Tweeting About Floods of Messinia (Greece, September 2016) - Towards a Credible Methodology for Disaster Management Purposes

Stathis G. Arapostathis[✉]

Harokopio University, Venizelou El. Ave. 71, 70 176 Kallithea, Athens, Greece
sarapos@hua.gr

Abstract. As repeatedly stated by many researchers, the phenomenon of volunteered geographic information (VGI) is an important new technological tool that can contribute to disaster management (DM). Specifically regarding floods, despite that there is very interesting published research work that assesses VGI and flood event management in various areas of the world, related research in Greece, is still limited. Moreover, the necessity of creating meaningful maps and graphs is still a significant challenge globally, along with the general need to reduce the time needed for data processing. Current research presents steps towards those issues, by advancing various innovative initial methods and techniques that have been presented previously. The floods of Kalamata, Messinia (Greece, September 2016) were selected for case study, while as a VGI source, a corpus of 111000 tweets, published within 120 h since the flood event occurrence, was used. As far as the methodology is concerned, the main steps, included: a. the classification of the tweets in categories, the main of which were: (i) flood tracking, (ii) tracking of consequences, (iii) simple identification of rain and (iv) irony/emotions. The next step was related to geo-referencing the classified information, by using an R script, while the final steps were linked to generating maps and graphs, from subsets of the processed data. The outcome of this research is based on the analysis of about 45% of the total dataset.

Keywords: Disaster management · Credible methodology · Tweeting · Volunteered Geographic Information · Floods of Messinia - Greece

1 Introduction

It is generally accepted that Volunteered Geographic Information (VGI) is one of the most important technological tools that can drastically contribute to having in-depth and rapid information regarding the management of a disaster event [11–14]. Particularly regarding flood events, there is a plethora of published research related to extracting, analyzing and visualizing important information from sources of VGI, most of them, related to Social Media.

Specifically about floods, various case studies have been used in order to discover the potentials of VGI in flood event management. From Australia and the Queens-land

floods [11] up to the Colorado floods in the USA [5]. Some significant findings of published research include the high productivity rhythm of information that in the case study of the Queensland flood reached about 15000 messages per hour along with the general assumption that VGI provides significant contribution regarding to the immediacy and the depth of information. [5, 14]. Moreover, the correctly tracking of a flood event that a VGI source provides, has been characterized as vital for studying, identifying weaknesses and limiting thus the negative impact of similar events that may potentially occur in the future [5].

In addition, during the latest years, there is some significant contribution by the European Union to provide innovative solutions that aim to track effectively the consequences along with the unfoldness of a disastrous event. The Copernicus emergency service provides tracked information which is automatically extracted through imagery. Although this service is very useful, it sometimes misses disastrous events; the case study used in this research for instance, (flood of Kalamata, Greece, 2016) is not present as the Copernicus database.

Another important aspect of VGI is related to the so-called Participatory GIS (PGIS) activities, which utilize volunteers who take actions in the physical environment for reducing the vulnerability of a geographic area in disastrous events. A nice approach was related to reducing flood vulnerability in the city of La Paz, in Baja California, Mexico [3], while another inspiring PGIS approach was the one that took place in Brazil, where a community of more than 117 Brazilian Experts from NGOs and private companies worked together in order to define the most suitable criteria that emerge areas, as vulnerable to floods, in South Brazil [4].

Especially regarding the data analysis part of VGI, there are various challenges. In author's opinion, those challenges are accumulated in four clusters. At first, it is a classification problem. While international research has proposed various classification structures [1, 6, 8] there is still room for improvement, as the most effective classification can lead to the most effective delivery of vital information. The second cluster of challenges is related to the precise georeferencing, as the only way to have meaningful maps is to have sufficient geo-referencing of the data. Next, the third cluster includes the selection and development of appropriate visualization techniques for delivering the processed information to the final receivers. The maps and graphs created, must make sense to people that could potentially, need to have access on information, but with zero knowledge in reading sophisticated scientific maps and schemas. Considering the latter, too complicated visualizations should be simplified. Finally, the last cluster of challenges is related to automation of the data analysis procedures for having deliverable results in real time. VGI data analysis is a time-consuming process [7–10].

Respecting the set of open challenges, the next sessions of current article describe a methodology that can be used for extracting, classifying and visualizing DM information.

As a case study, the floods of Kalamata, Messinia (September 2016, Greece) is used. The rain that resulted to flood, started in the 7th of September 2016, while a day after, the nearby area of Lakonia was also affected. The flood caused the death of three people. Many damages to the urban environment were reported while 34 families had no place to sleep. The motorway along with the airport of the city of Kalamata remained closed. Many politicians (Ministers of the government, the X-Prime Minister etc.) travelled to Kalamata for disaster management purposes.

2   Data Used

A dataset was acquired from Sifter containing a corpus of 111000 tweets that were published from 7th of September 00:00:01 GMT time, until the 11th of September 23:59:59 GMT time. This dataset included the total number of published tweets containing at least one of the following keywords: Flood (in Greek and English), Floods (in Greek), Rain (in Greek), Storm (in Greek), Damages (in Greek), Kalamata (in Greek and Latin). Those words exist in the tweet corpus either as a hashtag (#) or as a simple word. In current research about 45% of the total dataset was analyzed.

3   Methodology

The proposed methodology is consisted of few basic steps (Fig. 1). The first step is related to data preparation; the initial dataset was consisted of three csv files. Those files were loaded into R-Studio and various functions were applied for eliminating all null values and for removing all graphical characters. Next, the final dataset was exported in a csv format.
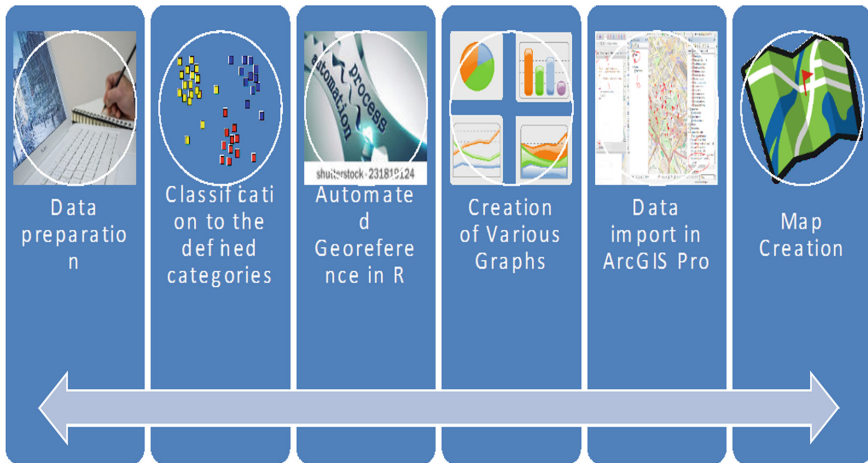
Fig. 1. Methodology

The next step of the methodology was related to classifying the tweets in categories that makes sense to visualize. Table 1 indicates all the categories that emerged through a conceptual mashup of the structures of other published research [1, 6, 8]. Classification was performed either by reading each tweet 1 by 1 or by using various text queries that contributed to accelerating the process.

Moreover, the category "consequences" was further sub-classified to 5 values, ranged from I to V. The first value (I) is related to simple identification of a rain or a storm while the value V is related to Loss of human life. A similar consequence scoring

Table 1. Classification categories

| Categories	
Not related to floods	Identification of other flood events
Identification of rain	Disaster management
Identification of the flood event of Kalamata	Consequences
Irony	Emotions
Effects on social life	Situation overview
Weather related information	Flood aid
Self-organizing of volunteers for flood related purposes	

was presented by the author previously [1]. However this version of scoring is enriched with more certain incidents (Table 2). Finally, the author created few scatterplots in order to assess the frequencies of each classified category, considering the time of publication.

Table 2. Description of consequence score values

Consequence score	Description
I	Simple identification of rain or storm
II	Torrential storm, human fear, terror
III	Damages, problems in the traffic network, minor human injuries, flooded streets, airport, school, or other public premises are closed, help tickets to fire brigade
IV	Huge damages, missing people, homeless people, serious danger to human life, emergency situation, no electricity or water at a city level. Buses change track due to flood, isolated hamlets, people that cannot escape a premise
V	Loss of human life

The third step of the methodology was consisted of georeferencing the information, by detecting the geolocations mentioned within the text of each tweet, and by multiplying each tweet N times where N is the number of detected geolocations. In the multiplied group of the tweets, the x y coordinates of each geolocation were added, in a way that each area was geo-referenced one time (Fig. 2). The whole procedure was done automatically by executing a script, developed by the author in R-programming language [2].

The visualization of the classified and geo-referenced information consists of the final part of the methodology. In specific, several graphs were created which utilize the frequency of tweets, classified to specific categories and published within different time periods. Finally, two maps were created, visualizing the quantity of tweets related to simple rain identification (Map I) and the consequence scores (Map II). Both maps, visualize extracted information, that was posted in Twitter, within the first 48 h of the start of the flood event.
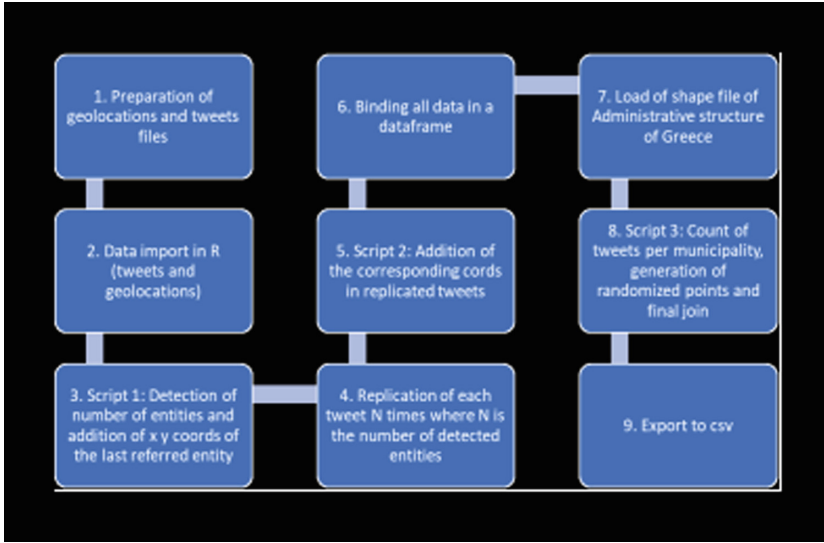
Fig. 2. Tweet geo-referencing method developed in R [2]

4   Results

In Fig. 3 (the left part) and Table 3 the words with the highest frequency in the analyzed text corpus, are displayed. The word "Kalamata" which is the name of the city that was devastated mostly by the floods, is highly mentioned, while the words bad weather and woman along with disasters and dead/deadly are ranked just after Kalamata. It can be assumed that within the first 6 h, since the start of the rain, twitter provides sufficient information. The popular words, in terms of frequency are multiplied within 24 h (Fig. 3 (the right part), Table 4). Kalamata also appears at a high frequency in Latin, which can lead to the assumption that within 24 h the news regarding the floods of Kalamata were spread at an international level. Moreover, apart from the disseminated information regarding the human losses, the names of various Greek politicians such as Samaras and Tsipras are mentioned frequently.

Figures 4 and 5 visualize various aspects of the classified information that was extracted from twitter. In particular on those graphs, the author focuses on the information regarding flood identification, consequence score values (Fig. 4), emotional expression and irony (Fig. 5). It can be assumed that, within the first hours of the flood, there is a plethora of disseminated information regarding the flood identification and the consequences to the urban and rural environment, while the expression of emotions and the ironic tweets started to be published later.

Table 3. Word frequencies within 6 h

Keyword	Frequency within 6 h
Kalamata	198
Bad weather	88
Woman	65
Disaster(s)	68
Dead/deadly	158
Rain	29
Act of God	23
Thouria	15
Problems	15
Michaniona	13
Floods	12
Missing	22

Table 4. Word frequency within 24 h respectively

| Keyword	Frequency within 24 h
Kalamata (Greek)	896
Kalamata (Latin)	246
Dead/deadly (Greek)	311
Damages (Greek)	167
Bad weather (Greek)	161
Greece (English)	136
Samaras (Greek)	132
Floods (Greek)	115
Floods/flooding (English)	114
Grecia (Spanish)	93
Messinia (Greek)	78
Tsipras (Greek)	67
Lluvias (Spanish)	66
Dead (English)	65
Lakonia (English)	59
Rain (English)	55
Flood affected (Greek)	54
Sparta (English)	53
Kill (English)	51
Thermaikos (Greek)	50
Aftermath	48
Act of God	48
Intense	48
Woman (Greek)	45
Leventis (Greek)	40
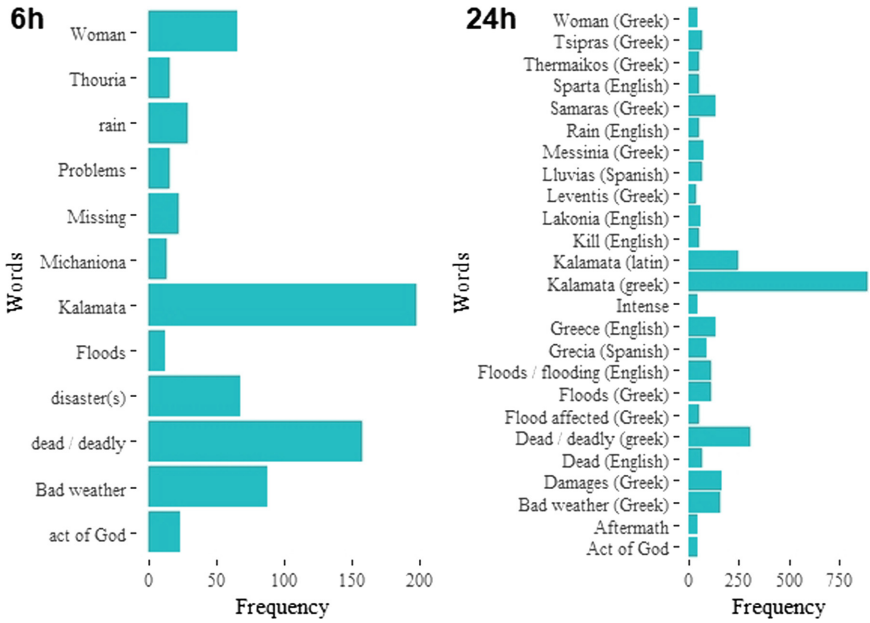
Fig. 3. Word frequency of tweet corpus, published within 6 h (left) and 24 h (right).
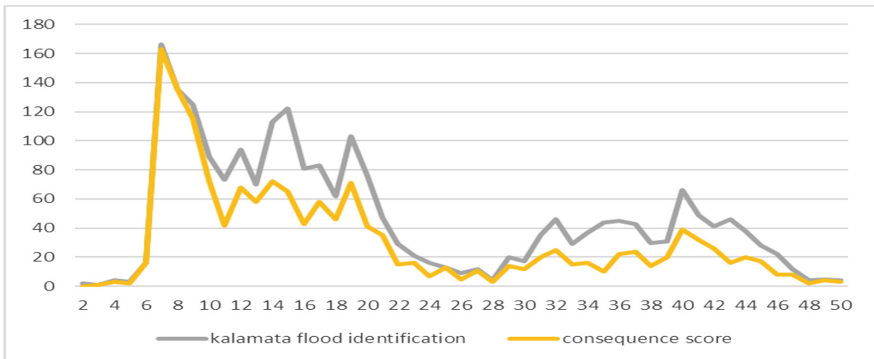
Fig. 4. Frequency of tweets related to flood identification and consequences

Figure 6 is interesting for assessing the information published hourly, considering also if it is day or night. As it can be seen, early in the day the spread of information is higher that later in the night. Moreover, there is a tendency that twitter users post more related information during the first day of the flood event occurrence, while the info related to disaster management and the expression of emotions is widely posted during the second day as well.
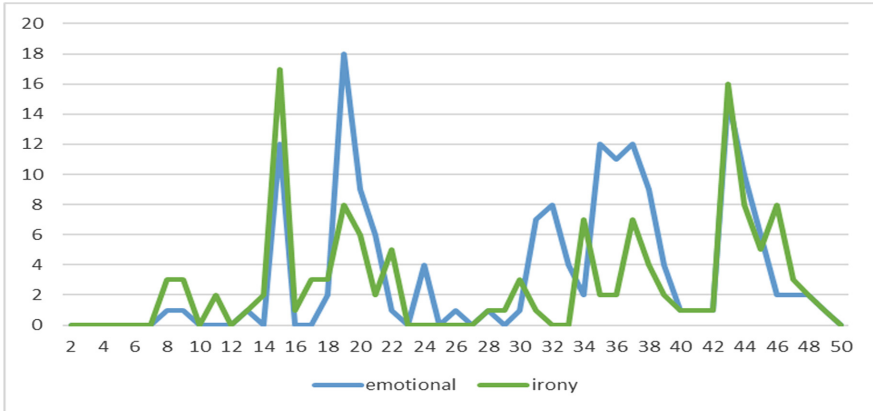
Fig. 5. Frequency of tweets related to ironic and emotional tweets

On Fig. 7 (Map I) the frequency of tweets that are related to rain identification, is visualized, spatially. As it can be easily seen, the municipality of Kalamata is referred mostly in this thematic twitter dataset. According to author's opinion a map like this could be used as an indicator that something extraordinary is happening, making sense in this way, to the stakeholders of disaster management as it could demonstrate an empiric evidence that in areas where there are too many tweets that report rain, there is a danger of a flood event occurrence.

Finally, Fig. 8 (Map II) displays the consequence score values that were extracted from the thematic twitter text corpus. Each bullet represents four reports; the red color is associated with the score value V, which represents human loss while bullets in orange, represent major damages and human injuries (score value IV). The author supports that this map, could also make sense to the authorities of disaster management and to people that have non-trained to science eyes.
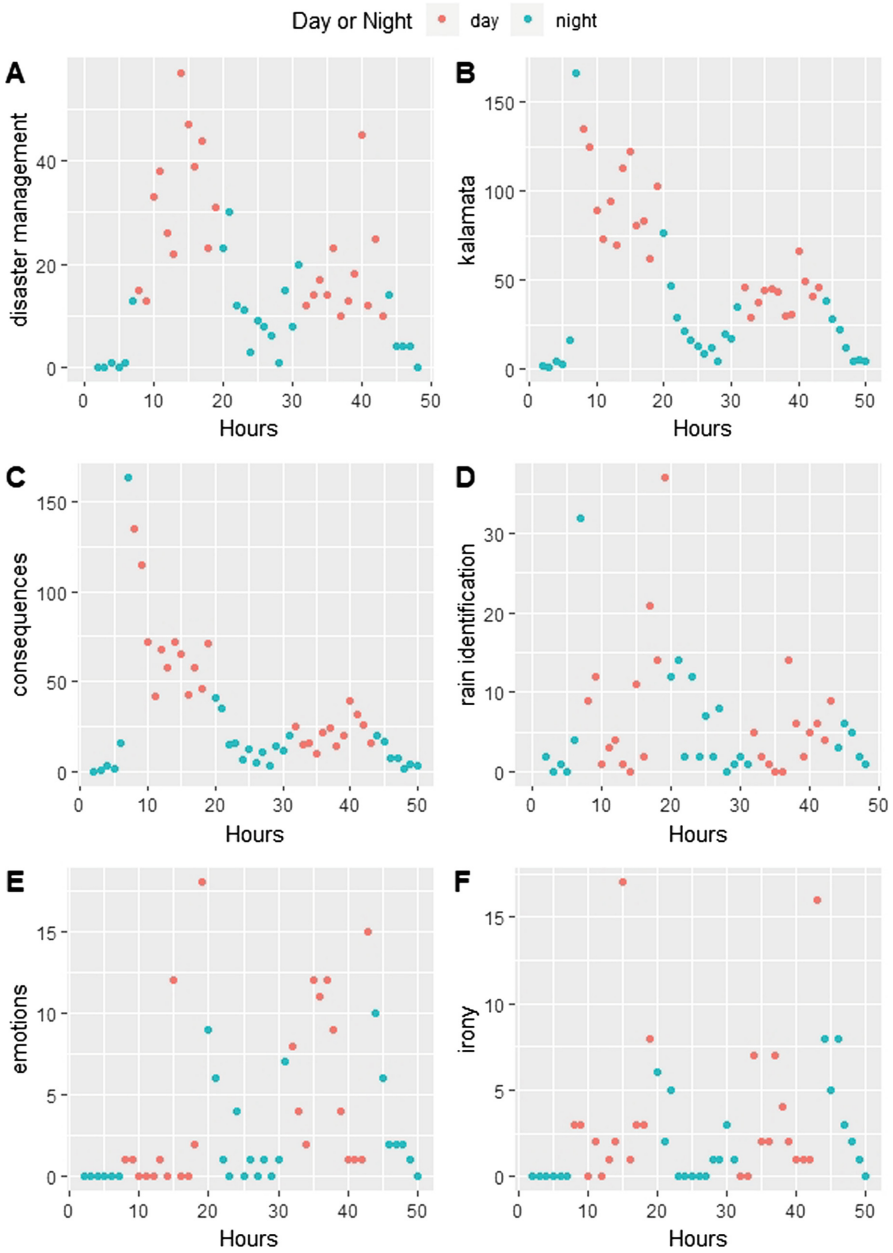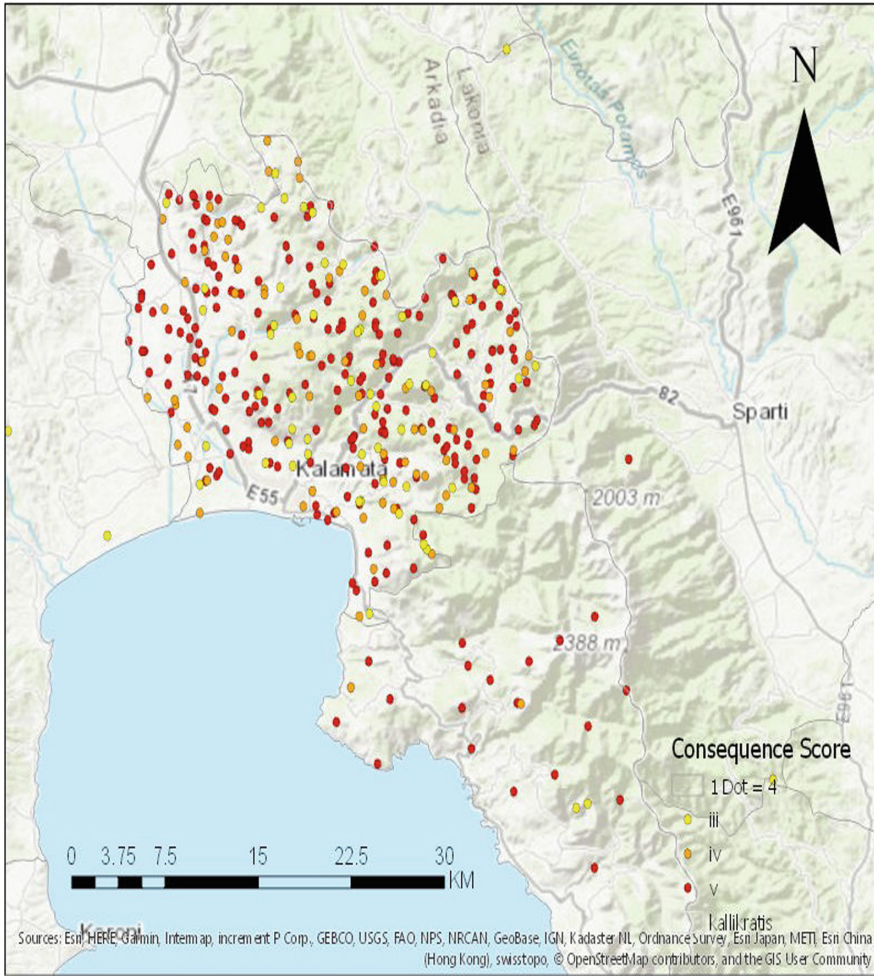
Fig. 6. Scatter Plots of classification categories vs time period

Fig. 7. Map I: Count of tweets related to rain identification

Fig. 8. Map II: Map of consequence score values extracted from the tweet corpus

5   Conclusions

In general, it is assumed that social media, as a source of VGI, can contribute to disaster management, even in areas of medium-sized population while, by creating various simple graphs and maps, the extracted information can make sense to the DM stakeholders. Towards that step though, there are quite a few challenges; appropriate classification, correct and precise geo-referencing, meaningful graphs and automation are the most important of those. Considering the above, next steps of cur-rent research will be focused on classifying the tweets automatically and effectively, probably by the use of Machine Learning. Moreover, another vital aspect of VGI is linked to the quality and credibility of information that neo-geographers or social media users post. The author,

in current research, accepts the validity of the Law of Linus in VGI, according to which, the more users post information about an incident in an area, the more credible and accurate the total information becomes. And there empiric evidence, that especially regarding DM information, social media users do not tend to post lies. After all, the description of consequences of a natural event, or the identification of a rain, is not a subjective or a controversial matter that different people may have different notions about it. However, as a future step in current re-search, various quality-related validations should be applied for ensuring the credibility of the outcome.

References

1. Arapostathis, S.G., Spyrou, N., Drakatos, G., Kalabokidis, K., Lekkas, E., Xanthopoulos, G.: Mapping information related to floods, extracted from VGI sources, for effective disaster management within the Greek territory; the floods of West Attica (November 2017 Greece) case study. In: 11th International Conference of the Hellenic Geographical Society (2018)
2. Arapostathis S.G.: Automated methods for effective geo-referencing of tweets related to disaster management. In: International Geomatics Applications Conference – Book of Abstracts (online), p. 62 (2018)
3. Bello, G.M.C., Cohen, M.A., Angelica, N., Zaragoza, M., Corona, A.H.L., Perez, J.R.: Flood vulnerability reduction using a partial participatory GIS approach. In: Proceedings of the Gi4DM Congress, Istanbul, 18–21 March 2018 (2018)
4. De Brito M., Evers, M., Almoradie A.D.S.: Participatory flood vulnerability assessment: a multi-criteria approach. Copernicus Publications, European Geosciences Union (2018)
5. Dashti S., Palen L., Heris M.P., Anderson K.M., Anderson S., Anderson T.J.: Supporting disaster reconnaissance with social media data: a design-oriented case study of the 2013 Colorado Floods. In: Proceedings of the 11th International ISCRAM Conference, University Park, Pennsylvania, USA (2013)
6. De 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. Int. J. Geogr. Inf. Sci. **29**(4), 667–689 (2015)
7. Ghavipour, M., Meybodi, M.R.: Trust propagation algorithm based on learning automata for inferring local trust in online social networks. Knowl. Based Syst. **143**, 307–316 (2018)
8. Gründer-Fahrer, S., Schlaf, A., Wustmann, S.: How social media text analysis can inform disaster management. In: Rehm, G., Declerck, T. (eds.) GSCL 2017. LNCS (LNAI), vol. 10713, pp. 199–207. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73706-5_17
9. He, W., Zha, S., Li, L.: Social media competitive analysis and text mining: a case study in the pizza industry. Int. J. Inf. Manag. **33**(3), 464–472 (2013)
10. Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., Meier, P.: Extracting information nuggets from disaster-related messages in social media. In: ISCRAM (2013)
11. McDougall K.: Using volunteered information to map the queensland floods. In: Proceedings of the Surveying and Spatial Sciences Biennial Conference, Wellington, New Zealand, 21–25 November 2011 (2011)
12. Rahman, M.M., Barua, U., Khatun, F., Islam, I., Rafiq, R.: Participatory Vulnerability Reduction (PVR): an urban community-based approach for earthquake management. Nat. Haz. **93**(3), 1479–1505 (2018)

13. Smith, L., Liang, Q., James, P., Lin, W.: Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. J. Flood Risk Manag. **10**(3), 370–380 (2017)
14. Zhong, X., Duckham, M., Chong, D., Tolhurst, K.: Real-time estimation of wildfire perimeters from curated crowdsourcing. Sci. Rep. **6**, 24206 (2016). To Enhance Emergency Situation Awareness IEEE Computer Society