 2007 (Goodchild, 2007) to emphasize the concept of user-generated content attached with geospatial information (e.g. location, place names) in the form of geotags or coordinates. No matter what people, services, devices, and sensors are sensing (e.g. noise, air quality), spatio-temporal context is a must to help in the understanding and interpretation of the collected data (Sagl, Blaschke, Beinat, & Resch, 2012). Various sources of VGI have been used in many contexts and diverse application scenarios, leading to a family of terms which highlight slightly different characteristics on the level of user participation, on whom produces the data (citizens, sensors, etc.), or on the particularities of data collection processes. Some authors have tried to create taxonomies for user-generated geographic content (Craglia, Ostermann, & Spinsanti, 2012). In this paper we continue to use the term VGI, albeit in an inclusive manner that also encompasses data that was not explicitly volunteered.

Contrary to the BBC case, where each picture was manually processed to help find the final location in Normandy, one of the determining factors today is the existence of enabling analytics technology (Chen, Chiang, & Storey, 2012) to quickly process and analyze huge amounts of data. The actual data-rich context reduces emphasis on the accuracy and exactness of data in favor of allowing some degree of inaccuracy, uncertainty and noise in return for capturing a far more comprehensive, larger set of data (Mayer-Schönberger & Cukier, 2013). The current state of the art in data analysis techniques enables the rapid discovery of correlations in large data. More data might also help to deal with inaccuracy of individual bits of data. In this context, we assume that data analysis techniques will gain traction over data quality and precision as long as more and more VGI is captured and used in projects and applications.

Recent works in the literature (Neis & Zielstra, 2014; Roick & Heuser, 2013) examine the nature of VGI itself (e.g. quality, accuracy, precision),

* Corresponding author.
E-mail addresses: carlos.granell@uji.es (C. Granell), f.o.ostermann@utwente.nl (F.O. Ostermann).

[1] http://en.wikipedia.org/wiki/Operation_Overlord.

http://dx.doi.org/10.1016/j.compenvurbsys.2016.01.006
0198-9715/

focus on the role of the contributor (e.g. gender, motivating factors to contribute), or explore the defining capabilities in terms of reliability, documentation, and easy-to-use of current VGI sources. This study focuses on the utilization and analysis of VGI in the domain of natural and man-made disasters management. While recent reviews question current data practices on using VGI during disaster management scenarios (Haworth & Bruce, 2015), this study takes a complementary view assuming that the detection of hidden and emerging patterns on the utilization of VGI could pave the way to advancing the VGI research field beyond its predominant focus in data collection to a level where rich spatiotemporal contexts, and advanced geospatial analysis techniques play a dominant role (Crampton et al., 2013).

In this sense, we expect that VGI research is now a stage where a body of knowledge and best practices should emerge. This is a difficult task because research on VGI is mostly technology-driven and changes at a fast pace. We aim to contribute to an ongoing consolidation by identifying important analytical trends and use patterns on the utilization of VGI, in order to shape future research and applications in the field.

To do so, we conducted a systematic mapping study that aims to find and classify the primary studies in the VGI and disaster management field. Engström and Runeson (2011) summarized a systematic mapping study as a useful tool looking "at a higher granularity level with the aim to identify research gaps and clusters of evidence in order to direct future research". In this work, we present the results of a systematic mapping study based on a sample of VGI-related studies published since the term's inception in 2007. To do so, we first designed a classification schema that allowed us to explore systematically the set of eligible papers through a "purpose" dimension. This enables us to investigate relationships between purposes, VGI sources and analysis methods, and reveal hidden and emerging patterns to expose novel, innovate purposes for VGI beyond data collection in disaster management situations. We argue that the outcomes of the present research are valuable also for researchers in other applications domains by taking the analysis framework as reference for subsequent studies.

In particular, this systematic mapping study addresses the following questions in the context of the disaster management domain:

- Q1: Which is a suitable classification schema for VGI research to enable a systematic mapping study?
- Q2: What are the most investigated intended uses, and in what user case scenarios?
- Q3: What are the most frequently VGI data sources, and in what context/application?
- Q4: What are the most frequently (spatial) data analyses methods, and in what context/application?

Sections 2 and 3 cover the systematic mapping review itself. We describe the methodology to obtain the set of eligible studies and the subsequent analysis for identifying main intended uses, data sources employed, and key enabling analysis methods. In Section 4, we interpret the results, and point out our own observations; whereas in Section 5 we discuss opportunities and challenges.

2. Method

A systematic mapping study is a successful tool in research fields such as software engineering (Petersen, Feldt, Mujtaba, & Mattsson, 2008) and rich web applications (Casteleyn, Garrigós, & Mazón, 2014). We adopt the methodology from Petersen et al. (2008) for this study. In short, the study answers specific research questions (Q2–4), related to the identification and coverage of the field of study, by first identifying eligible and primary studies, then classifying them in a newly developed schema (Q1), and analyzing the results. The following subsections detail each step of the approach.

2.1. Search and selection criteria

Initially, we obtained 426 papers as a result of several bibliographic search queries in major specialized and general databases engines such as ISI Web of Science, Scopus, ACM, IEEE, and DBLP, as well as thematic repositories like the Humanitarian Computing Library.[2] We also sought for relevant conferences and workshops for which VGI4DM (VGI for Disaster Management) was a central topic (See Annex A, Supplementary data). In order to better understand the resulting set, we performed an initial, three-phased exploratory analysis, using the following eligibility criteria:

1. Publication in scientific journals, magazines, conferences, symposia or workshops (excluding review and survey papers, editorials, comments and prefaces) with full text being accessible.
2. Written in English.
3. The title, abstract or keywords explicitly mention the utilization of VGI sources (including geo-social media) in a disaster management or crisis response context.
4. The publication was published in or after 2007, when the term VGI was coined.

During the first phase, a screening of titles and abstracts led to the removal of duplicates and those that clearly did not fulfill all eligibility criteria, reducing the set to 119.

A second in-depth analysis of the resulting set focused on criterion 3, the relevance and use of VGI or geo-social media in a disaster management context, resulting in a final set of 59 relevant and representative papers for the systematic mapping study. These papers appear in Annex C (see Supplementary data). In total, 35 unique publication venues were identified, of which about two thirds (25) features only one publication (see Annex A, Supplementary data). This high dispersion reflects the diversity of VGI research that spans many diverse disciplines and scientific fields. However, GIS journals are well represented in our selection and are identified as core journals in a recent bibliographic study about GIS journals (Scarletto, 2014).

Third and finally, we conducted an extensive qualitative analysis[3] in order to extract and synthesize relevant data about the remaining papers (N = 59). For each paper, we tried to answer: What is the main focus and intended use? What kind of data do the authors employ for their study? What data analysis and visualization techniques do the authors utilize? Do the authors compare or integrate their results with official/reference datasets? What is the use case or scenario? Who are the target user(s)? As a result, we extracted a list of 30 variables grouped on four thematic clusters:

- Article's bibliographic details: unique identifier; title; year of publication; the name and type of the publication venue, keywords as they appeared in the paper; DOI and abstract.
- Article's focus and intended use: the main category of a paper, its focus, and intended uses (the combination of these three variables contributed to the build-up of the classification schema described in the following section); the most relevant application domain (natural disaster, man-made disaster, both; use case or scenario (e.g. floods, earthquake, and riots); a concrete disaster, if any (e.g. 2010 Haiti Earthquake); and type of stakeholders or end users if explicitly mentioned.
- Data sources employed: (geo)social media/VGI sources used; additional comments/notes on how the VGI sources were used; use

[2] http://humanitariancomp.referata.com/wiki/Welcome.

[3] The resulting data set from the analysis, plus documents explaining all the variables of the data set, along with the subsequent data analysis (R scripts) for Section 3 are publicly available in the following Github repository: https://github.com/cgranell/paper-vgi-science.